



СОФИЙСКИ УНИВЕРСИТЕТ “СВ. КЛИМЕНТ
ОХРИДСКИ”

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

Методология за предсказване на успеха за технологични
стартиращи компании в България

Модели и софтуер за прогнозиране на успеха на стартиращи компании

Methodology for Bulgarian Technological Start-ups Success
Prediction

Models and a software application for success prediction of start-up companies

ДИСЕРТАЦИОНЕН ТРУД

за присъждане на образователна и научна степен "Доктор"
по научната специалност "Информатика",
професионално направление „Технологично предприемачество”

Докторант: Боян Бориславов Янков

Научен ръководител: доц. д-р Петко Русков Русков

София, 2015 г.

СЪДЪРЖАНИЕ

СЪДЪРЖАНИЕ.....	2
СПИСЪК НА ФИГУРИТЕ	6
СПИСЪК НА ТАБЛИЦИТЕ.....	9
ИЗПОЛЗВАНИ ТЕРМИНИ, СЪКРАЩЕНИЯ И КОНВЕНЦИИ.....	10
ГЛАВА 1: ВЪВЕДЕНИЕ	11
1.1 Дефиниция на проблема	11
1.2 Цели и задачи на дисертационния труд	12
1.3 Структура на дисертацията.....	12
ГЛАВА 2: ОБЗОР НА ПРОБЛЕМА И РЕЗУЛТАТИ ОТ ОПИТИТЕ ЗА РЕШАВАНЕТО МУ ДО СЕГА.....	14
2.1 Факти, тенденции, причини и актуалност.....	14
2.2 Дефиниции на основни понятия	16
2.2.1 Дефиниция за технологично предприемачество	17
2.2.2 Дефиниция за стартираща компания	19
2.2.3 Дефиниция за успешна стартираща компания.....	19
2.3 Избор на методология за изследването	23
2.4 Опити за решение, достижения и дефицит	25
2.4.1 Сравнение на съществуващи теоретични модели за прогнозиране успеха на стартиращи компании.....	26
2.4.2 Съществуващи теоретични модели за прогнозиране успеха на стартиращи компании, базирани на модела на Сендбърг	30
2.4.3 Проучване и анализ на софтуерни продукти за прогнозиране успеха на стартиращи компании.....	32
2.5 Обзор на софтуерни продукти за извличане на знания от данни	38
2.5.1 Обзор на софтуерния продукт IBM SPSS Statistics	38
2.5.2 Обзор на софтуерния продукт IBM SPSS Modeler	38
2.5.3 Обзор на софтуерния продукт Weka	39
2.5.4 Обзор на софтуерния продукт HP Vertica	39
2.5.5 Обзор на софтуерния продукт Microsoft SQL Server.....	40
2.5.6 Изводи и заключение.....	40
2.6 Избор на технологии и инструменти за разработка на информационна система ..	41
2.6.1 Избор на тип приложение	41
2.6.2 Избор на операционна система.....	41

2.6.3 Избор на приложение за уеб сървър	42
2.6.4 Избор на програмни средства за разработка от страна на сървъра.....	42
2.6.5 Избор на програмни средства за разработка от страна на клиента.....	43
2.6.6 Избор на бази данни	44
2.6.7 Избор на среда за разработка.....	44
2.7 Изводи и заключение	45
ГЛАВА 3: РАЗРАБОТВАНЕ НА МОДЕЛ ЗА ПРЕДСКАЗВАНЕ НА УСПЕХА.....	46
3.1 Модел на процеса на създаване на фирма.....	46
3.2 Абстрактен модел за предсказване на успеха на стартиращи фирми	47
3.3 Детайлизиран модел за предсказване на успеха на стартиращи фирми в България	49
3.4 Изводи и заключение	52
ГЛАВА 4: ИЗСЛЕДВАНИЯ, ПРИЛОЖЕНИЯ И ЕКСПЕРИМЕНТИ	53
4.1 Качествено изследване на факторите за успех на стартиращи фирми.....	53
4.2 Количествено изследване на факторите за успех на стартиращи фирми	56
4.2.1 Анализ на данните от количественото изследване.....	59
4.2.2 Подготовка на данните от количественото изследване	62
4.2.3 Групиране на факторите за успех.....	68
4.2.4 Извличане на знания от данни и създаване на модел за предсказване на успеха с IBM SPSS Modeler.....	74
4.2.5 Извличане на знания от данни и създаване на модел за предсказване на успеха с Weka	87
4.2.6 Синтезиране на модели за предсказване на успеха за технологични компании	93
4.2.7 Сравнение на получените модели	99
4.3 Резултати, изводи и заключение	100
ГЛАВА 5: РЕАЛИЗАЦИЯ НА СОФТУЕР ЗА ПРОГНОЗИРАНЕ НА УСПЕХА.....	102
5.1 Изисквания към софтуера за прогнозиране успеха на стартиращи компании	102
5.2 Проектиране на софтуера за прогнозиране успеха на стартиращи компании	103
5.2.1 Обща архитектура на приложението	104
5.2.2 Модули на приложението	106
5.3 Разработка на софтуер за прогнозиране на успеха	108
5.3.1 Алгоритъм за прогноза.....	108
5.3.2 Графика на резултата за прогноза	112

5.3.3 Програмни интерфейси на Google.....	114
5.3.4 Съхранение на данни	115
5.3.5 Функционалности, използващи програмни интерфейси на Google.....	118
5.3.6 Дизайн шаблони	121
5.3.7 Сигурност и надеждност	121
5.4 Тестване на софтуера за прогнозиране на успеха	123
5.4.1 Вход в системата	123
5.4.2 Настройки на приложението.....	125
5.4.3 Преглед на съдържанието на анкетен файл.....	126
5.4.4 Прогноза за успех.....	127
5.4.5 Административна част	129
5.4.6 Резултати, изводи и заключение.....	131
5.5 Апробиране и валидиране на софтуерния продукт	132
5.6 Резултати, изводи и заключение	133
ЗАКЛЮЧЕНИЕ	135
ОСНОВНИ ПРИНОСИ НА ДИСЕРТАЦИОННИЯ ТРУД	136
ПУБЛИКАЦИИ, СВЪРЗАНИ С ТЕМАТА НА ДИСЕРТАЦИЯТА.....	138
ИЗПОЛЗВАНА ЛИТЕРАТУРА	139
ДЕКЛАРАЦИЯ ЗА ОРИГИНАЛНОСТ	146
ПРИЛОЖЕНИЯ.....	147
ПРИЛОЖЕНИЕ 1: РЕЧНИК.....	147
ПРИЛОЖЕНИЕ 2: РЕЗУЛТАТИ ОТ АНАЛИЗИ СЪС СОФТУЕРНИ ПРИЛОЖЕНИЯ ЗА ИЗВЛИЧАНЕ НА ЗНАНИЯ ОТ ДАННИ.....	151
Резултати от факторен анализ със софтуерния продукт IBM SPSS Statistics	151
Резултати от синтезиране на модели за предсказване на успеха със софтуерния продукт IBM SPSS Modeler.....	160
Резултати от синтезиране на модели за предсказване на успеха със софтуерния продукт Weka	172
ПРИЛОЖЕНИЕ 3: ПРОГРАМЕН КОД НА ИНФОРМАЦИОННАТА СИСТЕМА IZSP	175
Програмен код на клас „GoogleService“:	175
Програмен код на клас „AnalyseService“:.....	178
Програмен код на клас „UserService“:	186
Програмен код на фронтенд шаблон „\templates\admin_manageusers.tpl“:	191

Програмен код на фронтенд шаблон „\templates\survey_data_analyse.tpl“:.....	192
Изнесена информация за структурата на базата данни във формат .SQL:.....	195

СПИСЪК НА ФИГУРИТЕ

Фиг. 1 Брой публикации и интерес на тема предприемачество (създадена от автора)...	15
Фиг. 2 Времева рамка на развитието на методологията Lean Startup	18
Фиг. 3 Връзката между технологичната промяна и предприемачество (създадена от автора по описания от Тони Байлетти)	19
Фиг. 4 Навлизане на нов продукт на пазара - прескачане на „пропастта“ между ранния и основния пазар (адаптирана по Тони Байлетти)	21
Фиг. 5 J-крива на паричния поток на технологична стартираща компания (адаптирана по Грабенвартер)	22
Фиг. 6 Декомпозиция на четири нива на методологията CRISP-DM (преведена от автора по фигура на Чапман)	24
Фиг. 7 Жизнен цикъл на проект за извличане на знания от данни по методологията CRISP-DM (преведена от автора по фигура на Чапман)	25
Фиг. 8 Резултати от инструмента за прогноза „Калкулатор на риска за стартиращи компании“ (създадена от автора чрез прилагане на инструмента)	34
Фиг. 9 Фрагмент от резултати за прогноза на успеха на инструмента „Startup Compass“ (създадена от автора чрез прилагане на инструмента)	35
Фиг. 10 Резултат на прогноза за успех инструмента на „Think Blueprint“ (създадена от автора чрез прилагане на инструмента)	36
Фиг. 11 Избор на приложение от графичния потребителски интерфейс на Weka (екранна снимка, създадена от автора)	39
Фиг. 12 Модел на процеса на създаване на нова фирма (преведена от автора по Джоан и Джеймс Карлънд)	46
Фиг. 13 Допълнен модел на процеса на създаване на нова фирма (създадена от автора)	47
Фиг. 14 Модел за предсказване на успеха на стартиращи компании (създадена от автора)	48
Фиг. 15 Част от анкетата за валидиране на модела за предсказване на успеха на стартиращи компании (създадена от автора)	59
Фиг. 16 Размер на компаниите (създадена от автора)	60
Фиг. 17 Възраст на компаниите (създадена от автора)	60
Фиг. 18 Индустриален сектор на компаниите (създадена от автора)	61
Фиг. 19 Пол на респондентите (създадена от автора)	62
Фиг. 20 Възраст на респондентите (създадена от автора)	62
Фиг. 21 Извадка от данните от анкетата, прегледани в Google диск (създадена от автора)	63
Фиг. 22 Извадка от данни от анкетата, след обработка в Microsoft Excel (създадена от автора)	66
Фиг. 23 Извадка от данни от анкетата, въведени в IBM SPSS Statistics (създадена от автора)	66
Фиг. 24 Извадка на променливи, към които се отнасят данните от анкетата, в IBM SPSS Statistics (създадена от автора)	67

Фиг. 25 Извадка от данни, заредени в софтуерния продукт Weka (създадена от автора)	68
Фиг. 26 Последователност от действия в IBM SPSS Modeler за получаване на класификационни модели – графично представяне (създадена от автора)	75
Фиг. 27 Избор на полета за автоматичния класификатор – входни данни и цели (създадена от автора)	76
Фиг. 28 Най-добри получени модели за предсказване на успеха при автоматична класификация (създадена от автора)	77
Фиг. 29 Оценка на точността на модел, получен чрез алгоритъма C5.0 и прилагане на кръстосана валидация (създадена от автора)	78
Фиг. 30 Разпределението на коректно и некоректно класифицираните от модела компании (създадена от автора)	79
Фиг. 31 Цялостно графично представяне (карта) на класификационното дърво (създадена от автора)	81
Фиг. 32 Класификационно дърво – първо ниво: Наличие на ясно конкурентно предимство (създадена от автора)	82
Фиг. 33 Класификационно дърво – клон 2: външната среда е ключов фактор за успеха на компанията (създадена от автора)	82
Фиг. 34 Класификационно дърво – клон 2.2: нематериални активи – изградена бизнес репутация на компанията (създадена от автора)	83
Фиг. 35 Класификационно дърво – клон 2.2.2: начин на навлизане на стартиращата компания на пазара (създадена от автора)	83
Фиг. 36 Графично представяне на класификационно дърво (преведена от автора)	84
Фиг. 37 Алгоритми за класификация в Weka (създадена от автора)	88
Фиг. 38 Настройка на алгоритъм J48 в Weka и създаване на класификационен модел (създадена от автора)	89
Фиг. 39 Схематична визуализация на класификационното дърво, генерирано с алгоритъм J48 в Weka (създадена от автора)	93
Фиг. 40 Най-добри получени модели за предсказване на успеха на технологични компании при автоматична класификация (създадена от автора)	94
Фиг. 41 Оценка на точността на модел за предсказване на успеха на технологични компании, получен чрез алгоритъма C5.0 и прилагане на кръстосана валидация (създадена от автора)	94
Фиг. 42 Цялостно графично представяне (карта) на класификационното дърво за технологични компании (създадена от автора)	96
Фиг. 43 Класификационно дърво – първо ниво: Повечето компании в индустрията имат добра печалба (създадена от автора)	97
Фиг. 44 Класификационно дърво – клон 1: ключови фактори за успеха: цена (създадена от автора)	97
Фиг. 45 Класификационно дърво – клон 1.2: ключови фактори за успеха: маркетинг и реклама (създадена от автора)	98
Фиг. 46 Класификационно дърво – клон 1.2.2: избраната стратегия е агресивна. (създадена от автора)	98
Фиг. 47 Диаграма на потока от данни в I3SP (създадена от автора)	103

Фиг. 48 Трислоен модел Модел – Изглед - Контролер при уеб базирано приложение, реализирано с програмния език PHP и база данни MySQL (адаптирана от автора по фигура от betterexplained.com)	104
Фиг. 49 Схематично представяне на архитектурата и функционалностите на приложението (създадена от автора).....	105
Фиг. 50 UML диаграма на действията при прогнозиране успеха на стартираща компания с ISP (създадена от автора)	109
Фиг. 51 Схема на алгоритъма за прогноза на приложението (създадена от автора).....	110
Фиг. 52 Фрагмент от файла-речник, използван в приложението (създадена от автора)	111
Фиг. 53 Регулярен израз за извличане на данни от файл, съдържащ класификационно дърво (създадена от автора)	112
Фиг. 54 Фрагмент от класификационното дърво, представено в страницата за анализ (създадена от автора)	113
Фиг. 55 Диаграма на структурата на базата данни на приложението (създадена от автора)	116
Фиг. 56 Фрагмент от изходен файл на IBM SPSS Modeler (създадена от автора).....	117
Фиг. 57 Фрагмент от Анкетен файл (създадена от автора)	118
Фиг. 58 Фрагмент от кода на дизайн шаблон в Smarty (създадена от автора).....	121
Фиг. 59 Вход в приложението с Google акаунт (създадена от автора)	123
Фиг. 60 Диалогов прозорец за разрешаване на достъп до Google диск (създадена от автора)	124
Фиг. 61 Изчакване на одобрение след изпратена заявка за достъп до системата (създадена от автора)	124
Фиг. 62 Страница с настройки в приложението (създадена от автора)	125
Фиг. 63 Диалогов прозорец за избор на анкетен файл от Google диск (създадена от автора)	126
Фиг. 64 Страница с визуализация на съдържанието на анкетен файл (създадена от автора)	127
Фиг. 65 Фрагмент от страницата с резултати от прогноза за успеха на компания (създадена от автора)	128
Фиг. 66 Запазени резултати от анализи (създадена от автора)	129
Фиг. 67 Управление на потребителите (създадена от автора).....	130
Фиг. 68 Качване на нов алгоритъм (класификационно дърво) в системата (създадена от автора)	131

СПИСЪК НА ТАБЛИЦИТЕ

Таблица 1 Дефиниции на използваните термини и съкращения.....	10
Таблица 2 Заявки за търсене и резултати в научни бази данни.....	14
Таблица 3 Брой публикации на тема предприемачество и предсказване на успеха на стартиращи компании.....	15
Таблица 4 Сравнение на модели за предсказване на успеха на стартиращи компании..	26
Таблица 5 Сравнение на модели за прогноза на успеха, базирани на модела на Сендбърг	31
Таблица 6 Обобщена съпоставка на анализирани инструменти за предсказване на успеха	36
Таблица 7 Сравнение на операционни системи за уеб сървър	41
Таблица 8 Обобщени резултати от качествено изследване на мненията във връзка с предложения модел за предсказване на успеха.....	54
Таблица 9 Тест на Кайзер-Майер-Олкин и Барлет	68
Таблица 10 Корелационна матрица, след прилагане на ротация	69
Таблица 11 Сравнение на категориите от теоретичния модел за предсказване на успеха и получените категории от факторния анализ.....	72
Таблица 12 Значение на компонентите от корелационната матрица.....	73
Таблица 13 Сравнение на класификационни модели, получени с Weka.....	89
Таблица 14 Съпоставка на модели за предсказване успеха на стартиращи компании ...	99
Таблица 15 Резултати от апробиране на софтуерния продукт I3SP.....	132
Таблица 16 Основни приноси на дисертационния труд.....	136

ИЗПОЛЗВАНИ ТЕРМИНИ, СЪКРАЩЕНИЯ И КОНВЕНЦИИ

Таблица 1 Дефиниции на използваните термини и съкращения

Термин / съкращение	Дефиниция
I3SP (Information System for Start-ups Success Prediction)	В превод „информационна система за предсказване успеха на стартиращи компании“, софтуерна информационна система, разработена към настоящата дисертация, която служи за прогнозиране на вероятността за успех на стартиращи компании и визуализация на класификационно дърво с факторите за успех.
Lean Startup	Методология за развитие на стартиращи компании, въведена от Ерик Райз, която цели да акцентира върху процесите в компанията, чрез итеративно изграждане на продуктите и задоволяване на нуждите на различни потребителски сегменти
NCIIA (The National Collegiate Inventors and Innovators Alliance)	В превод „Национален университетски съюз на изобретателите и иноваторите“, организация в САЩ, която обучава предприемачи и финансира студентски компании.
NVP (New Venture Performance)	В превод „производителност на стартираща компания“, крайният резултат от процеса на създаване и организиране на нов бизнес, който разработва, създава и маркетира продукти и услуги, за да задоволи пазарни потребности с цел печалба и растеж.

Често използвани популярни термини и съкращения, които се срещат в дисертацията, за изнесени в **Приложение 1: Речник**.

ГЛАВА 1: ВЪВЕДЕНИЕ

В условията на икономика на прехода, характерни за Централна и Източна Европа, се разчита в голяма степен на стартиращите компании, за да генерират условия за икономически растеж. В приоритетите на програмния период на Европейския Съюз от 2014 до 2020 година „Хоризонт 2020” е залегнала необходимостта от развитието на иновативния малък и средния бизнес, с цел генериране на икономически растеж и работни позиции (European Commission, 2014). Предприемачите и техните стартиращи компании увеличават конкурентността на икономиките на прехода чрез своите иновативни стратегии и бизнес модели, понякога противоречиви практики и нестихващата си енергия за успех (Peng M. W., 2001). Бизнесът също има нужда от стартиращи компании, които предлагат иновативни решения, за да подпомогнат развитието на различните индустриални сектори в българската икономиката. Освен това, те са добра инвестиционна възможност с потенциално висока възвръщаемост.

Присъствието на малки технологични компании, както и на по-големи корпоративни предприемачески екипи, на пазара, е свързано с технологична промяна (Jelinek, 1996). За това допринасят както малките компании, които са преобладаващ брой и се стремят към ефективност, така и средните и големи компании, развивайки корпоративно предприемачество и разработвайки големи и амбициозни проекти.

Към момента в България има създадени няколко бизнес инкубатора, фондове за рискови капитали, както и програми, финансирани от правителството и европейския съюз за подпомагане на малките и средни предприятия. Тази тенденция показва важността и фокуса на правителството, Европейския съюз и бизнеса към развитието на стартиращи компании в България. Резултатите от тази дейност все още са в малък мащаб и повечето предприемачи стартират самостоятелно, често без да разполагат с необходимия опит, информация и подкрепа, което понякога води до неуспехи и често е причина за неефективност, пропуснати възможности и загуба на направените инвестиции (Avramov, 2012). В България представата за предприемачите е изкривена, от тях се очаква да създават работни места, а развитието на стойностен и конкурентен продукт или услуга остава на заден план. Едва един от всеки пет иска да стане предприемач, което е двойно по-малко от средното за Европа (Стойчева, 2013).

1.1 Дефиниция на проблема

Предсказването на успеха на стартиращи компании е възможност за увеличаване на ефективността на процеса на стартиране на бизнес и за намаляване на риска и разхода на ресурси. Успехът на дадена компания е обвързан с множество фактори, като икономика, трудова заетост, образование, иновативност и добавена стойност на новите продукти или услуги, които компанията ще разработва, фирмената стратегия, качествата на екипа и други. Създадени са модели за предсказване на успеха на стартиращи компании, които са специфични за различни пазари, държави, икономически особености, както и понякога за моментната ситуация на тяхното създаване. За съжаление, в България няма много разработени модели за прогнозиране на успеха на стартиращи компании и почти никакъв софтуер, който да прилага тези модели и да ги прави достъпни до заинтересованите страни. България е малка екосистема и бизнес климатът е специфичен със своите етап на развитие, регулации, локация, психология и др. Това поражда нуждата от създаване на адаптиран специално за местния пазар модел за прогноза, който да служи като основа на софтуер за прогнозиране на успеха на стартиращи компании в България.

Модел за предсказване на успеха на стартиращи български технологични компании би бил от полза на предприемачи, собственици на бизнес, бизнес

инкубатори, училища и университети, обучаващи студенти по предприемачество, бизнес консултанти, бизнес анализатори, инвеститори и фондове за рисков капитал. Моделът може да бъде използван както за предсказване на успеха на стартиращи компании, така и за идентифициране на факторите довели до този резултат, идентифициране на възможните силни и слаби страни на компанията, за подобряване общото знание на заинтересованите страни за компанията, за измерване на прогреса ѝ за период от време, и за правене на сравнения между отделни компании.

На базата на модел за предсказване на успеха на стартиращи компании може да се разработи **софтуер, който да прилага модела**. Такъв софтуер би бил от полза както на самите предприемачи, започващи своя нов бизнес, така и на всички заинтересовани в процеса страни. С използването на софтуера се очаква рисковете за инвеститори в стартиращи компании, произтичащи от загубата на капитали при евентуален неуспех, да са по-малки, тъй като по-лесно ще може да се анализира състоянието на конкретен стартираш бизнес. На база на получени от специална анкета данни за компанията, софтуерът ще дава прогноза на ранен етап за шансовете за успех и неуспех. Той би могъл да се използва и за анализ на стартираща компания или съществуваща такава, като ще дава индикации къде и какво трябва да се промени в компанията, за да се получат по-високи прогнозни резултати, които на по-късен етап да доведат и до по-добър бизнес резултат.

1.2 Цели и задачи на дисертационния труд

Основната цел на настоящия дисертационен труд е да се създаде модел, методика и прототип на информационна система за предсказване на успеха на стартиращи компании и на технологични стартиращи компании от България, което да спомогне вземането на управленски и инвестиционни решения.

За постигане на поставената цел се дефинират следните **задачи**:

1. Да се проучи, изследва и анализира съвременното състояние у нас и в света в библиотеки, научни бази данни, научни трудове и интернет на моделите за предсказване успеха на стартиращи компании, факторите за успех на стартиращи компании, софтуерните продукти за извличане на знания от данни и тяхното приложение при класификационни модели за предсказване на успеха.
2. Да се изготви теоретичен модел за предсказване на успеха на стартиращи компании от България, който да бъде адаптиран за средата и за българските предприемачи посредством качествено изследване.
3. Теоретичният модел да се валидира чрез количествено изследване, и чрез софтуерни продукти за извличане на знания от данни да се изготвят модели за предсказване на успеха, адаптирани за компании от България. Да се изготви и модел за предсказване на успеха на технологични стартиращи компании.
4. Да се проектира и реализира прототип на информационна система за предсказване на успеха на стартиращи компании, която да позволява прилагане на получените в изследването модели и бъдещо надграждане и подобряване на алгоритъма на база събраните данни.

1.3 Структура на дисертацията

Текстът на настоящата дисертация е организиран в 5 глави, списък на фигурите, списък на таблиците, основни използвани термини, заключение, библиография и 3 приложения.

Глава 1 въвежда накратко в проблемната област и запознава най-общо с проблема за липсата на модели за предсказване на успеха на стартиращи компании, адаптирани за българския пазар. В тази глава се изяснява важността на проблема. Определят се целите и задачите на настоящата дисертация - създаване на методология, модели и софтуерен инструмент за предсказване на успеха на стартиращи технологични компании от България.

Глава 2 представлява подробен обзор на проблемната област и на вече известните методи, подходи и модели за решаване на задачи, подобни на задачата от настоящата дисертация. В тази глава се дефинират понятията технологично предприемачество, стартираща компания и успешна стартираща компания. Разгледани са и са сравнени софтуерни продукти за предсказване на успеха на стартиращи компании, които са мотивация за разработка на софтуера към настоящата дисертация. Проучени са и са избрани софтуерни продукти за извличане на знания от данни, които ще бъдат приложени в изследването. Анализирани са и са избрани технологиите за разработка, които ще бъдат използвани при реализацията на софтуерна система.

Глава 3 представя разработените от автора методология и модели за предсказване на успеха на компании от България. Предложени са теоретични абстрактни модели на процеса на създаване на компания и на успешна компания. Моделът на успешна компания е разработен в детайли, като са описани конкретните фактори за успех. Предложените модели представляват хипотези, които са проверени чрез качествено и количествено изследване, описани в Глава 4.

Глава 4 представя изследвания и експерименти, базирани на изготвения теоретичен модел на успеха на стартираща компания от Глава 3. Моделът е подобрен с помощта на провеждане на качествено изследване – интервюта с предприемачи и собственици на компании. След това, той е валидиран с помощта на изследване с количествен метод и прилагане на методология за извличане на знания от данни, като за целта са използвани различни софтуерни продукти. Направен е факторен анализ, с който са потвърдени предложените категории фактори за успеха на стартиращи компании. Създадени са модели за предсказване на успеха, чиято точност е проверена посредством метода на кръстосана валидация. Получените модели са сравнени и са разгледани алгоритмите, използвани за тяхното синтезиране. Идентифицирани са ключовите фактори за успех на стартиращи компании от България.

Глава 5 представлява подробно описание на реализирания софтуер за предсказване на успеха на стартиращи компании. Разгледани са изискванията към софтуерния продукт и са избрани технологиите за разработка. Представена е цялостната архитектура на приложението и в детайли са описани основните модули от нея и използваните външни библиотеки. Представени са резултати от тестването и приложението на софтуера на практика.

ГЛАВА 2: ОБЗОР НА ПРОБЛЕМА И РЕЗУЛТАТИ ОТ ОПИТИТЕ ЗА РЕШАВАНЕТО МУ ДО СЕГА

Настоящата глава представя подробен литературен обзор на проблемната област и на вече известните методи, подходи и модели за решаване на задачата за предсказване на успеха на стартиращи компании. Дефинирани са понятията технологично предприемачество, стартираща компания и успешна стартираща компания. Разгледани са и са сравнени софтуерни продукти за предсказване на успеха на стартиращи компании. Проучени са и са избрани софтуерни приложения за извличане на знания от данни, които ще бъдат използвани в изследването. Анализирани са и са избрани технологиите за разработка, които ще бъдат приложени при реализацията на софтуерна система.

2.1 Факти, тенденции, причини и актуалност

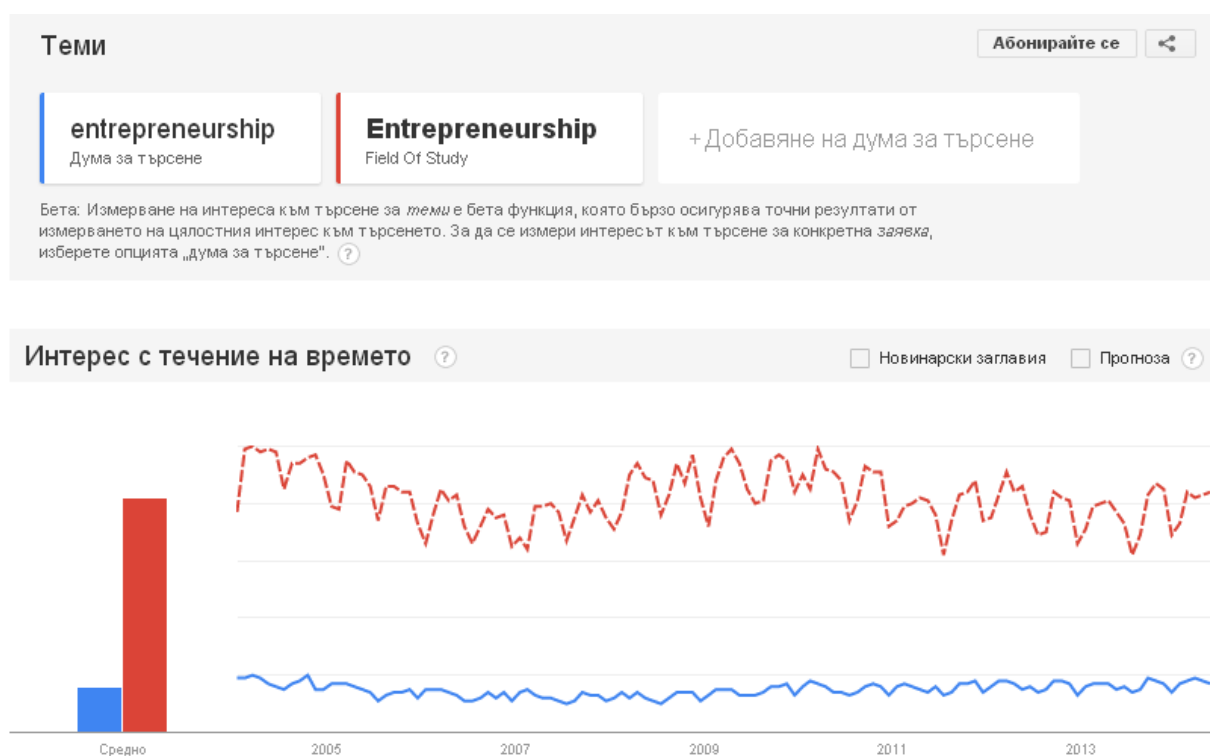
Настоящото изследване започва с литературен обзор. Проучени са литературни източници чрез търсения в научни бази данни (Emerald, 2014), (Elsevier, 2014), (ProQuest, 2014), (Google Inc., 2014), както и чрез конкретни търсения в интернет търсачки. Прегледани са общо над 1500 документа, отговарящи на ключовите думи и заявките за търсене (Таблица 2).

Таблица 2 Заявки за търсене и резултати в научни бази данни

Заявка за търсене	emeraldinsight.com	www.sciencedirect.com	dissexpress.umi.com/dxweb	scholar.google.bg
technology entrepreneurship prediction	425 резултата	3 146 резултата	40+ резултата	36 200 резултата
entrepreneurship success prediction	426 резултата	2 765 резултата	40+ резултата	41 900 резултата
В заглавието: entrepreneurship success	0 резултата	7 резултата	14 резултата	-
В заглавието: entrepreneurship prediction	0 резултата	1 резултата	2 резултата	-

На база разгледаните литературни източници, бяха анализирани и съпоставени хипотези, модели, методологии, изследвания и инструменти за предсказване на успеха на стартиращи компании.

На база направеното проучване, установих, че **предприемачеството** е популярна научно-приложна тема и област на изследване с над 40 000 публикации годишно в световен мащаб (Google Inc., н.д.). То запазва своята популярност стабилна в периода 2005-2014 година, което се забелязва от горната крива на приложената графиката – Фиг. 1, генерирана чрез услугата Google Тенденции. На долната крива се вижда обемът на търсенето на предприемачеството като ключова дума в периода 2005-2014 година. Кривата показва, че общественият интерес към темата е стабилен.



Фиг. 1 Брой публикации и интерес на тема предприемачество (създадена от автора)

Годишният брой публикации на тема предприемачество в световен мащаб бързо нараства в периода 1990-2007 година, като се увеличава над 10 пъти (Таблица 3). В периода 2008-2013 година, броят публикации на тема предприемачество се стабилизира между 40 и 50 хиляди годишно, като между 2011-2013 година отбелязва малък спад, който вероятно се дължи на вече достигнато насищане, тъй като интересът към темата остава относително постоянен. Темата за предприемачеството ще продължи да бъде актуална в Европа, тъй като заляга в приоритетите на Европейския Съюз за програмния период от 2014 до 2020 година „Хоризонт 2020“. Акцентът ще бъде върху иновациите в рамките на малкия и среден бизнес.

Таблица 3 Брой публикации на тема предприемачество и предсказване на успеха на стартиращи компании

Година	Брой публикации за предприемачество	Брой статии за предсказване на успеха на стартиращи компании
1990	3290 ↑	349 ↑
1991	3500 ↑	425 ↑
1992	3870 ↑	499 ↑
1993	4310 ↑	512 ↑
1994	4760 ↑	621 ↑
1995	4970 ↑	670 ↑
1996	5740 ↑	785 ↑
1997	6440 ↑	832 ↑
1998	7340 ↑	951 ↑

1999	8520 ↑	1140 ↑
2000	10700 ↑	1590 ↑
2001	12400 ↑	1510 ↓
2002	15400 ↑	1570 ↑
2003	18200 ↑	1860 ↑
2004	22500 ↑	1940 ↑
2005	27100 ↑	2150 ↑
2006	33400 ↑	2460 ↑
2007	37800 ↑	2670 ↑
2008	42600 ↑	2940 ↑
2009	46600 ↑	3150 ↑
2010	48500 ↑	3650 ↑
2011	47100 ↓	4050 ↑
2012	43600 ↓	4440 ↑
2013	41600 ↓	4630 ↑

Темата за **предсказване на успеха на стартиращи компании** не е нова, тъй като съществуват публикации от 70те години на миналия век (Wells, 1974). Въпреки това, тя бързо се развива и набира все по-голяма популярност. От 1990 до 2013 година броят публикации на годишна база се е увеличил над 10 пъти. Тенденцията за нарастване на интереса към темата за предсказване на успеха на стартиращи компании остава стабилна през последните години. В периода от 2008 до 2013 година броят публикации по темата е нараснал един път и половина, в сравнение с относително постоянния брой публикации на тема предприемачество.

Моделите за предсказване на успеха на стартиращи компании се развиват заедно с развитието на бизнеса и икономиката. Съществуват различни типове модели в зависимост от приложението: общи за оценка на стартиращи компании; за оценка на компании, участващи в състезание за бизнес план; за оценка на компании, опериращи на определен локален пазар; за оценка на технологични стартиращи компании и други. В зависимост от конкретната среда и условия, моделите притежават различни характеристики.

Настоящата дисертация има за цел да създаде цялостна методология и модел за предсказване на успеха на стартиращи компании в България, като се вземе предвид технологичния фактор. Проблемът за създаването на такъв тип модел е актуален, тъй като въпреки интереса към темата за предсказване на успеха, няма изготвени модели за България. Такъв модел е необходим на заинтересованите страни – собственици на стартиращи компании, бизнес анализатори, фондове за рисков капитал и инвеститори, учебни заведения - обучаващи студенти за стартиране на компании, организатори на състезания за бизнес план. На базата на този модел може да се създаде софтуер, който прогнозира успеха на стартиращи компании от България, и дава индикации какво трябва да се промени в компанията, за да се получат по-добри прогнозни резултати.

2.2 Дефиниции на основни понятия

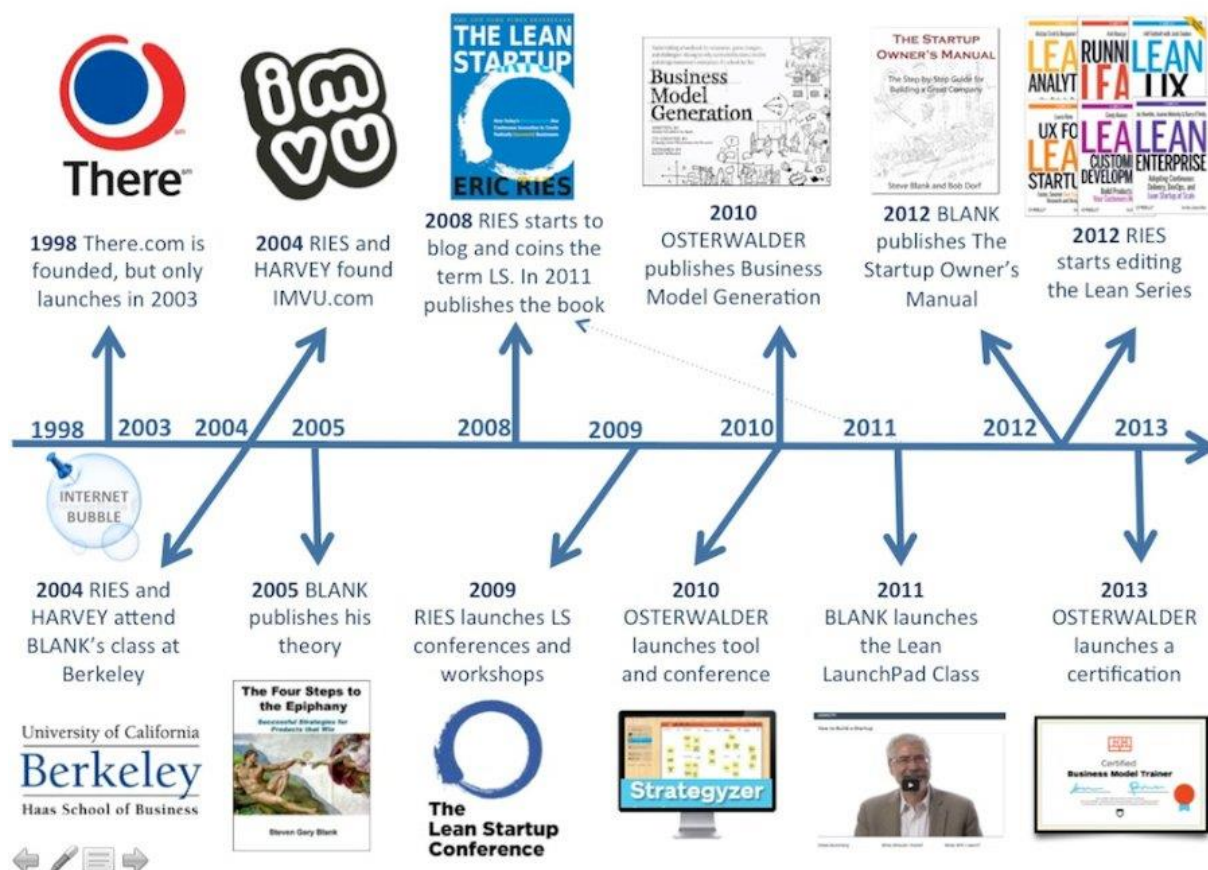
В настоящата точка са разгледани дефинициите на основни понятия, използвани в дисертационния труд, като технологично предприемачество, стартираща компания, успешна стартираща компания. Направен е обзор на определения, предложени от различни автори и са избрани значенията, които ще бъдат възприети в дисертацията.

2.2.1 Дефиниция за технологично предприемачество

НСПА дефинира предприемачеството (НСПА, 2007) като процеса на организиране и поемане на риск при стартиране на бизнес. Технологичното предприемачество е комбинация от стартирането на бизнес и прилагането на технологии и изследователска дейност.

Корените на технологичното предприемачество могат да се проследят до Древногръцката цивилизация (1450 г. пр.н.е.) (Nelson, 1999). Древните гърци са били едни от първите, които са строили мостове и сложни водопроводни и напоителни системи. Чрез технологичния напредък те са демонстрирали развитите си инженерни умения. Смята се, че са първите, чиято селскостопанска дейност е генерирала повече продукция, отколкото могат да оползотворят, а остатъкът се е използвал за износ.

Понятието технологичното предприемачество е споменато за пръв път на симпозиум през 1970 (Blank, The Four Steps to the Epiphany Steve Blank, 2005), (Bailetti, 2012) и се е превърнало в актуална тема с нарастващ брой публикувани статии през годините. Технологичното предприемачеството включва няколко области, и една от тях е създаването на нови фирми. Една от най-влиятелните тенденции в технологично предприемачество в момента е “Lean Startup” (Ries, The Lean startup - The Movement That is Transforming How New Products Are Built and Launched, 2014), (Ries, The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses, 2011) (Фиг. 2) – методология за развитие на стартиращи компании, която цели да акцентира върху процесите в компанията, чрез итеративно изграждане на продуктите и задоволяване на нуждите на различни потребителски сегменти.



Фиг. 2 Времева рамка на развитието на методологията Lean Startup

Съществуват множество съвременни дефиниции за технологично предприемачество. Според Тони Байлетти, технологичното предприемачество е инвестицията в проект, който обединява специалисти и разнородни активи, което довежда до научно и технологично познание за целите на създаване на стойност за дадена компания.

Според NCIP, технологично предприемачество са действията на предприемача и екипа за довеждането на откритието по пътя към пазара.

Инженерният департамент на университета McGill дефинира технологичното предприемачество като иновативното приложение на научни и технически познания от един или няколко души, които започват и развиват бизнес и поемат финансови рискове, за да постигнат своята визия или цел (McGill, 2014).

Според някои автори, технологичното предприемачество е процесът на създаване на нови технологични компании, който включва организиране, мениджмънт и справяне с риска на технологичната фирма, довеждайки изобретението на пазара (Nicholas, 2003), (Florida, 2007). Технологичното предприемачество изисква съвместни дейности за интерпретиране на неясни данни, за общо разбиране и поддържане на технологични дейности и постоянно и координирано усилие за постигане на технологична промяна.

На Фиг. 3 са показани дейности и теми, касаещи технологичното предприемачество, описани от Тони Байлетти.



Фиг. 3 Връзката между технологичната промяна и предприемачество (създадена от автора по описания от Тони Байлетти)

Базирайки се на горните дефиниции, за целите на настоящата дисертация приемаме, че **технологичното предприемачество** е свързано с **научно-развойна или изследователска дейност, или иновативно приложение** на съществуваща технология или научно техническо познание, в резултат на което се **създава продукт или услуга**, който добавя **стойност** за заинтересованите страни и спомага за **целите на бизнеса**.

2.2.2 Дефиниция за стартираща компания

За понятието стартираща компания (start-up) съществуват множество дефиниции, повечето от които не дават конкретни измерители относно в кой точно момент компанията престава да бъде стартираща и се превръща в зряла компания.

Според Ерик Райз, стартиращата компания е човешка институция, проектирана за да създаде нов продукт или услуга в условията на изключителна несигурност (Ries, The Lean Startup, 2011). Според някои автори, основаването на стартиращи компании е резултат от процеса на развитие, производство и предлагане на продукти или услуги, които задоволяват определена пазарна нужда с цел печалба и растеж (Gartner, 1988). Според Стив Бланк стартиращата компания е временна организация, създадена в търсене на бизнес модел, който може да бъде изпълнен и развит.

За целите на настоящата дисертация, на база изброените дефиниции, приемаме, че **стартираща компания** е такава, която е **в началото на своето развитие и по размер е SME** (микро, малка или средно голяма). За малки и средно големи компании приемаме тези с 0 до 250 наети лица (според дефиницията, отнасяща се за Европейския Съюз).

2.2.3 Дефиниция за успешна стартираща компания

Определението за успех на стартираща компания е необходимо, за да се различат успешните стартиращи компании от неуспешните. Успехът на стартиращите компании е термин с широка интерпретация, който в литературата е дефиниран по множество начини.

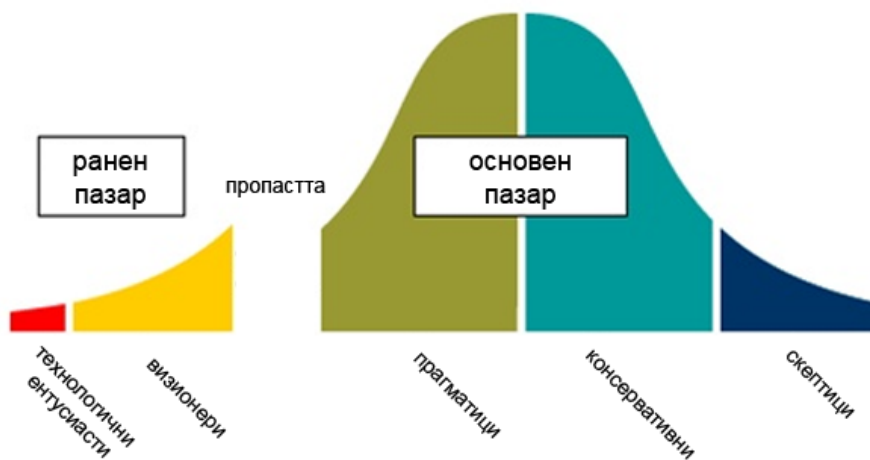
Ако потърсим в българския тълковен речник (БАН, 2012), значението на думата успех е дефинирано като „1. Постигане на цел; прогрес, сполука. Успех в бизнеса. Имам успех. Нямам успех в любовта. Успехи в кариерата. 2. Обществено признание, одобрение, прослава. Филмът има голям успех. Певецът гастролира с голям успех. Жъна успехи.

3. Степен на овладяване на знания, изразена в оценки. Слаб успех. Висок успех. 4. Сполука победа, слава.“. В първото значение на думата, успехът представлява постигане на определена цел, което съвпада и с дефиницията на Брокхаус (Brockhaus, 1982): изпълнението на определени цели (приходи, растеж на фирмата, лични постижения, възвращаемост на инвестицията и други). Това показва, че **успехът зависи от това, чии цели разглеждаме** – на предприемача, който стартира бизнес; на инвеститора, който е вложил пари; на банката, която е отпуснала кредит; на университета, който подготвя предприемачи; на правителството, което стимулира икономиката; на широката общественост или нечии други цели. При определянето на успеха, Кийли например взема страната на инвеститорите и определя успеха като вътрешната норма на възвръщаемост, реализирана от всички акционери (Keeley, 1989).

Когато избираме подходяща дефиниция за успеха на стартиращите компании, трябва да имаме предвид техните особености. **Стартиращите компании се различават от вече утвърдените на пазара** по своята по-младша възраст, по-малък, но често бързо нарастващ мащаб, както и по несигурността, с която се свързват. За вече утвърдена компания, успехът би могъл да се измери чисто финансово, например от печалбите или оборотите от годишните финансови отчети, чрез притежаваните активи и други, като тези измерители вероятно биха удовлетворили повечето заинтересовани страни. При една стартираща компания, в ранните етапи на нейното развитие обикновено се налагат множество инвестиции за развитие на продукта, а няма никаква печалба и възвръщаемост, тъй като все още няма продажби. Разработва се продуктът или услугата, екипът натрупва опит и контакти, проучва се пазарът, и стъпка по стъпка компанията се развива, но това не дава положително отражение върху отчета за приходи и разходи.

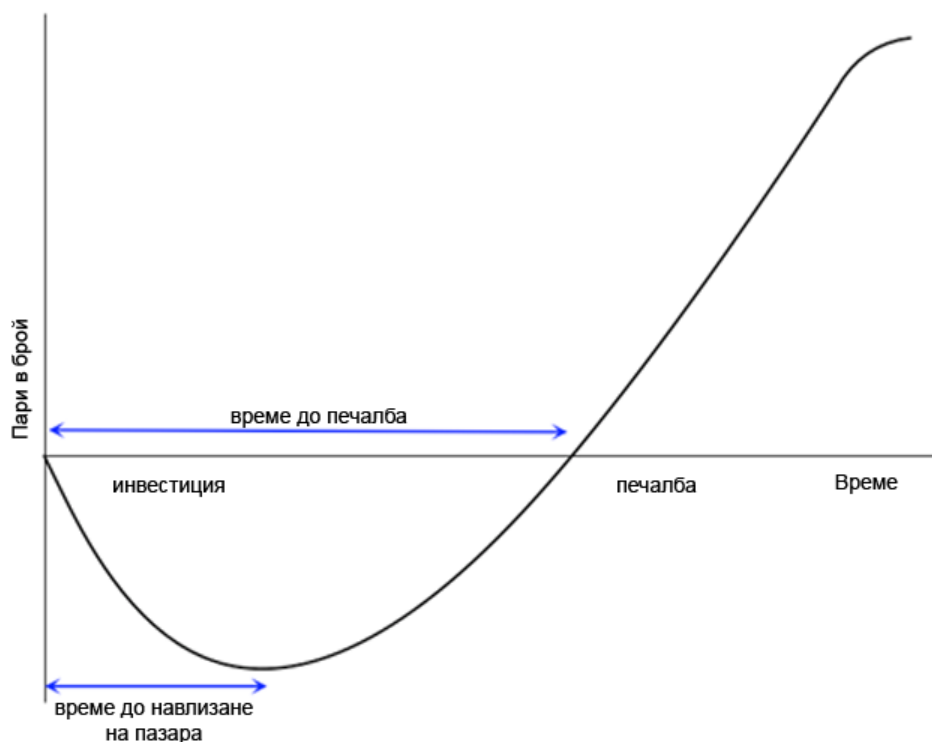
Развитието на една стартираща компания преминава през различни фази (Moore, 1991), от гледна точка на комерсиализацията на продукта (Фиг. 4). Постепенно се достигат различни пазарни сегменти, които отговарят на нови типове потребители, чиито изисквания към продукта нарастват. За да бъде достигнат основния пазар (“mainstream market”), е необходимо продуктът да бъде завършен, приет от пазара и неговите недостатъци от гледна точка на потребителите да бъдат отстранени.

Крива на възприемане на нов продукт на пазара (прескачане на пропастта)



Фиг. 4 Навлизане на нов продукт на пазара - прескачане на „пропастта“ между ранния и основния пазар (адаптирана по Тони Байлетти)

Навливането на нов продукт на пазара е дълъг процес, свързан с несигурност и много инвестиции. Това се отразява на паричния поток (Grabewarter, 2005) на стартиращите компании (Фиг. 5), и в най-голяма степен на технологичните компании. Поради изброяните особености на стартиращите компании, измерителите за успех, които се отнасят за зрели компании, не са пряко приложими за стартиращи.



Фиг. 5 J-крива на паричния поток на технологична стартираща компания (адаптирана по Грабенвартер)

За да се определят критериите за успех на стартиращи компании, бяха проведени **интервюта със собственици на бизнес и предприемачи**. На респондентите бяха поставени въпросите „Как определяте коя стартираща компания е успешна? Как различавате успешната стартираща компания от неуспешната?“.

Един от интервюираните от автора е Боб Дорф – един от най-популярните експерти по стартиращи компании и корпоративно предприемачество в света, преподавател в Колумбийски Бизнес Университет, ментор, бизнес консултант и предприемач. В рамките на проведения разговор, Боб Дорф отговори следното: „Много трудно е да се постави точна граница коя стартираща компания е успешна и коя е неуспешна и няма всеобщо признат критерий. За мен успешна компания е тази, която ако миналото тримесечие е имала 10 запитвания от клиенти и 2 продажби, то това тримесечие има 10 запитвания и 3 продажби. Без значение са етапът на развитие на компанията и паричният поток в момента, тъй като все още компанията може да е в много ранен етап. Важно е да има развитие“.

Бяха проведени интервюта и с петима основатели и собственици на български стартиращи компании. Техните отговори могат да бъдат обобщени с тезата, че за успеха на стартиращите компании е важно те да имат развитие и растеж. Докато при зрелите компании, за успех може да се счита чисто финансовият резултат, без увеличение на размера на компанията, при стартиращите компании развитието и растежът са ключови за тяхното по-нататъшно оцеляване и съществуване.

Над 50% от стартиращите компании в Европа (както и в България) имат между 0 и 4 служители (Eustat, European Commission, 2009). Повечето стартиращи компании не успяват да се развият до етап, който да осигури оцеляването им през периода на първите 5 години от дейността и от това логично следва, че са неуспешни.

На база на проведеното проучване и интервюта, **за успешна стартираща компания се счита тази, която е увеличила своя размер през последните пет години.** За неуспешна се счита компания, която е прекратила своята дейност. А за нито успешна, нито неуспешна се приема компания, която продължава дейността си, но не е увеличила своя размер.

2.3 Избор на методология за изследването

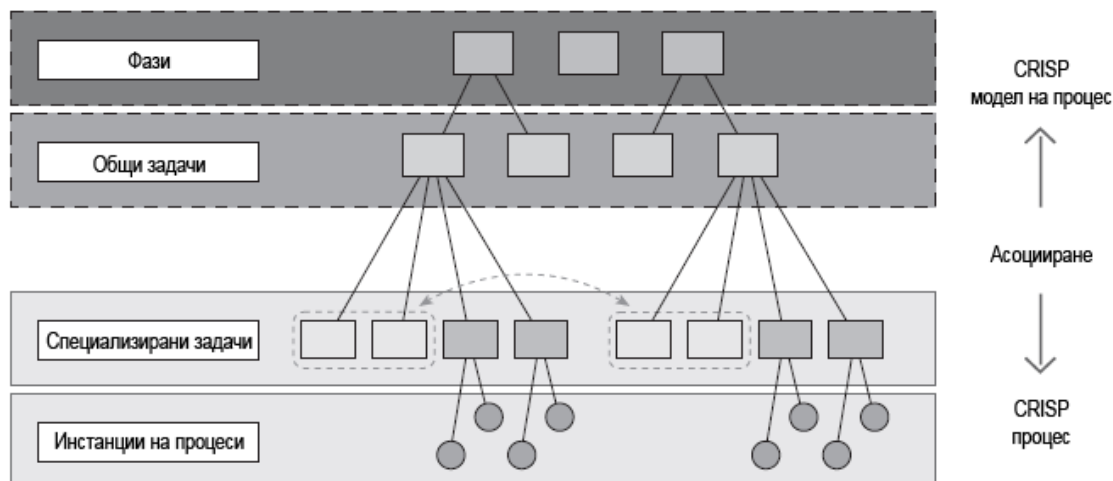
За целите на дисертационния труд са използвани различни методологии на научно-изследователската работа. Техният избор се основава на характера на решаваните задачи.

За решаване на задачите, свързани с проучване, изследване и анализ на проблемната област - съвременното състояние на научните трудове и постижения в областта на предсказване успеха на стартиращи компании, факторите за успех на стартиращи компании в света и България, е използван методът на **систематичното литературно проучване** (Tranfield, 2003). Това е вид литературен обзор, при който съществуващото знание по отношение на конкретен въпрос или проблемна област (под формата на т.нар. първични изследвания) бива идентифицирано, оценено и синтезирано по начин, който осигурява максимална обективност, прозрачност и възпроизводимост.

Предложеният в Глава 3 детайлизиран модел за предсказване на успеха на стартиращи компании е валидиран чрез прилагане на методите качествено и количествено изследване.

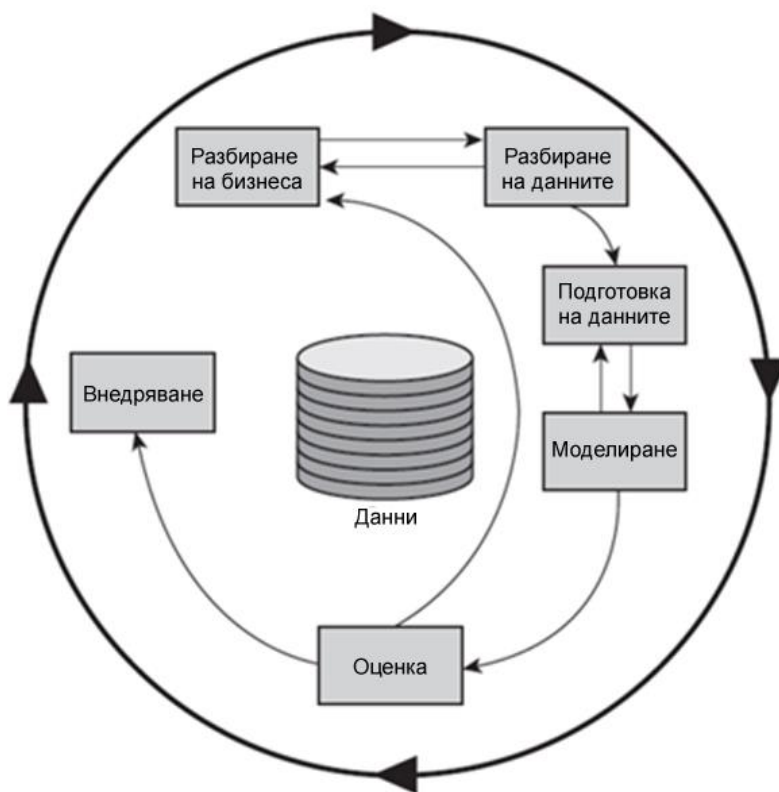
Качественото изследване представлява провеждане на задълбочени интервюта с малък брой респонденти с цел събиране на първоначална обратна връзка за модела и извършване на подобрения.

На база детайлизирания модел за предсказване на успеха, е проведено **количествено изследване** с цел определяне на променливите, които имат най-голямо влияние за успеха на стартиращите компании, както и зависимостите между отделните променливи. Количественото изследване се базира на методологията CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman P., 2000), която описва стандарт на процеса на извличане на знания от данни (data mining) и е създадена от консорциум от компании като проект на Европейския Съюз през 1997 година. CRISP-DM е описана като йерархичен модел на процес, съдържащ групи от задачи, описани на четири нива на абстракция (от общо към частно): фази, общи задачи, специализирани задачи и инстанции на процеси, както е показано на Фиг. 6.



Фиг. 6 Декомпозиция на четири нива на методологията CRISP-DM (преведена от автора по фигура на Чапман)

Жизненият цикъл на проект за извличане на знания от данни съдържа фазите на проекта, техните задачи и връзки. Жизненият цикъл се състои от шест фази, както е показано на Фиг. 7. Последователността на фазите е гъвкава, позволявайки придвижването напред и назад между различните фази. Резултатът от всяка фаза определя коя фаза или задача ще бъде изпълнена след това. Стрелките на фигурата показват най-важните и често срещани зависимости между фазите.



Фиг. 7 Жизнен цикъл на проект за извличане на знания от данни по методологията CRISP-DM (преведена от автора по фигура на Чапман)

Външният кръг на фигурата показва цикличната природа на процеса за извличане на данни, който не приключва с внедряването на решение. Получените изводи по време на изпълнение на методологията, и от внедреното решение, могат да отключат нови, по-фокусирани бизнес въпроси. Последващите процеси на извличане на данни ще бъдат подпомогнати от опита, получен от предходните.

Моделирането на данните се осъществява посредством софтуерните продукти IBM SPSS Modeler и Weka. Създадени са модели за предсказване на успеха на стартиращи компании от България, чрез използване на пълния събран набор от данни. Освен това, е създаден и модел за предсказване на успеха на технологични стартиращи компании, чрез използване само на данните за технологични компании, които са подмножество на всички данни.

2.4 Опити за решение, достижения и дефицит

В настоящата точка са разгледани и анализирани съществуващи модели за прогнозиране на успеха на стартиращи компании. Идентифицирани са тенденциите в създаване на модели за предсказване на успеха, както и утвърдените предпочитани модели.

Проучени са и са анализирани софтуерни продукти за предсказване на успеха на стартиращи компании, като е обърнато внимание на техните силни и слаби страни, както и пригодността им да решат задачата за предсказване на успеха на технологични стартиращи компании от България.

2.4.1 Сравнение на съществуващи теоретични модели за прогнозиране успеха на стартиращи компании

Съществуват множество научни разработки, които са концентрирани върху изясняването на характеристиките, описващи успеха на стартиращи компании. Данни за такива разработки са налични от далечната 1974 година, когато Уелз (Wells, 1974) предлага модел на успеха на стартираща компания. Моделът е базиран на променливите ангажираност на ръководството, продукт, маркетинг, умения (по маркетинг, инженерни, финансови, производствени), маркетингов план, референции, други участници в сделките, метод за излизане от пазара.

Детайлна класификация на моделите за предсказване на успеха е направена от Ензли и Тийл (Teal E. J., 1998), (Ensley, 1999), (Teal E. a., 2003) и тя включва 21 модела от 1976 до 1999 година. Подобни класификации са правени и от Карлънд и Крисман (Carland, A new venture creation model, 2000), (Chrisman, 1998). Класификациите са обединени от автора в една, която е допълнена с множество липсващи и по-актуални модели, които са описани, сравнени и анализирани (Abbas, 2008), (Astebro, 2000), (Biggadike, 1976), (Davis, 2005), (Gartner, 1988), (Shepherd D. , 1999), (Shirokova, 2010). Резултатът съдържа 44 модела за прогнозиране на успеха, които са представени и съпоставени в Таблица 4.

Таблица 4 Сравнение на модели за предсказване на успеха на стартиращи компании

Изследване	Основни фактори
Уелз, 1974	Ангажираност на ръководството, продукт, маркетинг, умения (по маркетинг, инженерни, финансови, производствени), маркетингов план, референции, други участници в сделките, метод за излизане от пазара
Бигъдайк, 1976	Предприемач и стратегия
Пойндекстър, 1976	Качество на управлението, очаквана норма на възвръщаемост, очакван риск, Процент на инвестициите от общия капитал, финансови разпоредби за правата на инвеститорите, темп на развитие на фирмата, процент капитализация, контрол на инвеститорите, прилагане на методи за намаляване на данъците
Фон Хипел, 1977	Предприемач и капитализация
Веспър, 1980	Познания, идея, връзки, ресурси, реализация
Хобсън и Морисън, 1983	Предприемач и стратегия
Тайбджи и Бруно, 1984	Препоръки на предприемача, марж на печалбата, уникалност на продукта, патентоспособност на продукта, наличност на суровини, достъп до пазара, пазарна нужда от продукта, размер на пазара, потенциал за растеж на пазара, свобода от регулации, защита от навлизане на конкуренти, устойчивост спрямо икономическите цикли, защита на остаряване на продукта, защита срещу неблагоприятно развитие, възможности за изход, потенциал за сливания и придобивания, хедж към настоящата инвестиция, данъчно предимство
Милър и Кемп, 1985	Стратегия

Гартнър, 1985	Личност, среда, организация, реализация
Сендбърг, 1986	Предприемач, стратегия и структура на индустрията
Муур, 1986	Личност, среда, иновации, реализация
МакМилиън и Дей, 1987	Стратегия
Романели, 1987	Стратегия и състояние на индустрията
МакДагъл, 1987	Стратегия, произход, структура на индустрията и взаимодействия
Стюарт и Абетти, 1987	Обкръжаваща среда, стратегия, структура и предприемач
Хол, 1989	Субективна оценка
Байгрейв, 1989	Личност, Среда, иновация, отключващо събитие, реализация
Гартнър, 1989	Психологическа предразположеност, събитие - предизвикало стартирането на бизнес, разпознаване на възможността, обкръжаваща и финансова подкрепа, реализация
Манинг, Бърли и Норбърн, 1989	Характеристики на компанията, връзки, детерминанти, реализация
Руур и Кийли, 1990	Мениджмънт, стратегия на компанията, конкурентна среда
Канкел, 1991	Стратегия, структура на индустрията и взаимодействия
Дженсен, 1991	Предприемач и индустрия
МакКарти, 1992	Стратегия, обкръжаваща среда, ресурси и промяна
Мак Дагъл, Робинсън и ДеНиси, 1992	Произход, стратегия, структура на индустрията и взаимодействия между стратегията и структурата
Брюдерл, Прайзендорфер и Циглер, 1992	Човешки капитал, отнасящ се до основателя, характеристики на организацията, условия на обкръжаващата среда
Боланд, 1993	Произход, стратегия и структура на индустрията
Бландън, 1993	Капитализация и тип на финансирането
Ниауали и Фогъл, 1994	Предприемачески и бизнес умения, социо-икономически фактори, правителствени политики, финансови ресурси, реализация
Стърнс, Картър, Рейналд и Уилям, 1995	Локация, стратегия, структура на индустрията и взаимодействия
МакГий, Даулинг и Мегисън, 1995	Експертиза на екипа, стратегия и договорености за сътрудничество
Робинсън, 1995	Стратегия, структура на индустрията и взаимодействия
Ласиър и Корман, 1995	Етап на навлизане на пазара, планиране, използване на външни съветници, образование на предприемача
Стюарт, 1996	Мотивация за постижения, склонност към поемане на риск, предпочитания към иновациите, психологическа предразположеност
Тийл, 1998	Предприемачески екип, стратегия и структура на индустрията
Ензли, 1999	Предприемачески екип, стратегия, структура на индустрията

	и взаимодействия между стратегията и структурата
Шепърд, 1999	Стабилност, момент на навлизане на пазара, времево предимство пред конкуренцията, конкурентност на средата, образователен капацитет, компетенция - свързана с индустриалния сектор
Парк и Лий, 2000	Стратегия, ресурси, обкръжаваща среда, взаимодействие между факторите
Карлънд и Карлънд, 2000	Идея, предприемаческа нагласа, знания, креативност, опит, интуиция
Тийл и Хофър, 2003	Стратегия, структура на индустрията, взаимодействие между стратегията и структурата
Астеброу, 2004	Очаквана печалба, технологична възможност, риск на разработката, подходящи условия
Дейвис и Цвайг, 2005	Пазар, продукт или услуга, финансови ресурси, управленски екип
Абас, 2008	Технологична иновация, концепция за бизнес, технологичен предприемач
Ширкова и Шаталов, 2010	Фактори на средата, стил на управление, стратегическо поведение
Манев, Манолова, Гьошев, Харкинс, 2012	Нетуъркинг (социални мрежи на основателя) и диференцираща стратегия

Разгледаните модели (Yankov, Overview of Success Prediction Models for New Ventures, 2012) се различават според своето предназначение и според условията, при които са създадени и могат да се групират в следните типове:

- общи - за оценка на стартиращи компании;
- за оценка на компании, участващи в състезание за бизнес план;
- за оценка на компании, опериращи на определен локален пазар;
- за оценка на технологични стартиращи компании;
- други.

Характерно за моделите, описващи успешното създаване на нови компании от далечните години е, че те основно разглеждат отделни фактори за успех. Някои от факторите, разглеждани в ранните модели са знания, бизнес идея, реализация на идеята, бизнес контакти, темперамент на предприемача, външна среда. При някои от тези модели компанията се разглежда като независима единица и не се вземат предвид външните фактори, като например бизнес средата, в която ще се развива дадената фирма. Моделите за прогнозиране на успеха, създадени на по-късен етап, постепенно започват да придобиват структурираност, чрез категоризация на факторите за успех в групи. Някои от тях (например моделите на Парк и Тийл) (Park, 2000) определят влиянието на външната среда като основен фактор в процеса на създаване на нови компании. Така, към основните фактори, определящи успеха на даден бизнес, се добавят фактори, характеризиращи средата в която компанията оперира: пазар, конкуренция, структура на индустрията и други.

Брюдерл, Презендорфер и Зиглер (Brüderl, 1992) определят шансовете за оцеляване на новосъздадените компании с помощта на модел, в който са използвани три основни категории фактори. Тези категории са: длъжностните и личностни характеристики на основателя на компанията, организационни характеристики и условия на средата, в която компанията развива своята дейност. Пълният списък на факторите от модела включва:

- *Личностна и длъжностна характеристика на основателя на компанията*
 - *Основни – години на обучение, научни степени, години трудов стаж*
 - *Специфични – натрупан опит в индустрията в която компанията ще развива дейността си, опит придобит от самостоятелни дейности, управленски опит*
- *Организационни характеристики*
 - *Новаторство*
 - *Големина на компанията*
 - *Управленска стратегия на компанията*
- *Характеристики на средата*
 - *Местоположение на компанията*
 - *Индустриален клъстер*
 - *Пазарни условия*

Лусиер и Корман през 1995 представят свой анализ за успеха на малки стартиращи компании (Lussier, 1995). Те определят четири основни фактора за успех. Тези фактори не формират завършен модел за прогноза, но могат да се използват като полезни насоки при стартиране на нов бизнес:

- *Несъвместимост на продукт/услуга* - компании, която разработват продукт или услуга, която е прекалено нова или прекалено предлагана на пазара, имат по-големи шансове за неуспех от компании, които разработват продукт или услуга, която а вече представена на пазара и има нарастващо търсене.
- *Планиране* – компании, които разработват бизнес планове имат по-голям шанс за успех от останалите.
- *Наличие на консултанти* – компании, който не използват помощта на професионални бизнес консултанти имат по-голям шанс за успех.
- *Образователна степен* – предприемачи, започващи своя бизнес без университетско образование имат по-малки шансове за успех от тези, които са завършили университет.

Шепърт представя друг интересен модел през 1999 година (Shepherd D. , 1999), който оценява шансовете за успех на стартираща компания според следните фактори:

стабилност, момент на навлизане на пазара, конкурентна среда, образованост и компетенции за индустрията в която ще се развива компанията.

През 2008 година Абас предлага своя модел за прогнозиране на успеха на студентско състезание за най-добър бизнес план (Abbas, 2008). В неговото предложение ресурсите и средата, в която ще се развива дадена компания, не се определят като основни фактори, тъй като студентските компании започват в една и съща среда и се борят за финансиране. Основните фактори на този модел са технологичните иновации, бизнес идеята и предприемача.

Друг модел, който заслужава внимание, е представен от Давис и Зуейг през 2005 година (Davis, 2005). Според тях, факторите, които най-силно повлияват успеха на стартиращите фирми, са: подходящ продукт/услуга, финансовите ресурси и управленския екип. Този модел е един от малкото, които включват финансовия ресурс като основен фактор за успеха.

В България няма много разработени модели за прогнозиране на успеха на стартиращи компании и почти никакъв софтуер, който да прилага тези модели и да ги прави достъпни до заинтересованите страни. Едно от изследванията в тази насока (Manev, 2012), на Манев, Манолова, Гьошев и Харкинс касае определени фактори за успех на стартиращите компании в България. То се базира на хипотезата, че нетуъркинг възможностите на стартиращите компании (зависещи пряко от контактите и социалните мрежи на основателя) и диференциращата стратегия са свързани с успеха на стартиращите компании. В изследването се допуска, че успехът на компаниите в условията на икономика на прехода зависи от различни фактори, в сравнение с компаниите, развиващи своята дейност в държави с развита икономика. За да докажат хипотезите, прилагат количествено изследване на 334 български компании. Изследването не включва цялостен набор от фактори, които биха повлияли успеха на стартиращите компании.

2.4.2 Съществуващи теоретични модели за прогнозиране успеха на стартиращи компании, базирани на модела на Сендбърг

Популярен базов модел, използван като основа от множество съвременни модели за предсказване на успеха на стартиращи компании е този на Сендбърг (Sandberg, 1986). В своето изследване, Сендбърг за първи път добавя и характеристики, които описват предприемача, бизнес стратегията на дадената компания и структурата на индустрията, в която тя се развива. Моделът може да бъде описан с формулата:

$$NVP = f(E, IS, BS) \quad (1)$$

Където NVP е продуктивността на новата фирма, E е предприемача, IS е структурата на индустрията, а BS е бизнес стратегията.

Следващи изследвания, базирани на Сендбърг, включват и други фактори, като взаимодействието между стратегията на фирмата и структурата на индустрията, наличните ресурси, а човешкият фактор (предприемаческият екип) често се пропуска. Таблица 5 сравнява модели за прогноза, базирани на основните фактори, предложени от Сендбърг: бизнес стратегия и индустриална среда. В нея са показани приликите и разликите между различните модели, като със знак плюс „+“ са показани включените

към съответния модел фактори, а допълнителните фактори са показани в последната колона.

Таблица 5 Сравнение на модели за прогноза на успеха, базирани на модела на Сендбърг

Изследване	Основни фактори			
	Стратегия	Индустриал на структура	Взаимодействие	Други
Сандбърг, 1986	+	+	-	Предприемач
Романели, 1987	+	+	-	
Мак Дагъл, 1987	+	+	+	Произход
Стюърт и Абетти, 1987	+	+	-	Предприемач, Среда
Канкел, 1991	+	+	+	
Мак Дагъл, Робинсън и ДеНиси, 1992	+	+	+	Произход
Боланд, 1993	+	+	-	Произход
Стърнс, Картър, Рейналд и Уилиям, 1995	+	+	+	Местоположение
Робинсън, 1995	+	+	+	
Крисман, 1998	+	+	-	Предприемач, ресурси, организационна структура
Тийл, 1998	+	+	-	Предприемачески екип
Ензли, 1999	+	+	+	Предприемачески екип
Тиъл и Хофър, 2003	+	+	+	
Ширкова и Шаталов, 2010	+	+	+	Глобални фактори на средата, управленски стил

Мак Дагъл (McDougall, 1987) разработва подобен на Сендбърг модел, в който включва факторите бизнес стратегия и индустриална среда, но новото при него е, че добавя и характеристики, които описват връзките на избраната бизнес стратегия с индустриалната среда, в която дадена компания ще развива своята дейност. По-късни разработки също потвърждават, че връзката на бизнес стратегията на дадена компания с индустриалната среда е ключова и има важна роля в определянето на успеха на стартиращите компании.

Значимостта на фактора „предприемачески екип” е спорна и той често се пропуска в моделите. При анализ на различни модели, базирани на линейния модел на продуктивността на новата фирма, предложен от Сендбърг, Джоан и Джеймс Карлънд наблюдават, че резултатите се различават. Те забелязват, че Сендбърг не успява емпирично да свърже характеристиките на предприемача с продуктивността на фирмата. Повечето изследвания на нови фирми са насочени към тяхното представяне

като функция на предприемача, и следователно, въпреки своите открития, Сендбърг не желае да заличи характеристиките на предприемача, като мениджмънт компетенции и опит в индустрията, от модела, тъй като инвеститорите и фондовете за рисков капитал ценят тези характеристики.

Крисман предлага разширен модел на продуктивността на новата компания (Chrisman, 1998), базиран на модела на Сендбърг. Освен предприемача (E), структурата на индустрията (IS) и стратегията на бизнеса (BS), моделът включва още ресурсите (R) и организационната структура (OS). Моделът е представен с формулата:

$$NVP = f(E, IS, BS, R, OS) \quad (2)$$

С доразвиването на моделите за предсказване на успеха и валидирането им с данни, при някои изследвания резултатите показват, че човешкият фактор (обикновено характеризиращ предприемаческия екип) няма връзка с успеха на компанията. Тиъл и Хофер, например, предлагат модел, който пропуска факторите отнасящи се до предприемаческия екип, а включва само бизнес стратегията на компания, индустриалната среда и връзката между двата фактора.

Други важни фактори, открояващи се като основни според по-новите модели, са предприемаческият екип, местоположението на компанията, средата и стила на управление. Ресурсите (като например активите) често се пропускат, тъй като не са важни, в случай че е осигурено финансиране (например в състезания за стартиращи компании). При модела на Ширкова и Шаталов от 2010 година (специално разработен за спецификите на Руския бизнес климат), факторът индустриална среда е разширен с фактори, описващи по-общо цялостната среда, а не само конкретната индустрия.

От разгледаните до момента модели за прогноза, може да се заключи, че **моделите, които са базирани на основни фактори като стратегия и индустриална структура са удачен вариант** за създаване на практическо приложение за прогноза на успеха. Тези модели могат да бъдат допълнени с още фактори, характеризиращи успешните компании, и да бъдат тествани със статистическа информация. След валидация, те могат да послужат като основа на практическо приложение за прогноза на успех на стартиращи компании.

2.4.3 Проучване и анализ на софтуерни продукти за прогнозиране успеха на стартиращи компании

Направеното проучване на софтуерни продукти за прогнозиране успеха на стартиращи компании показва, че инструменти с подобно предназначение се съдържат като част от портали за групово финансиране на бизнес идеи (crowdfunding).

Един от популярните портали за групово финансиране е [“EquityNet”](#) (EquityNet - The Leading Equity Crowdfunding Platform, 2014). Според статистиката, обявена на уеб сайта, той е използван от над 20 000 предприемачи и 20 000 инвеститори, а финансирането, предоставено от инвеститорите към предприемачите в портала възлиза на повече от 240 милиона долара. Други по-известни портали за групово финансиране са [„Sellaband”](#) (Sell a Band - Where Fans Invest in Music, 2014), [„IndieGoGo”](#) (Indiegogo - Crowdfunding for What Matters to You, 2014), [„GiveForward”](#) (GiveForward - The Number 1 Way to Raise Money for a Loved One, 2014), [„Kickstarter”](#) (KickStarter - The World's Largest Funding Platform for Creative Projects, 2014) - с отпуснато финансиране за повече от 1 милиард долара, [„RocketHub”](#) (RocketHub - Crowdfunding, 2014), [„Fundly”](#) (Fundly -

Online Fundraising Websites to Raise Money for Anything, 2014), „[GoFundMe](#)” (GoFundMe - Number One for Crowdfunding and Fundraising Websites, 2014) с над 350 милиона долара инвестиции, „[Appsplit](#)” (AppSplit - Crowdsourcing for Apps, 2014).

Порталите за групово финансиране доста се различават от идеята на софтуера за прогнозиране на успеха, разработен в настоящата дисертация. Съдържащите се в тях инструменти, служещи за анализ на риска, се градят върху представен бизнес план от страна на предприемача. Предприемачите трябва да представят публично (на потребителите на порталите) своите бизнес идеи. Недостатъкът на този подход е, че предприемачите рискуват бизнес идеите им да бъдат заимствани, или дори откраднати, преди изобщо да получат някакво финансиране за тяхната реализация.

Подбрани са и са разгледани в детайли и три приложения за прогнозиране на успеха, които разчитат на входни данни от анкета, използват статистическа информация и правят оценка на шансовете за успех и неуспех на дадена компания, попълнила анкетата. Тези приложения, като принцип на работа са близки до софтуерния продукт, разработен към настоящата дисертация.

2.4.3.1 Калкулатор на риска за стартиращи компании от StartupNation

Най-добре развитият инструмент за прогнозиране успеха на стартиращи компании е създаден от компанията “StartupNation”. Инструментът носи името „Калкулатор на риска за стартиращи компании” ([Startup Risk Calculator](#)) (Startup Risk Calculator, 2014). Този инструмент има връзка с “EquityNet”, защото използва техни статистически данни, за да създаде модела си за прогноза. Едва година след създаването си, той започва да се предлага като инструмент от портфолиото на EquityNet. Инструментът представлява уеб приложение със свободен достъп в интернет и неговата цел е да определи шансовете за успех на стартиращи компании на база на осем фактора:

- размер на инвестициите, направени до момента в дадената компания,
- достъп до нови средства на финансиране,
- управление на финансите,
- задълбоченост на бизнес планирането от страна на мениджмънта на компанията,
- ръст на компанията на годишна база,
- управленски опит,
- опит на компанията в индустриалния сектор, в който тя развива своята дейност,
- времеви интервал на оценката.

На потребителя на приложението се предоставя възможност да отговори на осем въпроса, свързани с изброените фактори. След оценка на отговорите на анкетата, приложението дава прогноза в проценти за вероятността за успех на компанията, попълнила анкетата. Като част от резултата са представени и факторите, които повлияват резултата (Фиг. 8). Поради използването на прекалено малко фактори за оценка на успеха, както и на малкото информация, която потребителят трябва да въведе, за да опише своя бизнес, генерираната прогноза от този инструмент не изглежда много надеждна. Като метод за комерсиализация на инструмента, страницата, показваща резултатите, е разработена като маркетингов инструмент, представящ оферти за бизнес съвети, анализи и консултации.

YOUR ODDS OF SURVIVABILITY

40%

WHAT FACTORS RAISE YOUR ODDS

- ↑ Regularly managing cash flow is a key practice of successful businesses. Your answer to this question (**every 2 months**) suggests that you keep updated and accurate records and are disciplined about managing expenses and cash resources.
- ↑ Planning doesn't make perfect, but it sure does help. Businesses like yours that exhibit a relatively high degree of planning (**5 on a scale of 1-7**) are more likely to have well-thought-out, specific strategies for operational success and are more likely to execute on those strategies.
- ↑ Management has a huge impact on likelihood of survival. The above average management experience (**5 years**) you have gives you a higher chance of survival than businesses managed by people without as much management experience.
- ↑ Your business is managed by people with above average industry experience (**5 years**), and accordingly, you have a higher chance of survival than those businesses managed by people with lesser industry experience.

WHAT FACTORS LOWER YOUR ODDS

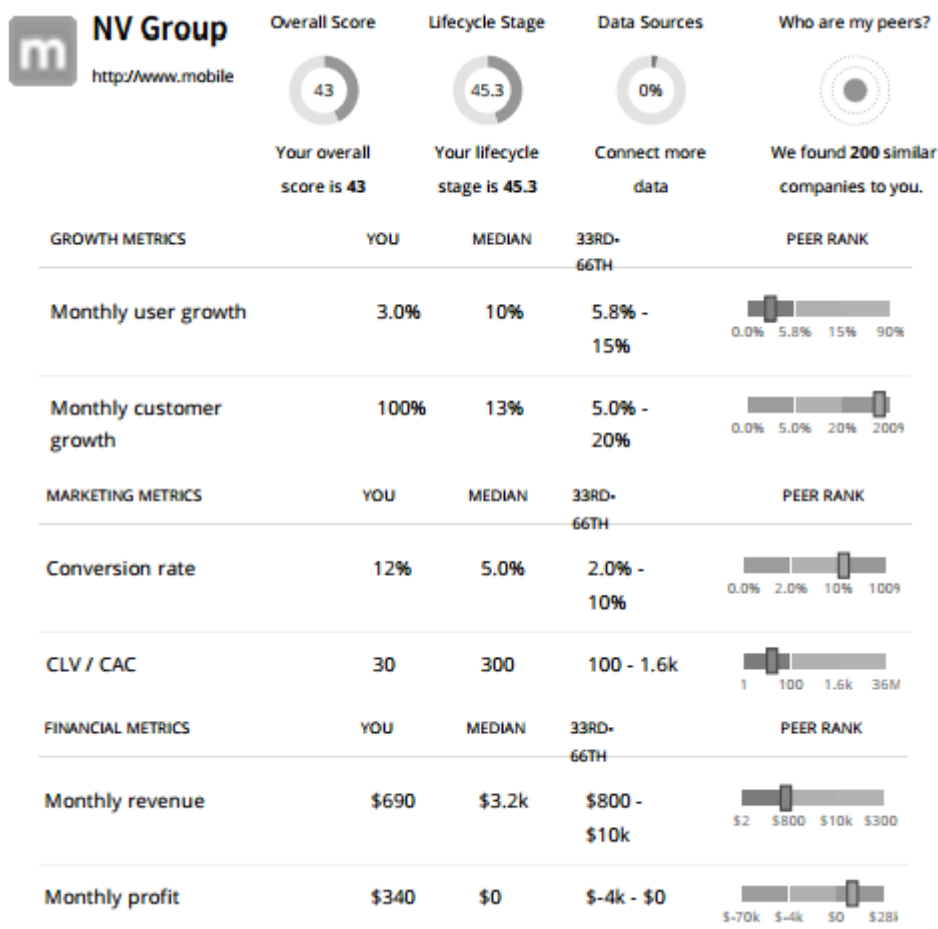
- ↓ You indicated that a below-average amount of debt and/or equity capital (**\$10,000**) has been received by your business. The amount of capital infused impacts your likelihood of survival. The less you get, the less likely you will be to endure difficulties and to be able to utilize that money for revenue-generating initiatives. That means you have fewer resources to reach profitability or a liquidity event.
- ↓ Healthy, growing markets provide an environment in which your chances of survival are improved. The opposite is true, too. The unappealing growth rate (**5% - low growth**) in your market negatively impacted your odds of survival.
- ↓ In this calculator, survival is defined as your business being financially viable. You asked the calculator for your odds of survival over a relatively long timeframe (**5 years**). Because businesses encounter "hazard" risks that occur every year and that accumulate over time, your odds were negatively impacted. Enter a shorter timeframe into the calculator to see how it positively impacts your odds of survival.

Фиг. 8 Резултати от инструмента за прогноза „Калкулатор на риска за стартиращи компании“ (създадена от автора чрез прилагане на инструмента)

Инструментът на StartupNation може да бъде от полза на предприемачи без никакъв опит, тъй като дава основни насоки за част от важните фактори, които могат да доведат до успех даден бизнес.

2.4.3.2 Компас за нови компании

Друго приложение за прогнозиране на успеха на стартиращи компании е „Компас за нови компании“ ([Startup Compass](#)) (Compass - Benchmark Your Business Metrics From 30+ Data Sources, 2014). Приложението представлява инструмент за оценка на стартиращи технологични компании, който дава възможност за попълване на анкета. Потребителят може да актуализира ежемесечно попълнените данни. След попълване на анкетата, се извеждат резултати, оценяващи компанията според нейния тип и етап на развитие - под формата на различни графики, показващи текущите резултати на компанията на базата на останалите компании, използващи инструмента (Фиг. 9). Идеята на създателите на приложението е то да послужи като персонална оценка на даден бизнес и с негова помощ да бъдат по-лесно забелязани потенциални проблеми, да бъдат изяснени приоритетите, да бъде подпомогнат процеса по оформяне на предприемаческия екип и по-лесно да се измери напредъка на дадена компания.



Фиг. 9 Фрагмент от резултати за прогноза на успеха на инструмента „Startup Compass” (създадена от автора чрез прилагане на инструмента)

Инструментът използва статистически данни от над 600 стартиращи технологични компании предимно от Силиконовата долина, за да определи ключовите фактори за успеха и да изгради своя модел на работа. „Startup Compass” използва следните фактори, за да направи своя анализ:

- клиенти, продукт / услуга,
- предприемачески екип,
- бизнес модел и начини на финансиране.

„Startup Compass” съветва как да бъдат оценени тези фактори в подходящото време. Например, компании, които не отделят време да идентифицират какъв ще бъде пазарът на техния продукт или услуга, не анализират пазара в етапа на планиране, а преминават директно към създаването на продукта, се смятат за непоследователни и шансовете им за успех са малки. Според инструмента, 80% от този тип компании са неуспешни, докато 45% от компаниите, които първоначално са идентифицирали своите потенциални клиенти и едва след това са разработили своя продукт или услуга, са успешни. В началото предприемачите могат лесно да се заблудят и да концентрират всичките си усилия в създаването на своя продукт, без дори да са оценили истинската нужда от ползването му. Създателите на приложението отбелязват също, че колкото повече компании използват техния инструмент, толкова по-задълбочени и верни

анализи ще прави той, и ще дава по-тясно свързани с конкретна компания препоръки за развитие.

Инструментът „Startup Compass” е създаден за да анализира само технологични компании. Използването му е безплатно, но се изисква регистрация.

2.4.3.3 Инструмент за проверка на стартиращ бизнес от Blueprint

Друго уеб приложение за прогнозиране успеха на стартиращи компании се нарича Инструмент за проверка на стартиращ бизнес ([Blueprint New Business Startup Tester](#)) (Blueprint New Business Startup Tester, 2014). То е създадено от австралийската компания „Think Blueprint” и представлява цялостно решение за бизнес планиране, включващо инструменти за създаване на бизнес планове, оценка на риска и открояване на основните дейности при създаване на нова компания. Използването на самия инструмент е безплатно, но за да се приложи целият набор от инструменти за планиране, се заплаща такса. Инструментът, представен от „Think Blueprint”, използва кратка анкета (около 10 въпроса) за получаване на данни за стартиращите компании. Резултатите за шансовете за успех се изпращат на имейл. Съдържанието на имейла е показано на Фиг. 10.

Title of Your New Business: NV Group

Your Preparation Index: 40% - 60%

Blueprint Preparation Index: 60%

Your Success Probability Score: 40% - 60%

Blueprint Success Probability Score: 63%

Фиг. 10 Резултат на прогноза за успех инструмента на „Think Blueprint” (създадена от автора чрез прилагане на инструмента)

Работа на приложението не може да бъде проследена и не се дава подробна информация как е получен резултатът. Това приложение е с най-базова функционалност от всички подобни приложения, разгледани в проучването.

2.4.3.4 Изводи

Разгледаните приложения за предсказване на успеха на стартиращи компании не изпълняват в максимална степен задачата за предсказване успеха на технологични стартиращи компании от България. Те са предпоставка и мотивация за разработката на софтуера към настоящата дисертация. Основните прилики са в попълването на данни за бизнеса от основателя, собственика или мениджъра на компанията, която ще бъде анализирана и използването на статистически данни за съпоставка и получаване на прогноза за успех. При някои от приложенията, за по-лесно възприемане, към резултата се добавят и графични изображения.

В Таблица 6 е представена обобщена съпоставка на анализирани инструменти за предсказване на успеха, и съответно техните силни и слаби страни.

Таблица 6 Обобщена съпоставка на анализирани инструменти за предсказване на успеха

Инструмент	Силни страни	Слаби страни
------------	--------------	--------------

Калкулатор на риска за стартиращи компании от StartupNation	<ul style="list-style-type: none"> • използва огромна база от статистически данни получени от другите приложения на компанията, което спомага за по-голяма точност • средата, в която компанията развива своята дейност не е определяща за точността, тъй като резултатите в базата данни са събрани от компании от различни държави 	<ul style="list-style-type: none"> • броят на входящите параметри е малък (осем променливи) и това предполага малка прецизност и точност на прогнозата
Компас за нови компании	<ul style="list-style-type: none"> • добър дизайн на приложението и ясни графично представени резултати • възможност за ежемесечна актуализация на резултата и за измерване на напредъка на компанията • предложения за безплатни софтуерни продукти, подпомагащи бизнеса 	<ul style="list-style-type: none"> • тясно специализиран за работа със стартиращи технологични компании • статистическите данни, които се използват, са от САЩ и инструментът вероятно няма да даде точни резултати за други държави.
Инструмент за проверка на стартиращ бизнес от Blueprint	<ul style="list-style-type: none"> • има опростена (базова) функционалност 	<ul style="list-style-type: none"> • липсва информация за статистическите данни, които са използвани • липсват данни за начина на изчисление на резултата • броят на входящите параметри е малък (10 променливи) и това предполага малка прецизност и точност на прогнозата

От прегледа на съществуващите приложения може да се заключи, че софтуерът за прогнозиране на успеха на стартиращи компании няма да бъде първото подобно приложение. Най-известните приложения от този тип всъщност представляват портали и включват в себе си както потребители, които имат ролята на предприемачи, така и инвеститори, готови да финансират някои от идеите на предприемачите. При регистрация в тези портали, предприемачът обикновено трябва да предостави голямо количество информация за своя бизнес или идея. В повечето случаи е необходимо да създаде бизнес план и да го направи достъпен в съответния портал. На базата на този бизнес план, порталът дава възможност за оценка на бизнес идеята по ключови фактори и предлага инструменти за оценка на риска. Данните от тези анализи са видими както за предприемача, регистрирал своята идея в портала, така и за потенциалните инвеститори.

Инвеститорите, като потребители на такъв портал, най-често получават възможност да преглеждат бизнес планове и да се запознават с нови бизнес идеи на предприемачи, регистрирани в портала и търсещи финансиране. С помощта на инструментите за анализ и на базата на представения бизнес план, даден инвеститор, или няколко заедно могат да решат да направят предложение за финансиране на определен бизнес, като в замяна най-често получават дял от компанията. Този метод на финансиране е по известен като групово финансиране (crowdfunding). Приложението, обект на настоящата дисертация, би могло също да се развие в бъдещ момент като уеб портал, който среща интересите на предприемачи и инвеститори.

Софтуерът, който е разработен към дисертацията, превъзхожда съществуващите в момента с това, че базира прогнозата си на значително повече входящи данни (около 100 променливи, в сравнение с около 10 при конкурентните инструменти), което спомага за по-точна прогноза. Моделът за предсказване на успеха и софтуерът са създадени специално за българската бизнес среда. Статистическите данни, които са използвани за създаване на модела за прогноза, са събрани от български стартиращи компании. В следствие на това, приложението за прогнозиране на успеха дава най-точни резултати за български компании.

2.5 Обзор на софтуерни продукти за извличане на знания от данни

В настоящата точка се разглежда направено от автора проучване на софтуерни продукти за извличане на знания от данни и на техните характеристики. Съпоставени са различни продукти и са избрани тези, които са подходящи за прилагане в изследването към настоящата дисертация с цел потвърждение на хипотезите чрез количествени методи и синтезиране на модели за предсказване на успеха на стартиращи компании.

2.5.1 Обзор на софтуерния продукт IBM SPSS Statistics

IBM SPSS Statistics е интегриран набор от софтуерни продукти, които адресират целия аналитичен процес на извличане на знания от данни – от планирането, до събирането на данни, отчитането и внедряването (IBM, 2014). Продуктът, който първоначално е бил разработен от SPSS Inc, в момента се поддържа от IBM и разполага с повече от десет напълно интегрирани модула, позволяващи провеждане на изследвания и вземане на решения. Името на софтуера „SPSS“ произхожда от абревиатура, означаваща „Статистически пакет за социалните науки“, но днес продуктът е популярен в различни области.

2.5.2 Обзор на софтуерния продукт IBM SPSS Modeler

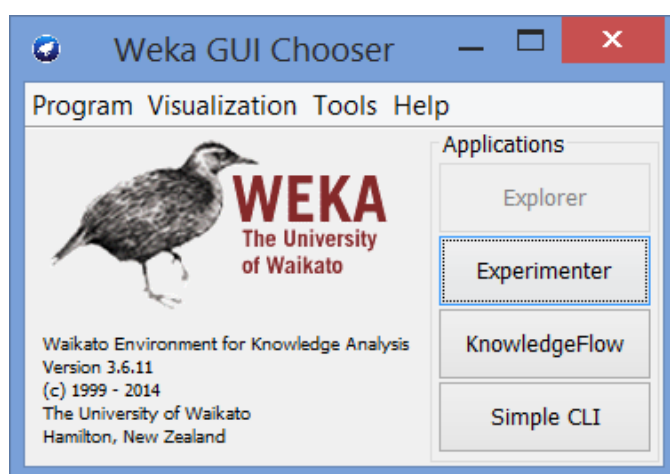
IBM® SPSS® Modeler е набор от софтуерни инструменти за извличане на знания от данни (data mining), които позволяват бърза разработка на модели за предсказване чрез използване на бизнес експертиза и тяхното внедряване в дейността на бизнеса за подобряване процеса на вземане на решения. Проектиран около възприетата като индустриалния стандарт методология CRISP-DM, IBM SPSS Modeler поддържа цялостния процес на извличане на знания от данни. Той предлага разнообразие от методи за моделиране, заимствани от машинното учене, изкуствения интелект и статистиката. Те позволяват извличането на нова информация от наличните данни и разработката на модели за предсказване. Всеки от наличните в продукта методи има определени предимства и е най-подходящ за конкретен тип задачи.

IBM® SPSS® Modeler е комерсиален продукт и може да бъде закупен като самостоятелен продукт, или да бъде използван като приложение-клиент в комбинация с

SPSS Modeler Server. За целите на настоящото изследване продуктът IBM® SPSS® Modeler е предоставен за безплатно използване от IBS Bulgaria.

2.5.3 Обзор на софтуерния продукт Weka

Weka е набор от алгоритми за машинно учене, прилагани при задачи за извличане на знания от данни (Bouckaert, 2013), (Waikato, 2014). Алгоритмите могат да бъдат приложени директно към набор от данни или извикани от Java код. Weka съдържа инструменти за предварителна обработка на данни, класификация, регресия, разделяне на клъстери, правила за асоцииране и визуализация. Продуктът е подходящ и за разработка на нови правила за машинно учене. Той съдържа инструменти, изпълнявани от командния ред, както и графичен потребителски интерфейс. Графичният потребителски интерфейс включва четири различни приложения: Explorer, Experimenter, KnowledgeFlow, Simple CLI (Фиг. 11), подходящи за различни задачи.



Фиг. 11 Избор на приложение от графичния потребителски интерфейс на Weka (екранна снимка, създадена от автора)

Weka е безплатен продукт и се разпространява с отворен код под лиценза GNU General Public License.

2.5.4 Обзор на софтуерния продукт HP Vertica

HP Vertica Analytics Platform е платформа за анализ на големи обеми от данни (HP Vertica, 2015). Тя позволява на големи организации да анализират масивни обеми от структурирани, полу-структурирани и неструктурирани данни, възползвайки се от предлаганите скорост и аналитичност на платформата.

Основните предимства на HP Vertica са по-бързото изпълнение на заявки към базата данни (от 50 до 1000 пъти по-бързо в сравнение с традиционните бази данни), от 10 до 30 пъти повече съхранявана информация на всеки сървър от инфраструктурата и възможност за прилагане на специализирани инструменти за бизнеса.

HP Vertica Analytics Platform поддържа концепцията за извличане на знания от данни (data mining) чрез прилагането на “R” - статистически софтуер с отворен код, включващ ефикасен език за програмиране (The R Foundation, 2015). Платформата HP Vertica е подходяща за изпълнение на сложни алгоритми, програмирани на R, които се изпълняват върху множество сървъри с данни. Примери за такива алгоритми са

предложения за продукти в онлайн магазини с милиони артикули, изчисление на PageRank от търсещи машини (Lawrence, 1998), изчисление на цени на финансови инструменти, разпознаване на важни хора в социални мрежи.

HP Vertica е комерсиален продукт, чието използване е обвързано със заплащане на лиценз.

2.5.5 Обзор на софтуерния продукт Microsoft SQL Server

Microsoft SQL Server е платформа за бази данни за управление на голям брой онлайн транзакции, съхранение на данни, електронна търговия, както и BI (business intelligence) (Microsoft, 2015).

Един от компонентите на Microsoft SQL Server 2005 се нарича Analytical Services и служи за анализ на бизнес данни, онлайн аналитична обработка и извличане на знания от данни. Analytical Services съдържа следните възможности и инструменти, чрез съвместното прилагане на които, могат да бъдат решавани сложни проблеми за извличане на знания от данни:

- набор от алгоритми за извличане на знания от данни, следващи индустриалните стандарти в областта;
- инструмента Data Mining Designer, който позволява създаване, поддържане, преглед и изготвяне на прогнози от модели за извличане на знания от данни;
- езика DMX, който може да се приложи за управление на модели за извличане на знания от данни и създаване на сложни заявки за предсказване.

Microsoft SQL Server 2005 е комерсиален продукт, чието използване е обвързано със заплащане на лиценз.

2.5.6 Изводи и заключение

От направения обзор може да се заключи, че повечето продукти за извличане на знания от данни са комерсиални и за тяхното прилагане в изследването е нужно закупуване на лиценз.

Ще бъде използван продукта IBM SPSS Modeler, тъй като той е специализиран в решаването на задачи с извличане на знания от данни, разработва се от утвърдена компания (IBM) и е лесен за прилагане. За целите на настоящото изследване, продуктът IBM® SPSS® Modeler е предоставен за безплатно използване от IBS Bulgaria. За подготовка на данните и за някои конкретни задачи ще бъде използван в допълнение продуктът IBM SPSS Statistics, който е съвместим с IBM SPSS Modeler.

Тъй като различните софтуерни продукти за извличане на знания от данни създават класификационни модели чрез различни алгоритми, може да се очаква, че резултатите, получени чрез прилагане на повече от един продукт върху един и същи данни ще бъдат различни. По тази причина ще бъде приложен още един продукт, за да бъдат класифицирани данните и това е Weka – популярен софтуер с отворен код. Допълнителен принос на прилагането на още един продукт и създаването на по-разнообразни модели е, че в софтуера за предсказване на успеха към дисертацията ще бъде разработена възможност за обновяване и избор на модела за предсказване на успеха, и потребителите ще могат да избират между алтернативни модели.

2.6 Избор на технологии и инструменти за разработка на информационна система

В настоящата точка са разгледани инструментите, програмните езици и технологиите и са съпоставени различни възможни решения за разработката на софтуерната система, обект на дисертацията. Избрани са технологиите и инструментите, които най-добре пасват на целите, и които ще бъдат приложени.

2.6.1 Избор на тип приложение

Преди да започне разработката на софтуерът I3SP, трябва да се вземе решение дали той да се реализира като десктоп приложение или като клиент-сървър уеб приложение.

Десктоп приложенията имат основен недостатък в това, че са платформено зависими и изискват инсталация. За всяка различна операционна система (Windows, Linux, OS X, Android), трябва да се създават различни приложения. Основно предимство на десктоп приложенията е, че те имат директен достъп до ресурсите на системата, в резултат на което могат да работят по-бързо.

Клиент-сървър уеб приложенията са платформено независими от гледна точка на крайния клиент. Потребителят има нужда само от уеб браузър, за да ги използва, а такъв се съдържа във всяка съвременна операционна система. Като недостатък на уеб приложенията може да се отбележи, че клиентът има нужда от постоянна свързаност с интернет и бързината на приложението е зависима от скоростта на пренос на данни между сървъра и клиентския браузър. Постоянната свързаност е сред предварителните изисквания към I3SP, поради използване на данни от сървърите на Google.

Поради описаните причини, е взето решение за разработката на I3SP като клиент-сървър уеб приложение.

2.6.2 Избор на операционна система

След като беше взето решение I3SP да се разработи като клиент-сървър уеб приложение, трябва да се избере операционна система, на която да бъде инсталирана сървърната част. В Таблица 7 са представени основните характеристики на двата най-популярни типа операционни системи (Surveys, 2014).

Таблица 7 Сравнение на операционни системи за уеб сървър

	Linux и Unix-базирани	Microsoft Windows
Пазарен дял	67,6% пазарен дял при уеб сървърите	32,5% пазарен дял при уеб сървърите
Цена	Голяма част от дистрибуциите се разпространяват безплатно под лиценза GPL. Сървърните услуги (виртуален сървър и уеб хостинг) , работещи с Linux се предлагат на по-ниски цени.	Разпространява се срещу заплащане. Нужно е закупуването на лиценз за ползване. Сървърните услуги (виртуален сървър и уеб хостинг) , работещи с Windows се предлагат на по-високи цени.
Софтуер	Linux се предлага в голям брой дистрибуции: Ubuntu, OpenSUSE, Fedora, Debian, CentOS, Mandriva, Gentoo и кодът му е отворен. Повечето приложения за уеб сървъри	Част от приложенията са платени (напр. Microsoft SQL Server), въпреки че съществуват и безплатни. При създаването на платените приложения, често са вложени повече

	са безплатни (напр. Apache, MySQL, PHP, Perl). Linux и повечето приложения за него често са по-сложни за ползване и настройка от тези за Windows.	усилия и средства за разработка и тестване, а при някои от тях се предлага и поддръжка.
Сигурност и стабилна работа	Доказана сигурност, възможности за различни конфигурации, гъвкавост и стабилност на работа, без честа нужда от рестартиране, пълна независимост от графичния интерфейс.	Добра сигурност и стабилна работа, големи възможности на графичния интерфейс, по-лесно първоначално конфигуриране, по-малко възможности за настройки и адаптация.

За целите на приложението, като сървърна операционна система ще бъде използвана Linux, а основни фактори за избора са възможността за легално безплатно ползване и по-достъпните уеб хостинг услуги.

2.6.3 Избор на приложение за уеб сървър

Клиент-сървър уеб приложението се нуждае от софтуер за уеб сървър, който да се грижи за заявките, постъпващи през HTTP протокол от клиентския браузър и да връща подходящото съдържание.

Един от най-разпространените и известни уеб сървъри е **Apache** – най-често инсталиран под Linux, но с версии и за други операционни системи. Той е с отворен код, разпространява се безплатно, има репутация на сигурен и добре скалируем уеб сървър, особено подходящ за малки и средни по мащаб проекти.

Със сървърните версии на операционната система Windows безплатно се предлага уеб сървърът **IIS**, който също предлага достатъчно стабилна работа и скалируемост, както и голям брой настройки.

За целите на приложението I3SP, ще бъде използван Apache уеб сървър, тъй като той е мултиплатформен и безплатен за ползване.

2.6.4 Избор на програмни средства за разработка от страна на сървъра

Езиците за програмиране от страната на уеб сървъра дават възможност за динамично генериране на уеб страници. Следните програмни езици от страната на сървъра се използват най-често:

- **PHP** – популярен скриптов програмен език с общо предназначение, специално пригоден за бърза разработка на уеб приложения (Achour, 2014). Основните му предимства са, че е безплатен, работи бързо, интуитивен е за ползване (при умерена сложност на проекта) и позволява сравнително бърза разработка на динамични уеб сървърни приложения. Съществуват множество модули за PHP, които добавят различни функционалности. PHP се разпространява под отворен лиценз и работи на различни платформи.
- **ASP.NET** – платформа за разработка на уеб сървърни приложения, създадена от Microsoft (Microsoft, 2014). ASP.NET позволява програмирането на различни обектно ориентирани езици, като C#, Visual Basic, C++. Платформата ползва виртуална машина наречена CLR (Common Language Runtime), която с помощта на CLI (Common Language Infrastructure), изпълнява кода, написан на различните езици. ASP.NET работи под Windows и не изисква заплащане на допълнителен лиценз.

- **JSP (Java Server Pages)** – технология за разработка на уеб сървърни приложения, използваща езика за програмиране Java (Oracle Corp., 2014). Кодът на приложенията се компилира, като технологията използва Java виртуална машина, за да позволи изходният код да бъде транслиран към собствен (native) код, което позволява работа под различни платформи. Използването на Java не е обвързано със заплащане на лицензи.
- **PERL** - скриптов програмен език с общо предназначение. Разпространява се под GNU General Public License, който е лиценз за свободен софтуер. Работи под множество платформи.

За целите на разработка на софтуера I3SP е избран програмен език PHP, тъй като той е безплатен за използване, работи под Linux и е пригоден за сравнително бързо разработване на уеб приложения.

2.6.5 Избор на програмни средства за разработка от страна на клиента

Езикът, който уеб браузърите интерпретират, за да визуализират информацията, е **HTML** (в превод „език за маркиране на хипертекст“). HTML се състои от думи, наречени тагове, заградени в триъгълни скоби, които описват семантиката и в някои случаи външния вид на документа. HTML документите могат да включват вграден код на скриптов език, като например JavaScript. HTML кодът се интерпретира от уеб браузърите и, в комбинация с езика CSS, задава начина, по който информацията между HTML таговете се визуализира при потребителя. **xHTML** е вариант на езика HTML, който притежава характеристиките на езика XML и е възприет като добра практика за изграждане на HTML документи. Тъй като xHTML е най-често използваният език за разработка на фронтенд на уеб приложения и се поддържа от всички уеб браузъри, той ще бъде използван за разработката на софтуера за предсказване на успеха.

JavaScript е скриптов език за програмиране, с чиято помощ се създават фрагменти код, които се вграждат в HTML документа. JavaScript кодът се интерпретира от клиентския браузър и не може да бъде ползван директно за запис на данни на сървъра, но с негова помощ може да се изпраща информация към сървъра и там тя да се обработва от сървърните скриптове. Значението на JavaScript през последните години нарасна значително с широкото навлизане на AJAX, JSON и jQuery.

jQuery е JavaScript библиотека, публикувана в началото на 2006 от Джон Резик. jQuery опростява достъпа до всеки елемент на дадена уеб-страница, чрез синтаксис, който не зависи от използвания уеб браузър, като по този начин позволява лесно изграждане на динамична функционалност в страниците. jQuery е безплатен софтуер с отворен код, лицензиран под MIT лиценз или GNU GPL.

CSS (Cascading Style Sheets) е програмен език, използван за описание на визуалното представяне на документ, написан на език за форматиране (HTML, xHTML, XML). Каскадните стилове позволяват по-добър контрол върху елементите на страницата и дават допълнителни възможности за визуализацията им.

Поради поставените изисквания за връзка с Google диск, I3SP има нужда от асинхронно предаване на данни между клиентския браузър и сървъра. Ще бъдат разгледани две технологии за асинхронно предаване на данни между клиента и сървъра:

- **AJAX (Asynchronous JavaScript and XML)** е технология за асинхронно предаване на данни, която работи със сървърните езици PHP, ASP.NET, PERL, JSP, както и други и се поддържа от почти всички уеб браузъри, налични към момента. Базира се на JavaScript. AJAX е похват в уеб разработките за създаване на

интерактивни уеб приложения чрез асинхронен обмен на малки порции данни без да се презарежда цялата уеб страница. За разработката на приложения с AJAX не се изисква заплащане на лиценз.

- **Adobe Flex** е набор от технологии, предоставящи възможност за изграждане на интерактивни уеб приложения работещи върху Adobe Flash платформата. Adobe Flex представлява цялостна рамка за разработка на RIA (Rich Internet Application, в превод интернет приложения с богата функционалност).

За целите на разработката на I3SP, е избрана технологията AJAX и често са използвани асинхронни заявки. Използвана е и техниката JSON-P, която позволява заявките да се насочват към различни интернет домейни.

2.6.6 Избор на бази данни

Базата данни на приложението ще бъде избрана съобразно очакваното натоварване, съхраняваните типове данни, наличието на клиентски софтуер за администриране и писане на заявки и дали избраният език за програмиране има добра поддръжката на базата данни. Ще бъдат разгледани някои от най-популярните бази данни:

- **MySQL** – популярна релационна база данни, която се разпространява под GNU General Public License. Повечето версии могат да се ползват безплатно. MySQL притежава всички важни елементи, които се очакват от една съвременна база данни (съхранени процедури, тригери, изгледи, курсори, транзакции, пълнотекстов индекс и други). Най-често, приложенията, написани на PHP, използват MySQL база данни. Тя може да се инсталира под Windows, Linux и OS X. Допълнително към базата данни се разпространява безплатен клиентски софтуер за администриране и писане на заявки – MySQL Workbench и phpMyAdmin.
- **Microsoft SQL Server** - система за управление на релационни бази данни. MSSQL е добре развита база данни, предлагаща отлични възможности за съхранение и достъп до големи обеми от данни. Приложенията, написани на ASP.NET и .NET framework, най-често използват тази база данни. MSSQL Server работи само под операционна система Windows и се разпространява под лиценза Microsoft EULA (със затворен код). Обикновено тази база данни се ползва за много големи приложения (enterprise) - поради високите цени на лицензиите за базата данни и за операционната система.
- **PostgreSQL** – релационна база данни с отворен код, която не отстъпва по възможности на платените продукти. Езикът на заявките, който ползва тази база данни е PL/SQL. PostgreSQL може да бъде инсталирана на различни платформи.

За целите на разработка на приложението I3SP, ще бъде използвана MySQL база данни, тъй като MySQL се поддържа добре от PHP, има много добра производителност и е безплатна за ползване.

2.6.7 Избор на среда за разработка

За среда на разработка на софтуера I3SP ще бъдат използвани NetBeans IDE и PSPad. **NetBeans IDE** е безплатен софтуер с отворен код, който спомага разработването на уеб приложения, като предлага специализирани текстови редактори, които разпознават PHP, HTML, CSS и JavaScript код. Той поддържа оцветяване на кода, което значително подобрява четимостта. Средата за разработка също така предлага бързи описания на методи и класове (предназначение, очаквани параметри и др.). **PSPad** е лек текстов редактор за разработка на софтуер, който също предлага оцветяване на кода и описания

на PHP функциите и техните параметри. PSpad предлага и допълнителното удобство на работа директно през FTP протокол, което е полезно при разработка на уеб приложения в отдалечен уеб хостинг сървър.

2.7 Изводи и заключение

От направения обзор на проблемната област може да се заключи, че технологичното предприемачество е актуална тема, а проблемът за предсказване на успеха на стартиращи компании набира все по-голяма популярност през последните 25 години и става все по-актуален. Поради факта, че развитието на предприемачеството е заложено като основен приоритет за Европейския съюз за новия програмен период Хоризонт 2020, а и представлява силен интерес за инвеститорите, се очаква тенденциите за развитие в тази област и нарастващият брой публикации да се запазят.

Разгледани са дефинициите за технологично предприемачество, стартираща компания и за успех на стартираща компания в литературата. Успехът на стартиращите компании е разгледан като многостранен проблем с различните тенденции и варианти за неговото дефиниране. Проведени са интервюта с предприемачи, основатели и собственици на компании, на базата на които е избрана дефиниция за успешна стартираща компания за целите на настоящата дисертация.

Разгледани са различни модели за предсказване на успеха, създадени от 1974 година до сега. Анализирани е развитието на моделите и са съпоставени предложените в тях фактори за успех на компаниите. Разгледан е по-детайлно модел, създаден от Сендбърг, който служи като основа на множество следващи модели, чиито фактори са анализирани в различни проучвания. Моделът на Сендбърг може да послужи като основа за изграждане на нови модели за предсказване на успеха на стартиращи компании от България.

Избрана е методика за отделните етапи на изследването.

Проучени и анализирани са различни софтуерни продукти за прогнозиране успеха на стартиращи компании. Идентифицирани са три продукта, които са най-близки по тип до софтуера, който ще бъде разработен към дисертацията. Трите продукта са сравнени на база техните предимства и недостатъци. Разгледани са уеб портали за групово финансиране, които свързват предприемачите с инвеститорите и в някои случаи предлагат функционалност за прогнозиране на успеха на компаниите. Идентифицирани са основните предимства и разлики на софтуера, който ще бъде разработен, спрямо съществуващите.

Съпоставени са различни технологични решения, инструменти и програмни езици, и са избрани тези, които ще бъдат приложени при разработката на софтуерния продукт към дисертацията – система за прогнозиране на успеха на стартиращи компании от България.

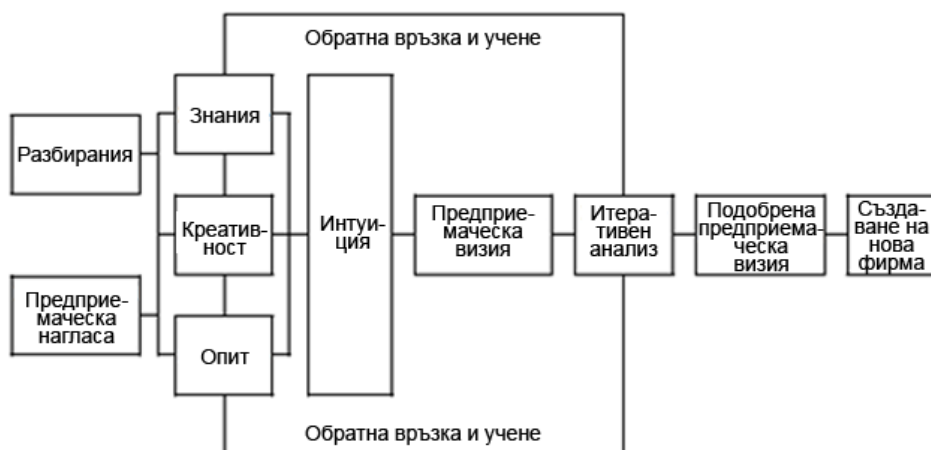
ГЛАВА 3: РАЗРАБОТВАНЕ НА МОДЕЛ ЗА ПРЕДСКАЗВАНЕ НА УСПЕХА

В настоящата глава се разработват и предлагат хипотези за:

- модел на процеса на създаване на компания,
- абстрактен модел за предсказване на успеха на стартираща компания, съдържащ категориите и подкатегориите на факторите за успех,
- детайлизиран модел на успеха на компания, съдържащ предложение за отделните фактори за успех.

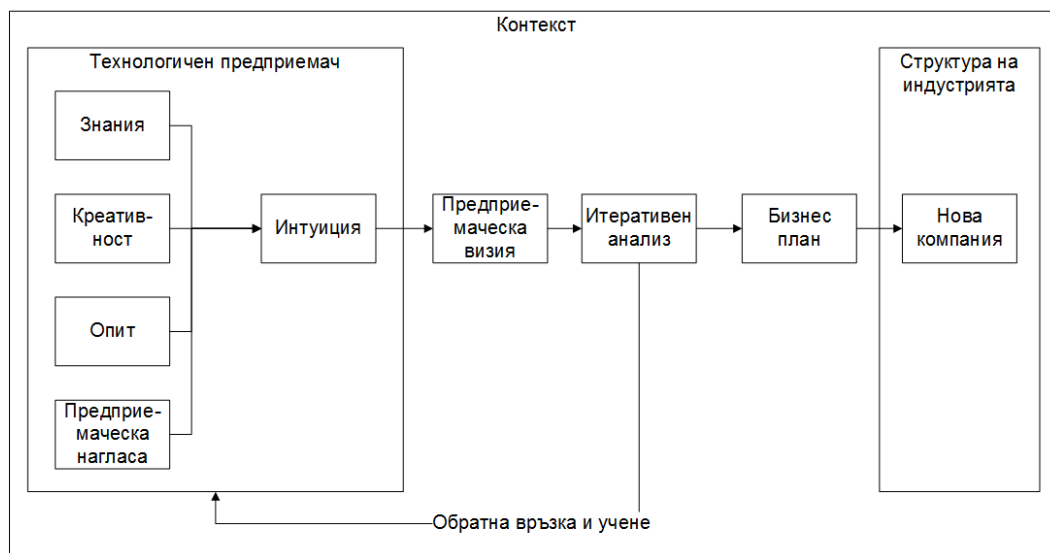
3.1 Модел на процеса на създаване на фирма

Джоан и Джеймс Карлънд (Carland, A new venture creation model, 2000) твърдят, че успехът на новите фирми може да бъде предсказан, ако преди това моделираме процеса на създаване на компания. Те анализират 8 модела на процеси на създаване на фирма и предлагат подобрен модел (Фиг. 12), в който предприемачът разработва визия за фирмата, базирана на неговите разбирания, предприемаческа нагласа, знания, креативност, опит и интуиция. Тази ранна визия се разработва допълнително, в резултат на което може да бъде отхвърлена или да се развие в нов бизнес.



Фиг. 12 Модел на процеса на създаване на нова фирма (преведена от автора по Джоан и Джеймс Карлънд)

Авторът е развил и допълнил модела на процеса на създаване на нова фирма, като са добавени контекста, включващ стратегия на компанията и структура на индустрията (от модела на Сендбърг) и подобрената предприемаческа визия е заместена с по-формалното определение - бизнес план (Фиг. 13). Целият процес се случва в и зависи от контекста. Първоначалната част от процеса – формирането на предприемаческата визия се случва в мисловната рамка на технологичния предприемач. След анализ, промени и преосмисляне, визията може да се развие в (писмен или неформален) бизнес план. Тогава новата фирма се образува в рамките на съответната индустрия.



Фиг. 13 Допълнен модел на процеса на създаване на нова фирма (създадена от автора)

Така формулиран, моделът на процеса служи за осмисляне на етапите, действията и предпоставките, които довеждат до създаването на нова компания.

3.2 Абстрактен модел за предсказване на успеха на стартиращи фирми

На база разгледаните предишни опити за създаване на модели за предсказване на успеха на стартиращи компании, се достигна до заключението, че моделите за предсказване на успеха, основани върху модена на Сендбърг, са добър шаблон за проектиране на нови модели. Моделът за предсказване на успеха на стартиращи компании трябва да включва следните фактори: стратегията на фирмата, структурата на индустрията, и в някои случаи взаимодействието между стратегията и структурата. Включването на други фактори в модела, като например предприемаческия екип, би било от полза, тъй като те могат да се окажат важни при валидация на модела. Всеки фактор в модела трябва да бъде измерим и ясно дефиниран.

Чрез анализиране на изискванията към моделите за предсказване на успеха, авторът предлага разширен модел за предсказване успеха на нови компании, базиран на Сендбърг. Моделът е представен с формулата:

$$NVP = f(E, IS, BS, R) \quad (3)$$

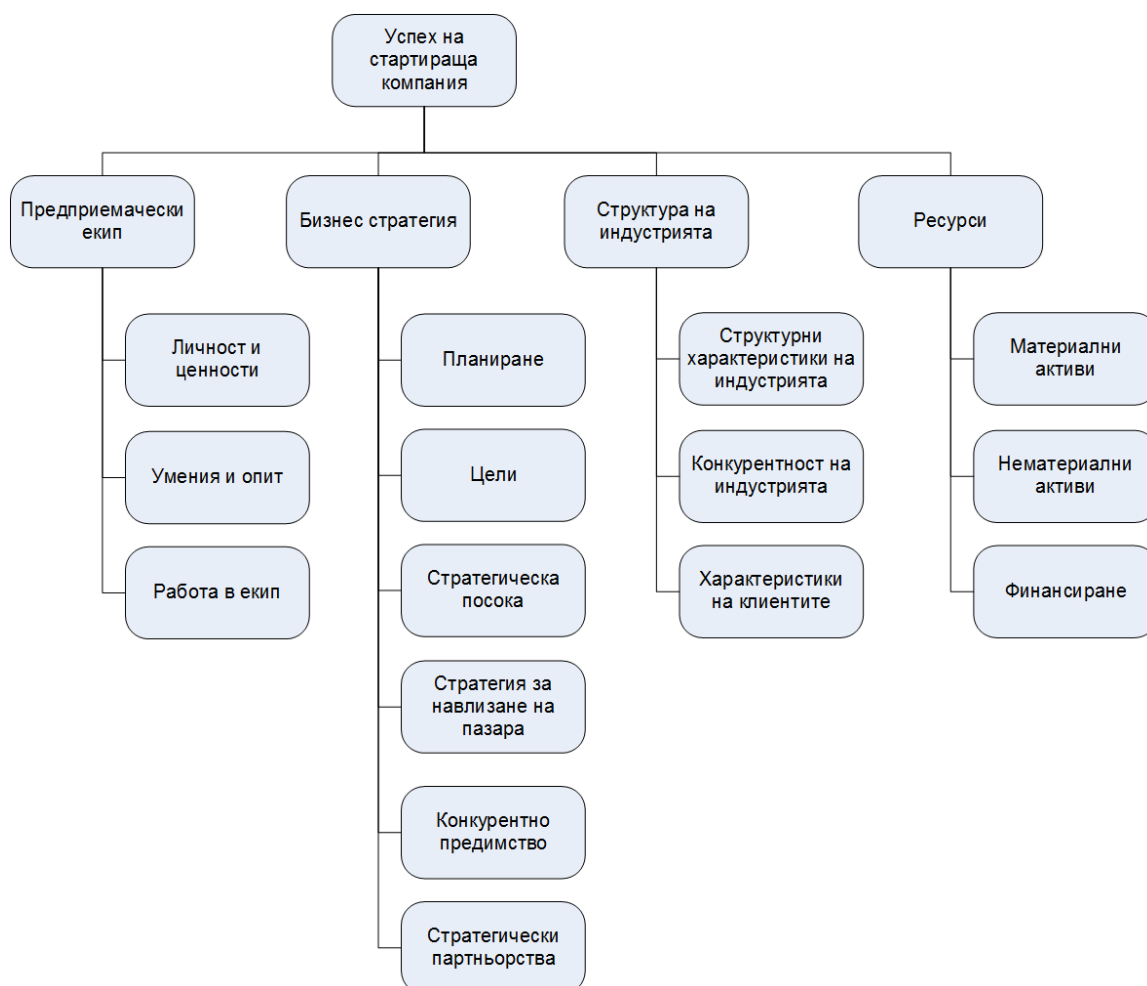
Където NVP означава продуктивността на новата фирма, E е предприемачът, IS е структурата на индустрията, BS е бизнес стратегията, а R са ресурсите.

В сравнение с модела от Крисман, Бауершмид и Хофър, в предложения модел липсва организационната структура, която е дефинирана като процесите и системите, разработени от фирмата, за имплементирането на нейната стратегия и постигането на целите ѝ. Разработените процеси и системи се приемат за част от нематериалните активи на фирмата и следователно – част от ресурсите.

Екипът се причислява към човешките ресурси и би могъл да попадне в категорията ресурси, но в контекста на предприемачеството и стартиращите компании,

предприемаческият екип е основен фактор за успеха, който има ключова роля за създаване и развитие на компанията и по тази причина е обособен в отделна категория.

Всяка от основните категории в модела за предсказване на успеха на компаниите се състои от подкатегории, както е показано на Фиг. 14 (Yankov, A Model for Predicting the Success of New Ventures, 2013).



Фиг. 14 Модел за предсказване на успеха на стартиращи компании (създадена от автора)

Категорията предприемачески екип включва личността и ценностите на основателя на компанията (напр. увереност, локус на контрол, толерантност към рискове), уменията и опита на мениджмънт екипа (напр. технически умения, умения по мениджмънт, по маркетинг, опит на подобна позиция) и работата в екип (напр. възможността на екипа да работи заедно и да генерира стойност).

Категорията бизнес стратегия включва планирането (напр. задълбоченост и обширност на планирането), целите, стратегическата посока (напр. агресивност на фирмената стратегия), етапа на навлизане на пазара (напр. иноватор, рано или късно), конкурентното предимство и стратегическите партньорства (напр. партньорства с конкуренти, с клиенти, с правителството).

Категорията структура на индустрията включва характеристиките на индустрията (напр. бариери, размер и предвидимост), конкуренцията в индустрията и характеристиките на клиентите (напр. концентрация, хетерогенност).

Категорията ресурси включва материалните активи (напр. пари в брой, складови наличности), нематериалните активи (напр. изградена бизнес репутация, патенти, социални мрежи) и използвания източник на финансиране (напр. със собствени средства, рисков капитал или банков заем).

За гарантиране на пълнота и изчерпателност за модела за българската предприемаческа и бизнес среда, беше проведено качествено изследване с предприемачи и собственици на бизнес (описано в Глава 4) и техните препоръки и предложения бяха включени в предложения модел. Описаните категории и подкатегории представляват хипотеза, която ще бъде проверена и с помощта на количествено изследване и прилагане на техниките факторен анализ и моделиране.

3.3 Детайлизиран модел за предсказване на успеха на стартиращи фирми в България

Детайлизираният модел за предсказване на успеха на стартиращи компании в България е разработен, базирайки се на предложения абстрактен модел за предсказване на успеха. Всяка от четирите основни категории и техните подкатегории от абстрактния модел са представени чрез конкретни променливи. Променливите се базират както на моделите за предсказване на успеха, приложени за други държави от други автори, които са разгледани в литературния обзор, така и на адаптации и нововъведения, специфични за България.

Едно от нововъведенията в модела е включването на нематериалните активи към ресурсите на компанията, което след проведеното количествено изследване действително се оказва ключов фактор за успеха на компаниите. Други нововъведения в модела са добавянето на променливи, определящи качествата на личността и ценностите на предприемача, наличието на криза в индустриалния сектор (фактор, който е актуален през последните години в България), източникът на финансиране на компанията (фактор, отнасящ се към стартиращи и млади компании).

Някои от променливите в детайлизирания модел, като възрастови групи на предприемачите, завършено образование и индустриални сектори на компаниите са адаптирани според данните от НСИ, което прави възможна съпоставимостта на данните от проучването със средните нива за България.

Представеният детайлизиран модел е предварително ревизиран и подобрен спрямо първоначалната му версия, посредством качествено изследване – интервюта с предприемачи и собственици на бизнес, на които е представен модела. Извършени са адаптации и подобрения на база отговорите и получените препоръки. Детайли и резултати, касаещи качествено изследване са представени в Глава 4.

Детайлизираният модел за предсказване на успеха, създаден от автора, включва следните категории, подкатегории и променливи (въпроси, определящи успеха на стартиращите компании):

Предприемачески екип

- Личност и ценности
 - Мога да се справям в работата си без нужда от съдействие.
 - Аз съм уверен в себе си.

- Аз съм предприемчив, инициативен.
- Това, което ми се случва в живота зависи от мен самия.
- Аз съм амбициозен.
- Склонен съм да поемам рискове.
- Не се притеснявам да вземам решения без да разполагам с необходимата информация.
- Умения и опит
 - Имам предприемачески умения (умения в стартирането на бизнес).
 - Имам умения по мениджмънт.
 - Имам умения по маркетинг.
 - Имам технически умения (за технологии, свързани с моя бизнес).
 - Имам умения по човешки ресурси (HR).
 - Имам умения по инвестиране.
 - Имам предишен опит в стартирането на бизнес.
 - Имам опит на подобна позиция (спрямо заеманата в момента).
 - Имам опит в същата сфера на индустрията.
 - Имам управленски опит.
 - Какво образование сте завършили?
 - В коя област е завършеното от Вас образование?
 - Каква е Вашата възрастова група?
 - Предприемачи / собственици на бизнес ли са Вашите родители?
- Работа в екип
 - Фирмата разполага със завършен екип.
 - Екипът притежава необходимите знания.
 - Екипът притежава необходимите умения.
 - Екипът притежава необходимата положителна нагласа.
- **Бизнес стратегия**
 - Планиране
 - Планирам, обхващайки широка област.
 - Планирам задълбочено.
 - Правя формален (писмен) стратегически план.
 - Планирам често.
 - Планирам по функционални области / отдели.
 - При планирането разглеждам множество алтернативи.
 - Често правя анализ на конкуренцията.
 - Цели
 - Целите, които поставям пред бизнеса си, са амбициозни.
 - Целевият пазарен дял на фирмата е:
 - Целевата годишна печалба на фирмата е (в лв)
 - Стратегическа посока
 - Успявам да се придържам към първоначално избраната стратегия.
 - Стратегията, която съм избрал, е агресивна.
 - Набелязаната стратегия е подробна.
 - Имам ясна стратегия за бизнеса си.

- Стратегия за навлизане на пазара
 - Начин на навлизане на пазара
 - Бизнесът ми навлиза на пазара на следния етап.
 - Продуктът, който предлагам, е възприет от пазара в следната степен.
- Конкурентно предимство
 - Моят бизнес притежава ясно конкурентно предимство.
 - Бизнесът ми има следното основно конкурентно предимство.
 - Смятам, че следните фактори са ключови за успеха за моя бизнес.
- Стратегически партньорства
 - Фирмата участва в партньорства с конкуренти.
 - Фирмата участва в партньорства с клиенти.
 - Фирмата участва в партньорства с правителството.
 - Фирмата участва в партньорства с доставчици.
 - Фирмата участва в партньорства с други заинтересовани страни.
- **Структура на индустрията**
 - Характеристики на индустрията, в която фирмата оперира
 - Има бариери (пречки) за навлизане на нови фирми в индустрията.
 - Има бариери (пречки) за излизане от бизнеса.
 - Повечето компании в индустрията имат добра печалба.
 - Индустрията, в която фирмата оперира, генерира голяма част от brutния вътрешен продукт (БВП).
 - В индустрията има много фалити на фирми.
 - В индустрията има стабилно търсене.
 - Индустриалният сектор, в който фирмата оперира, е голям в сравнение с останалите сектори.
 - Индустрията е в период на растеж.
 - В индустрията има криза.
 - Индустрията е предсказуема.
 - Фирмата работи в следния индустриален сектор.
 - Индустрията, в която работи фирмата, се намира в стадий на.
 - Конкурентност на индустрията
 - В индустрията има силна конкуренция.
 - Конкурентите в индустрията са агресивни.
 - Компанията е изложена на конкурентни атаки.
 - В индустрията присъстват и малки фирми.
 - Характеристики на клиентите
 - В индустрията има концентрация на клиенти.
 - Клиентите са различни типове.
- **Ресурси**
 - Материални активи
 - Компанията притежава оборудване / машини.
 - Компанията притежава складови наличности.
 - Компанията притежава недвижими имоти.

- Компанията притежава пари в брой.
- Нематериални активи
 - Компанията притежава патенти / търговски марки.
 - Компанията притежава разпознаваем бранд.
 - Компанията притежава софтуерни системи за управление на бизнес процесите (ERP, CRM).
 - Компанията притежава изградена бизнес репутация.
 - Компанията притежава разработени собствени продукти.
 - Компанията притежава лицензи.
 - Компанията притежава изградени бизнес контакти (социална мрежа)
 - Компанията има достъп до канали за дистрибуция.
 - Компанията има достъп до пазари на труда.
 - Компанията има достъп до капиталови пазари.
 - Компанията има достъп до материали и доставчици.
- Финансиране
 - Основният източник, използван за първоначално финансиране на фирмата е.

3.4 Изводи и заключение

На база на литературния обзор на съвременното състояние на проблемната област, проведеното проучване и качественото изследване, включващо интервюта с български предприемачи, следва заключението, че разработката на нов и актуален модел за предсказване на успеха на стартиращи компании трябва да бъде **базирана на предишни изследвания** на утвърдени автори и по-конкретно на модела на **Сендбърг**, както и на специфичните условия на **средата**, описани от българските предприемачи.

В настоящата глава е представено предложение на **модел на процеса на създаване на нова компания**. Предложен е **абстрактен модел за предсказване успеха на стартиращи компании от България**, включващ основните категории и подкатегории на факторите за успех. Абстрактният модел е подробно разработен с отделни фактори за успех, формирайки **детайлизиран модел за предсказване на успеха**.

Предложените модели представляват **хипотези**, които са валидирани посредством провеждане на качествено и количествено изследване, подробно описани в Глава 4.

ГЛАВА 4: ИЗСЛЕДВАНИЯ, ПРИЛОЖЕНИЯ И ЕКСПЕРИМЕНТИ

В настоящата глава са описани проведените изследвания, приложения и експерименти. Проверени са представените в Глава 3 хипотези. Чрез факторен анализ е валидирано групирането на факторите в предложените модели за предсказване на успеха. Чрез статистически методи и методи за извличане на знания от данни са синтезирани различни модели за предсказване на успеха, които са анализирани и сравнени. Идентифицирани са най-удачните модели, които могат да бъдат използвани за създаване на софтуерен продукт за предсказване на успеха на стартиращи компании.

4.1 Качествено изследване на факторите за успех на стартиращи фирми

Проведено е качествено изследване за събиране на мнения, откриване на тенденции и по-добро разбиране на мотивацията на българските предприемачи. Използваната техника е провеждане на дълбочинни интервюта с малък брой непредставителни случаи, чрез задаване на въпроси спрямо предварително подготвено указание за интервю.

На интервюираните предприемачи предварително са представени абстрактният и детайлизираният модел за предсказване на успеха на стартиращи компании. Целта е техните отговори и коментари, свързани с модела да помогнат за неговото развитие, преди да се пристъпи към валидация с количествени методи и по този начин да се получат по-добри резултати при прогнозиране на успеха на стартиращи компании чрез използването на модела.

Интервюто беше проведено на база следното указание за интервю:

1. Коя е основата сфера на дейност на Вашия бизнес?
2. Каква е позицията Ви във фирмата?
3. От колко време се занимавате с Вашия бизнес?
4. Какво е мнението Ви за модела за предсказване на успеха на стартиращи фирми?
5. Смятате ли, че в модела има пропуски? Какви?
6. Смятате ли, че в модела има излишни критерии? Кои са те?
7. Смятате ли, че софтуер за предсказване на успеха на стартиращи и млади фирми би бил полезен за Вашия бизнес? С какво?
8. Какво смятате за отделните критерии от модела: разбираеми ли са, измерими ли са, бихте ли могли да ги оцените, бихте ли добавили нещо?
9. Как определяте коя стартираща компания е успешна? Как различавате успешната стартираща компания от неуспешната?

Респондентите са петима български предприемачи и собственици на млади компании в областите информационни технологии, медицина и земеделие. Компаниите са създадени преди от 1 до 5 години и всички от тях са действащи. Всички компании използват собствени средства на техните основатели за първоначално финансиране и са основани без помощта на бизнес инкубатори, фондове за рисков капитал, бизнес ангели или правителството.

Продължителността на всяко интервю е от 0:30 до 2:30 часа, в зависимост от отговорите и демонстрирания интерес от страна на интервюираните. Резултатите от интервютата са представени в обобщен вид в Таблица 8.

Таблица 8 Обобщени резултати от качествено изследване на мненията във връзка с предложения модел за предсказване на успеха

Въпрос №	Резултати от качествено изследване
1	Сферите на дейност на компаниите са информационни технологии, медицина и земеделие.
2	Всички интервюирани са собственици на бизнес.
3	Компаниите са създадени преди 1 до 5 години и всички от тях са действащи.
4	Всички интервюирани описват модела за предсказване на успеха на нови компании като логичен и правилен.
5	Респондентите не откриха пропуски в модела, но предложиха множество подобрения и допълнения към индивидуалните критерии.
6	Респондентите не откриха излишни данни в модела.
7	Всички интервюирани смятат, че софтуер за предсказване на успеха на стартиращи и млади компании би бил от полза. Само някои от тях са готови да заплатят за подобен софтуер, но всички биха го използвали при условие, че е безплатен. Респондентите биха използвали софтуера в помощ на техните стартиращи компании, първоначални проучвания, за оценка на компанията, за локализиране и подобрене на слаби места. Някои от респондентите биха се доверили на софтуера, но други се нуждаят от доказателства: информация за алгоритъма и логиката на работа, детайли и обяснения относно модела.
8	Респондентите описват повечето критерии като логични, ясни и измерими. Те са способни да разберат и отговорят на повечето въпроси от представената чернова на анкетата без трудности. Въпреки това, респондентите се чувстват несигурни спрямо множество въпроси, касаещи екипа, личността на предприемача и индустрията. Никой от тях не знае своя тип темперамент според теста на Майерс-Бригс. Всички интервюирани поискаха дефиниция на мениджмънт или маркетинг термини, които не успяха да разберат или разбраха погрешно. Те предложиха множество подобрения и допълнения към индивидуалните критерии.
9	Респондентите определиха успешните стартиращи компании като такива, които имат растеж и развитие. Според тях в началните етапи на развитие на компанията, когато се развива продукт или услуга и се сформира екип, не са важни толкова финансовите показатели, колкото постигнатия напредък.

Всички респонденти описват модела като логичен и не откриват значителни пропуски. И.М. – собственик на фирма за уеб дизайн каза „Моделът е логичен и няма

никакви значителни пропуски. Той се е подобрил чрез добавянето на допълнителни категории към традиционния модел”. В.В., собственик на компания за интернет търговия смята, че „Моделът покрива основните области”. Респондентите идентифицираха всички категории като важни за успеха на фирмата и не откриха излишни данни. Л.М., собственик на фирма за био земеделие смята, че „Няма нищо за махане”. На база на тези отговори беше взето решение да се запазят предложените категории в абстрактния модел.

На респондентите бяха показани отделните критерии, представящи подкатегориите в диаграмата на модела, както и чернова на анкетата. Те успяха да разберат и отговорят на повечето въпроси без затруднения. Въпреки това, се чувстваха несигурни за множество въпроси, касаещи екипа, личността на предприемача и индустрията. И.М. сподели, че за стартиращите компании екипът не е важен за успеха на фирмата, а само основният мениджмънт екип е от значение и предложи да се предостави дефиниция за предприемачески екип. Интервюираните не разполагаха с информация за техния тип темперамент (според теста на Майерс-Бригс и класификацията на Карл Юнг) (Jung, 1971), (Briggs, 1995). Някои от тях не разбираха или интерпретираха грешно термини от маркетинга и мениджмънта, като например „продуктово портфолио”. Те поискаха да **бъдат предоставени определения** на всички термини, за да могат да разберат правилно въпросите. Респондентите идентифицираха човешкия ресурс (предприемаческия екип), клиентите, уникалния продукт и конкурентите (а също и анализа на конкуренцията) като най-важните фактори за успех на младата компания. Те **предложиха множество подобрения и допълнения** към черновата на анкетата, които са взети предвид в детайлизирания модел и в количественото изследване.

На база предложенията на предприемачите за критериите в детайлизирания модел за предсказване на успеха, по него са **направени множество промени**. Всички термини, които затрудниха предприемачите, са преведени на български език и обяснени в анкетата, за да могат те да отговорят правилно на въпросите. Премахнат е въпроса относно темперамента според класификацията на Майерс-Бригс и е отхвърлена първоначалната идея за провеждане на тест за темперамент като част от анкетата. Вместо това, са предложени други въпроси, определящи психологическия профил на предприемача. Много от въпросите са преработени, за да звучат по-конкретно и да бъдат ясно измерими.

Всички респонденти **демонстрират интерес към изследването** и споделят, че **софтуер за предсказване на успеха на стартиращи и млади компании би бил полезен за тях**. Въпреки това, някои от тях остават резервирани относно практическата реализация, използваемостта и логиката. И.М., основател на фирма за уеб дизайн смята, че „Резултатите няма да бъдат приети на сериозно, докато не бъде предоставено логично обяснение и списък от слаби страни. Имам нужда от слабите места, за да направя подобрения в бизнеса”. П.П., собственик на фирма за дистрибуция на специализирана медицинска апаратура твърди следното: „Софтуерът би бил полезен. Ще фокусира вниманието на предприемача върху детайли, като екипа и стратегията”. П.М., собственик на компания за разработка на софтуер сподели: „Добре е да има такъв продукт на пазара и ще го използвам, ако е безплатен, но ако е платен, по-скоро бих провел собствено проучване”.

Интервюираните собственици на бизнес изразиха ясен интерес спрямо изследването и необходимост от софтуер за предсказване успеха на млади компании и за предлагане на възможни подобрения в бизнеса. Това е доказателство, че **българските предприемачи се нуждаят от подобен инструмент**, който да им помогне в развитието на техните компании. Собствениците на бизнес притежават положителна нагласа и са про-

активни, понеже виждат потенциални ползи в изследването. Те с удоволствие споделят опита си и желаят да помогнат за подобряването на бизнес климата в България.

На база на тези отговори, при проектирането и разработката на софтуера за предсказване на успеха на стартиращи компании, са предвидени редица подобрения. Предвидена е възможност за графично показване на модела, по който е направена прогнозата. Моделът за предсказване на успеха, получен чрез прилагането на количествени методи, е лесно разбираем и представлява класификационно дърво и по този начин предприемачите могат сами да анализират алгоритъма, върху който се гради прогнозата за техния бизнес. Върху класификационното дърво са маркирани конкретните фактори, които са довели до съответния прогнозен резултат и могат да се разгледат като възможности за подобрение, ако резултатът от прогнозата не е удовлетворяващ. Предвидено е софтуерът да даде процентна прогноза за вероятността за успех на компанията, а не само крайно решение.

Изводът от проведеното качествено изследване е, че предложеният модел за предсказване на успеха на нови компании е подходящ според интервюираните български предприемачи и се приема от тях като логичен и завършен. Те идентифицираха предложената категория „ресурси” като неразделна и важна част от модела. Предприемачите предложиха подобрения в модела, които са приложени както в него, така и в количественото изследване. Софтуерът за прогнозиране на успеха е съобразен с получените препоръки и изисквания.

4.2 Количествено изследване на факторите за успех на стартиращи фирми

Изследването е базирано на методологията за извличане на знания от данни CRISP-DM. Данните от анкетата са анализирани с помощта на софтуерните продукти IBM SPSS Modeler и Weka, които генерират модели за предсказване на успеха (напр. класификационно дърво), които дават практическо измерение на теоретичния модел за предсказване. Чрез IBM SPSS Statistics е изготвен факторен анализ, чрез който са валидирани групите фактори от предложената хипотеза за модел за предсказване на успеха на стартиращи компании.

С цел валидиране на модела за предсказване на успеха на стартиращи компании, е проведено количествено изследване с помощта на онлайн анкета, чрез която са събрани данни. Анкетата е изготвена, като за всеки от факторите от абстрактния модел за предсказване на успеха е подбран определен набор от подходящи въпроси, които най-добре характеризират този фактор. Отговорите на тези въпроси съставляват данните на количественото изследване. Обработването и анализът на данните от анкетата ще послужат за валидиране на модела за предсказване на успеха и за извличане на знания от данни – ключови фактори за успех и модели, базирани на данните от анкетата.

Имайки предвид обема на набора от данни от проучванията в други държави, както и факта, че предвидената анкета е подробна и отнема значително време за попълване (около 20 минути), **първоначално поставените цели** бяха следните:

- Извършване на проучване чрез анкета, съдържаща около 100 въпроса,
- Провеждане на анкетата в рамките на 6 месеца,
- Попълване на анкетата от поне 100 представители на целевата група – предприемачи, собственици на бизнес, мениджъри на стартиращи малки и средно големи компании.

Предвидената анкета подробно анализира множество фактори за успех на стартиращи компании. Въпросите са маркирани като незадължителни за попълване с цел

получаване на максимално много отговори и възможността респондентите да пропускат трудни или нежелани от тях въпроси. Времето за попълване на анкетата от един респондент е около 15-20 минути.

Целевата група (респондентите на анкетата) са собственици, управители и мениджъри на малки и средно големи компании, които са стартиращи или млади. Според предложените в Глава 2 дефиниции, за стартиращи и млади компании се приемат такива, които са **в началото на своето развитие и по размер са SME** (микро, малки или средно големи компании). За малки и средно големи компании приемаме тези с 0 до 250 наети лица (според дефиницията за Европейския Съюз). От важност за изследването би бил броят на компаниите, попадащи в целевата група, но за съжаление НСИ не предоставя такава извадка за България.

Ще бъде определен приблизително **броят на компаниите, попадащи в целевата група** от наличните данни от НСИ. Според демография на предприятията (НСИ, Национален Статистически Институт - Република България, 2013), публикувана от НСИ, през 2012 в България са действали 323 745 активни предприятия. За съжаление, към момента (2014 година) не са налични по-актуални данни. От всички действащи предприятия, само 54% (или около 174 500) имат поне едно наето лице. Този факт, възможността някои предприятия да не са активни, както и наличието на сива икономика, поражда съмнение за реалния брой действащи предприятия в България. 91% от предприятията имат до 10 наети лица, а само 0,1% имат над 250 наети лица. За съжаление не се съобщава броят предприятия, попадащи в целевата група. През 2012 година са били създадени 13% от общия брой предприятия, от които 81% оцеляват година по-късно. От тези данни очевидно не може да се определи точно броя на предприятията, попадащи в целевата група, а само може да се изчисли приблизително. Според отчетите от НСИ изчислих, че **броят предприятия, попадащи в целевата група в края на 2012 година, е между 106 000 и 193 000.**

Според направено от автора проучване, при подобни анкети за валидиране на модели за предсказване на успеха на стартиращи компании, броят респонденти варира между 33 и 334, като зависи от спецификите на проучването и размера на целевата група. Следва списък с подобни проучвания, проведени в различни държави, подреден в хронологичен ред:

- 46 венчър капиталисти от САЩ, (Tyebjee, 1984),
- 97 фирми за рисков капитал от САЩ, (Poindexter, 1975),
- 155 производствени предприятия (Chandler G. H., 1994),
- 225 фирми от САЩ, (Carland, The Theoretical Bases and Dimensionality of the Carland Entrepreneurship Index, 1996),
- 63 фирми от Китай, (Luo, 1999),
- 137 фирми от Корея, (Park, 2000),
- 56 фирми от Китай, (Peng M. L., 2000),
- 334 малки производствени предприятия от САЩ, (Ebben, 2005),
- 33 фирми от САЩ (Abbas, 2008),
- 162 фирми от Русия (Shirokova, 2010),

- 334 фирми от България – проучване на факторите социален капитал и конкурентни стратегии, (Manev, 2012).

Посредством анкета, за период от 12 месеца (през 2013 - 2014 година) бяха събрани **данни за 142 компании от България**, което е в рамките на поставените цели.

За събиране на данни за анкетата, проведена от автора, бяха използвани множество източници и канали:

- лични бизнес контакти,
- онлайн социални мрежи (Facebook, LinkedIn, Google+),
- бизнес и предприемачески общности (студенти, завършили обучение по технологично предприемачество, участници в състезания за учебни компании и други),
- имейл маркетинг до действащи български компании,
- личен уеб сайт <http://byankov.com/>,
- рекламна кампания в уеб сайтове чрез платформата Google AdWords.

Анкетата, проведена от автора, е интернет базирана, за да позволи максимална достъпност за респондентите. Реализирана е чрез инструмента Google Forms (Google Inc., 2014), който позволява създаване на анкети, събиране и съхранение на данни от тях и експортиране на данните във файлов формат .CSV. Анкетата се състои от 5 на брой интернет страници с около 100 въпроса за попълване от типовете избор на един от множество варианти, избор на няколко от множество варианти, текстово поле. Откъс от първата страница на анкетата е показан на Фиг. 15.

Анкета - Модел за предсказване на успеха на стартиращи фирми

Настоящото изследване е с научна цел във връзка с дисертация по "Технологично предприемачество" към Софийски Университет "Св. Климент Охридски" и има за цел да провери модел за предсказване на успеха на стартиращи компании в помощ на българския бизнес. Анкетата е предназначена за мениджъри, управители, предприемачи и собственици на фирми. Всички отговори са анонимни и няма да бъдат обработвани индивидуално, а само съвкупно.

Предприемачески екип

Личност и ценности

Моля, на всеки ред изберете в каква степен сте съгласен с всяко едно от следните твърдения

	Напълно несъгласен	По-скоро несъгласен	Нито съгласен, нито несъгласен	По- скоро съгласен	Напълно съгласен
Мога да се справям в работата си без нужда от съдействие.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Аз съм уверен в себе си.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Аз съм предприемчив, инициативен.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Това, което ми					

Фиг. 15 Част от анкетата за валидиране на модела за предсказване на успеха на стартиращи компании (създадена от автора)

За отговор на повечето въпроси от анкетата е избрана стандартна ликертова скала (Likert, 1932) с 5 нива. Въпросите са формулирани като твърдения, а възможните отговори са степените на съгласие: „Напълно несъгласен“, „По-скоро несъгласен“, „Нито съгласен, нито несъгласен“, „По-скоро съгласен“ и „Напълно съгласен“.

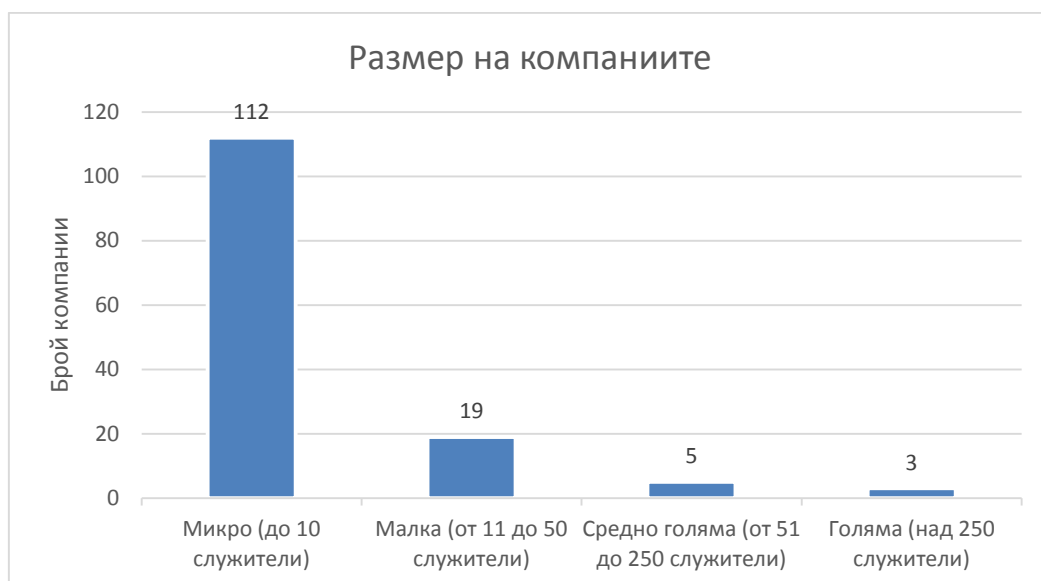
Ликертовата скала трябва да бъде симетрична и е предпочетен вариантът с 5 нива, тъй като вариантът с 3 нива е неточен (не изразява степените на съгласие), а вариантът със 7 нива е по-сложен за респондентите и може да предизвика затруднение за тях. Освен това, ликертовата скала с 5 нива е най-популярна.

4.2.1 Анализ на данните от количественото изследване

За да се определят типовете и профилите на компаниите, попълнили анкетата, ще бъдат разгледани и анализирани събраните данни.

4.2.1.1 Размер на компаниите

Преобладаваща част от компаниите (136) са микро, малки и средни предприятия (SME), а само 3 са големи компании (Фиг. 16). Това е очаквано разпределение на компаниите по размер, като се имат предвид разгледаните данни от НСИ. Като профил, преобладават стартиращите микро компании. Общият брой е по-малък от 142, тъй като част от респондентите не са посочили размер на компанията.



Фиг. 16 Размер на компаниите (създадена от автора)

4.2.1.2 Възраст на компаниите

От компаниите, участващи в анкетата, преобладават тези на възраст между 1 и 5 години (общо 62), но са представени и компании на възраст над 5 години (41), както и компании, които сега започват своята дейност (33). Общият брой е по-малък от 142, тъй като част от респондентите не са посочили възраст на компанията (Фиг. 17).

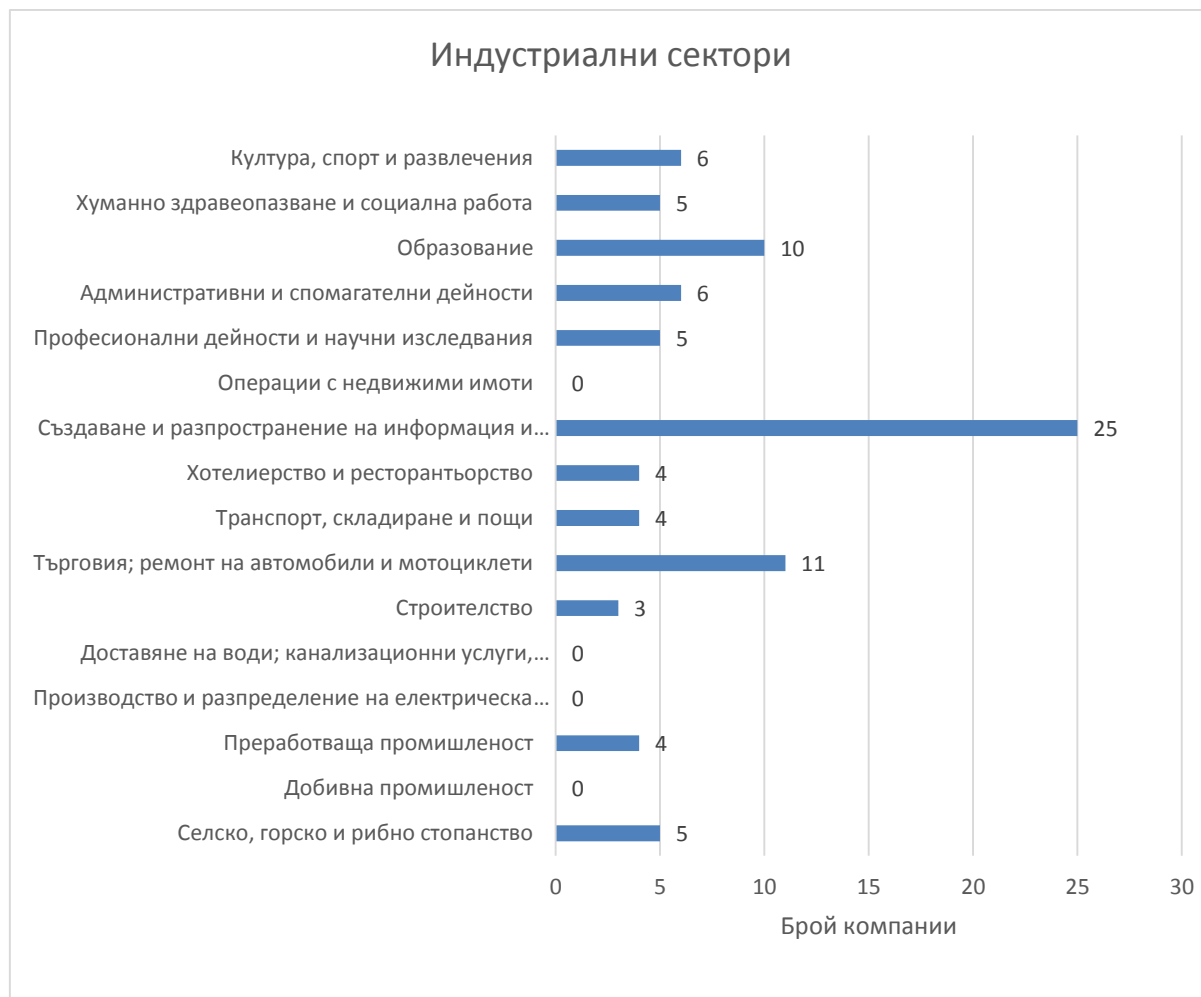


Фиг. 17 Възраст на компаниите (създадена от автора)

4.2.1.3 Индустиален сектор на компаниите

Индустиалните сектори на компаниите бяха проверени според класификацията на НСИ (Национален Статистически Институт - Република България, 2014), за да могат при нужда данните да бъдат сравнявани. От представените компании, най-голяма част

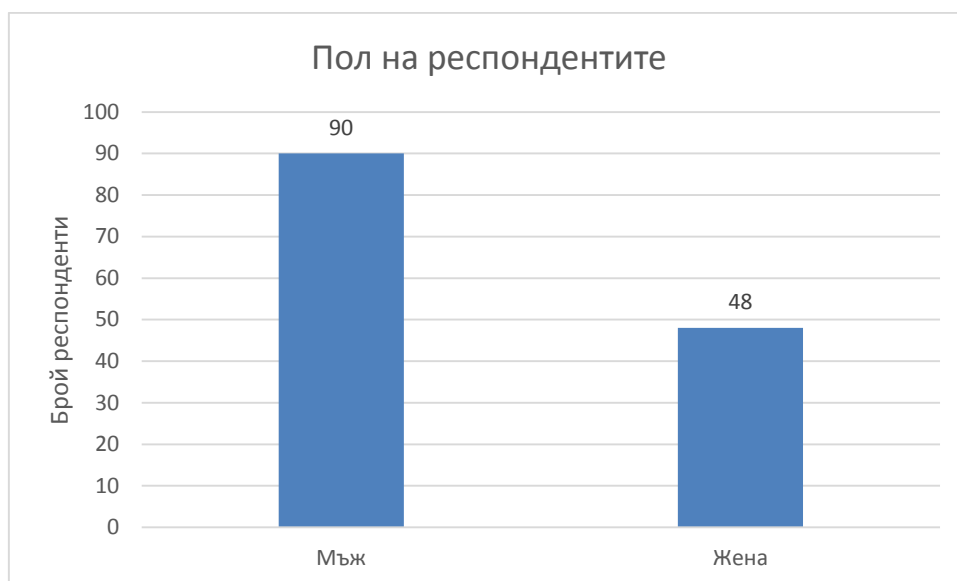
работят в сферите „Създаване и разпространение на информация и творчески продукти; далекосъобщения“ (25 компании), „Търговия; ремонт на автомобили и мотоциклети“ (11 компании), „Образование“ (10 компании). Точното разпределение на индустриалните сектори е показано на Фиг. 18.



Фиг. 18 Индустриален сектор на компаниите (създадена от автора)

4.2.1.4 Пол на респондентите

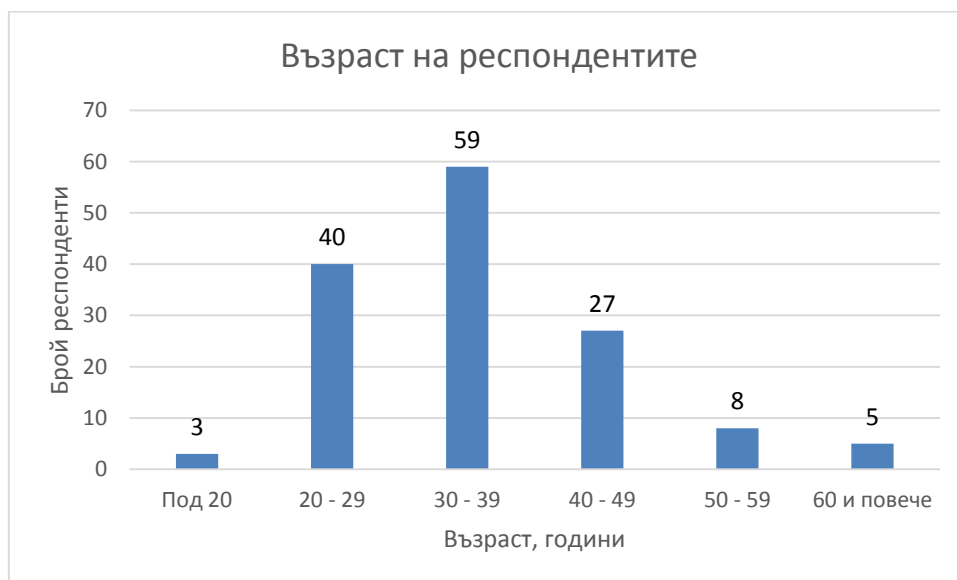
В данните са включени попълвания на анкетата от предприемачи и от двата пола, като преобладават мъжете (90) спрямо жените (48) (Фиг. 19).



Фиг. 19 Пол на респондентите (създадена от автора)

4.2.1.5 Възраст на респондентите

В набора от данни преобладават предприемачите на възраст 30-39, 20-29 и 40-49 години, но са представени предприемачи от всички възрастови групи (Фиг. 20). Разделението на възрастови групи е организирано по същия начин, като извадките на НСИ, за да бъдат възможни сравнения.



Фиг. 20 Възраст на респондентите (създадена от автора)

4.2.2 Подготовка на данните от количественото изследване

Събраните чрез анкетата данни за стартиращи български компании включват набор валидни попълвания от 142 компании. При получаването им от попълнените анкети, те се запазват в електронен файл, съхранен в услугата Google диск. Данните в получения

вид не са годни за анализ и се нуждаят от предварителна подготовка. За да бъдат разбираеми за софтуер за извличане на знания от данни и статистически анализ, те трябва да бъдат подготвени по специфичен начин. На полетата в базата данни трябва да бъде зададен определен тип (номинални, ординални, непрекъснати) и данните в тях трябва да бъдат преобразувани в числови стойности.

От Google диск има възможност за изнасяне (експорт) на данните във файлов формат CSV, който съдържа таблични данни. CSV форматът е текстов и съдържа по един ред с данни за всеки ред от таблицата (отговарящ на едно попълване на анкетата). В рамките на всеки ред, данните в отделните колони са разделени със специален символ. Първият ред във файла съдържа заглавията на колоните. Данните от анкетата, прегледани в Google диск, изглеждат по начина, показан на Фиг. 21.

	A	B	C	D	E	F	G
1	Timestamp	Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]	Личност и ценности [Аз съм уверен в себе си.]	Личност и ценности [Аз съм предприемчив, инициативен.]	Личност и ценности [Това, което ми се случва в живота зависи от мен самия.]	Личност и ценности [Аз съм амбициозен.]	Личност и ценности [Склонен съм да поемам рискове.]
2	6/23/2013 9:05:51	По-скоро съгласен	По-скоро съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	По-скоро съгласен
3	6/27/2013 6:16:47	По-скоро съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен
4	6/27/2013 23:10:36	По-скоро съгласен	Нито съгласен, нито несъгласен	По-скоро съгласен	Напълно съгласен	По-скоро съгласен	По-скоро съгласен
5	6/28/2013 11:06:40	Напълно съгласен	Напълно съгласен	По-скоро съгласен	По-скоро съгласен	По-скоро съгласен	По-скоро съгласен

Фиг. 21 Извадка от данните от анкетата, прегледани в Google диск (създадена от автора)

4.2.2.1 Подготовка на данните от анкетата в Excel

Изнесенният файл с данни от анкетата е във формат CSV и може да се разчете от софтуерния продукт Microsoft Excel (Microsoft Excel - Софтуер за електронни таблици, 2014), който е използван за основната част на подготовката на данните. Посредством формули в Excel, данните са обработени, така че отделните стойности да имат числов вид, като всяко число има определено значение. Преобладаваща част от въпросите са в скала на съгласие с 5 стойности от „Напълно съгласен“ до „Напълно несъгласен“, които са представени по следния начин в числов вид:

1 => Напълно несъгласен

2 => По-скоро несъгласен

3 => *Нито съгласен, нито несъгласен*

4 => *По-скоро съгласен*

5 => *Напълно съгласен*

Тъй като в анкетата отговорите са представени във вид на текст, преобразуването към числови данни се извършва с помощта на следната Excel формула (Форм. 4):

```
=SUBSTITUTE((4)  
SUBSTITUTE(  
SUBSTITUTE(  
SUBSTITUTE(  
SUBSTITUTE(  
"Form Responses"!B2;  
"Напълно съгласен";  
"5");  
"По-скоро съгласен";  
"4");  
"Нито съгласен, нито несъгласен";  
"3");  
"По-скоро несъгласен";  
"2");  
"Напълно несъгласен";  
"1")
```

В анкетата се съдържат и други типове въпроси, за чието преобразуване се използват други формули – общо 49 различни формули, които се прилагат към около 100 променливи в набора от данни.

В данните присъства и малък брой въпроси, чиито отговори съдържат свободен текст (отворени въпроси). Това се налага, за да се проучи точното мнение на предприемачите по важни въпроси, без да се поставят предварителни ограничения. Пример за такъв въпрос е „Смятам, че следните фактори са ключови за успеха за моя бизнес:“, при който предприемачите имат свободата да изброят важните за тях фактори.

Подготовката на тези данни е по-трудна, тъй като трябва всички отговори да се прочетат и анализират, след което да се групират по подходящ начин. Тъй като един респондент може да посочи нито един, един или повече от един фактори, всяка от групите фактори се представя като отделна променлива в обработените данни.

Наличието на този фактор за съответния респондент се отбелязва с една стойност на променливата - в случая с „1“, а липсата на този фактор – с „2“. Стойността нула е избегната като вариант, тъй като някои софтуерни продукти за извличане на знания от данни и техните алгоритми биха могли да я интерпретират като липсващи данни. За да се автоматизира поне частично обработката на данните, както и обработката на бъдещи

попълвания на анкетата, всяка група фактори е представена с ключови думи и фрази, за чието наличие в данните се проверява чрез формула. Например за фактора екип се използва Формула 5:

$$=IF(\tag{5}$$

$$AND($$

$$ISERROR(SEARCH("екип";'Form Responses'!BI2));$$

$$ISERROR(SEARCH("хора";'Form Responses'!BI2));$$

$$ISERROR(SEARCH("служител";'Form Responses'!BI2));$$

$$ISERROR(SEARCH("човешки";'Form Responses'!BI2));$$

$$ISERROR(SEARCH("персонал";'Form Responses'!BI2));$$

$$ISERROR(SEARCH("професионалист";'Form Responses'!BI2))$$

$$);$$

$$"2";$$

$$"1"$$

$$)$$

Анкетата съдържа и два въпроса, по които се определя дали дадена стартираща компания е успешна или не. Формулировката на тези въпроси се базира на възприетата в настоящата дисертация дефиниция за успешна стартираща компания. Това са въпросите дали фирмата продължава дейността си в настоящия момент и дали е регистрирала растеж през последните 5 години. Отговорите им разделят фирмите на три категории: „успешни“, „нито успешни, нито неуспешни“ и „неуспешни“. Всички останали въпроси и отговори в анкетата описват характеристиките на дадена компания, чрез независими променливи.

След обработка на данните в Excel редовете на таблицата, съответстващи на попълвания на анкетата, съдържат само числа, съответстващи на отговори на въпросите (Фиг. 22).

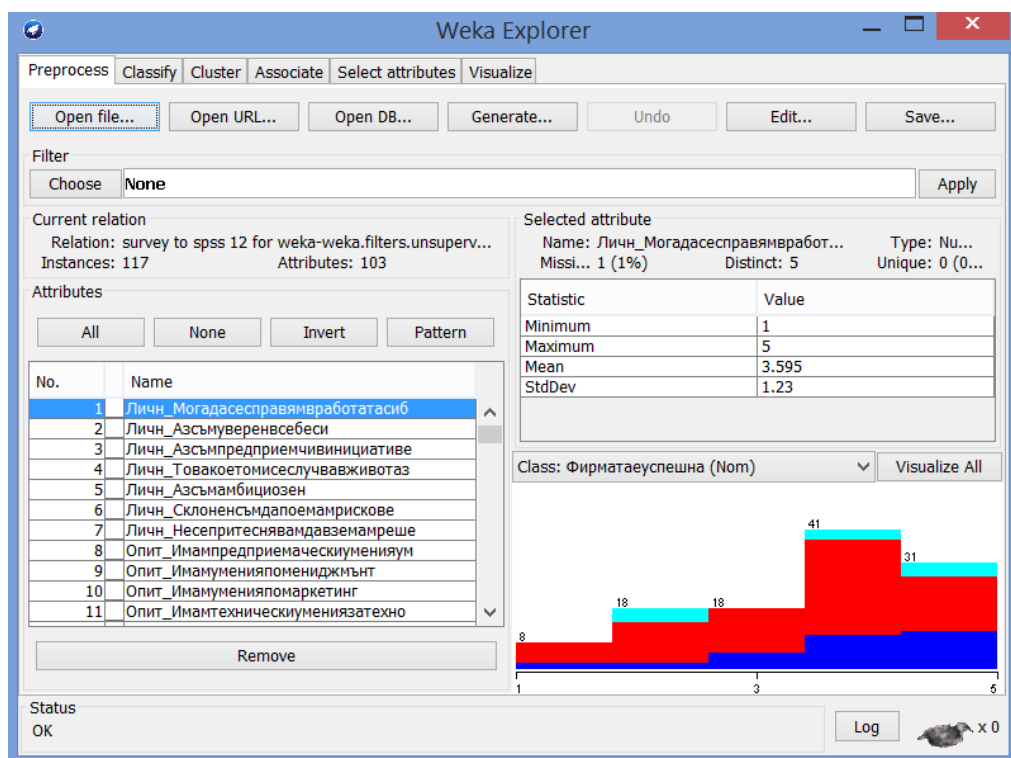
потребителя. На всяка променлива се задава и роля: “input” – променливи, използвани, като входни данни за анализа или “target” – целеви променливи, чиито стойности анализираме. Изгледът на колоните с променливи е показан на Фиг. 24.

	Name	Type	Width	Deci...	Label	Values	Missing	Columns	Align	Measure	Role
91	НМАktiv_материалиидо...	Numeric	1	0	Нематериални акти...	{1, Напълно н...	0	1	≡ R...	Ordinal	Input
92	Съссобственисредства...	Numeric	1	0	Със собствени сре...	{1, Да}...	None	1	≡ R...	Nominal	Input
93	Средстваотсемеystoп...	Numeric	1	0	Средства от семей...	{1, Да}...	None	1	≡ R...	Nominal	Input
94	Стипендияфондзамалък...	Numeric	1	0	Стипендия / фонд з...	{1, Да}...	None	1	≡ R...	Nominal	Input
95	ЗаемкредитОсновниятиз...	Numeric	1	0	Заем / кредит - Осн...	{1, Да}...	None	1	≡ R...	Nominal	Input
96	БизнесинкубаторОсновн...	Numeric	1	0	Бизнес инкубатор - ...	{1, Да}...	None	1	≡ R...	Nominal	Input
97	Индивидуалениинвестит...	Numeric	1	0	Индивидуален инве...	{1, Да}...	None	1	≡ R...	Nominal	Input
98	Инвестиционенфондзар...	Numeric	1	0	Инвестиционен фо...	{1, Да}...	None	1	≡ R...	Nominal	Input
99	БартерОсновниятизточн...	Numeric	1	0	Бартер - Основният...	{1, Да}...	None	1	≡ R...	Nominal	Input
100	ПартньорствоОсновният...	Numeric	1	0	Партньорство - Осн...	{1, Да}...	None	1	≡ R...	Nominal	Input
101	Работасединоснoвенкли...	Numeric	1	0	Индивидуален инве...	{1, Да}...	None	1	≡ R...	Nominal	Input
102	КакваепозициятаВивъв...	Numeric	1	0	Каква е позицията ...	{1, Служител...	None	1	≡ R...	Ordinal	Input
103	КакъвeВашиятпол	Numeric	1	0	Какъв е Вашият пол?	{1, Мъж}...	None	1	≡ R...	Nominal	Input
104	Каквавъзрасттанакомпа...	Numeric	1	0	Каква възрастта н...	{1, Стартира...	None	1	≡ R...	Ordinal	Input
105	Какъвесегашниятразме...	Numeric	1	0	Какъв е сегашният ...	{1, Микро (до ...	None	1	≡ R...	Ordinal	Input
106	Фирматапродължаваде...	Numeric	1	0	Фирмата продължа...	{1, Да}...	None	5	≡ R...	Nominal	Target
107	Фирматаеувеличилама...	Numeric	1	0	Фирмата е увеличи...	{1, Да}...	None	5	≡ R...	Nominal	Target
108	Фирматаеуспешна	Numeric	1	0	Фирмата е успешна	{1, Не}...	0	9	≡ R...	Ordinal	Target

Фиг. 24 Извадка на променливи, към които се отнасят данните от анкетата, в IBM SPSS Statistics (създадена от автора)

4.2.2.3 Подготовка на данните от анкетата в Weka

За да бъдат прочетени данните от софтуерния продукт Weka (Bouckaert, 2013), те трябва да бъдат преобразувани във файлов формат „arff“. Този формат съдържа данните в числов вид, разделени със запетаи („.csv“), което е директно поддържан формат в Microsoft Excel. Освен това, „arff“ файлът съдържа мета данни, описващи променливите – техните имена, типове и допустими стойности. На Фиг. 25 са показани данните от анкетата, заредени в софтуерния продукт „Weka“.



Фиг. 25 Извадка от данни, заредени в софтуерния продукт Weka (създадена от автора)

4.2.3 Групиране на факторите за успех

Факторният анализ (Харалампиев, 2012) е статистически метод, създаден от Чарлс Спирмън и служещ за открояване на факторите, които са общи за множество от променливи, между които съществуват корелации. Факторният анализ ще бъде приложен върху променливите, които представляват факторите за успех на стартиращите компании, за да се получат групи от фактори. Софтуерният продукт, който ще се използва за анализа е IBM SPSS Statistics.

Като първа стъпка от анализа, част от променливите бяха изключени, тъй като методът работи само с ординални индикатори. Липсващите стойности, са малък брой, но се срещат в почти всички индикатори, по тази причина беше избран методът на замяна на липсващите стойности със средни стойности, за да може да се изпълни факторният анализ. Беше проверена извадковата адекватност за всички фактори, както и за всеки фактор по отделно. Извадковата адекватност беше неприемлива, поради наличието на голям брой фактори, някои от които с лоша извадкова адекватност. Тези фактори бяха изключени при следващата итерация на анализа. Всички изключени от анализа фактори са изброени в списък в Приложение 2.

При следващата итерация на анализа, извадковата адекватност според теста на Кайзер-Майер-Олкин (Kaiser, 1974) е приемлива (Таблица 9).

Таблица 9 Тест на Кайзер-Майер-Олкин и Барлет

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.697
Bartlett's Test of Sphericity	Approx. Chi-Square	3154.260
	Df	1431
	Sig.	.000

След настройка на 11 групи и прилагане на ротация върху получената корелационна матрица, се получава разпределението на факторите по групи (Таблица 10). Факторите се намират в лявата колона на таблицата, а групите са отбелязани като компоненти (Components) и номерирани от 1 до 11. Числата, които свързват факторите с компонентите показват в каква степен факторът се отнася до съответната група, като по-големите числа означават по-силна корелация. Корелациите със стойност под 0,5 не са показани в таблицата.

Единадесетте компонента на корелационната матрица са генерирани от IBM SPSS Statistics и представляват групи, които не са озаглавени, тъй като софтуерът не индикира автоматично техния бизнес смисъл, а единствено групира представените фактори. На база включените фактори в компонентите, ще бъде анализирано и тяхното значение.

Таблица 10 Корелационна матрица, след прилагане на ротация

Rotated Component Matrix ^a											
	Component										
	1	2	3	4	5	6	7	8	9	10	11
Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]											
Личност и ценности [Аз съм уверен в себе си.]		0.754									
Личност и ценности [Аз съм предприемчив, инициативен.]		0.761									
Личност и ценности [Аз съм амбициозен.]		0.683									
Личност и ценности [Склонен съм да поемам рискове.]		0.643									
Умения и опит [Имам предприемачески умения (умения в стартирането на бизнес).]											
Умения и опит [Имам умения по мениджмънт.]				0.775							
Умения и опит [Имам умения по маркетинг.]				0.784							
Умения и опит [Имам умения по човешки ресурси (HR).]				0.777							
Умения и опит [Имам умения по инвестиране.]				0.546							
Умения и опит [Имам предишен опит в стартирането на бизнес.]					0.645						

Умения и опит [Имам опит на подобна позиция (спрямо заеманата в момента).]					0.782						
Умения и опит [Имам опит в същата сфера на индустрията.]					0.748						
Умения и опит [Имам управленски опит.]					0.695						
Работа в екип [Фирмата разполага със завършен екип.]			0.743								
Работа в екип [Екипът притежава необходимите знания.]			0.883								
Работа в екип [Екипът притежава необходимите умения.]			0.886								
Работа в екип [Екипът притежава необходимата положителна нагласа.]			0.775								
Планиране [Планирам, обхващайки широка област.]	0.635										
Планиране [Планирам задълбочено.]	0.768										
Планиране [Правя формален (писмен) стратегически план.]	0.653										
Планиране [Планирам често.]	0.633										
Планиране [Планирам по функционални области / отдели.]	0.659										
Планиране [При планирането разглеждам множество алтернативи.]	0.659										
Цели [Целите, които поставям пред бизнеса си, са амбициозни.]											
Целевият пазарен дял на фирмата е:											
Стратегическа посока [Успявам да се придържам към първоначално избраната стратегия.]											
Стратегическа посока [Стратегията, която съм избрал, е агресивна.]											
Стратегическа посока [Набелязаната стратегия е подробна.]	0.704										
Стратегическа посока [Имам ясна стратегия за бизнеса си.]	0.653										
Бизнесът ми навлиза на пазара на следния етап:										- 0.515	
Конкурентно предимство [Моят бизнес притежава ясно конкурентно предимство.]											
Стратегически партньорства [Фирмата участва в партньорства с конкуренти.]											
Стратегически партньорства [Фирмата участва в партньорства с клиенти.]						0.689					

Стратегически партньорства [Фирмата участва в партньорства с доставчици.]						0.728					
Стратегически партньорства [Фирмата участва в партньорства с други заинтересовани страни.]						0.699					
Характеристики на индустрията, в която фирмата оперира [Повечето компани в индустрията имат добра печалба.]									0.684		
Характеристики на индустрията, в която фирмата оперира [Индустрията, в която фирмата оперира, генерира голяма част от brutния вътрешен продукт (БВП).]									0.609		
Характеристики на индустрията, в която фирмата оперира [В индустрията има стабилно търсене.]									0.634		
Характеристики на индустрията, в която фирмата оперира [Индустрията е в период на растеж.]									0.704		
Конкурентност на индустрията [Конкурентите в индустрията са агресивни.]											0.821
Конкурентност на индустрията [Компанията е изложена на конкурентни атаки.]											0.7
Материални активи [Компанията притежава складови наличности.]								0.77			
Материални активи [Компанията притежава недвижими имоти.]								0.685			
Материални активи [Компанията притежава пари в брой.]								0.504			
Нематериални активи [Компанията притежава разпознаваем бранд.]						0.736					
Нематериални активи [Компанията притежава софтуерни системи за управление на бизнес процесите (ERP, CRM).]						0.647					
Нематериални активи [Компанията притежава изградена бизнес репутация.]						0.619					
Нематериални активи [Компанията притежава лицензи.]						0.636					
Нематериални активи [Компанията притежава изградени бизнес контакти (социална мрежа).]											

Нематериални активи [Компанията има достъп до канали за дистрибуция.]											
Нематериални активи [Компанията има достъп до пазари на труда.]										0.536	
Нематериални активи [Компанията има достъп до капиталови пазари.]											
Нематериални активи [Компанията има достъп до материали и доставчици.]								0.539			
Extraction Method: Principal Component Analysis.											
Rotation Method: Varimax with Kaiser Normalization.											
a. Rotation converged in 10 iterations.											

Получените компоненти от факторния анализ в голяма степен съответстват на групите фактори от предложения абстрактен модел за предсказване на успеха (Глава 3). В първата колона на Таблица 11 са изброени категориите, а във втората колона са изброени подкатегиите от теоретичния модел за предсказване на успеха. В третата колона са изброени всички групи, получени чрез факторния анализ, които имат общи фактори със съответната подгрупа от теоретичния модел. Както се вижда от таблицата, има ясно съответствие и съвпадение за почти всички от подкатегиите от теоретичния модел с групи от факторния анализ. Ясно се вижда разграничението на отделните подкатегории и това, че те съответстват на различни групи от анализа. При разглеждане на главните категории от теоретичния модел, се вижда, че на всяка от тях съответстват по няколко групи от факторния анализ.

Таблица 11 Сравнение на категориите от теоретичния модел за предсказване на успеха и получените категории от факторния анализ

Категория фактори от теоретичния модел	Подкатегория фактори от теоретичния модел	Съответстващи групи фактори от факторния анализ
Предприемачески екип	Личност и ценности	Компонент 2
	Умения и опит	Компонент 4 Компонент 5
	Работа в екип	Компонент 3
Бизнес стратегия	Планиране	Компонент 1
	Цели	-
	Стратегическа посока	Компонент 1
	Стратегия за навлизане на пазара	Компонент 10
	Конкурентно предимство	-
	Стратегически партньорства	Компонент 6
Структура на индустрията	Характеристики на индустрията, в която фирмата оперира	Компонент 9

	Конкурентност на индустрията	Компонент 11
	Характеристики на клиентите	-
Ресурси	Материални активи	Компонент 8
	Нематериални активи	Компонент 7
		Компонент 8
		Компонент 10
	Финансиране	-

Подкатегориите от теоретичния модел, за които няма съответствие с групи от факторния анализ са:

- Цели
- Конкурентно предимство
- Характеристики на клиентите
- Финансиране

Причините за липсата на съответствие за изброените подкатегории могат да се търсят в няколко хипотези:

- фактът, че от анализа бяха изключени част от променливите, които може да съдържат важна информация,
- възможността при анализа да се създадат повече на брой групи, които е възможно да съответстват на подкатегориите,
- факторите от подкатегориите, всъщност може да имат корелация с някои от вече създадените групи, но тя да е по-слаба и затова да не е посочена в корелационната матрица,
- неправилна хипотеза за групиране на факторите от теоретичния модел в подкатегории, за които не съответстват групи от анализа.

Съпоставянето на единадесетте компоненти на корелационната матрица, генерирани от IBM SPSS Statistics, със съответстващите им променливи, позволява да се определи тяхното значение (Таблица 12). Компонентите в горната част на таблицата (с по-малки номера) представляват категориите, които са по-ясно разграничени. Може да се направи извода, че има разграничение между експертните умения (Компонент 4) и уменията по предприемачество и ръководене на бизнес (Компонент 5). Като цяло смисълът на компонентите от корелационната матрица съвпада с хипотезата за абстрактен модел за предсказване на успеха (Глава 3).

Таблица 12 Значение на компонентите от корелационната матрица

Компоненти от корелационната матрица	Значение
Компонент 1	Планиране

Компонент 2	Личност и ценности
Компонент 3	Работа в екип
Компонент 4	Умения и опит – експертни
Компонент 5	Умения и опит – предприемачество и управление на бизнес
Компонент 6	Стратегически партньорства
Компонент 7	Нематериални активи
Компонент 8	Материални активи и достъп до материали
Компонент 9	Характеристики на индустрията, в която фирмата оперира
Компонент 10	Стратегия за навлизане на пазара
Компонент 11	Конкурентност на индустрията

На база на изготвения факторен анализ и сравнението на получените резултати с теоретичния модел за предсказване на успеха от Глава 3, може да се заключи, че предложеният модел е адекватен и в голяма степен отговаря на резултатите, получени от анализа. Пълните изходни данни от анализа са представени в Приложение 2.

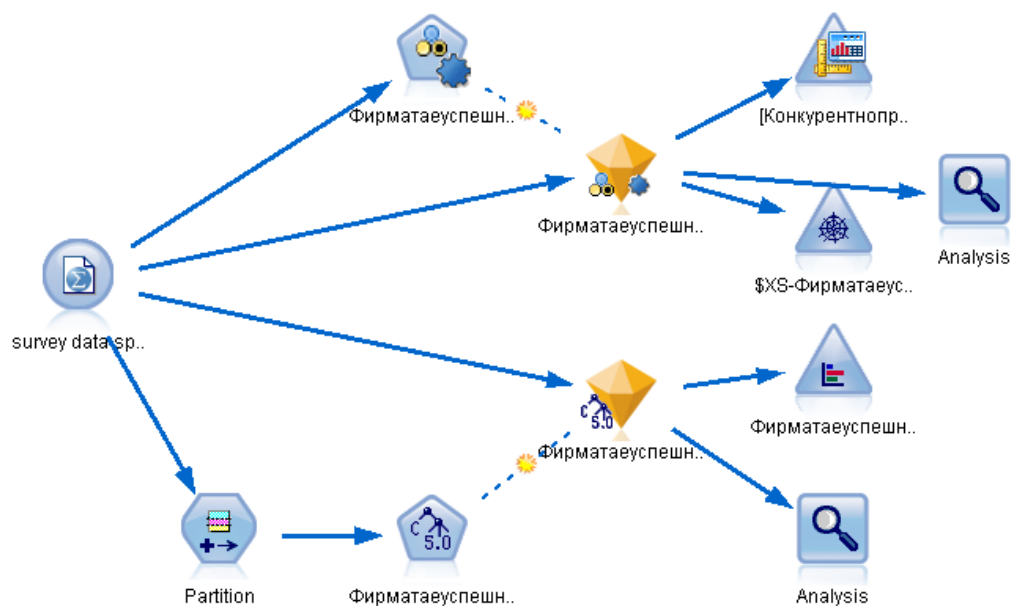
4.2.4 Извличане на знания от данни и създаване на модел за предсказване на успеха с IBM SPSS Modeler

В тази секция е разгледан процесът на извличане на знания от данни и създаване на модели за предсказване на успеха на стартиращи компании със софтуерния продукт IBM SPSS Modeler (IBM Corporation, 2012). Създадените модели са анализирани и сравнени от гледна точка на техните принципи на работа и ефективност. Разгледани са софтуерните алгоритми, чрез които са получени най-ефективните модели.

4.2.4.1 Синтезиране на модели за предсказване на успеха

С помощта на софтуерния продукт IBM SPSS Modeler са създадени модели за предсказване на успеха, които извършват класификация на компаниите - предвиждат в коя категория попадат те: „успешни“, „нито успешни, нито неуспешни“ или „неуспешни“. Софтуерният продукт предлага инструмент, който автоматично генерира множество модели, които решават поставената задача, сравнява ги и предлага най-добрите от тях.

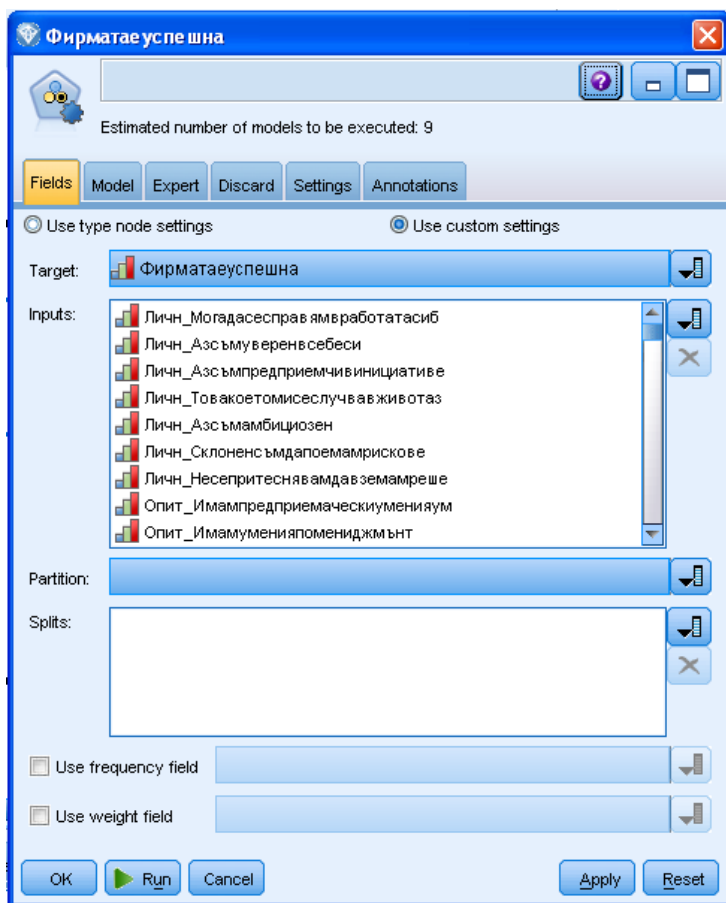
В IBM SPSS Modeler последователностите от действия, които се изпълняват от софтуера могат да бъдат представени графично (Фиг. 26). Последователността в горната част на фигурата показва получаване на модели чрез автоматичния класификатор, а последователността в долната ѝ част – получаване на модел чрез алгоритъма C5.0 с ръчни настройки и прилагане на кръстосана валидация.



Фиг. 26 Последователност от действия в IBM SPSS Modeler за получаване на класификационни модели – графично представяне (създадена от автора)

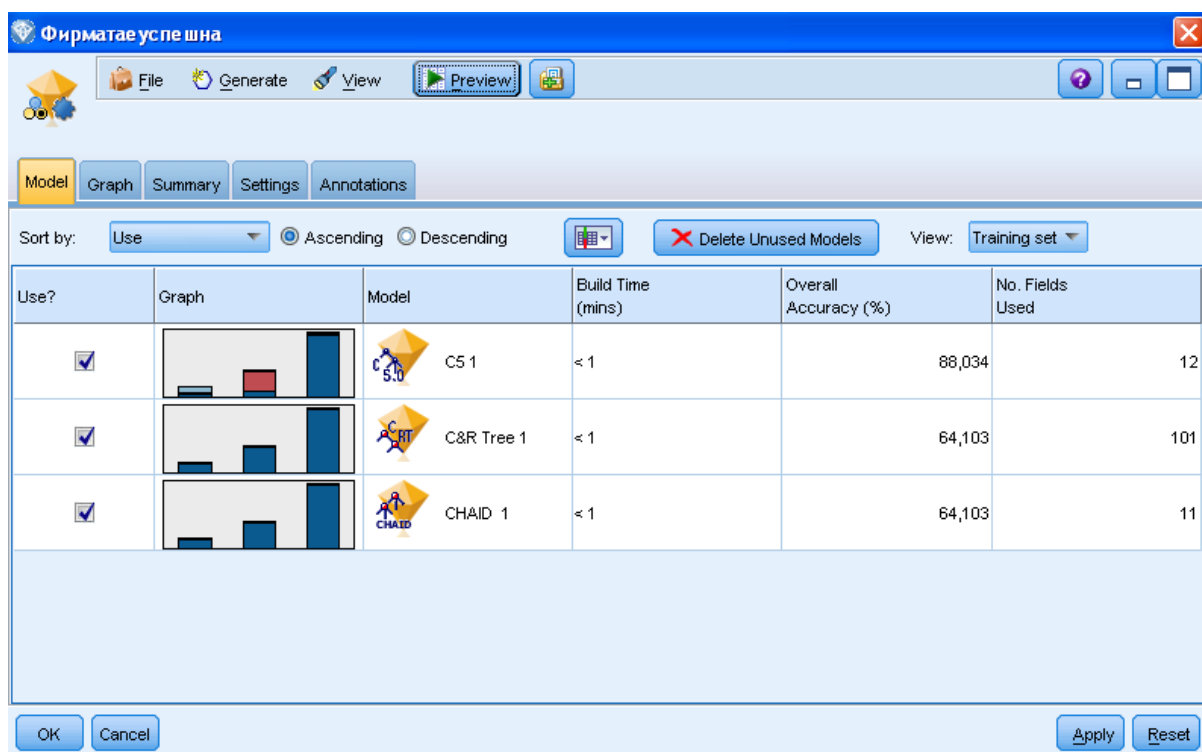
Данните от анкетите се съхраняват във файл от IBM SPSS Statistics (файлов формат .sav) и се зареждат в IBM SPSS Modeler чрез звеното „Statistics File“, разположено в лявата част на фигурата и маркирано с име „survey data spss“. Звеното за автоматична класификация „Auto Classifier“ (с форма на петогълник в горната част на фигурата) синтезира, оценява и сравнява класификационни модели за избраната цел, използвайки различни алгоритми. Получените модели за предсказване на успеха могат да бъдат разгледани в звено с форма на диамант (model nugget), разположено в горната част на фигурата.

Звеното за автоматична класификация се настройва, като се избират цел и входни данни (Фиг. 27). При синтезирането за модели за предсказване на успеха, за цел избирам променливата успех на компанията, показваща в каква степен компанията е успешна. Входните данни са останалите независими променливи, чиито типове и данни са предварително подготвени в Excel и IBM SPSS Statistics.



Фиг. 27 Избор на полета за автоматичния класификатор – входни данни и цели (създадена от автора)

Автоматичният класификатор синтезира и сравнява получените модели и запазва най-добрите три модела за предсказване на успеха (тези с най-висока точност) за по-нататъшен анализ (Фиг. 28).



Фиг. 28 Най-добри получени модели за предсказване на успеха при автоматична класификация (създадена от автора)

Получените модели са съпоставени по своята точност (overall accuracy), като в случая не се прилага кръстосана валидация и посочената точност е по-скоро максимална, отколкото реалистична. Най-точен е моделът, получен чрез алгоритъма C5.1 – с точност 88,03%. Останалите модели са получени чрез алгоритмите C&R Tree 1 (дърво за класификация и регресия) и CHAID 1, като точността и на двата е 64,10%. Класификационните дървета започват с най-важните фактори, предсказващи успеха и разделят случаите в групи (представени от възли) в зависимост от отговорите. Процесът продължава до достигането на краен възел (листо), който показва предсказаната стойност на целта – успеха на компанията.

Моделът, получен чрез алгоритъма C5.1 използва 12 променливи от входните данни, за да предскаже успеха на компанията. Моделът, получен чрез алгоритъма C&R Tree 1 използва 101 променливи, а полученият чрез CHAID 1 използва 11 променливи. При еднаква точност се предпочитат модели, използващи по-малък брой променливи.

При първоначалното синтезиране на модели чрез IBM SPSS Modeler и автоматичния класификатор, не е активиран методът на кръстосаната валидация. Този метод е от значение за прогнозната точност на модела и без неговото прилагане, прогнозите за точност на моделите са завишени.

Базирайки се на получените модели, избираме алгоритъма C5.1, тъй като чрез него е получен модел с най-висока точност. Алгоритъмът ще бъде използван като отправна точка за синтезиране на модел с висока точност, чрез прилагане на кръстосана валидация. По този начин ще бъде създаден модел, чиято точност ще бъде оценена по-реалистично.

4.2.4.2 Синтезиране на модел с кръстосана валидация чрез алгоритъма C5.0

При създаване на модели за предсказване на успеха, може да се приложи методът кръстосана валидация. При кръстосаната валидация данните се разделят на два набора: набор за обучение (training set), служещ за създаване на модела и набор за тестване (testing set), служещ за проверка на модела. Предимството на този метод е, че оценката за точността на получения модел е значително по-реалистична. Недостатък на метода е, че когато разполагаме с ограничен обем от данни, се налага да заделим част от тях за валидация на модела. В следствие на това по-малък набор от данни се използва за синтезиране на модела, което води до по-малко точен модел.

Прилагам алгоритъма C5.0, който е имплементация на изчисления в предната точка алгоритъм C5.1. C5.0 поддържа множество ръчни настройки и кръстосана валидация на данните. След настройване на алгоритъма за да използва кръстосана валидация и след прилагането му върху данните, се получава модел с точност 83,76% (Фиг. 29).

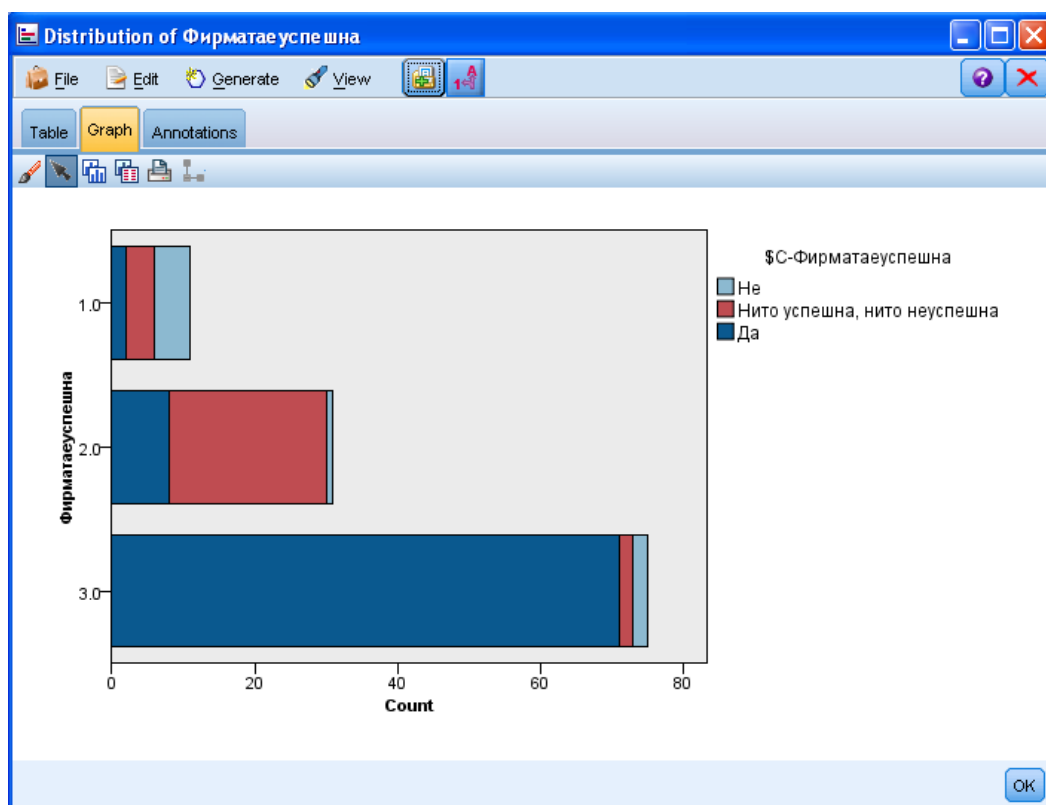
Results for output field Фирматаеуспешна

Comparing \$C-Фирматаеуспешна with Фирматаеуспешна

Correct	98	83,76%
Wrong	19	16,24%
Total	117	

Фиг. 29 Оценка на точността на модел, получен чрез алгоритъма C5.0 и прилагане на кръстосана валидация (създадена от автора)

Разпределението на коректно и некоректно класифицираните от модела данни е показано на Фиг. 30. Горната лента на фигурата показва неуспешните компании, като коректно класифицираните са отбелязани в светло синьо. Средната лента показва компаниите, които са нито успешни, нито неуспешни, като коректно класифицираните в тази група са отбелязани в червено. Долната лента показва успешните компании, като правилно класифицираните са показани в тъмно синьо. Долната лента е с най-голям размер, тъй като успешните компании са най-голям брой.



Фиг. 30 Разпределението на коректно и некоректно класифицираните от модела компании (създадена от автора)

Полученият чрез кръстосана валидация модел представлява класификационно дърво с 9 променливи. За да се запознаем в детайли с модела, ще разгледаме неговото текстово представяне, което изчерпателно описва дървото. **При текстовото представяне**, всеки възел от дървото е представен на отделен ред. Йерархичността в дървото е показана с отстояния (табулации) в началото на редовете. Колкото повече отстояния има пред даден възел, толкова повече възли има над него в дървото. На всеки ред първо е изписано съкратеното име на променливата, а след него (във фигурни скоби) са показани стойностите от данните, които водят до текущия възел. След стрелката е показана прогнозната стойност за целта (успеха на стартиращата компания), следваща от текущия възел. Накрая на реда, в скоби са показани броят случаи от данните, които попадат във възела и вероятността за успех – в абсолютни единици (от 0 до 1). Текстовото представяне, генерирано от софтуера за извличане на знания от данни, е следното:

КонкурентнопредимствоМоятбизнес in [1 2] [Mode: 2] => 2,0 (3; 0,667)

КонкурентнопредимствоМоятбизнес in [3 4 5] [Mode: 3] (94)

ФУспех_Външнасреда = 1 [Mode: 2] (9)

Опит_Имамопитнаподобнапозициясп in [1 2 3 4] [Mode: 2] => 2,0 (6; 0,833)

Опит_Имамопитнаподобнапозициясп in [5] [Mode: 3] => 3,0 (3; 0,667)

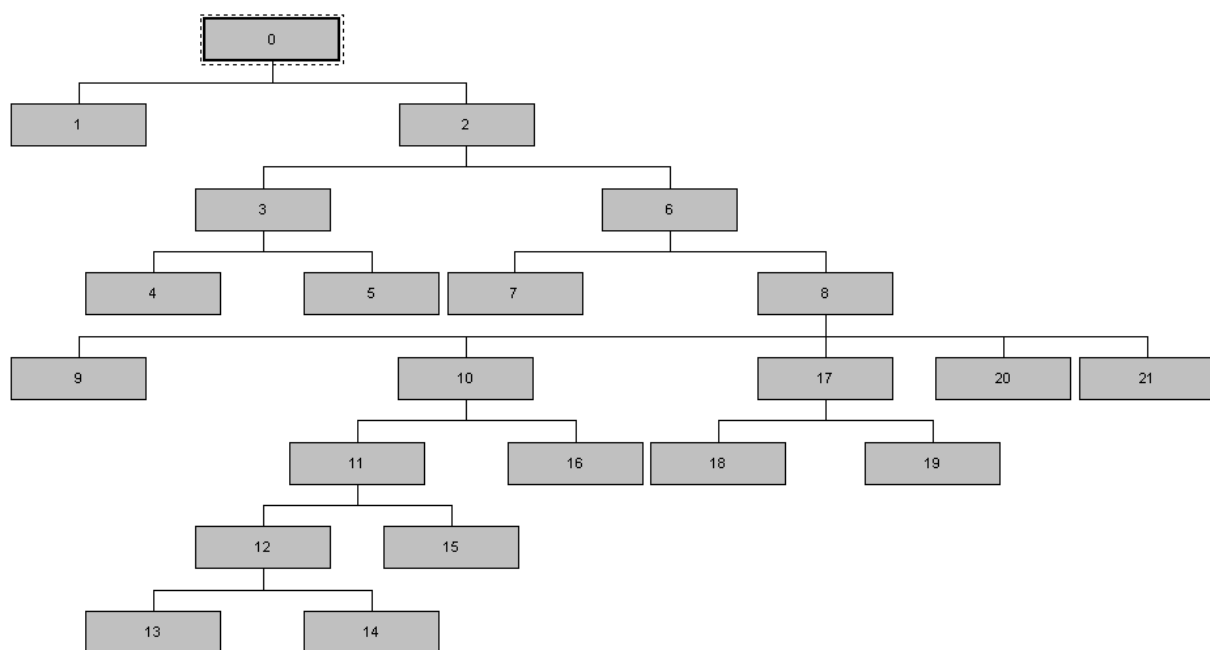
$\Phi \text{Успех_Външна среда} = 2 \text{ [Mode: 3] (85)}$
 $\text{НМАктив_изградена бизнес репутация in [1] [Mode: 1]} \Rightarrow 1,0 \text{ (8; 0,625)}$
 $\text{НМАктив_изградена бизнес репутация in [2 3 4 5] [Mode: 3] (77)}$
 $\text{Начин на навлизане на пазара} = 0 \text{ [Mode: 2]} \Rightarrow 2,0 \text{ (3; 0,667)}$
 $\text{Начин на навлизане на пазара} = 1 \text{ [Mode: 3] (30)}$
 $\text{Индустрия_Повечето печалба in [1 2 3 4] [Mode: 3] (27)}$
 $\text{Партньор_други in [1 2 3] [Mode: 2] (8)}$
 $\text{Клиент_В индустрията има концентрация in [1 2 3 4] [Mode: 2]} \Rightarrow 2,0 \text{ (5; 1,0)}$
 $\text{Клиент_В индустрията има концентрация in [5] [Mode: 3]} \Rightarrow 3,0 \text{ (3; 1,0)}$
 $\text{Партньор_други in [4 5] [Mode: 3]} \Rightarrow 3,0 \text{ (19; 0,947)}$
 $\text{Индустрия_Повечето печалба in [5] [Mode: 2]} \Rightarrow 2,0 \text{ (3; 0,667)}$
 $\text{Начин на навлизане на пазара} = 2 \text{ [Mode: 3] (25)}$
 $\text{НМАктив_разпознаваем бранд in [1] [Mode: 2]} \Rightarrow 2,0 \text{ (2; 1,0)}$
 $\text{НМАктив_разпознаваем бранд in [2 3 4 5] [Mode: 3]} \Rightarrow 3,0 \text{ (23; 0,957)}$
 $\text{Начин на навлизане на пазара} = 3 \text{ [Mode: 3]} \Rightarrow 3,0 \text{ (0)}$
 $\text{Начин на навлизане на пазара} = 4 \text{ [Mode: 3]} \Rightarrow 3,0 \text{ (19; 0,947)}$

Моделът съдържа следните 9 променливи, които описват успешната стартираща компания, като след всяка променлива в скоби е изписан съответстващият номер на възел в дървото, чиито разклонения зависят от променливата:

- Конкурентно предимство: Моят бизнес притежава ясно конкурентно предимство. (възел 0)
- Смятам, че следните фактори са ключови за успеха за моя бизнес: Външна среда. (възел 2)
- Умения и опит: Имам опит на подобна позиция (спрямо заеманата в момента). (възел 3)
- Нематериални активи Компанията притежава изградена бизнес репутация. (възел 6)
- Начин на навлизане на пазара (възел 8)
- Нематериални активи: Компанията притежава разпознаваем бранд. (възел 17)
- Характеристики на индустрията, в която фирмата оперира: Повечето компании в индустрията имат добра печалба. (възел 10)

- Стратегически партньорства: Фирмата участва в партньорства с други заинтересовани страни. (възел 11)
- Характеристики на клиентите: В индустрията има концентрация на клиенти. (възел 12)

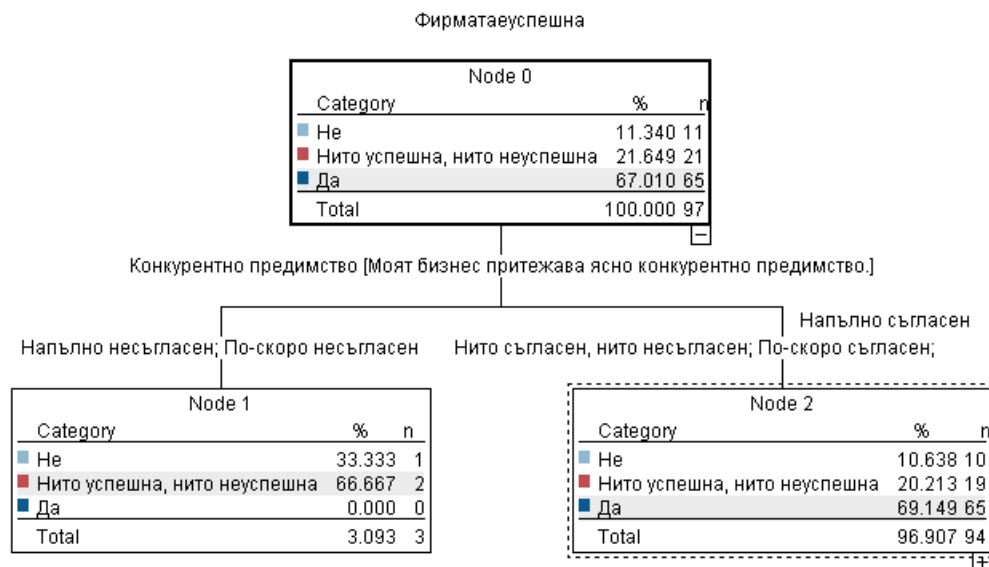
Цялостно графично представяне на класификационното дърво (карта на дървото) е показано на Фиг. 31, като във всеки от възлите е изписан неговият номер. Най-горният възел (0) е основен и се разклонява на два възела (1 и 2) в зависимост от стойността за основния фактор за успех на компанията – наличието на конкурентно предимство. След това възел 2 се разклонява в зависимост от следващия фактор и така до достигането на краен възел.



Фиг. 31 Цялостно графично представяне (карта) на класификационното дърво (създадена от автора)

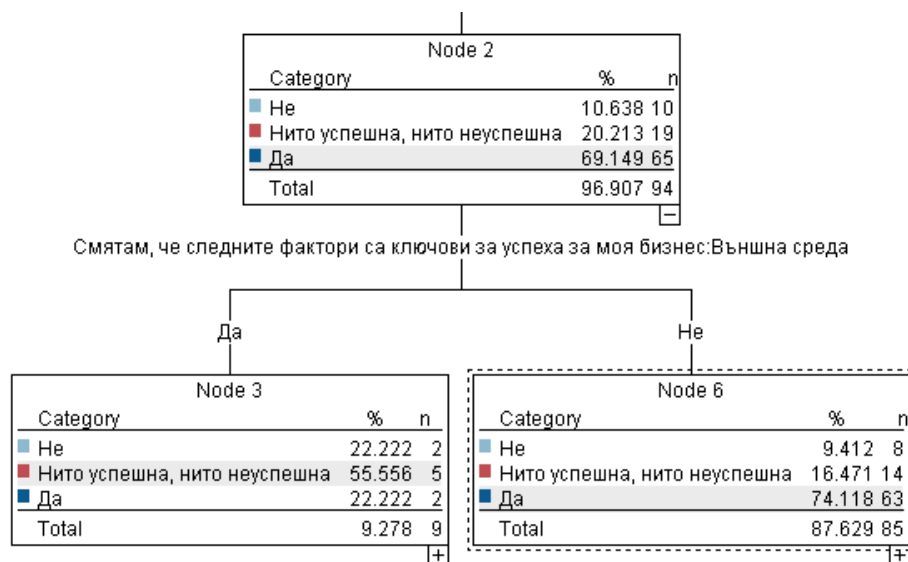
Ще бъдат разгледани някои от възлите на класификационното дърво, които съдържат най-важните фактори за успех, както и техните разклонения. Чрез разглеждане на структурата на класификационното дърво, можем да установим важността на факторите за успех, тъй като най-важните фактори се намират в горните възли на дървото и определят първоначалното разделяне на данните.

Първото ниво на дървото (възел 0) има две разклонения, зависещи от променливата „наличие на ясно конкурентно предимство“ (Фиг. 32). Тази променлива е най-важна за предсказване на успеха на компаниите от анализирания набор от данни. Стартиращите компании, които имат ясно **конкурентно предимство** (възел 2), са по-успешни от тези, които нямат (възел 1). Потвърждава се тезата, която се среща и в други изследвания (Ruskov, 2012), че наличието на конкурентно предимство е ключов фактор за успеха на стартиращите компании.



Фиг. 32 Класификационно дърво – първо ниво: Наличие на ясно конкурентно предимство (създадена от автора)

Възел 2 има две разклонения в зависимост от следващия фактор за успех на компанията „**околната среда е ключов фактор за успеха**“ (Фиг. 33). Тези компании от анализирания набор от данни, които считат околната среда за ключов фактор за успеха (възел 3), са по-малко успешни от останалите (възел 6). Възел 3 има малък брой елементи от набора от данни и по тази причина няма да анализирам в дълбочина неговите наследници.



Фиг. 33 Класификационно дърво – клон 2: външната среда е ключов фактор за успеха на компанията (създадена от автора)

Възел 6 има две разклонения в зависимост от следващия фактор за успех на компанията „**наличие на изградена бизнес репутация**“ (Фиг. 34). Тези компании от анализирания набор от данни, които нямат изградена бизнес репутация (възел 7), са по-малко успешни от останалите (възел 8).



Фиг. 34 Класификационно дърво – клон 2.2: нематериални активи – изградена бизнес репутация на компанията (създадена от автора)

Възел 8 има множество разклонения в зависимост от следващия фактор за успех на компанията „**начин на навлизане на пазара на стартиращата компания**“ (Фиг. 35). Промяната на съществуващ продукт или услуга е най-успешният начин на навлизане на пазара за компаниите от анализирания набор от данни. Фирмите, които развиват нов продукт или услуга, както и тези, които навлизат на пазара чрез паралелна конкуренция също много често са успешни.



Фиг. 35 Класификационно дърво – клон 2.2.2: начин на навлизане на стартиращата компания на пазара (създадена от автора)

По-нататъшен анализ на структурата на дървото показва разположението по възли, взаимовръзките и относителната важност на останалите фактори за успех на компанията. Пълните изходни данни от анализа са представени в Приложение 2.

4.2.4.3 Сравнение на алгоритмите за получаване на модели

Предложените от IBM SPSS Modeler алгоритми за генериране на модели (C5.1, C&R Tree и CHAID) използват техниката индукция на правила (rule induction). Софтуерните алгоритми синтезират класификационно дърво, представляващо набор от правила, които описват определени сегменти в данните във връзка с целта, която в нашия случай е успеха на компанията. Те обхождат набора от входни данни и последователно го разделят на подгрупи на базата на връзки между данните и целта.

Получените чрез алгоритмите модели ясно представят логиката за всяко правило и следователно могат да се използват, за да се разбере процесът на вземане на решение, който води до съответния резултат (успех на компанията). Друго предимство на индукцията на правила, в сравнение с други методи, като например невронните мрежи, е че процесът автоматично елиминира полетата, които не са важни за вземане на решение.

За да обясня **как работи алгоритъмът за индукция на правила**, ще използвам за пример вземането на решение за покупка на жилище. Най-важният фактор може да бъде цената, тъй като той отразява възможността да си позволим жилището. Следващият фактор би могъл да бъде типа имот, който търсим – къща или апартамент. Следващият фактор би могъл да бъде местоположението и т.н. Чрез представяне на факторите във вид на въпроси и чрез йерархичното им комбиниране, получаване класификационно дърво (Фиг. 36).



Фиг. 36 Графично представяне на класификационно дърво (преведена от автора)

Алгоритмите C5.1, C&R Tree и CHAID се различават по начина, по който генерират резултат (класификационно дърво).

C5.1 използва за **цел** само полета от тип категория (флаг, номинални и ординални), докато C&R Tree и CHAID поддържат цели от тип категория и непрекъснати. В нашия

случай целта е от тип категория. Когато алгоритмите обхождат рекурсивно данните, те създават различни видове разделяния на данните (split) в подгрупи (възли) на база променливи от данните (предиктори). C&R Tree поддържа само двоични разделяния (на две подгрупи), докато CHAID и C5.1 поддържат разделяне на повече от две подгрупи.

Алгоритмите се различават по **критерия, използван за разделяне** на данните. При C5.1 (съответно и C5.0) се използва измерител от информационната теория – коефициентът на извличане на информация (information gain ratio). Когато алгоритъмът C&R Tree предсказва поле от тип категория, се използва дисперсионен измерител (по подразбиране коефициент на Джини). А алгоритъмът CHAID използва критерий хи-квадрат (chi-square).

Разгледаните алгоритми **допускат липсващи стойности** в полетата с данни, въпреки че прилагат различни методи за тяхната интерпретация. C5.1 използва метод на разпределение, който заделя част от записа с липсващи данни към всяко от анализирания разклонения на дървото. C&R Tree използва заместване на полетата, за да продължи с генериране на дървото. CHAID създава отделна категория за липсващите стойности и ги използва при изграждане на дървото.

В анализирания данни, получени чрез анкети, има малък брой полета с липсващи стойности. Не е препоръчително пълното премахване на записите, които имат малък брой липсващи полета. Освен това, пропускането на отговора на определен въпрос може да е от значение. По тази причина липсващите стойности са маркирани като такива във файловете и тяхната интерпретация е поверена на съответните софтуерни алгоритми.

Дърветата, които различните алгоритми генерират не са обезателно идентични, тъй като те прилагат различни **критерии за избиране на променливите**, които предсказват стойността на целта. Алгоритмите CHAID и QUEST използват по-стандартни статистически методи, докато C5.0 и C&R Tree използват не-статистически метрики.

Чрез алгоритъма C5.0 е получен най-добър модел за предсказване на успеха на стартиращи компании и по тази причина, **C5.0 ще бъде разгледан в детайли**.

Разглеждането касае и версията на алгоритъма при прилагане на автоматична класификация, която е обозначена като C5.1. Трябва да се има предвид, че алгоритъмът C5.0 е със затворен код и точният му начин на работа при Windows версията му, приложена в IBM SPSS Modeler, не е общодостъпен. Въпреки това, основните му принципи на работата са официално документирани (IBM, 2010). Алгоритъмът C5.0 е разработен от компанията RuleQuest (RuleQuest, 2014), която предлага негова версия за операционната система Linux под отворения лиценз Gnu GPL. Основната част от **програмния код** на Linux версията на алгоритъма е включена към Приложение 2, а пълният код е достъпен за изтегляне на уеб сайта на производителя. Известно е, че C5.0 се базира на алгоритъма за класификация с отворен код C4.5 (Learning, 1993), чиято дефиниция е напълно достъпна.

Алгоритъмът C5.0 работи на база принципа на селекция въз основа на **съотношението на извличане на информация (Gain Ratio Selection)**, чрез който се взема решение как да бъдат разделени данните, и който се основава на теорията на информацията. Избира се такова разделение на данните в клоновете на класификационното дърво, че извличането на информация да бъде максимално. Информацията, извлечена чрез разделяне на данните на база на категориите на поле X (променлива, предсказваща целта) се измерва чрез Формула 6.

$$GAIN(X) = INFO(DATA) - INFO_X(DATA) \quad (6)$$

С $GAIN(X)$ е означена извлечената информация. $INFO(DATA)$ представлява усреднената информация, необходима за идентифициране на класа (категорията на резултата) на запис в набора от данни. А $INFO_X(DATA)$ представлява очакваната изисквана информация след като веднъж данните са били разделени според всеки резултат за текущото поле.

Според теорията на информацията, **критерият за извличане на информация е дефиниран** по следния начин: “Информацията, предадена чрез съобщение зависи от неговата вероятност и може да бъде измерена в битове като минус логаритъм при основа 2 от тази вероятност.” Ако например има 8 еднакво вероятни съобщения, информацията, предадена чрез което и да е от тях е $-\log_2(1/8)$, или 3 бита.

Въпреки, че критерият за извличане на информация дава добри резултати, той има недостатъка, че дава превес на разделяния, които имат голям брой случаи. По този начин категорийна променлива от входните данни, която има много стойности, има предимство спрямо друга с по-малък брой стойности.

Алгоритъмът C5.0 решава този проблем, чрез използване на **критерия съотношение на извличане на информация**. Очакванията в критерия за извличане на информация, могат да бъдат коригирани чрез вид нормализация, при която е регулирана извлечената информация при входни данни с множество стойности. По тази причина C5.0 използва съотношение на извличане на информация, което представлява пропорция на информацията, генерирана чрез разделяне на данните в родителския възел във всяка от категориите на поле X, която е полезна при класификацията (Формула 7).

$$GAIN\ RATI O(X) = GAIN(X) / SPLIT\ INFO_X(DATA) \quad (7)$$

Където $GAIN\ RATI O(X)$ е съотношението на извличане на информация. С $GAIN(X)$ е означена извлечената информация. $SPLIT\ INFO_X(DATA)$ представлява потенциалната информация, генерирана чрез разделяне на данните в n случая, като се има предвид, че извличането на информация измерва информацията, релевантна към класификацията.

Алгоритъмът C5.0 избира да раздели данните на база полетата, при които е максимално съотношението на извличане на информация. Това правило е ограничено от допълнително правило: извлечената информация трябва да бъде много, или поне толкова, колкото е средно извлечената информация от всички разгледани случаи. По този начин се **избягва нестабилност на критериите** на алгоритъма в случаи, когато извлечената информация е малко.

При **оценка на запис от алгоритъма C5.0** (класификация и изчисляване на увереност), записът от входните данни се оценява чрез класа и увереността на правилото, което го генерира.

Ако набор от правила **директно се генерира от възел на дървото**, то увереността за правилото се изчислява по следния начин (Формула 8):

$$< \text{увереност за правилото} > = \frac{< \text{брой правилно класифицирани записи в крайния възел} > + 1}{< \text{общ брой записи в крайния възел} > + 2} \quad (8)$$

Ако набор от правила се **генерира от класификационно дърво, генерирано от възела**, то увереността за правилото се изчислява по следния начин (Формула 9):

$$< \text{увереност за правилото} > = \frac{< \text{брой правилно класифицирани записи в крайния възел} > + 1}{< \text{общ брой записи в крайния възел} > + < \text{брой категории в целта} >} \quad (9)$$

Когато се прави **оценка на база набор от правила**, за всеки запис се разглеждат всички правила, които се отнасят за него. За всяко правило се генерират и асоциират стойности за прогноза (класификация) и увереност. За всяка стойност на резултатите от прогнозата се изчислява сума на увереността, и като краен резултат от прогнозата се избира този, за който е изчислена най-висока стойност от тази сума. Увереността за крайния резултат от прогнозата се изчислява като сума от увереността за тази стойност, разделена на броя приложени правила за съответния запис.

4.2.4.4 Сравнение на получените модели за предсказване на успеха

Преди да бъдат сравнени получените модели е необходимо да се отбележи, че инструментът за автоматична класификация в IBM SPSS Modeler е приложен без да се използва кръстосана валидация на данните. Това означава, че едни и същи данни се използват за синтезиране на модела и за проверка на неговата точност. В резултат на това посочените прогнози за точност са завишени. Получените чрез автоматичния класификатор модели са използвани за идентифициране на най-подходящия алгоритъм – C5.1. След което е генериран модел чрез прилагане на кръстосана валидация чрез алгоритъма C5.0 с използване на ръчни настройки, чиято точност е оценена.

Ако вземем предвид единствено фактора най-висока прогнозна точност, най-точен би бил модела, генериран чрез C5.1. Но имайки предвид, че трябва да се приложи кръстосана валидация, за да бъде точността пресметната по-обективно, най-добрият модел е подобен на първия, но генериран с алгоритъма C5.0 и с прилагане на кръстосана валидация.

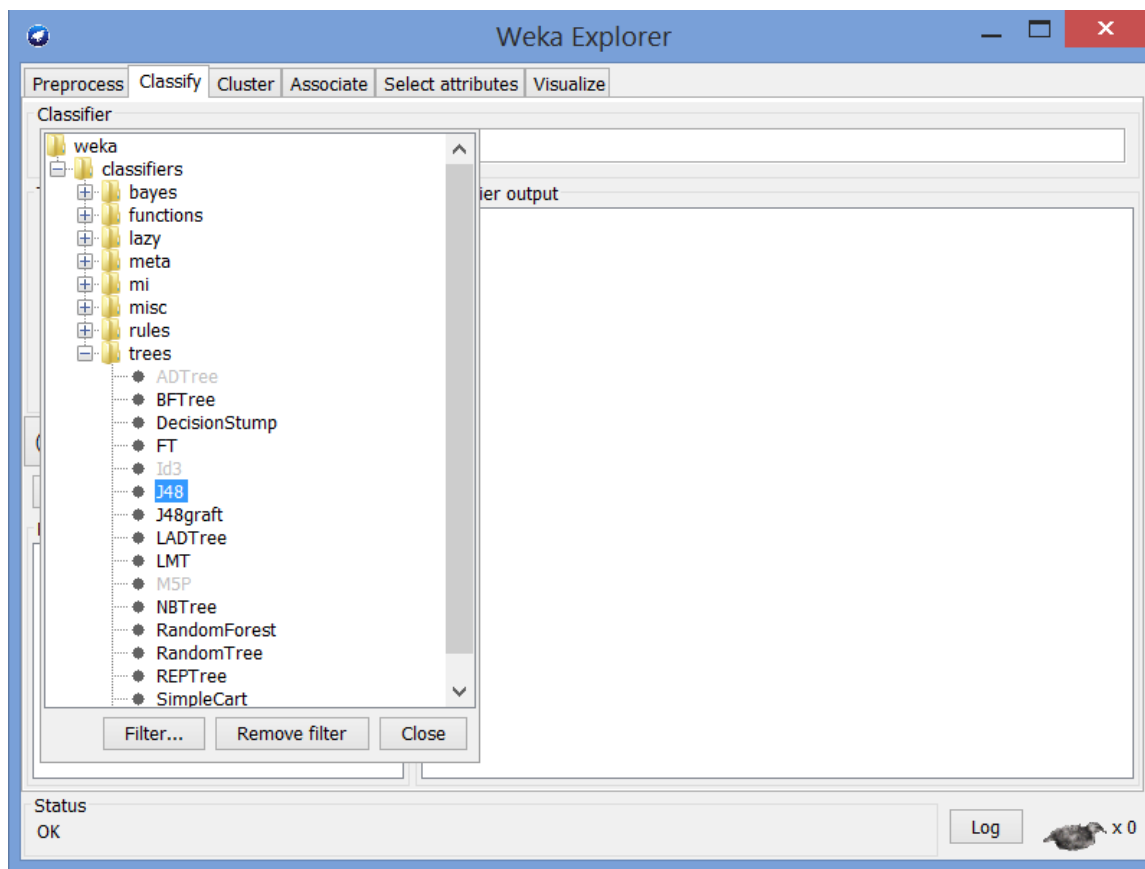
4.2.5 Извличане на знания от данни и създаване на модел за предсказване на успеха с Weka

В тази секция е разгледан процеса на извличане на знания от данни и създаване на модели за предсказване на успеха на стартиращи компании със софтуерния продукт Weka (Hall, 2009). Създадените модели са анализирани и сравнени от гледна точка на техните принципи на работа и ефективност. Разгледани са софтуерните алгоритми, чрез които са получени най-ефективните модели.

4.2.5.1 Синтезиране на модели за предсказване на успеха

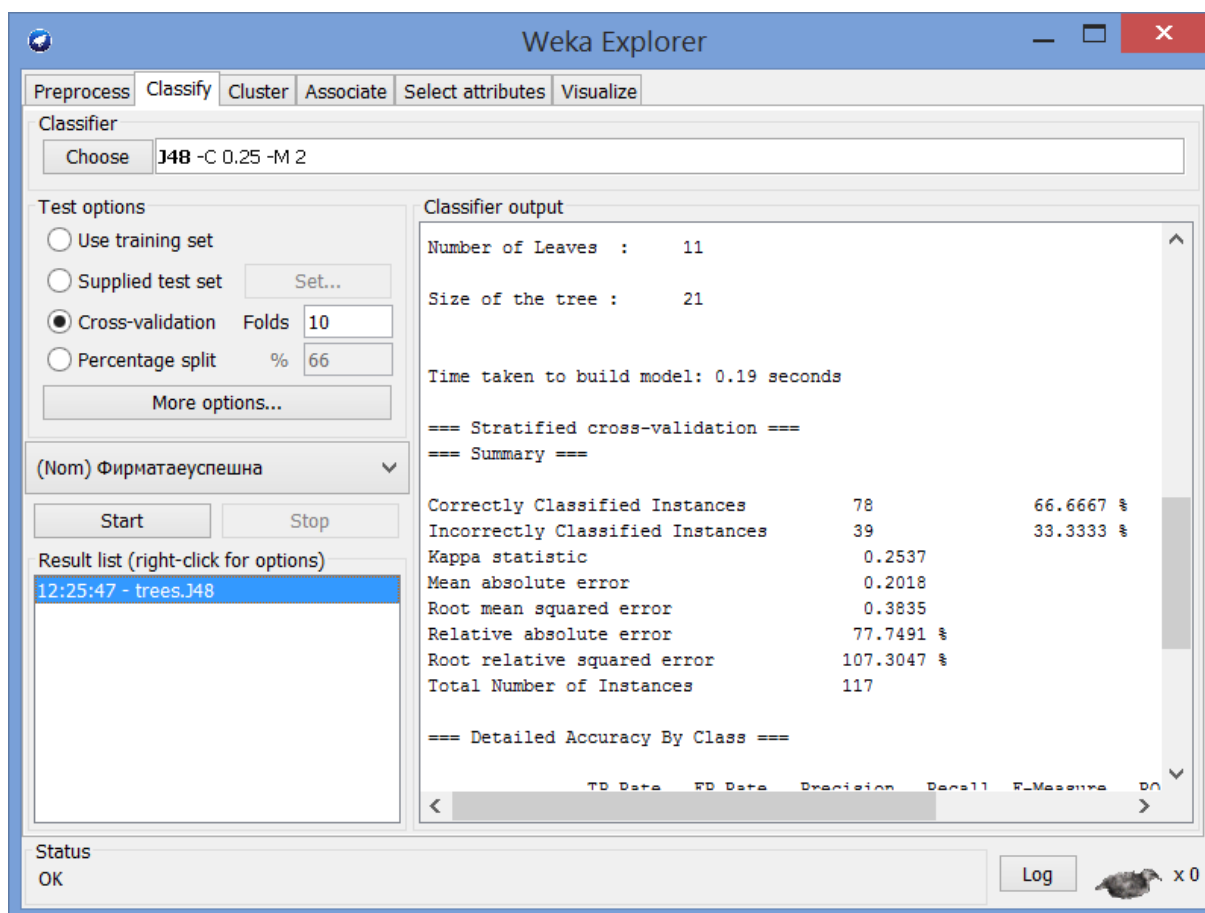
С помощта на софтуерния продукт Weka са създадени модели за предсказване на успеха, които извършват класификация на компаниите - предвиждат в коя категория попадат те: „успешни“, „нито успешни, нито неуспешни“ или „неуспешни“.

Weka предлага множество алгоритми за създаване на модели за класификация (Фиг. 37), които могат да бъдат избрани и стартирани от графичния потребителски интерфейс.



Фиг. 37 Алгоритми за класификация в Weka (създадена от автора)

След избор на алгоритъм, се избира **цел за модела** – променливата, чиято стойност се предсказва, в случая – успеха на компанията. Освен това, може да се настрои алгоритъма за генериране на модел – дали да се приложи **кръстосана валидация** на модела и други настройки, специфични за конкретния алгоритъм.



Фиг. 38 Настройка на алгоритъм J48 в Weka и създаване на класификационен модел (създадена от автора)

След настройка на алгоритъма и натискане на бутона „Start“, в полето „Classifier output“ се получава текстова хронология на данните (Фиг. 38), получени от класификационния алгоритъм. В тази хронология се съдържа информация за модела и неговото синтезиране. Част от получените данни съдържат и прогнозата за точност на модела. Тези действия се повтарят до генериране на желаните модели, които след това се анализират и сравняват.

Изпробвани бяха различни алгоритми за класификация, като всички бяха изпълнени с и без прилагане на кръстосана валидация. Резултатите са показани в Таблица 13, където моделите са подредени в низходящ ред според точността при прилагане на кръстосана валидация.

Таблица 13 Сравнение на класификационни модели, получени с Weka

Алгоритъм за класификация	Тип на модел	Точност (без кръстосана валидация)	Точност (с кръстосана валидация)
J48	Tree	83.76%	66.67%
J48graft	Tree	83.76%	66.67%
DecisionTable	Rules	68.38%	64.10%

LMT	Tree	66.67%	64.10%
BayesNet	Bayes	65.81%	64.10%
BFTree	Tree	64.10%	64.10%
REPTree	Tree	69.23%	63.25%
SimpleCart	Tree	64.10%	63.25%
FT	Tree	100.00%	60.68%
DecisionStump	Tree	64.95%	60.68%
RandomForest	Tree	98.29%	59.83%
NBTree	Tree	91.45%	54.70%
NaiveBayes	Bayes	80.34%	52.14%
LADTree	Tree	83.76%	48.72%
RandomTree	Tree	99.15%	47.86%

В таблицата са представени модели, които имат разнороден принцип на работа (запазена е терминологията, използвана в Weka): класификационно дърво (Tree), индукция на правила (Rules), байесов класификатор (Bayes).

Пълните изходни данни от анализа за модела J48 са представени в Приложение 2.

4.2.5.2 Сравнение на алгоритмите за получаване на модели

Предложените от Weka алгоритми, които генерират модели с най-висока точност при използване на кръстосана валидация са J48 и J48graft. Генерираните модели са **класификационни дървета**, представляващи набор от правила, които описват определени сегменти в данните във връзка с целта, която в нашия случай е успехът на компанията.

J48 е имплементация с отворен код на алгоритъма C4.5, реализиран на програмния език Java. Алгоритъмът е реализиран като клас „public class J48 extends AbstractClassifier implements OptionHandler, Drawable, Matchable, Sourcable, WeightedInstancesHandler, Summarizable, AdditionalMeasureProducer, TechnicalInformationHandler, PartitionGenerator “.

Поради сходството си с C4.5, J48 наподобява алгоритъма C5.0 от IBM SPSS Modeler. J48 изгражда класификационно дърво от набор от данни, чрез прилагане на концепцията за **информационна ентропия**. Данните за обучение са набор от вече класифицирани случаи (семпли) $S = s_1, s_2, \dots$. Всеки семпъл s_i се състои от p -мерен вектор $(x_{1,i}, x_{2,i}, \dots, x_{p,i})$, където x_j представлява атрибутите на семпла, както и класа, в който s_i попада.

Във всеки възел от дървото, алгоритъмът избира атрибута от данните, който най-ефективно разделя наборът от семпли в подсемпли. Критерият за разделяне е **нормализираната извлечена информация** (разликата в ентропията), като атрибута с най-висока нормализирана извлечена информация е избран за вземане на решение (и за разделяне на данните). След това алгоритъмът се повтаря рекурсивно върху получените подсемпли.

Алгоритъмът J48 (и C4.5) има следните основни принципи:

- Ако всички семпли от набора от данни принадлежат на един и същи клас, алгоритъмът просто създава единичен краен възел (листо) за класификационното дърво.
- Ако никой от атрибутите не предоставя извлечена информация, алгоритъмът създава възел на решението, високо в дървото, използвайки очакваната стойност на класа.
- Ако срещне инстанция на клас, който не е срещан преди, алгоритъмът създава възел на решението, високо в дървото, използвайки очакваната стойност на класа.

4.2.5.3 Сравнение на получените модели за предсказване на успеха

От получените класификационни модели, най-висока точност при прилагане на кръстосана валидация има моделът, получен с алгоритъма J48: 66,67%. Тази точност е получена чрез използване на различни данни за обучение и тестване на модела, и е сравнително реалистична. Точността на модела е като цяло относително ниска, както и по-ниска от точността, постигната, чрез прилагане на софтуерния продукт IBM SPSS Modeler.

Ще разгледам в по-голяма дълбочина модела, получен чрез J48 при прилагане на кръстосана валидация, с цел идентифициране на факторите, довели до успех на компанията и тяхната относителна важност. В хронологията, генерирана от класифициращия алгоритъм е показана структурата на класификационното дърво в текстов вид:

```

КонкурентнопредимствоМоятбизнесп <= 2: 2.0 (6.0/1.0)
КонкурентнопредимствоМоятбизнесп > 2
| ФУспех_Цена <= 1
| | Опит_Имамуменияпомениджмънт <= 3: 1.0 (2.0)
| | Опит_Имамуменияпомениджмънт > 3: 3.0 (2.0)
| ФУспех_Цена > 1
| | ФУспех_Външнасреда <= 1
| | | Опит_Имамопитнаподобнапозицияспр <= 4: 2.0 (5.0)
| | | Опит_Имамопитнаподобнапозицияспр > 4: 3.0 (3.0/1.0)
| | ФУспех_Външнасреда > 1
| | | НМАктив_изграденабизнесрепут <= 1
| | | | Продуктъткойтопредлагамевъзприет <= 2: 3.0 (3.03/1.03)
| | | | Продуктъткойтопредлагамевъзприет > 2: 1.0 (4.04/0.04)
| | | | НМАктив_изграденабизнесрепут > 1
| | | | НМАктив_разпознаваембренд <= 2
| | | | | РаботасединосновенклиентОсновния <= 1: 3.0 (2.0)
| | | | | РаботасединосновенклиентОсновния > 1
| | | | | Личн_Несепритеснявамдавземамреше <= 3: 2.0 (9.15/1.0)

```

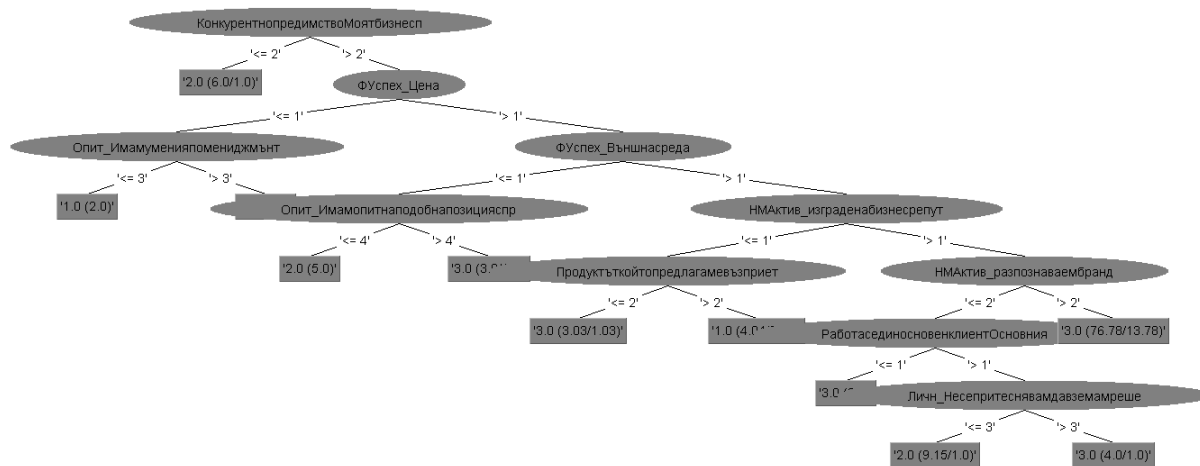
		Личн_Несепритеснявамдавземамреше > 3: 3.0 (4.0/1.0)
		НМАктив_разпознаваембренд > 2: 3.0 (76.78/13.78)

В текстовото представяне на дървото факторите са описани със съкратени имена. Най-важният фактор за успеха на компанията е наличието на ясно конкурентно предимство. Този резултат е консистентен и с резултатите, получени чрез алгоритъма C5.0 в IBM SPSS Modeler. Останалите фактори в модела, генериран чрез Weka и J48 са:

- Конкурентно предимство: Моят бизнес притежава ясно конкурентно предимство.
- Смятам, че следните фактори са ключови за успеха за моя бизнес: Цена
- Смятам, че следните фактори са ключови за успеха за моя бизнес: Външна среда
- Умения и опит: Имам умения по мениджмънт.
- Нематериални активи: Компанията притежава изградена бизнес репутация.
- Умения и опит: Имам опит на подобна позиция (спрямо заеманата в момента).
- Нематериални активи: Компанията притежава разпознаваем бранд.
- Продуктът, който предлагам, е възприет от пазара в следната степен.
- Основният източник, използван за първоначално финансиране на фирмата е: Индивидуален инвеститор (бизнес ангел)
- Личност и ценности: Не се притеснявам да вземам решения без да разполагам с необходимата информация.

Част от изброените фактори за успех на компанията са подобни на получените с помощта на IBM SPSS Modeler и алгоритъма C5.0: конкурентно предимство, външната среда като ключов фактор за успеха, изградена бизнес репутация, разпознаваем бранд. Останалите фактори са специфични за модела.

Weka не предлага подробна визуализация на класификационното дърво, за разлика от IBM SPSS Modeler. Има възможност единствено за схематична визуализация (Фиг. 39), която показва възлите на дървото, връзките между тях и съкратено наименование на променливите.



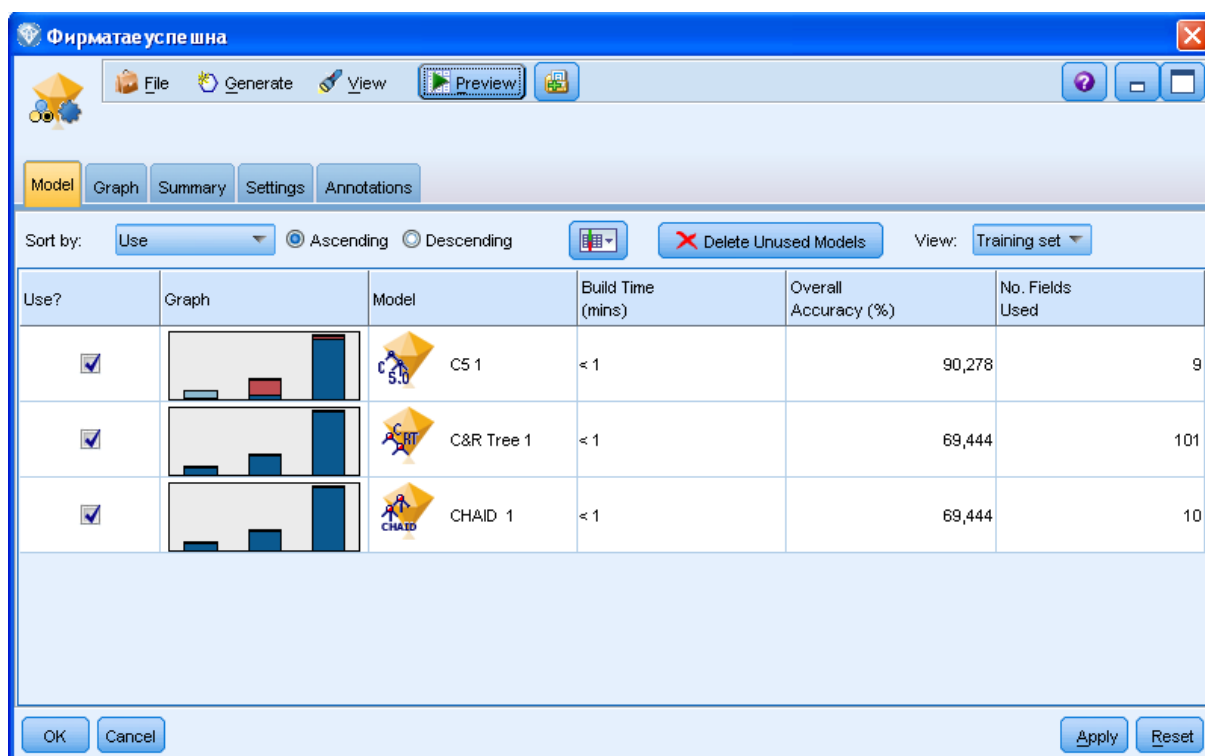
Фиг. 39 Схематична визуализация на класификационното дърво, генерирано с алгоритъм J48 в Weka (създадена от автора)

4.2.6 Синтезиране на модели за предсказване на успеха за технологични компании

На база дефиницията, че **технологичните компании** развиват нов или променят съществуващ продукт (или услуга), предложена в Глава 2, от набора от данни бяха избрани 72 технологични стартиращи компании от България.

Върху избраните данни за технологични стартиращи компании е направен анализ със софтуер за извличане на знания от данни IBM SPSS Modeler по методологията CRISP-DM. Опитната постановка е същата, както при извличането на знания от данни за всички стартиращи компании, като различен е само входният набор от данни. Приложените инструменти са отново автоматичен класификатор и класификационен алгоритъм C5.0 с прилагане на кръстосана валидация. По тази причина, опитната постановка и принципът на работа на алгоритмите няма да бъдат повтаряни.

Автоматичният класификатор класира получените модели и запазва най-добрите три модела за предсказване на успеха (тези с най-висока точност) за по-нататъшен анализ (Фиг. 40).



Фиг. 40 Най-добри получени модели за предсказване на успеха на технологични компании при автоматична класификация (създадена от автора)

Получените модели са съпоставени по своята точност (overall accuracy), като в случая не се прилага кръстосана валидация и посочената точност е по-скоро максимална, отколкото реалистична. Най-точен е моделът, получен чрез алгоритъма C5.1 – с точност 90,28%. Останалите модели са получени чрез алгоритмите C&R Tree 1 (дърво за класификация и регресия) и CHAID 1, като точността и на двата е 69,44%.

За получаване на модел с по-реалистична прогнозна точност, прилагам алгоритъма C5.0. Алгоритъмът C5.0 е избран, тъй като той е имплементация на алгоритъма C5.1, чрез който е получен най-добър модел при използване на автоматичния класификатор. C5.0 поддържа множество ръчни настройки и кръстосана валидация на данните. След настройване на алгоритъма с кръстосана валидация и прилагането му върху данните, се получава модел с точност 87,50% (Фиг. 41).

Results for output field Фирматаеуспешна

Comparing \$C-Фирматаеуспешна with Фирматаеуспешна

Correct	63	87,5%
Wrong	9	12,5%
Total	72	

Фиг. 41 Оценка на точността на модел за предсказване на успеха на технологични компании, получен чрез алгоритъма C5.0 и прилагане на кръстосана валидация (създадена от автора)

Полученият чрез кръстосана валидация модел представлява класификационно дърво с 8 променливи. За да се запознаем в детайли модела, ще разгледаме неговото текстово представяне, което изчерпателно описва дървото:

Индустр_Повечетопечалба in [1 2 3 4] [Mode: 3] (55)

ФУспех_Цена = 1 [Mode: 1] (4)

Опит_Имамуменияпомениджмънт in [1 2 3] [Mode: 1] => 1,0 (2; 1,0)

Опит_Имамуменияпомениджмънт in [4 5] [Mode: 3] => 3,0 (2; 1,0)

ФУспех_Цена = 2 [Mode: 3] (51)

ФУспех_Маркетингирек = 1 [Mode: 2] => 2,0 (4; 0,75)

ФУспех_Маркетингирек = 2 [Mode: 3] (47)

Страт_Стратегиятакоятосъмизбрал in [1 2 3] [Mode: 3] (32)

Конкур_Виндустриятамасилнаконку in [1 2] [Mode: 2]
=> 2,0 (2; 1,0)

Конкур_Виндустриятамасилнаконку in [3 4 5] [Mode: 3]
(30)

НМАктив_изграденабизнесрепут in [1 2 3] [Mode:
2] (9)

СтратегическипартньорстваФирмата in [1
] [Mode: 3] => 3,0 (3; 1,0)

СтратегическипартньорстваФирмата in [2
3 4 5] [Mode: 2] => 2,0 (6; 1,0)

НМАктив_изграденабизнесрепут in [4 5] [Mode: 3
] => 3,0 (21; 0,905)

Страт_Стратегиятакоятосъмизбрал in [4 5] [Mode: 3] =>
3,0 (15; 1,0)

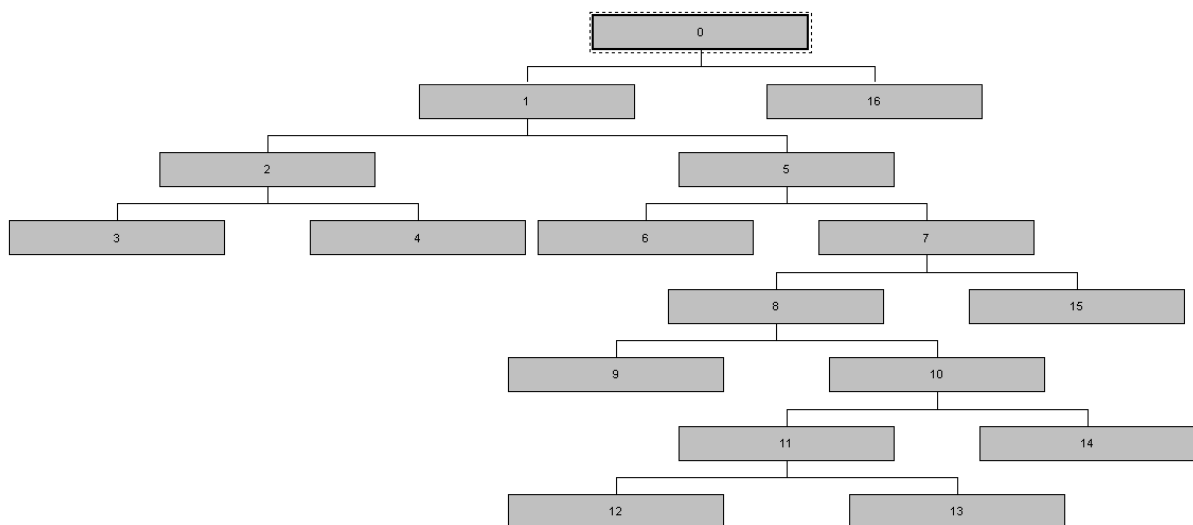
Индустр_Повечетопечалба in [5] [Mode: 1] => 1,0 (5; 0,6)

Моделът съдържа следните 8 променливи, които описват успешната технологична стартираща компания, като след всяка променлива в скоби е изписан съответстващият номер на възел в дървото, чиито разклонения зависят от променливата:

- Повечето компании в индустрията имат добра печалба. (възел 0)
- Смятам, че следните фактори са ключови за успеха за моя бизнес: Цена (възел 1)
- Смятам, че следните фактори са ключови за успеха за моя бизнес: Маркетинг и реклама (възел 5)
- Умения и опит: Имам умения по мениджмънт. (възел 2)
- Стратегията, която съм избрал, е агресивна. (възел 7)
- В индустрията има силна конкуренция. (възел 8)

- Нематериални активи: Компанията притежава изградена бизнес репутация. (възел 10)
- Стратегически партньорства: Фирмата участва в партньорства с конкуренти. (възел 11)

От изброените фактори за успех на технологични стартиращи компании се вижда, че те са много различни от факторите за успех на стартиращите компании от България като цяло. Много от факторите са свързани с външната среда. **Цялостно графично представяне** на класификационното дърво за предсказване на успеха (карта на дървото) е показано на Фиг. 42.



Фиг. 42 Цялостно графично представяне (карта) на класификационното дърво за технологични компании (създадена от автора)

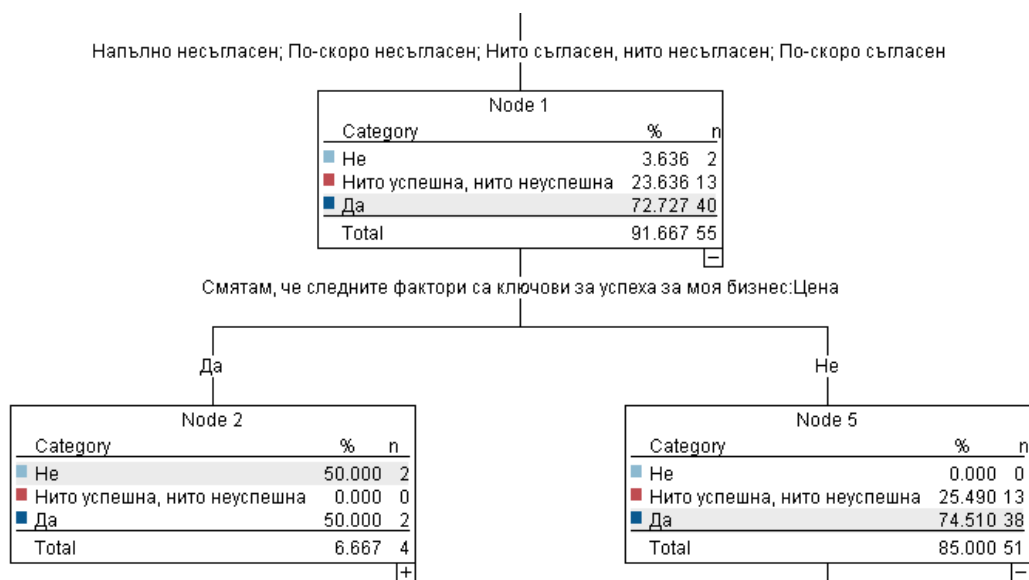
Ще бъдат разгледани някои от възлите на класификационното дърво, които съдържат най-важните фактори за успех, както и техните разклонения. Чрез разглеждане на структурата на класификационното дърво, можем да установим важността на факторите за успех, тъй като най-важните фактори се намират в горните възли на дървото и определят първоначалното разделяне на данните.

Първото ниво на дървото (възел 0) има две разклонения, зависещи от променливата „повечето компании в индустрията имат добра печалба“ (Фиг. 43). Тази променлива е най-важна за предсказване на успеха на компаниите от анализирания набор от данни. Стартиращите компании, които са напълно съгласни, че повечето компании в индустрията имат добра печалба (възел 16), са по-малко успешни от останалите (възел 1). На пръв поглед това заключение изглежда нелогично и по-всяка вероятност има скрита причина, която би могла да бъде връзката на печалбите на компаниите в индустрия с конкурентната среда и наличието на големи фирми на пазара.



Фиг. 43 Класификационно дърво – първо ниво: Повечето компании в индустрията имат добра печалба (създадена от автора)

Възел 1 има две разклонения в зависимост от следващия фактор за успех на компанията „цената като ключов фактор за успеха“ (Фиг. 44). Тези компании от анализирания набор от данни, които считат цената за ключов фактор за успеха (възел 2) са по-малко успешни от останалите (възел 5). Възел 2 има малък брой елементи от набора от данни и по тази причина няма да анализираме в дълбочина неговите наследници.



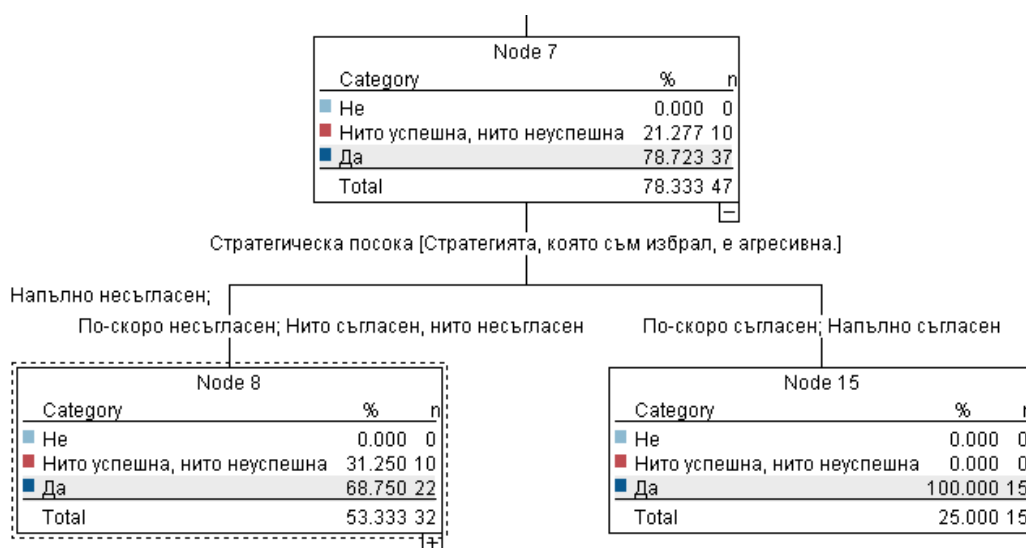
Фиг. 44 Класификационно дърво – клон 1: ключови фактори за успеха: цена (създадена от автора)

Възел 5 има две разклонения в зависимост от следващия фактор за успех на компанията „маркетинга и рекламата като ключови фактори за успеха“ (Фиг. 45). Тези компании от анализирания набор от данни, които считат маркетинга и рекламата за ключов фактор за успеха (възел 6) са по-малко успешни от останалите (възел 7).



Фиг. 45 Класификационно дърво – клон 1.2: ключови фактори за успеха: маркетинг и реклама (създадена от автора)

Възел 7 има две разклонения в зависимост от следващия фактор за успех на компанията „избраната стратегия е агресивна“ (Фиг. 46). Тези компании от анализирания набор от данни, които са избрали агресивна стратегия (възел 15) са по-успешни от останалите (възел 8).



Фиг. 46 Класификационно дърво – клон 1.2.2: избраната стратегия е агресивна. (създадена от автора)

По-нататъшен анализ на структурата на дървото показва разположението по възли, взаимовръзките и относителната важност на останалите фактори за успех на компанията. Пълните изходни данни от анализа са представени в Приложение 2.

4.2.7 Сравнение на получените модели

На база на направените анализи, **двата най-точни модела за предсказване на успеха на стартиращи компании** (при прилагане на метода кръстосана валидация) са:

- класификационно дърво, генерирано с помощта на продукта IBM SPSS Modeler и алгоритъма C5.0 с точност **83,76%**,
- класификационно дърво, генерирано с помощта на продукта Weka и J48 с точност **66,76%**.

Синтезиран е и модел, който предсказва успеха на **технологични стартиращи компании** (при прилагане на кръстосана валидация.):

- класификационно дърво, генерирано с помощта на продукта IBM SPSS Modeler и алгоритъма C5.0 с точност **87,50%**.

Съпоставка на факторите за успех на трите изброени модела е показана в Таблица 14.

Таблица 14 Съпоставка на модели за предсказване успеха на стартиращи компании

Класификационно дърво за стартиращи компании, получено чрез C5.0	Класификационно дърво за стартиращи компании, получено чрез J48	Класификационно дърво за технологични стартиращи компании, получено чрез C5.0
<ul style="list-style-type: none"> • Конкурентно предимство: Моят бизнес притежава ясно конкурентно предимство. • Смятам, че следните фактори са ключови за успеха за моя бизнес: Външна среда • Умения и опит: Имам опит на подобна позиция (спрямо заемащата в момента). • Нематериални активи: Компанията притежава изградена бизнес репутация. • Начин на навлизане на пазара • Нематериални активи: Компанията притежава разпознаваем бранд. 	<ul style="list-style-type: none"> • Конкурентно предимство: Моят бизнес притежава ясно конкурентно предимство. • Смятам, че следните фактори са ключови за успеха за моя бизнес: Цена • Смятам, че следните фактори са ключови за успеха за моя бизнес: Външна среда • Умения и опит: Имам умения по мениджмънт. • Нематериални активи: Компанията притежава изградена бизнес репутация. • Умения и опит: Имам опит на подобна позиция (спрямо заемащата в 	<ul style="list-style-type: none"> • Повечето компании в индустрията имат добра печалба. • Смятам, че следните фактори са ключови за успеха за моя бизнес: Цена • Смятам, че следните фактори са ключови за успеха за моя бизнес: Маркетинг и реклама • Умения и опит: Имам умения по мениджмънт. • Стратегията, която съм избрал, е агресивна. • В индустрията има силна конкуренция.

<ul style="list-style-type: none"> • Характеристики на индустрията, в която фирмата оперира: Повечето компании в индустрията имат добра печалба. • Стратегически партньорства: Фирмата участва в партньорства с други заинтересовани страни. • Характеристики на клиентите: В индустрията има концентрация на клиенти. 	<p>момента).</p> <ul style="list-style-type: none"> • Нематериални активи: Компанията притежава разпознаваем бранд. • Продуктът, който предлагам, е възприет от пазара в следната степен • Основният източник, използван за първоначално финансиране на фирмата е: Индивидуален инвеститор (бизнес ангел) • Личност и ценности: Не се притеснявам да вземам решения без да разполагам с необходимата информация. 	<ul style="list-style-type: none"> • Нематериални активи: Компанията притежава изградена бизнес репутация. • Стратегически партньорства: Фирмата участва в партньорства с конкуренти.
---	--	---

От таблицата е видимо, че двата най-добри модела за предсказване на успеха на стартиращи компании си приличат по включените в тях фактори. Моделът за предсказване на успеха на технологични стартиращи компании е специфичен и въпреки, че има някои общи фактори с останалите модели, повечето се различават.

4.3 Резултати, изводи и заключение

В настоящата глава е описано проведеното **качествено изследване** с цел предварителна проверка и подобрене на предложените в Глава 3 абстрактен и детайлизиран модел за предсказване на успеха на стартиращи компании от България. В резултат на събраните данни от качествено изследване и на направените подобрения и допълнения в моделите, те **отговарят на изискванията** и предпочитанията на **българските предприемачи**, взели участие в изследването. Това спомага за постигане на **пълнота и изчерпателност** на предложените модели.

Групирането на факторите от абстрактния модел за предсказване на успеха, предложен в Глава 3, е **валидирано** чрез статистическия метод **факторен анализ**, който показва много близко до хипотезата на автора групиране. От резултатите от факторният анализ може да се заключи, че предложеният абстрактен модел **групира коректно предложените фактори за успех**.

Чрез прилагане на методологията CRISP-DM и продукти за машинно учене и извличане на знания от данни, бяха **синтезирани модели за предсказване на успеха** на стартиращи компании от България, както и за технологични стартиращи компании. Получените модели бяха анализирани от гледна точка на алгоритмите, използвани за генерирането им, съдържащите се фактори, както и точността на моделите. Бяха съпоставени и избрани най-добрите модели за предсказване на успеха, които са от типа класификационно дърво, и които могат да послужат за изграждане на информационна система за предсказване на успеха на стартиращи компании. Най-добрите модели

предсказват успеха на стартиращите компании и на технологичните стартиращи компании с **висока точност**, от което може да се заключи, че те **успешно изпълняват** поставените цели.

ГЛАВА 5: РЕАЛИЗАЦИЯ НА СОФТУЕР ЗА ПРОГНОЗИРАНЕ НА УСПЕХА

В настоящата глава ще бъде разгледана реализацията на софтуер за предсказване на успеха на стартиращи компании. Софтуерът ще бъде наричан за краткост с неговото име I3SP (Information System for Start-ups Success Prediction), или в превод „информационна система за предсказване успеха на стартиращи компании“.

5.1 Изисквания към софтуера за прогнозиране успеха на стартиращи компании

Софтуерът за предсказване на успеха на технологични стартиращи компании трябва да отговаря на определени функционални изисквания, за да бъде едновременно полезен за потребителите и в синхрон с целите и резултатите от изследването.

Минималните изисквания към приложението са следните:

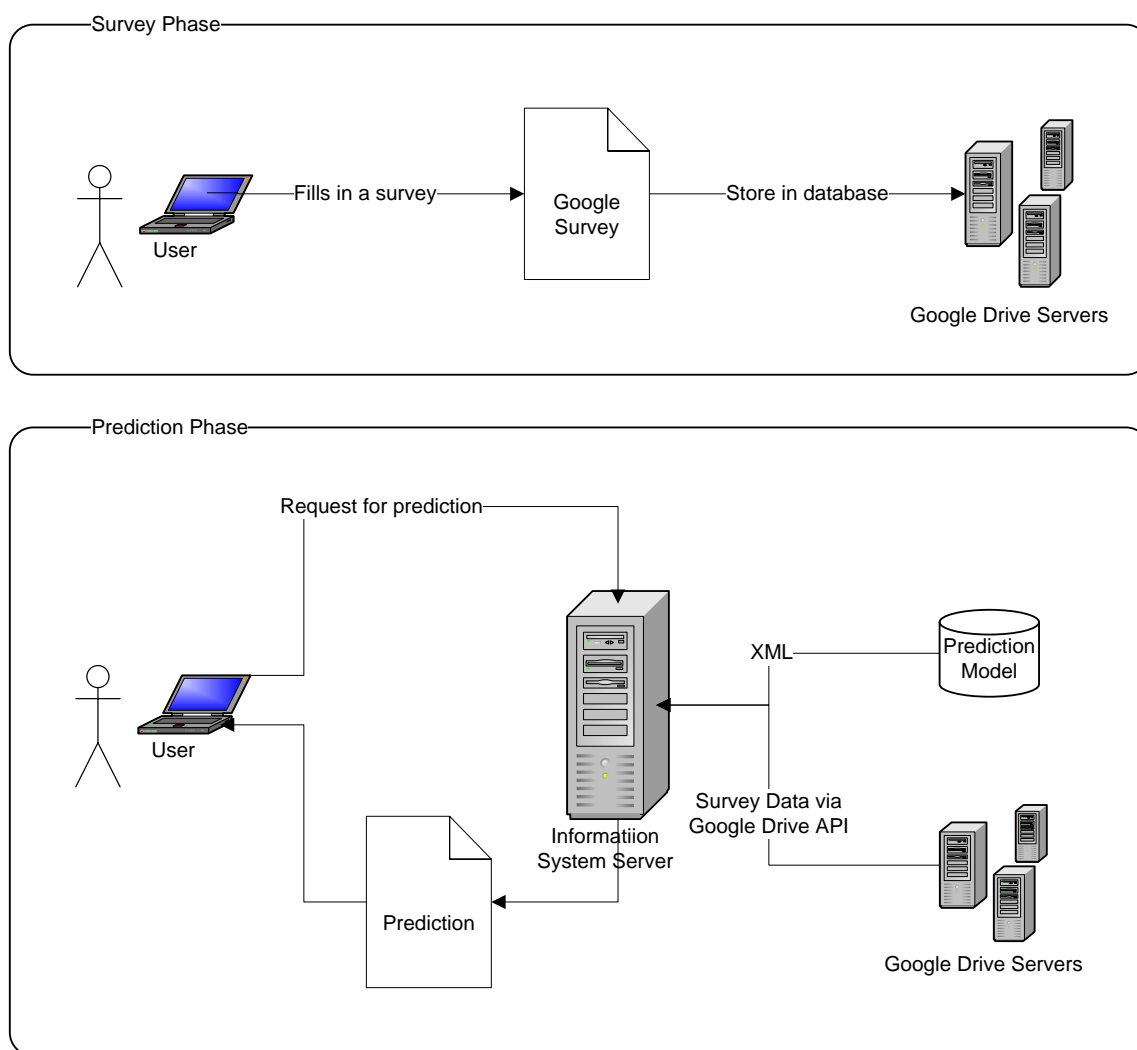
- Съвместимост с моделите за предсказване на успеха (от софтуерните продукти за извличане на знания от данни) и възможност за актуализация на модела. Като минимум, трябва да се поддържат моделите от тип класификационно дърво.
- Съвместимост с Google Диск. Данните от анкетите, които потребителите попълват, се съхраняват в Google Диск и се използват едновременно за прогнозиране на успеха и за бъдеща актуализация на моделите за предсказване. Поддръжката на интерфейсите на Google Диск ще позволи на приложението директно да изтегля актуалните данни от попълнените анкети, за да прогнозира шансовете за успех на компаниите.
- Многоплатформеност – приложението трябва да бъде съвместимо с различни устройства и различни операционни системи. Софтуерът трябва да бъде съвместим минимум със следните браузъри: Google Chrome (версия 30 и по-висока), Mozilla Firefox (версия 27 и по-висока) , Opera (версия 19 и по-висока) и Microsoft Internet Explorer (версия 9 и по-висока), което ще осигури съвместимост с бъдещи браузъри и версии.
- Ниски разходи за имплементация и поддръжка - това ще направи приложението достъпно за целевата група – стартиращи, най-често малки компании, които разполагат с ограничени бюджети за своята дейност.
- Софтуерът трябва да бъде лесен за работа - да не се изискват технически компетенции, за да се изготви прогноза за успеха на компаниите.
- Софтуерът трябва да изготвя прогноза за успеха на компаниите, която да бъде рационално и разбираемо обяснена на потребителя.
- Софтуерът би било добре да дава допълнителни препоръки, свързани с резултата от анализа – например фактори, които могат да бъдат подобрени, за да бъде стартиращата компания по-успешна.
- Софтуерът трябва лесно да се развива и актуализира. За тази цел той трябва да работи едновременно с анкетите от Google Drive и моделите от софтуера за извличане на знания от данни. При натрупване на достатъчно данни в Google Drive, ще се изготвя нов класификационен модел във формат на текстов сайл и той трябва лесно да се качи в софтуера за предсказване на успеха.

На база функционалните изисквания към софтуера, са избрани подходящи технологии и инструменти за неговата реализация, и е проектирана неговата архитектура.

5.2 Проектиране на софтуера за прогнозиране успеха на стартиращи компании

В тази точка ще бъде разгледано проектирането на софтуерния продукт – неговата архитектура, алгоритъм на работа и модули.

Информационната система за предсказване на успеха на стартиращи компании (I3SP) предоставя прогноза за успех на компания на база входни данни от анкета, попълнена от собственика на компанията и модел за предсказване на успеха (Фиг. 47).



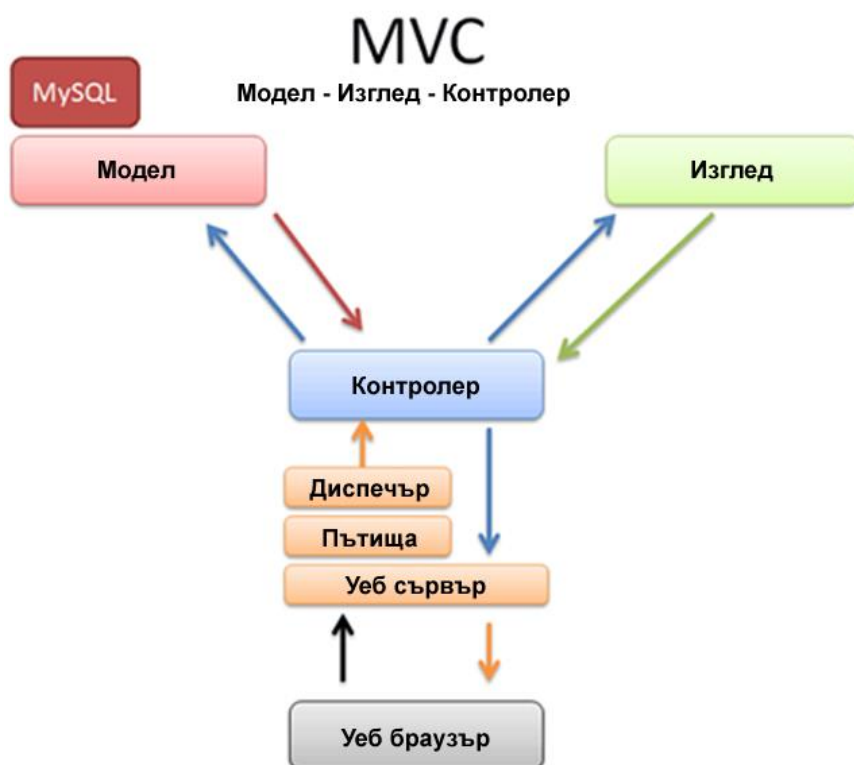
Фиг. 47 Диаграма на потока от данни в I3SP (създадена от автора)

Първо, потребителят попълва анкета и отговорите се запазват в Google диск (анкетна фаза). След това потребителят заявява анализ (фаза на предсказване) и I3SP генерира и връща резултат за вероятността за успех на компанията и попълнено класификационно дърво. I3SP намира съответствие между данните, попълнени в анкетата от потребителя

и класификационното дърво, предварително създадено от софтуер за извличане на знания от данни.

5.2.1 Обща архитектура на приложението

Софтуерът I3SP ще се придържа към архитектура, реализираща трислойния модел **MVC (Модел – Изглед - Контролер)** (Фиг. 48). MVC е шаблон за дизайн на софтуерна архитектура, използван при реализацията на потребителски интерфейси, който представя софтуерното приложение като три взаимно-свързани части, разделяйки вътрешното представяне на информацията от информацията, представена на потребителя. Приложенията, реализирани чрез прилагането на MVC са по-лесни за разработка и поддръжка, особено от екип - дизайнери и разработчици. Това води до по-добра координация в екипа и по-малко грешки.



Фиг. 48 Трислойен модел Модел – Изглед - Контролер при уеб базирано приложение, реализирано с програмния език PHP и база данни MySQL (адаптирана от автора по фигура от betterexplained.com)

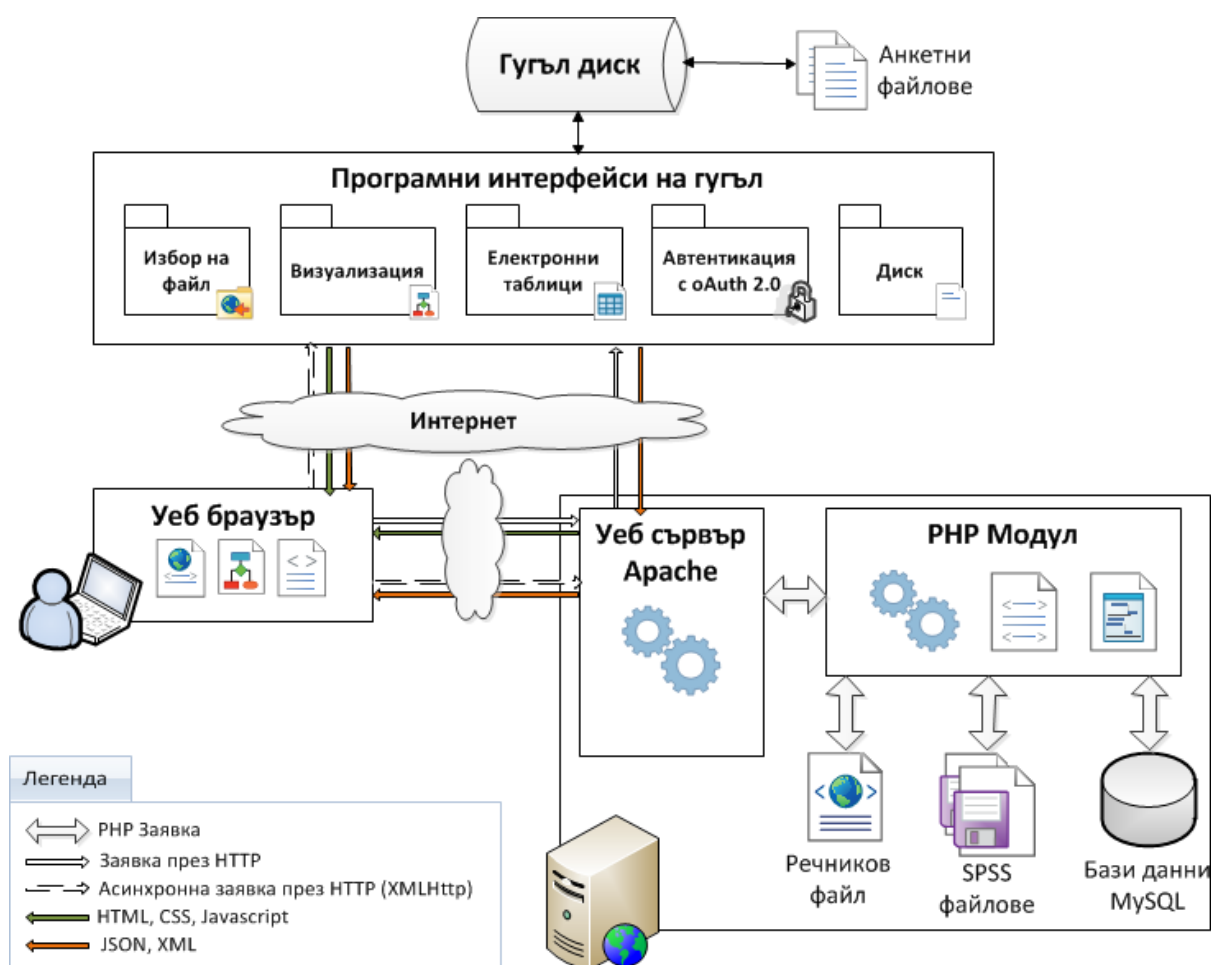
Контролерът (Controler) служи за приемане и обработка на входящите заявки към уеб приложението. Той комуникира както с изгледа, така и с модела. Софтуерът за прогнозиране на успеха използва техника, наречена една входна точка (single entry point), в резултат на което, всички заявки, изпращани от уеб браузъра на потребителя, се предават от уеб сървъра към един и същи програмен файл на приложението. Този файл съдържа логиката за разпределяне на заявките вътре в приложението и, на базата на подадените параметри в заявката, решава кой метод за обработка на данните да стартира. Архитектурата с една входна точка се счита за добра практика, тъй като всички файлове, съдържащи бизнес логиката на приложението, остават скрити от външните потребители. За правилната реализация на този похват се правят и някои

специфични настройки на веб сървъра. Създават се правила за пренаписване на заявките (rewrite rules) и се добавят декларации и файлове за ограничаване на достъпа до определени директории на веб сървъра.

Моделът (Model) съдържа основна част от бизнес логиката на приложението и изгражда функционалностите на отделните модули. Той комуникира с базата данни и с програмните интерфейси на трети страни.

Изгледът (View) се грижи за представянето на данните, създадени от модела, към потребителя. В него са включени презентационни и графични шаблони и стилове за представяне на данните в различни типове страници, и почти не се съдържа бизнес логика. В рамките на приложението I3SP, презентационният слой има и допълнителна много важна функция, която не се вписва стриктно в методологията на трислойната архитектура. При отделни модули на приложението, презентационният слой съдържа програмен код (JavaScript), който осъществява връзка с програмен интерфейс на трета страна и изпраща и получава данни от нея, което се налага поради специфично изискване към приложението - да работи с програмния интерфейс на Google диск.

На Фиг. 49 е показана схематично архитектурата на приложението за прогноза, връзките между отделните функционалности и взаимодействието с външни приложения и програмни интерфейси.



Фиг. 49 Схематично представяне на архитектурата и функционалностите на приложението (създадена от автора)

5.2.2 Модули на приложението

Програмната част на I3SP ще представлява клиент-сървър уеб приложение, което е изградено от няколко модула с различен тип функционалност. Всеки от тези модули се състои от един или няколко файла с програмен код. Връзката между отделните модули се осъществява чрез навигационен контролер. За реализация на части от кода с по-голяма сложност и необходимост от по-добро структуриране се ползва обектно ориентирания модел, реализиран чрез синтаксиса на PHP. За останалите части се ползва процедурния модел на PHP, който позволява по-бърза разработка на прости функционалности от приложението. Основните модули в I3SP са изброени в следващите подточки.

5.2.2.1 Модул за вход

Входът в система ще се осъществява чрез външен метод за аутентикация – чрез Google акаунт. За да може даден потребител да използва приложението I3SP, той ще трябва да има валиден Google акаунт, тъй като се изисква потребителят да има достъп до данни в анкетен файл, който се намира на сървърите на Google. Акаунтът на потребителя може да се ползва едновременно за достъп до приложението и до анкетния файл.

Решението, което ще бъде приложено в софтуера за прогноза е въвеждането на име и парола да става само веднъж – при първоначалното отваряне на приложението. След успешна аутентикация с Google акаунт, приложението ще получава от програмния интерфейс на Google ключ за достъп (access token), който в последствие ще се ползва за достъп до анкетния файл, без нуждата от въвеждане на потребителско име и парола отново.

Изисква се ограничаване на достъпа до приложението само за определени потребители. За да се постигне тази цел, след като потребителят бъде добавен в списък на регистрираните потребители и след одобрение от администратор, неговият профил се активира. Списъкът с потребители се управлява от потребителите, които притежават административни права. Те могат да одобряват, променят, изтриват и делегират права на останалите потребителски акаунти.

5.2.2.2 Модул за администрация

Потребителите в системата, които притежават административни права, имат достъп до разширена функционалност. Администраторите получават известие по пощата, когато нов потребител поиска достъп до приложението. Те имат възможност за преглед на списък с всички потребители, могат да активират чакащите одобрение и да деактивират тези, които вече нямат нужда да ползват приложението или са нарушили правилата за ползване. Активацията на потребители е необходима, за да се ограничи достъпа до приложението.

Потребителите с административни права се грижат за въвеждането и актуализацията на класификационните дървета, получени чрез софтуерни продукти за извличане на знания от данни. Класификационните дървета ще се актуализират периодично при събиране на повече данни за фирми, като тази промяна ще се осъществява ръчно. По този начин точността на прогнозата ще се увеличава. Администраторите имат възможност да актуализират речник-файла, който е един за цялото приложение и служи за осъществяване на връзка между данните от анкетите и класификационните дървета.

5.2.2.3 Модул за настройка

След като даден потребител бъде одобрен от администратор, той получава правото да използва системата. Първото нещо, което всеки потребител трябва да направи след успешен вход в системата, е настройка на двете основните конфигурации – избор на анкетен файл и алгоритъм за прогноза (класификационно дърво).

Анкетният файл се намира на отдалечен сървър – Google диск. Достъпът до този файл се осъществява само след успешна аутентикация на потребител с неговия Google акаунт. За да се изведе на екрана на потребителя списък с всички файлове от тип анкета (респективно електронна таблица) от неговия Google диск, ще се използва програмен интерфейс за работа с електронни таблици. Чрез този интерфейс ще бъде дадена възможността на потребителя да избере анкетен файл и да прави анализи (прогнози) върху избрани редове (отговори на анкетата) от този файл. Реализирана е функционалност, която проверява дали избраният файл е валиден.

Потребителят е необходимо да избере класификационно дърво (алгоритъм, по който да се прави прогноза) преди да започне работа с приложението. Приложението предлага избор между различни класификационни дървета, генерирани от софтуер за извличане на знания от данни IBC SPSS Modeler, Weka или друг. Това позволява на потребителите да сравняват резултатите, получени за едни и същи данни при прилагане на различни класификационни дървета. За всеки потребител ще има избрано класификационно дърво по подразбиране.

5.2.2.4 Модул за данни от анкетата

Преди да се изготви прогноза за успеха, потребителят трябва да избере кои данни ще бъдат анализирани посредством модулът за данни от анкетата. Това става чрез преглед на съдържанието на избрания анкетен файл (в табличен вид) и избор на ред за анализ с натискане върху него. Тъй като анкетният файл е записан на отдалечен сървър (Google диск) и приложението няма директен достъп до него, се използва програмният интерфейс на Google за работа с електронни таблици. Този интерфейс позволява директни заявки към отдалечения сървър и получаване на данни от избрания анкетен файл. Комуникацията става директно между клиентския браузър и Google диск чрез асинхронно извикване, като се разчита на факта, че потребителят е удостоверил, че разполага с достъп до съответния файл.

5.2.2.5 Модул за анализ и прогноза

Модулът за анализ и прогноза е най-важната част на софтуера за предсказване на успеха на стартиращи компании. Модулът ще приема като входни параметри:

- идентификационен ключ на анкетния файл,
- избран ред/индекс от същия файл
- параметър за класификационното дърво, което ще бъде използвано от изчисляващия алгоритъм.

За да функционира коректно модула, тези параметри трябва да съответстват на съществуващи обекти и да се отнасят до текущия потребител в системата. Няма да бъде позволен анализът на файл, който не принадлежи на текущия потребител. Данните от анкетния файл не са налични в локалната база данни и приложението генерира заявка към Google диск, за да ги достъпи. За целта се използва програмният интерфейс на Google за работа с електронни таблици. Тъй като данните, извлечени от Google диск

трябва да се обработят на сървъра на приложението, е нужен механизъм за извикване на програмния интерфейс на Google от сървърния код на приложението.

След като данните бъдат получени от отдалечения сървър, те трябва да се преобразуват във формат, удобен за работа на приложението и алгоритъма за изчисление.

Преобразуването на данни ще се случва с помощта на речник-файла. Алгоритъмът, който прави прогнозата за успех, се нуждае от преобразуваните данни, от ред на анкетата, и от данни от класификационното дърво.

Основното действие в работата на алгоритъма е обхождане на възел от класификационното дърво. На всяка стъпка, алгоритъмът взема решение кой възел да посети, в зависимост от данните постъпили от анкетния ред. Резултатът от работата на алгоритъма за прогноза е достигането на краен възел на класификационното дърво. Всеки краен възел на дървото съдържа процент вероятност за успех на стартиращата компания.

Получените данни, представляващи прогноза, ще бъдат представени на потребителя в графичен вид. Графиката, генерирана при приключване на работата на алгоритъма, дава възможност за проследяване работата на алгоритъма за прогноза, и с нейна помощ могат да се открият ключовите фактори за успех на стартираща компания. Графиката се начертава под формата на дърво с нанесени всички данни от анкетния ред. Всеки възел в дървото съдържа данните, отнасящи се до конкретен въпрос и отговор от анкетата и данни от избраното класификационно дърво.

Модулът за анализ съдържа и функционалност за запамятаване на резултатите от анализа, за да могат да се сравнят анализи на различни компании или анализи на една компания в различни моменти от нейното развитие. Има и възможност за преизчисление на вече направена прогноза при същите анкетни данни, но различно класификационно дърво, за да може да се прави сравнителен анализ на резултатите от работа на алгоритъма за прогноза с различни класификационни дървета.

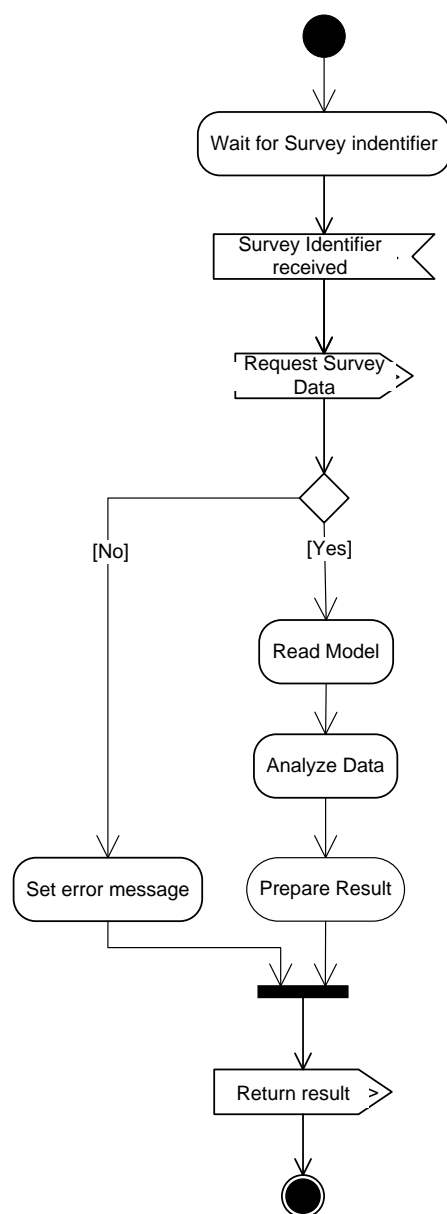
5.3 Разработка на софтуер за прогнозиране на успеха

В тази точка се разглежда същинската реализация на приложението. Представени са основни концепции при изграждането му и е обяснен начинът, по който са приложени избраните технологии и алгоритмите за осъществяване на проектираната функционалност. Цитирани са фрагменти от кода, за да се даде по-ясна представа за реализацията на конкретни функционалности. Софтуерът е разработен като дипломна работа под ръководството на автора.

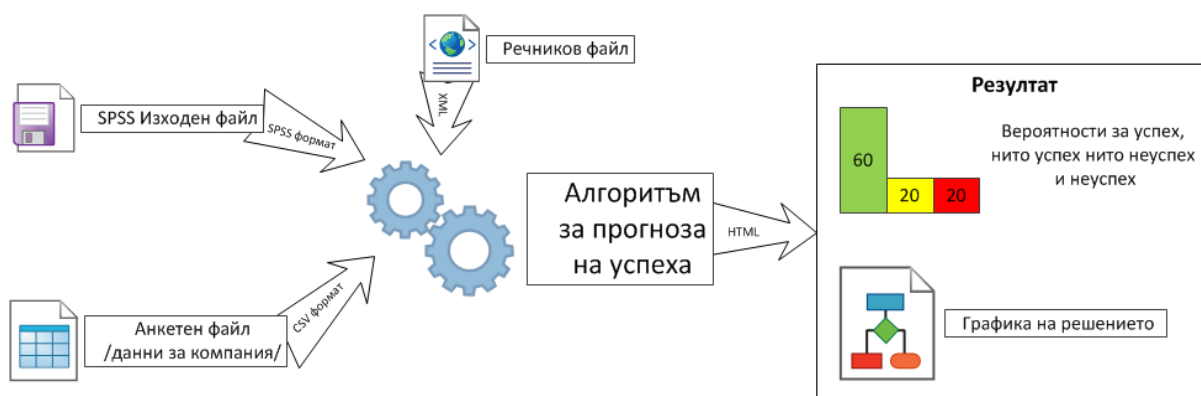
5.3.1 Алгоритъм за прогноза

Основната цел на I3SP е да предостави възможност за прогнозиране на вероятността за успех на стартиращи компании.

За да е възможно изготвянето на прогноза, трябва да разполагаме с попълнена анкета. След това, с помощта на класификационно дърво, създадено чрез софтуер за извличане на знания от данни, се прилагат правила, по които компанията се класифицира като „успешна“, „нито успешна, нито неуспешна“, или „неуспешна“. Класификационното дърво е записано и се съхранява във файл, като всеки ред от файла представлява възел от дървото. Софтуерът позволява актуализация на дървото. Алгоритъмът за прогноза работи, като взема данните от анкетния файл, прилага правилата, записани в класификационното дърво, прилага правилата за преобразуване на данните и връща резултат. Последователността на описания процес е илюстрирана чрез UML диаграма на действията (Фиг. 50)



Фиг. 50 UML диаграма на действията при прогнозиране успеха на стартираща компания с I3SP (създадена от автора)



Фиг. 51 Схема на алгоритъма за прогноза на приложението (създадена от автора)

Реализацията на алгоритъма за прогноза започва със зареждането на необходимите входящи параметри според показаната схема (Фиг. 51), което се осъществява чрез извикването на следните методи:

```

$survey_data = FileService::convertCSVStringToAssocArray($survey_string);
$dictionary = FileService::convertDictionaryToArray();
$prediction_tree_array =
FileService::convertPredictionTreeToArrayWithTranslation($prediction_file,
$dictionary);

```

Изброените методи преобразуват необходимата информация от файлове с данни в асоциативни масиви в PHP за по-удобна работа. При създаването на масива „\$prediction_tree_array“, отговарящ на файла с класификационно дърво от софтуера за извличане на знания от данни, се изпълняват и някои допълнителни дейности – създават се идентификационни номера за всеки възел и се изчислява и записва предшестващия възел, за да може да се получи дървовидна структура, представена в масив.

В алгоритъма се използва файл-речник, който служи за свързване на данните от анкетата с класификационното дърво, и който е специфичен за въпросите от анкетата. Към данните от класификационното дърво се добавят всички данни от файла-речник „\$dictionary“ за съответните възли от дървото. Към отговорите от анкетния файл също се добавят данните от файла-речник. Чрез тази операция данните се унифицират и се елиминира проблема с различното представяне на данните в класификационното дърво и анкетния файл.

На следващата стъпка се прави обхождане на класификационното дърво. Обхождането не се прави рекурсивно, а последователно - на нива. Всички възли на дадено ниво в дървото трябва да са с едно и също име (да съдържат един и същи анкетен въпрос), което е специфична особеност на файла с класификационно дърво. Възлите се различават по стойностите, които притежават за различните отговори на този въпрос. Ако допуснем, че на ниво 1 в дървото след въпроса „Фирмата притежава конкурентно предимство“ има два възела и единият от тях е разклонение за отговор „Да“, а другият - за отговор „Не“, то тогава, за класифициране на компанията е достатъчно да се

проверим кой възел съдържа същата стойност, каквато е дадена и в анкетата. След намирането на този възел, алгоритъмът елиминира останалите възли от същото ниво и продължава да обхождането в по-долно ниво, като се интересува само от възли на които избраният възел от предното ниво е предшественик. Процесът продължава до достигането на краен възел (лист).

Данните от крайния възел за процент вероятност за успех на компанията всъщност са резултат от прогнозата. Има и частен случай, в който при обхождането на дървото за прогноза може да се достигне до възел, който не съдържа валидна стойност според отговора от анкетата. Това е възможно например, когато в анкетата има пропуснат отговор за въпрос, участващ в класификационното дърво. В тези случаи, алгоритъмът връща резултат (прогноза за успеха) от последния обходен възел и дава индикация, че не е завършил изцяло обхождането.

5.3.1.1 Файл-речник

Софтуерът IZSP приема данни от различни източници: отговорите на анкетата от компанията (съдържащи се в анкетния файл от Google диск), класификационното дърво (което се генерира от софтуер за извличане на знания от данни). Данните в анкетния файл са под формата на текст, описващ всеки въпрос и прилежащия му отговор, а данните във файла с класификационното дървото са комбинация от числа и съкратени низове. Част от числата имат същото значение, като отговорите в анкетния файл. Например „1“ означава напълно несъгласен при въпросите, на които се отговаря със степен на съгласие или несъгласие. Всички съкратени низове са имена на променливи, използвани в софтуера за извличане на знания от данни и отговарят на въпроси на анкетата. Например низът „КонкурентнопредимствоМоятбизнесп“ се асоциира с въпроса „Конкурентно предимство [Моят бизнес притежава ясно конкурентно предимство.]“. За да се уеднакви представянето на данните и да се улесни работата на приложението, е създаден специален файл-речник във **формат XML**. Файлът се управлява от администраторите на софтуера, които могат да го актуализират. На Фиг. 52 е показан фрагмент от съдържанието на файла-речник.

```
<dictionary>
  <element index="R" type="scale">
    <short_name>Опит_Имамуправленскиопит</short_name>
    <question>Умения и опит [Имам управленски опит.]</question>
    <answers>
      <item key="1.000" value="Напълно несъгласен"/>
      <item key="2.000" value="По-скоро несъгласен"/>
      <item key="3.000" value="Нито съгласен, нито несъгласен"/>
      <item key="4.000" value="По-скоро съгласен"/>
      <item key="5.000" value="Напълно съгласен"/>
    </answers>
  </element>
</dictionary>
```

Фиг. 52 Фрагмент от файла-речник, използван в приложението (създадена от автора)

Файлът-речник е йерархично структуриран в XML формат, като всеки таг „<element>“ представлява въпрос от анкетния файл и има два атрибута: „index“ и „type“, които показват съответно номера на колоната в анкетния файл (представен чрез електронна

таблица) и типа на въпроса в анкетата (свободен текст, скала и др.). Към всеки елемент има вложени тагове. Един от тях е „<question>“ тагът, който съдържа текста на съответния въпрос от анкетата. Краткото представяне на въпроса, използвано в класификационното дърво, е записано в таг „<short_name>“.

Възможните отговори на въпросите от анкетата са записани в тага „<answers>“, съдържащ няколко елемента „<item>“. Атрибутите на таг „<item>“ съдържат възможните отговори, представени като низ (използван в анкетата) и числова стойност (използвана в класификационното дърво). Отговорите на някои въпроси от анкетата са в свободен текст, като в такъв случай не е практично за всеки възможен въведен низ да има представена стойност и се използва друг атрибут за „<item>“ тага, който съдържа шаблон, съответстващ на различни низове.

5.3.1.2 Регулярни изрази

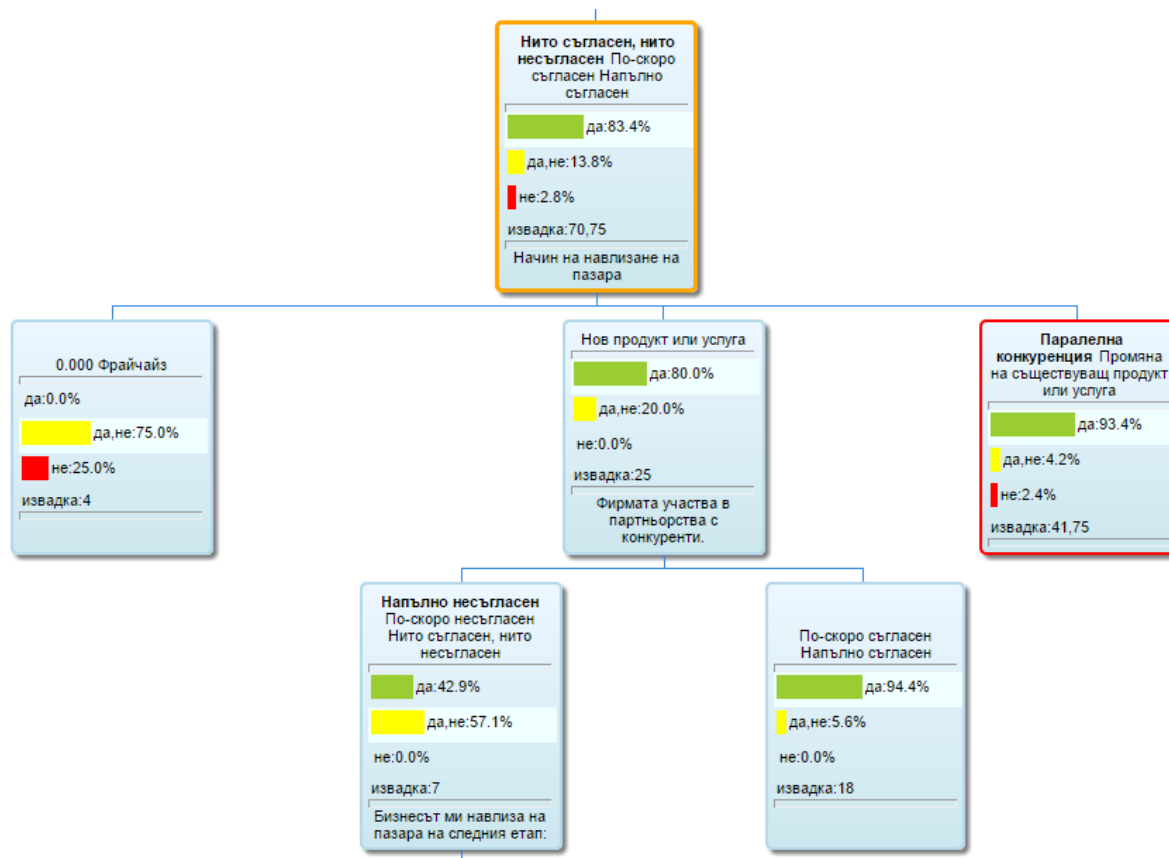
Файлът, генериран от софтуера за извличане на знания от данни, е текстов и съдържа данни за структурата и възлите на класификационното дърво, представени чрез специфичен синтаксис. Този файл се използва като входен параметър за алгоритъма за прогноза и е ключов за работата на I3SP. За разпознаване на данните от файла се използва регулярен израз (Фиг. 53), който се прилага чрез PHP функцията preg_match(). В променливата „\$parsedOK“ се записват обработените данни.

```
$parsedOK = preg_match('/(?P<offset>\s*)(?P<name>[a-zA-Z\p{Cyrillic}0-9_-\s]+)'
    . '( in \[ ]*( = )*(?P<values>[0-9. ,]+) ( \] )*( )*\[ Mode: (?P<currentMode>\d) \]'
    . '( => )?(?P<exitMode>[0-9,]*) \((?P<Ncount>[0-9,]+) [; ]*'
    . '(?P<Probability>[0-9,]*\)) \((?P<percent_no>[0-9. ,]*);'
    . '(?P<percent_yesno>[0-9. ,]*);(?P<percent_yes>[0-9. ,]*\)) [ $] */u'
, $line
, $matches);
```

Фиг. 53 Регулярен израз за извличане на данни от файл, съдържащ класификационно дърво (създадена от автора)

5.3.2 Графика на резултата за прогноза

Основната цел на I3SP е да изчисли вероятността за успех на стартиращи компании. Алгоритъмът за прогноза връща като резултат вероятността за успех, но резултатите, поднесени по-този начин на крайния потребител са неясни и непълни (според отговорите на респондентите в проведеното качествено изследване), тъй като не съдържат информация за работата на алгоритъма и за факторите, довели до решение. За да бъде приложението по-полезно за крайния потребител, е реализирана **графика тип класификационно дърво** (Фиг. 54), която показва нагледно как алгоритъмът обхожда възлите на дървото и защо са се получили резултатите.



Фиг. 54 Фрагмент от класификационното дърво, представено в страницата за анализ (създадена от автора)

За визуализацията на дървото е използван инструмент за визуализация на графика, който е част от програмния интерфейс Google Visualization API (Google Inc., 2014). Той е реализиран на езика JavaScript, изпълнява се в клиентския уеб браузър и служи за изчертаване на графика с дървовидна структура.

Интерфейсът очаква като параметър специално форматиран обект във формат JSON, който съдържа информацията за структурата на дървото (връзките между отделните възли), както и информация за съдържанието на конкретните възли. Всеки възел от класификационното дърво съдържа в себе си въпрос от анкетата, вероятността за успех на компанията и стойности на отговори на анкетата за предходния въпрос, при които алгоритъмът ще достигне до възела. Във всеки възел е показана и информация за това кой точно от възможните отговори за този възел е даден от потребителя (т.е. във възела е нанесен отговора от текущата анкета).

Алгоритъмът за прогноза обхожда класификационното дърво и при всеки посетен възел, който е част от решението, се записва информация за възела. В края на обхождането се получава колекция, съдържаща пътя на обхождане. Накрая всички данни се подават към шаблон за дизайн, който се интерпретира от Smarty, и който генерира HTML съдържанието на конкретния възел. Конструира се JSON низ, който съдържа информация за изчертаване на цялото дърво, като се вземат данните от обхождането на всички възли. JSON низът се записва директно в изходния код на страницата като JavaScript променлива. Следва фрагмент от JavaScript кода (от шаблонния файл "templates/survey_data_analyse.tpl"), който се грижи за изчертаване на графика:

```
var data = google.visualization.arrayToDataTable({$jsonTree});
var chart = new
google.visualization.OrgChart(document.getElementById('pnlChart'));
chart.draw(data, {allowHtml:true});
```

Създава се JavaScript обект “chart”, който получава JSON низ с инструкции за изчертаване на дървото и се извиква метод за изчертаване „chart.draw“.

5.3.3 Програмни интерфейси на Google

Програмните интерфейси на Google (Google APIs) (Google Inc., 2014) предоставят възможност на различни софтуерни приложения да взаимодействат с приложения, създадени от Google. Примери за такива програмни интерфейси са Google диск (Google Drive), Google карти (Google Maps), социална мрежа (Google+) и други. Спецификата на I3SP изисква работа с подобни програмни интерфейси, тъй като анкетните резултати, използвани от софтуера, се съхраняват на отдалечени сървъри на Google диск. Програмните интерфейси, които ще осигуряват достъп до тези данни са:

- **Google Drive API** - за работа с файловете съхранявани в Google диск,
- **Google Spreadsheet API** – за работа с електронни таблици съхранявани в Google диск,
- **Google Visualization API** – програмен интерфейс, който дава възможност за показване на данните от Google диск в различни графични формати.

5.3.3.1 Програмен интерфейс Google диск

Google диск (Google Inc., 2014) предоставя възможност на потребители на Google да създават, променят и съхраняват различни типове файлове в облачните масиви на Google, като предоставя удобни способи за споделяне, защита и резервни копия на тези данни. Анкетният файл, който I3SP обработва, е създаден с помощта на Google анкети (Google Forms) и се съхранява в пространството на Google диск.

Програмният интерфейс за работа с Google диск ще позволи работата на приложението с анкетния файл. Той предлага възможности за създаване на нови файлове в различни формати, за сваляне на файл, за достъп до съдържанието и мета данните на файл.

В I3SP има нужда от механизъм, който позволява на потребителите да преглеждат съдържанието на собствените си файлове в Google диск и да избират подходящ анкетен файл (създаден от тях или споделен от друг потребител). За реализирането на тази част от функционалността на приложението, ще бъдат използвани функционалностите за избор и преглед на файл на потребителския интерфейс Google диск.

Програмният интерфейс Google диск може да се ползва само след успешен вход с протокол OAuth 2.0 (OAuth community, 2014) и разрешаване на достъп.

5.3.3.2 Програмен интерфейс Google електронна таблица

Програмният интерфейс на Google за работа с електронни таблици (Google Inc., 2014) предоставя широк набор от функционалности, които могат да бъдат прилагани върху електронни таблици, съхранявани в Google диск. Този интерфейс дава възможност за извличане на съдържанието на цяла електронна таблица, извличане на съдържанието само на отделен работен лист (worksheet), извличане на данните за определени редове

от този работен лист, а също и извличане на данни от определена клетка в избран работен лист. Поддържат се също и заявки, написани на специален език, подобен на SQL за извличане и комбиниране на специфична информация от електронна таблица, сортиране и групиране по определени критерии.

Приложението за прогнозиране успеха на стартиращи компании има необходимост от извличане на определен ред от електронната таблица на анкетния файл и сортиране на редовете на тази таблица, което се постига с подходящи заявки в програмния интерфейс.

Програмният интерфейс Google електронна таблица може да се прилага само след успешен вход с протокол OAuth 2.0 и разрешаване на достъп.

5.3.3.3 Програмен интерфейс Google графични инструменти

Програмният интерфейс на Google за работа с графични инструменти приема данни, изпратени от клиентския браузър под формата на XML и JSON, обработва ги и изпраща готов HTML код който визуализира тези данни посредством клиентския браузър. Предоставя се възможност за визуализация на данните в различни графични обекти, като таблици, диаграми, дървета и много други. I3SP използва този програмен интерфейс за създаването на дървовидна графика, която изобразява работата на алгоритъма за прогноза.

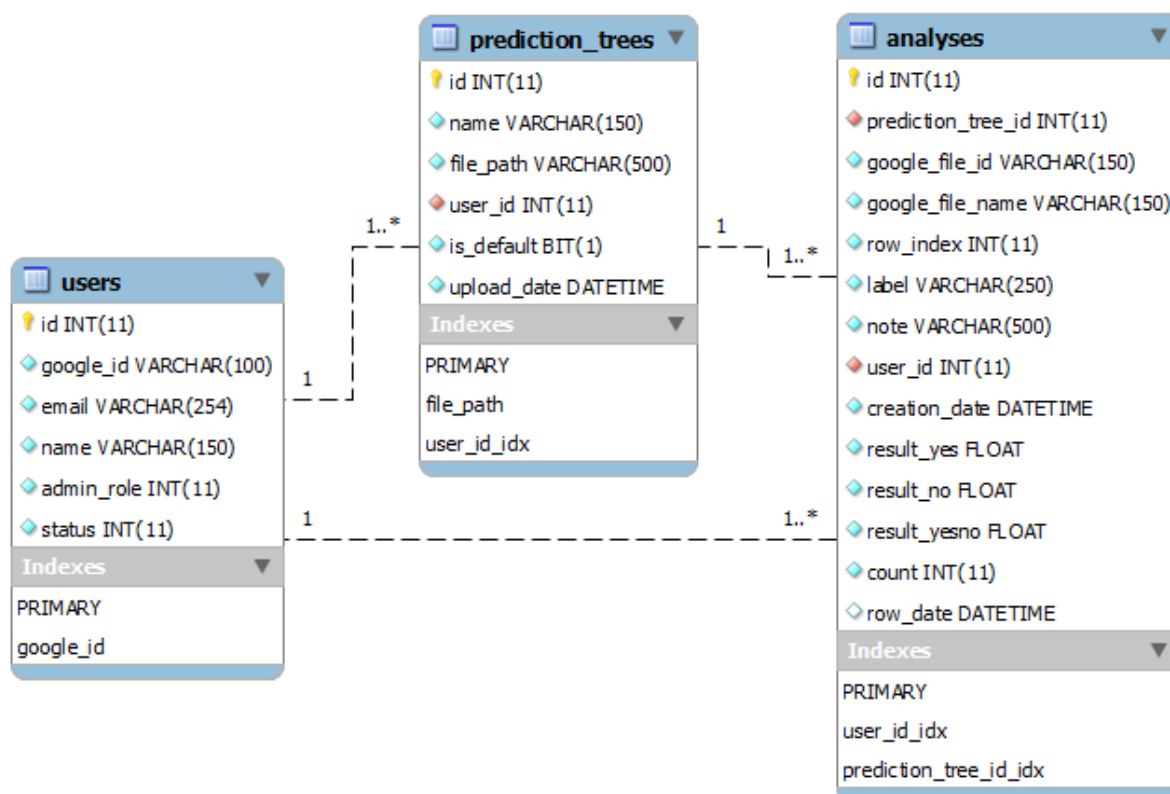
Ползването на програмния интерфейс на Google за работа с графични инструменти не изисква аутентикация.

5.3.4 Съхранение на данни

I3SP разчита на три различни източника на данни за своята работа. Това са данни, съхранявани в локална база данни, в локални файлове, запазени на файловата система на сървъра, както и данни, съхранявани на отдалечени сървъри и достъпни само чрез програмен интерфейс. Следва да бъде разгледана употребата и на трите различни метода за съхраняване на данни, използвани в приложението.

5.3.4.1 Структура на базата данни

Базите данни предоставят много бърз и удобен начин за съхранение и извличане на данни. За целите на I3SP, ще бъде използвана MySQL база данни. Данните, който ще бъдат записани в базата данни, са свързани с потребителите на приложението, класификационното дърво (генерирано от софтуер за извличане на знания от данни), както и данни за вече направени анализи. Информацията в базата данни ще бъде съхранявана на български и английски език, затова е избрана колация “utf8_general_ci“, която поддържа UTF8 кодиране на символите, включващо кирилица, латиница и други азбуки и символи. Фиг. 55 илюстрира структурата на базата данни, ползвана от приложението. От диаграмата се вижда, че приложението ще използва три таблици: users, prediction_tree и analyses.



Фиг. 55 Диаграма на структурата на базата данни на приложението (създадена от автора)

Таблица users съдържа локално съхранявана информация за данните на потребителите. Голяма част от тази информация ще бъде копирана от Google акаунтите. Таблицата има създаден първичен ключ (колоната „id“), който се използва в останалите две таблици и служи за връзка с тях (релация). В users има създадено ограничение имейлите да са уникални (unique constraint). Колоната admin_role съдържа информация за правата на даден потребител в системата – дали е администратор или обикновен потребител.

В **таблицата prediction_trees** се съхранява информация за различни класификационни дървета, които приложението използва. В таблицата се съхраняват само мета данни за дърветата, генерирани от IBM SPSS Modeler или от други приложения за извличане на знания от данни. Самите дървета се съхраняват под формата на текстови файлове във файловата система, а таблицата съхранява пътя до файловете. Тя притежава уникален първичен ключ, който ще бъде използван като връзка с таблица analyses, както и ключ към таблица users, показващ кой потребител е създал записа в таблица prediction_trees.

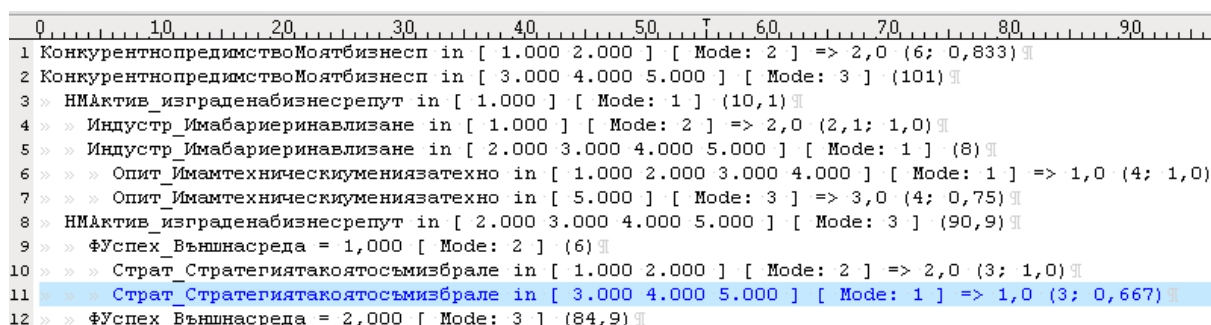
Таблица analyses съдържа архив на информацията за резултатите от прогнозите за успех, направени от различни потребители за различни компании. Тя съдържа полета за съхранение на име и идентификатор на Google диск файл (анкетен файл), за който се отнася прогнозата, както и индекс (номер) на ред от файла. В таблицата се съдържат още външни връзки с таблиците users и prediction_trees, които съхраняват данни съответно за потребителя, запазил прогнозата, и за алгоритъма, използван при генерирането ѝ. Таблицата съхранява само резултатите от прогнозите за успех (процент вероятност за успех) и не съдържа данни от анкетния файл и дървото.

5.3.4.2 Локални файлове

Приложението използва локални файлове за съхранение на някои специфични видове данни – речник-файл и класификационно дърво.

Речник-файлът е ключов за работата на приложението и е структуриран като XML документ, съдържащ информация, описваща правилата, по които данните от класификационното дърво са съпоставени с данните от анкетата. РНР предлага удобни и доста бързи програмни средства за работа с XML.

Класификационните дървета са алгоритми във вид на дърво, създадени чрез софтуер за моделиране - IBM SPSS Modeler или Weka. Това са текстови файлове, форматиран по специфичен начин, зададен от софтуера за извличане на знания от данни. Всеки нов ред от тези файлове представлява възел от класификационното дърво. Данните във всеки ред имат еднотипен формат и дават информация за конкретния възел от дървото, ниво на вложеност, и съответстващото на възела решение за успеха на компанията.



```
0 10 20 30 40 50 60 70 80 90
1 КонкурентнопредимствоМоятбизнес in [ 1.000 2.000 ] [ Mode: 2 ] => 2,0 (6; 0,833)
2 КонкурентнопредимствоМоятбизнес in [ 3.000 4.000 5.000 ] [ Mode: 3 ] (101)
3 » НМАктив_изграденабизнесрепут in [ 1.000 ] [ Mode: 1 ] (10,1)
4 » » Индустр_Имабариеринавливане in [ 1.000 ] [ Mode: 2 ] => 2,0 (2,1; 1,0)
5 » » » Индустр_Имабариеринавливане in [ 2.000 3.000 4.000 5.000 ] [ Mode: 1 ] (8)
6 » » » » Опит_Иматехническиумениязатехно in [ 1.000 2.000 3.000 4.000 ] [ Mode: 1 ] => 1,0 (4; 1,0)
7 » » » » Опит_Иматехническиумениязатехно in [ 5.000 ] [ Mode: 3 ] => 3,0 (4; 0,75)
8 » НМАктив_изграденабизнесрепут in [ 2.000 3.000 4.000 5.000 ] [ Mode: 3 ] (90,9)
9 » » ФУспех_Външнасреда = 1,000 [ Mode: 2 ] (6)
10 » » » Страт_Стратегиятакоятосъмизбрале in [ 1.000 2.000 ] [ Mode: 2 ] => 2,0 (3; 1,0)
11 » » » Страт_Стратегиятакоятосъмизбрале in [ 3.000 4.000 5.000 ] [ Mode: 1 ] => 1,0 (3; 0,667)
12 » » ФУспех_Външнасреда = 2,000 [ Mode: 3 ] (84,9)
```

Фиг. 56 Фрагмент от изходен файл на IBM SPSS Modeler (създадена от автора)

На Фиг. 56 е показан фрагмент от файл, съдържащ класификационно дърво, генерирано чрез IBM SPSS Modeler. Като част от анализа за прогноза на успеха на стартираща компания, дървото ще бъде визуализирано и ще бъде достъпно за преглед от потребителя. За целта ще бъде създаден механизъм, чрез който дървото ще се преобразува във формат, разбираем за софтуера. За реализирането на този механизъм ще се използва регулярен израз за извличане на данните от дървото и речник-файл, чрез който данните се преобразуват от един формат в друг, според нуждите на приложението.

5.3.4.3 Файлове, намиращи се на Google диск

I3SP има нужда от данни (анкетен файл), които да анализира и върху които да създаде прогноза за успех. Анкетата, създадена като част от модела за прогноза, е реализирана с помощта на инструмент на Google, наречен Google анкети (Google Forms) (Google Inc., 2014). Форматът, в който данните, подадени чрез анкетата, се записват, представлява редове от електронна таблица. Примерен фрагмент от анкетния файл е показан на Фиг. 57.

	A	B	C	D	E	F	G
1	Timestamp	Личност и ценности [Аз съм уверен в себе си.]	Личност и ценности [Аз съм предприемчив, инициативен.]	Личност и ценности [Аз съм амбициозен.]	Личност и ценности [Склонен съм да поемам рискове.]	Умения и опит [Имам умения по мениджмънт.]	Умения и опит [Имам умения по маркетинг.]
2	6/23/2013 9:05:51	По-скоро съгласен	Напълно съгласен	Напълно съгласен	По-скоро съгласен	По-скоро съгласен	По-скоро съгласен
3	6/27/2013 6:16:47	Напълно съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен
4	6/27/2013 23:10:36	Нито съгласен, нито несъгласен	По-скоро съгласен	По-скоро съгласен	По-скоро съгласен	Напълно съгласен	По-скоро съгласен
5	6/28/2013 11:06:40	Напълно съгласен	По-скоро съгласен	По-скоро съгласен	По-скоро съгласен	Нито съгласен, нито несъгласен	Нито съгласен, нито несъгласен

Фиг. 57 Фрагмент от Анкетен файл (създадена от автора)

Всеки ред от анкетния файл представлява отговор на компания, попълнила анкетата. Файлът съдържа повече от сто колони (съответстващи на въпроси от анкетата), а стойностите на редовете представляват данни за отговори на тези въпроси.

За да се прогнозира успеха на дадена компания, данните трябва да бъдат изтеглени от електронната таблица и обработени от приложението. Те не се записват локално, а се съхраняват в Google диск и се актуализират при попълване на анкети с Google анкети, и се достъпват от I3SP всеки път, когато са необходими. Директният достъп няма да генерира голям трафик, който да затрудни работата на приложението, тъй като обемът на предаваната информация не е голям. Механизмът на достъп до данните е посредством програмните интерфейси на Google.

За да бъдат преобразувани данните във формат, разбираем за софтуера, се използва речник-файл, който превръща данните от анкетата в удобни за ползване от приложението.

5.3.5 Функционалности, използващи програмни интерфейси на Google

Изискванията към I3SP налагат използването на анкетни файлове, които се намират на отдалечен сървър Google диск, аутентикация на потребителите с Google Account и графична визуализация на данните – функционалности, за които се прилагат програмни интерфейси на Google.

5.3.5.1 Аутентикация

За да се ползва даден програмен интерфейс на Google е необходимо да са изпълнени определени изисквания. Задължително е при заявка към интерфейса да се подава като параметър идентификационен ключ. Така интерфейсът удостоверява че даденото приложение (използващо потребителски Google акаунт) има права до поисканите ресурси (в случая – файл, разположен на Google диск).

Идентификационният ключ се получава след успешна аутентикация с протокол OAuth 2.0 и е валиден в рамките на 60 минути. Той се съхранява в сесийна променлива. Извикването на програмен интерфейс на Google се осъществява чрез заявяване на достъп до него, например „<http://spreadsheets.google.com/tq>“. Комуникацията между приложението и интерфейсите се осъществява чрез HTTP протокола.

За улеснение на разработчиците, с всеки програмен интерфейс се предоставят библиотеки и примери. Тези библиотеки са реализирани на различни програмни езици, като PHP, ASP.NET, Java, JavaScript. Използването на библиотеките не е задължително, но се препоръчва, тъй като позволява по-лесна и бърза разработка. При разработката на софтуера за прогнозиране на успеха е използвана библиотека за работа с Google диск, написана на PHP. Към програмните интерфейси на Google има обръщения както от

сървърната, така и от клиентската част. Сървърната част на приложението реализира клас „GoogleServices“ за работа с програмните интерфейси на Google и за извличане на данни от Google диск. Следващият фрагмент от програмен код показва конструктора на клас „GoogleServices“:

```
public function __construct() {  
    $this->client = new Google_Client();  
    $this->client->setApplicationName(APP_NAME);  
    $this->client->setClientId(CLIENT_ID);  
    $this->client->setClientSecret(CLIENT_SECRET);  
    $this->client->setRedirectUri(REDIRECT_URI);  
    $this->client->setScopes(array(  
        'https://www.googleapis.com/auth/drive',  
        'https://www.googleapis.com/auth/drive.file',  
        'https://www.googleapis.com/auth/userinfo.email',  
        'https://www.googleapis.com/auth/userinfo.profile',  
        'https://spreadsheets.google.com/feeds'));  
}
```

В конструктора се създава инстанция на клас „Google_Client“ (дефиниран в библиотеката на Google) и се задават всички необходими параметри за правилна работа с интерфейсите. Тази инстанция се явява поле в класа „GoogleService“ (\$this->client), чрез което се осъществява достъп до интерфейса в останалите методи на класа. Константите „CLIENT_ID“ и „CLIENT_SECRET“ се получават от разработчика при регистрация на приложението в Google и се съхраняват в конфигурационния файл на приложението.

5.3.5.2 Извличане на информация за потребителския профил

Следващият програмен фрагмент представлява метод от класа „Google_Client“, част от програмния код на I3SP, който се свързва с програмния интерфейс на Google за работа с потребители и извлича информация за Google профила на потребителя на приложението:

```
function getLoggedUserInfo() {  
    $token_decoded = json_decode($this->client->getAccessToken(),true);  
    $query = 'https://www.googleapis.com/oauth2/v1/userinfo?access_token=' .  
    $token_decoded["access_token"];  
    $user_info = file_get_contents($query);  
    return json_decode($json,false);  
}
```

5.3.5.3 Работа с Google диск

„GoogleServices“ е PHP клас, който притежава методи за извличане на данни (един ред или всички данни) от файл в Google диск. Следващият програмен фрагмент, показва метод от класа, извличащ цялото съдържание на файл от Google диск.

```
function getSurveyDataWhole($fileId){  
    $driveService = new Google_DriveService($this->client);  
    // Retrieve metadata for the file specified by $fileId.  
    $file = $driveService->files->get($fileId);  
    // Get the contents of the file.  
    $response = $this->client->getIo()->authenticatedRequest($request);  
    $Data = $response->getResponseBody();  
  
    return $Data;  
}
```

Другият начин, по който приложението се свързва с програмния интерфейс на Google е директно от клиентския браузър - с помощта на програмния език JavaScript. В страницата се зареждат един или няколко скриптови файлове директно от сървърите на Google - в зависимост от спецификите на програмния интерфейс. Това се осъществява във фронтенд частта на приложението, поради спецификите на програмния език JavaScript, който се интерпретира от клиентския браузър. Следва фрагмент от JavaScript код, който демонстрира работа с програмен интерфейс на Google чрез JavaScript.

```
this.picker = new google.picker.PickerBuilder().  
    addView(google.picker.ViewId.SPREADSHEETS).  
    setAppId(this.clientId).  
    setOAuthToken(this.accessT).  
    setCallback(this._pickerCallback.bind(this)).  
    build().  
    setVisible(true);
```

Цитираният програмен код извиква интерфейса за избор на файлове от Google диск (Google Picker API). При изпълнение на показания програмен код, се отваря диалогов прозорец на клиентския браузър и вътре се зареждат всички файлове (електронни таблици) от Google диск, които потребителят притежава.

I3SP работи и с интерфейса на Google за работа с електронни таблици чрез JavaScript. Характерно при него е, че има възможност за конструиране на заявки (подобни на

SQL), чрез които се извлича специфична информация от електронната таблица. Пример за такава заявка е:

```
query.setQuery('select * order by A desc limit ' + $('#last_rows').val());
```

Резултатът на тази заявка са последните редове от избрания анкетен файл.

Данните, получавани от програмните интерфейси на Google, могат да бъдат в различни формати, като най-често се използват JSON и XML.

5.3.6 Дизайн шаблони

В I3SP се използва инструментът за шаблони Smarty, който разделя кода, реализиращ бизнес логиката на приложението (разработен на PHP) от кода, реализиращ визуалната част от приложението (разработен на HTML, CSS, JavaScript). Това спомага за разделяне на слоя изглед от слоя модел от трислойната архитектура (MVC). Smarty е реализиран на програмния език PHP и кодът му се съхранява в директорията на основното приложение. Освен това, се създават и няколко служебни папки между които е и папката „templates“. В нея по-подробно се записват всички шаблони, използвани от приложението. Всички изгледи, които се подават към крайния потребител, са реализирани с шаблони за дизайн на Smarty. Данните, подавани от слоя реализиращ логиката към генератора за шаблони, се енкапсулират и са видими само в рамките на шаблона. Не се налага смесването на PHP и HTML код, което подобрява дизайна на приложението.

На Фиг. 58 е показан фрагмент от код на дизайн шаблон, който съдържа част от фронтенд кода на страницата за настройки на алгоритмите за прогноза в приложението.

```
{foreach name=outer item=prediction from=$predictions}
<tr>
  <td>{$prediction["name"]} </td>
  <td>{$prediction["file_path"]} </td>
  <td>{if $prediction["is_default"] eq 1}Да{else}Не{/if} </td>
  <td>
    <a target="_blank" href="{ $prediction["file_path"] }">преглед</a>
    {if $prediction["is_default"] eq 0}
    | <a href="{ $SCRIPT_NAME }?
    p=admin_prediction_tree&setdefault=&file_id={$prediction["id"] }">по подразбиране</a>
    <a href="{ $SCRIPT_NAME }?
    p=admin_prediction_tree&delete=&file_id={$prediction["id"] }">изтрий</a>
    {else}{/if}
  </td>
</tr>
{foreachelse}
```

Фиг. 58 Фрагмент от кода на дизайн шаблон в Smarty (създадена от автора)

5.3.7 Сигурност и надеждност

Сигурността и надеждността на I3SP са от съществено значение от потребителска гледна точка и стават все по-важни с популяризация на приложението. За да се осигури

стабилната и безпроблемна работа на приложението, както и сигурността на данните на потребителите, се изисква да бъдат взети специални мерки, които ще бъдат разгледани в настоящата точка.

5.3.7.1 Спазване на добри практики

Чести варианти за подаване на некоректни данни са:

- инжектиране на заблуждаващ SQL код през методите GET и POST на HTTP протокола (**SQL injection**). Най-често се осъществява чрез манипулирани стойности в текстови полета и URL параметри;
- качване на файлове с некоректни имена, което при определени обстоятелства може да доведе до запис и стартиране на файл извън регламентираните места на сървъра (**file injection**);
- инжектиране на скриптов код в текстовите полета, който нерегламентирано събира и изпраща данни от клиентския браузър към други сървъри (**XSS - cross site scripting**);
- извличане на информация за сесийни променливи (**session hijacking**) и/или манипулиране на техните стойности (**session fixation**), обикновено с цел получаване на неоторизиран достъп.

Приложението I3SP разполага с механизми, за превенция на подобни често срещани атаки. За защита от „SQL injection“ приложението използва стандартна PHP функция `mysqli_real_escape_string()`. За защитата от XSS данните от текстови полета винаги се филтрират през PHP функцията `htmlspecialchars()`, която замества чрез кодиране опасните HTML символи.

5.3.7.2 Валидация, филтриране и подготовка на данните

Част от данните, с които I3SP работи, включително файлове и съдържание в базата данни, се подават от потребителите на приложението. Поради различни причини - умишлени и неумишлени, тези данни могат да бъдат неверни, непълни, в некоректен формат, или манипулирани. В такива случаи, данните биха могли да предизвикат непредвидени последици при работата на приложението, ако не е спазена практиката те да бъдат проверени и подготвени преди използването им.

Всички данни, подавани от потребителя, преди тяхното използване, се обработват, чрез програмните средства, предвидени в езика PHP. Например, данните, които ще бъдат записвани в базата данни, се подготвят с PHP функцията `mysqli_real_escape_string()`, която ги преобразува в безопасни за работа. Потребителските данни, които се използват във фронтенда (генерираните HTML документи), се преобразуват с PHP функцията `htmlspecialchars()`, която замества служебните за езика HTML символи със специални кодове.

Качените от потребителите файлове се записват със служебно генерирано от системата име, за да се избегне нежелано изпълнение. Приложението разполага и с методи за валидация на съдържанието на анкетния файл и на файла с класификационно дърво, генериран от софтуера за извличане на знания от данни. Преди стартиране на алгоритъма за прогноза, винаги се проверява дали файлът- речник е валиден. За файла с класификационно дърво също се прилага валидация – при неговото качване се стартира метод, който с помощта на регулярни изрази удостоверява дали съдържанието на файла представлява информацията, необходима за работа на приложението.

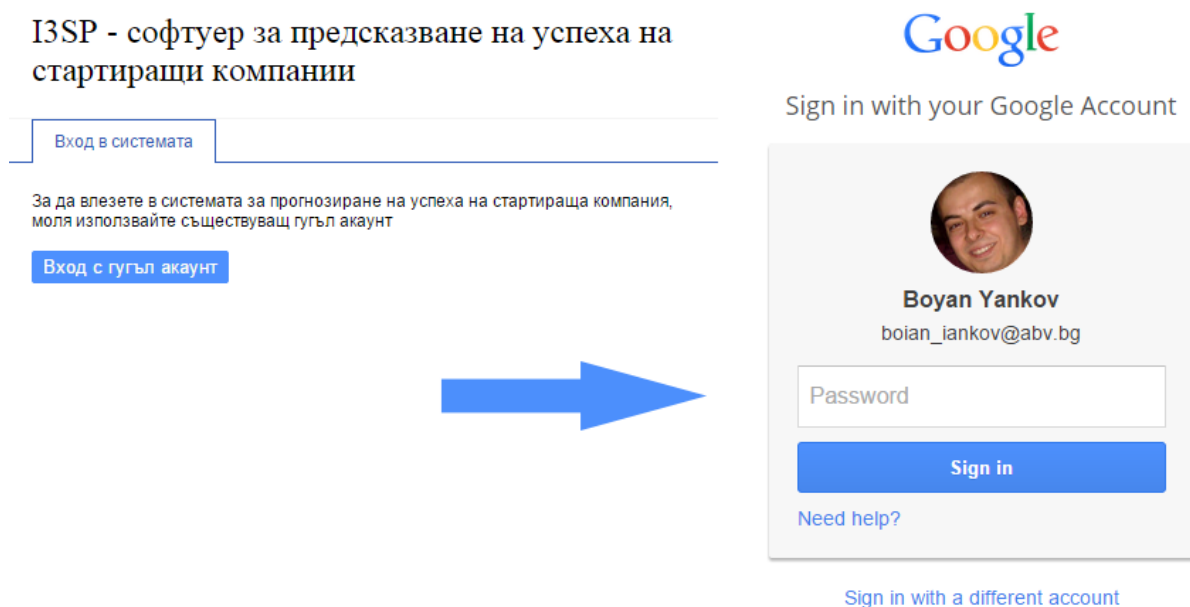
При некоректно подадени данни, системата обработва данните, и извежда на екрана на потребителя подходящо съобщение за грешка, което описва какъв проблем е настъпил.

5.4 Тестване на софтуера за прогнозиране на успеха

В тази точка е описана работата на софтуера за предсказване успеха на стартиращи компании чрез използване на примери, графики, тестови сценарии и коментари. Показани са снимки на екраните на основните функционалности, представящи потребителския интерфейс и начина на използване на приложението и модулите.

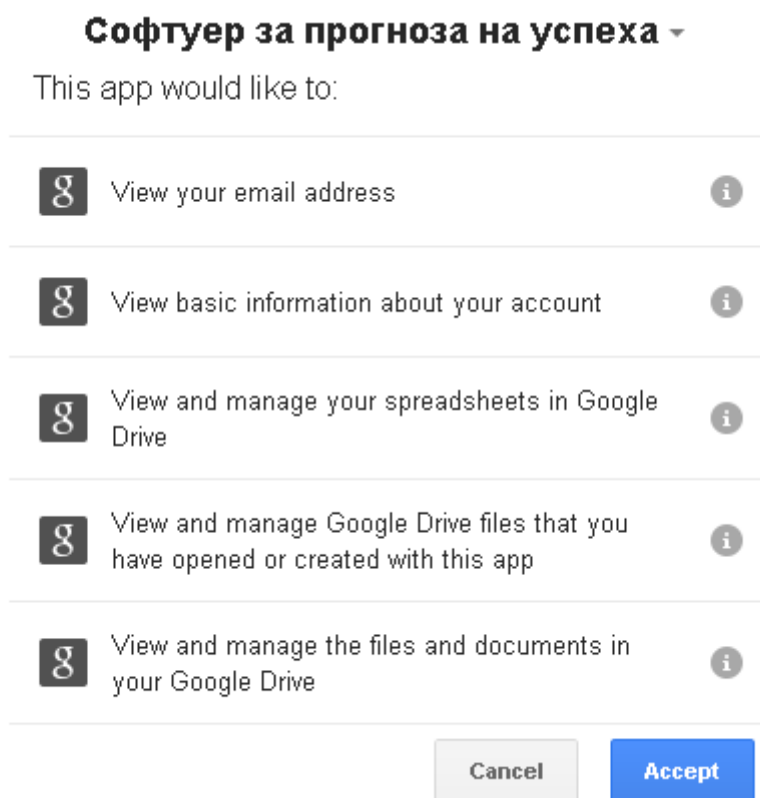
5.4.1 Вход в системата

Софтуерът може да бъде използван от потребители, които имат валидни Google акаунти. При първоначален достъп на даден потребител до приложението, се зарежда страница с кратко описание и бутон за вход (Фиг. 59). Не се предоставя традиционната функционалност за регистриране на потребители.



Фиг. 59 Вход в приложението с Google акаунт (създадена от автора)

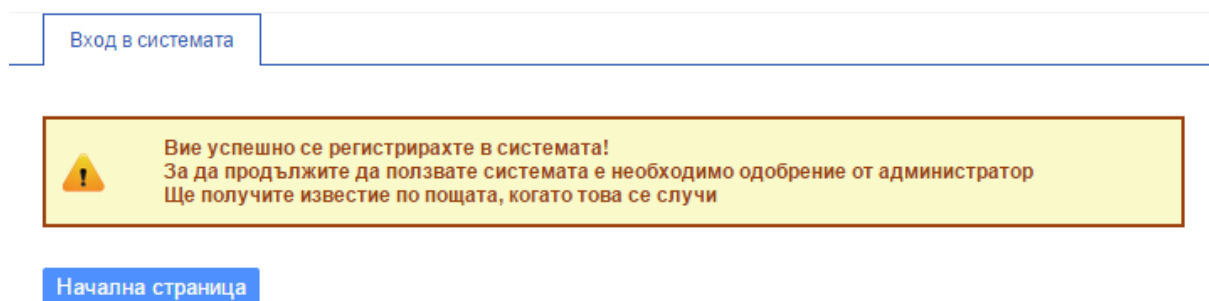
След натискането на бутона за вход, ако потребителят няма активна Google сесия за този браузър (session cookie), се отваря прозорец за въвеждане на име и парола. След въвеждането на правилни име и парола, или при наличие на активна сесия, потребителят се пренасочва към диалогов прозорец за разрешаване на достъп до Google диск (Фиг. 60). Този прозорец обяснява на потребителя, че приложението изисква достъп до определени ресурси от Google акаунта на потребителя: информация за акаунта и достъп до Google диск файловете му.



Фиг. 60 Диалогов прозорец за разрешаване на достъп до Google диск (създадена от автора)

При натискане на бутон „Асерт”, потребителят се съгласява с условията и разрешава достъп на приложението. Ако потребителят вече е активиран от администратор, той директно влиза в приложението и отива на страницата за настройки. Ако не е активиран, потребителят получава съобщение, обясняващо защо няма достъп до приложението към момента (Фиг. 61) и изчаква активация от администратор.

I3SP - софтуер за предсказване на успеха на стартиращи компании



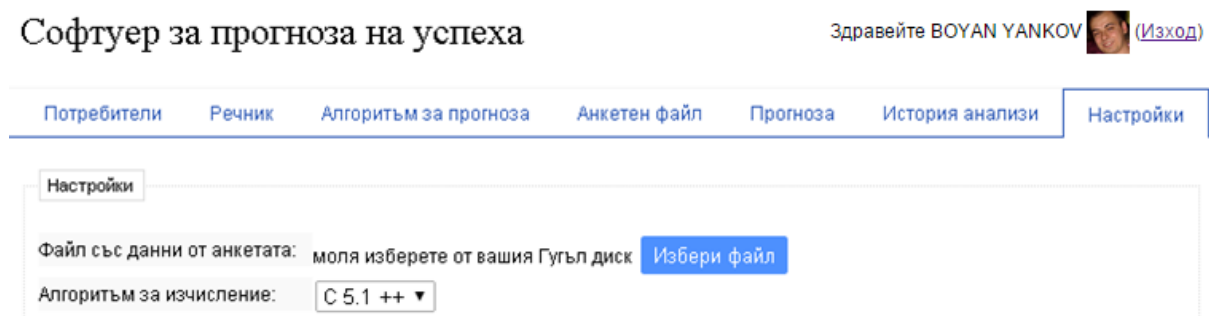
Фиг. 61 Изчакване на одобрение след изпратена заявка за достъп до системата (създадена от автора)

Бяха успешно извършени следните тестови сценарии:

1. Проверка за невалиден потребител: Опит да се влезе в системата с невалиден Google акаунт
2. Проверка за валиден потребител: Проверява се, че след въвеждане на валиден акаунт потребителят вижда прозореца за достъп.
3. Натискане на бутона „Cancel“ в прозореца за достъп: Проверява се, че след успешен вход в системата, ако потребителят натисне бутона „Cancel“, той не може да получи достъп до системата за прогноза.
4. Вход с вече активиран потребител: Проверява се, че след вход на вече активиран потребител, той влиза успешно в приложението.
5. Вход с неактивен потребител: Проверява се, че след вход на потребител за първи път, се показва съобщение за активация и се изпраща емайл до администратора.

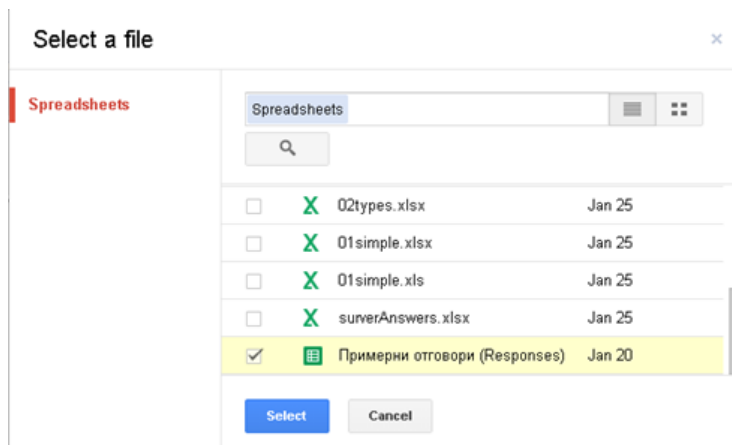
5.4.2 Настройки на приложението

Веднага след успешно влизане в системата, на потребителя се показва страницата за настройки, от която той може да избере анкетен файл и класификационно дърво (алгоритъм) за анализ (Фиг. 62). Алгоритъмът по подразбиране е избран предварително от администратора на системата, но текущия потребител може да промени тази настройка.



Фиг. 62 Страница с настройки в приложението (създадена от автора)

Настройката за избор на анкетен файл е задължителна за работата на приложението, за да бъде възможно изготвянето на анализи за успеха на компанията. При натискане на бутона „Избери файл“, се визуализира диалогов прозорец, съдържащ имената на всички файлове (от тип електронна таблица) от Google диск пространството на потребителя (Фиг. 63).



Фиг. 63 Диалогов прозорец за избор на анкетен файл от Google диск (създадена от автора)

След избора на файл, се показва името на избрания файл и линк, предоставящ опция за неговата смяна.

Бяха успешно извършени следните тестови сценарии:

1. Избор на файл: Проверка, че файл може да бъде избран след натискането на бутона „Анкетен файл” и избиране на файл от диалоговия прозорец
2. Смяна на анкетния файл: Проверка, че файлът може да бъде сменен, след като вече има избран файл.
3. Прекратяване избор на файл: Проверка, че след натискане на бутона „Cancel“ в диалоговия прозорец, анкетният файл не се сменя.
4. Смяна на изчислителния алгоритъм: Избор на нов алгоритъм, и проверка, че настройката за алгоритъм е запазена успешно.

5.4.3 Преглед на съдържанието на анкетен файл

След като потребителят е избрал анкетен файл от настройките на приложението, за него става достъпна страницата за преглед на съдържанието на избрания файл. Фрагмент от страницата е показан на Фиг. 64.

Потребители

Речник

Алгоритъм за прогноза

Анкетен файл

Прогноза

История анализи

Настройки

Анкетен файл

Съдържание на файл:

Анкета - Модел за предкоказване на успеха на стартиращи фирми (Responses)

В момента са показани последните 10 записа на файла. Ако искате видите повече ползвайте филтъра по долу:

Покажи последните:

10

Филтър

За да анализирате ред от анкетата моля изберете ред и натиснете върху него

Timestamp	Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]	Личност и ценности [Аз съм уверен в себе си.]	Личност и ценности [Аз съм предприемчив, инициативен.]	Личност и ценности [Това, което ми се случва в живота зависи от мен самия.]	Личност и ценности [Аз съм амбициозен.]	Личност и ценности [Склонен съм да поемам рискове.]	Личност и ценности [Не се притеснявам да вземам решения без да разполагам с необходимата информация.]	Умения и опит [Имам предприемачески умения (умения в стартирането на бизнес).]	Умения и опит [Имам умения по мениджмънт.]	Умения и опит [Имам умения по маркетинг.]	Умения и опит [Иматехнически умения / технологически умения / свързан с моя бизнес.]
9/15/2014 16:24:08	По-скоро съгласен	Напълно съгласен	По-скоро съгласен	Напълно съгласен	Напълно съгласен	По-скоро съгласен	По-скоро съгласен	Напълно съгласен	Напълно съгласен	Нито съгласен, нито несъгласен	Напълно съгласен
5/15/2014 14:47:51	Напълно съгласен	Напълно съгласен	По-скоро съгласен	Напълно съгласен	Напълно съгласен	Напълно съгласен	По-скоро съгласен	По-скоро съгласен	Нито съгласен, нито несъгласен	По-скоро несъгласен	Напълно съгласен

Фиг. 64 Страница с визуализация на съдържанието на анкетен файл (създадена от автора)

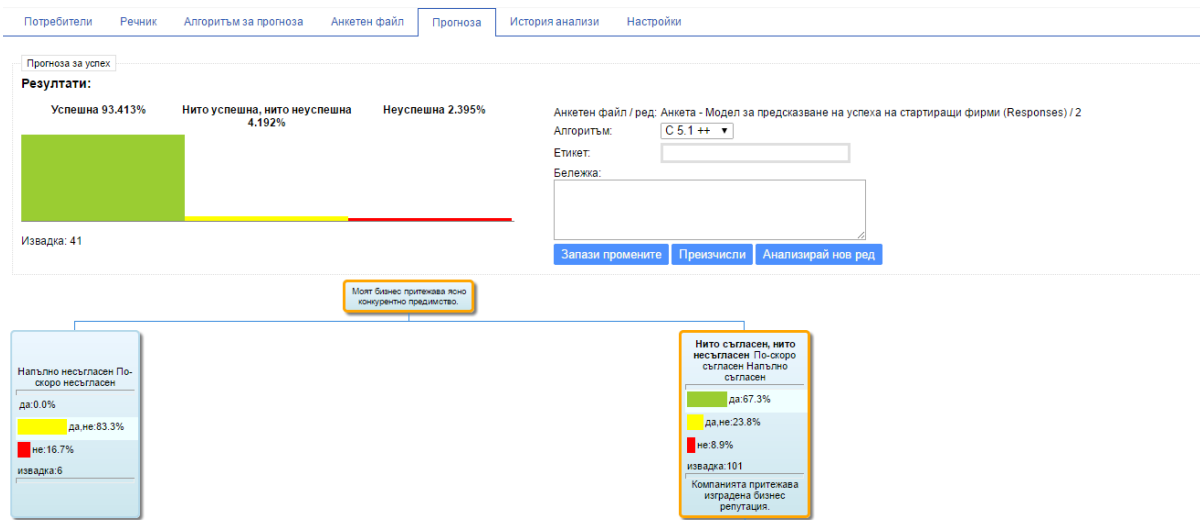
Данните от анкетния файл се показват в табличен вид. Тъй като файлът съдържа повече от 100 колони с данни (съответстващи на въпроси от анкетата), в повечето от които се съдържат множество думи, за да се визуализират целите редове се използва хоризонтален плъзгач (скрол). В страницата е предвиден филтър, който по подразбиране показва последните 10 реда от анкетния файл. След преглед на данните, потребителят може да избере желанния ред за анализ, като натисне един път с мишката върху него, което го препраща към страницата за прогноза.

Бяха успешно извършени следните тестови сценарии:

1. Проверка за показваните данни: Проверка, че се показват последните 10 реда от анкетния файл, и ако той е с по-малко от 10 реда, че се показват всички редове.
2. Прилагане на филтър: Проверка, че след прилагане на филтър, се показват точния брой редове и, че те съответстват на тези в анкетния файл.
3. Избор на ред от анкетния файл: Проверка, че след избор на ред от анкетния файл, потребителят е препратен на страницата за прогноза за съответното попълване на анкетата.
4. Коректно съобщение, когато няма избран анкетен файл: Проверка, че се показва подходящо съобщение, когато няма анкетен файл или той е с несъвместимо съдържание.

5.4.4 Прогноза за успех

Най-важната част от приложението I3SP е модулът за прогнозиране на успеха. Потребителят получава достъп до страницата за прогноза, след като избере ред от анкетния файл за анализ, който отговаря на попълване на анкетата от конкретна компания. Ако не е избран ред от файла, се показва поле, в което потребителят може да въведе номер на реда. След избор на ред, се извеждат и резултати от анализа за избраната фирма (Фиг. 65).



Фиг. 65 Фрагмент от страницата с резултати от прогноза за успеха на компания (създадена от автора)

Страницата с резултати съдържа вероятността за успех на компанията, показана във вид на бар диаграма с числови данни към нея. Изчертава се и персонализирано класификационно дърво за конкретната компания. На всеки възел от дървото се съдържа информация за възможните отговори на анкетата, чрез които се достига до него, вероятността за успех при достигане до този възел и въпроса за разклоняване към следващи възли. Пътят на обхождане е показан в дървото чрез ограждане на активните възли и крайния възел в цвят. Страницата съдържа и формуляр за запазване на данните от прогнозата.

При запазване на данните от прогноза, се показва страница със запазените данни (Фиг. 66), която може да бъде достъпена и от менюто на приложението. Данните за предходни анализи са представени в таблица, в която са отбелязани най-важните полета (етикет на анализа, процент вероятност за успех на компанията, анкетен файл, ред, използван алгоритъм за изчисление, дата). Вероятността за успех е представена чрез компактна графика. В края на всеки ред има бутон „детайли”, при натискане на който отново се зарежда страницата с анализа за компанията.

Потребители	Речник	Алгоритъм за прогноза	Анкетен файл	Прогноза	История анализи	Настройки
Анализи - история						
<div> <div>Успешна</div> <div>Нито успешна, нито неуспешна</div> <div>Неуспешна</div> </div>						
Етикет	Резултат	Анкетен файл/ред	Алгоритъм	Актуалност		
test прогнозиран успех: 0% извадка: 3	100%	Копие на Анкета - Модел за предсказване на успеха на стартиращи фирми (Responses) / 6 от:	C 5.1 ++	2014-02-08 10:06:47	детайли	
тест прогнозиран успех: 0% извадка: 4	75% 25%	Сору of Анкета - Модел за предсказване на успеха на стартиращи фирми (Responses) / 3 от:	C 5.1 ++	2014-02-08 11:03:59	детайли	
Грешка при избор на ред прогнозиран успех: 100% извадка: 4	100%	Копие на Анкета - Модел за предсказване на успеха на стартиращи фирми (Responses) / 105 от:	C 5.1 ++	2014-02-17 20:14:21	детайли	
тестова прогноза прогнозиран успех: 93.413% извадка: 41	93.413%	Анкета - Модел за предсказване на успеха на стартиращи фирми (Responses) / 138 от: 26/01/14 16:21	C 5.1 ++	2014-02-27 06:18:45	детайли	
прогноза за последното потърляване прогнозиран успех: 0% извадка: 2	100%	Анкета - Модел за предсказване на успеха на стартиращи фирми (Responses) / 0 от: 15/09/14 16:24	C 5.1 ++	2014-10-03 06:14:33	детайли	

Фиг. 66 Запазени резултати от анализи (създадена от автора)

Бяха успешно извършени следните тестови сценарии:

1. Проверка на класификационното дърво: Проверка за съвпадение на изчертаното в страницата с резултати класификационно дърво спрямо подадения текстов файл с дефиниция на дървото.
2. Проверка на интерпретацията на данните от анкетата: Избиране на ред от файла с данни и проверка, че данните и пътя в дървото съвпадат с тези от избрания ред, както и че прогнозата за успех на компанията съвпада с дадената в крайния възел на дървото.
3. Показване на дървото, когато анкетният ред е непълен: При липса на отговор на въпрос от алгоритъма, в дървото са оцветени само отговорите преди този въпрос, показват се вероятността за успех до този въпрос и се показва съобщение, че данните от анкетата са непълни.
4. История на анализите: Проверка, че данните от анализа могат да бъдат запазени, а запазените анализи могат да бъдат повторно прегледани в страницата за анализ.
5. Въвеждане на несъществуващ ред: Проверка, че след като се въведе несъществуващ ред, се показва подходящо съобщение.
6. Проверка за неизбран или невалиден анкетен файл или неизбран алгоритъм: Проверка, че ако не е избран анкетен файл, анкетният файл не отговаря на формата, или не е избран алгоритъм, се появява подходящо съобщение.

5.4.5 Административна част


За потребителите, притежаващи административни права, приложението предоставя в допълнение следните административни функционалности:

- Управление на потребителите на приложението,
- Настройки на файла-речник,

- Качване и настройка на файлове с класификационни дървета, получени от софтуер за извличане на знания от данни.

И трите вида функционалности са особено важни за правилната работа на приложението и за това за видими само за администратора, за когото се предполага, че има по-задълбочени познания за работа със системата.

Страницата за настройки на потребителите (Фиг. 67) съдържа списък на всички потребители в системата и дава възможност за разрешаване и прекратяване на достъпа на даден потребител с бутоните „Активирай” и “Забрани”. При първо посещение на потребител в приложението, на администратора на приложението се изпраща имейл, предлагащ одобрение на новия потребител с връзка към страницата за настройки на потребителите. При активиране на даден потребител, до него се изпраща и имейл, съобщаващ, че вече може да използва системата.

I3SP - софтуер за предсказване на успеха на стартиращи компании Здравейте BOYAN YANKOV  (Изход)

Потребители Речник Алгоритъм за прогноза Анкетен файл Прогноза История анализи Настройки

Потребители в приложението

Потребител	Емайл	Администратор	Статус
Nikolay Vitanov Diplom	nikolay.vitanov.diplom@gmail.com	Да	Активен
Nikolay Vitanov	nikolay.vitanov@gmail.com	Не	Забранен (Активирай)
Boyan Yankov	boian_yankov@abv.bg	Да	Активен
		Не	Изчаква одобрение (Активирай)
Nadia Nikolay	nadianikolaytania2013@gmail.com	Не	Изчаква одобрение (Активирай)
Nadia Nikolay	nadianikolay2012@gmail.com	Не	Изчаква одобрение (Активирай)
nadianikolay	nadianikolay@gmail.com	Не	Изчаква одобрение (Активирай)
S. Valentinova	s.valentinova@gmail.com	Не	Активен (Забрани)
Уеб Моушън	office@webmotion.bg	Не	Изчаква одобрение (Активирай)

Фиг. 67 Управление на потребителите (създадена от автора)

При работата с I3SP се натрупват анкетни данни, които ще бъдат използвани за бъдещи подобрения в модела за предсказване на успеха чрез използването на софтуер за извличане на знания от данни. Поради тази причина, I3SR включва функционалност за качване в системата на нови файлове на класификационното дърво (Фиг. 68). I3SP може да работи с повече от едно класификационно дърво, което позволява гъвкавост. Възможно е да се прилагат различни дървета за различни набори от данни – например за технологични компании и за всички стартиращи компании. Администраторът избира кой алгоритъм да бъде използван по подразбиране и той се прилага за всички потребители на приложението, освен ако те изрично не са избрали друг.

Потребители Речник **Алгоритъм за прогноза** Анкетен файл Прогноза История анализи Настройки

Настройки на алгоритъма за прогноза

Налични алгоритми:

Име	Файл	По подразбиране	Действия
C 5.1 ++	./files/phpdZhkuU	Не	преглед по подразбиране изтрий

Ако искате да качите нов алгоритъм за прогноза (генериран от IBM SPSS) моля ползвайте формата по-долу:

Работно име:

По-подразбиране: ☐

Качи файл: spss modelelet...chnical.txt

Фиг. 68 Качване на нов алгоритъм (класификационно дърво) в системата (създадена от автора)

Модулът за административни настройки съдържа и възможност на качване и преглед на файл-речник за приложението. Файлът-речник е един за приложението и качването на нов презаписва вече съществуващия. Той е наличен в приложението от самото начало, но е предоставена възможност за промяна, което е необходимо, ако се промени съдържанието на анкетата, която потребителите попълват.

Бяха успешно извършени следните тестови сценарии:

1. Активиране на нов потребител: Аутентикация на нов потребител в системата, активиране на потребителя от администратор, вход с вече активния потребителски профил.
2. Промяна на статуса на потребител: Проверка, че след като даден потребител е деактивиран, той не може повече да влезе в системата и неговия статус е „Забранен”, а след повторна активация на профила, неговият профил отново получава права за достъп.
3. Качване на алгоритъм за прогноза: Проверка, че нов алгоритъм за прогноза може да бъде успешно качен.
4. Изтриване на алгоритъм: Проверка, че даден алгоритъм за прогноза може да бъде изтрил.
5. Избор на алгоритъм по подразбиране: Проверка, че само един алгоритъм може да бъде избран по подразбиране от администратора и, че той се активира за останалите потребители.
6. Преглед на алгоритъм: Проверка, че даден алгоритъм се отваря в нов прозорец след натискане на бутона „преглед”
7. Смяна на файла-речник: Проверка, че файлът-речник може да бъде успешно актуализиран от администратора.

5.4.6 Резултати, изводи и заключение

От разгледаните тестови сценарии за различните функционалности на I3SP може да се заключи, че софтуерът функционира коректно и е готов за внедряване в работна среда.

5.5 Апробиране и валидиране на софтуерния продукт

Софтуерният продукт I3SP беше апробиран чрез прилагането му върху данни за 6 компании и интервюта с техните собственици. За всяка от компаниите бяха събрани данни през 2013 и отново през 2015 година, за да се проследи нейното развитие и дали направената прогноза е била вярна. Резултатите от валидацията на софтуера са показани в Таблица 15, където по искане на респондентите имената на компаниите са заменени с цифри, а имената на техните представители – с инициали, за да не бъде нарушена търговската тайна.

Таблица 15 Резултати от апробиране на софтуерния продукт I3SP

Компания, представител, сфера на дейност	Прогнозни резултати, получени от данните през 2013 година	Реален резултат през 2015 година
Компания 1, П.П., медицинска апаратура	Успешна	Успешна (увеличила е оборота)
Компания 2, И.М., уеб дизайн	Нито успешна, нито неуспешна	Успешна (увеличила е броя на служителите и оборота)
Компания 3, Л.М., био земеделие	Успешна	Успешна (увеличила е оборота)
Компания 4, П.М., разработка на софтуер	Успешна	Успешна (увеличила е броя на служителите и оборота)
Компания 5, В.В., интернет търговия	Нито успешна, нито неуспешна	Нито успешна, нито неуспешна (продължава дейност, но не е увеличила броя на служителите и оборота)
Компания 6, С.Ц., образование	Нито успешна, нито неуспешна	Нито успешна, нито неуспешна (продължава дейност, но не е увеличила броя на служителите и оборота)

Представените в таблицата резултати **потвърждават ефективността на софтуера**, като прогнозните резултати съвпадат с реалните в 5 от общо 6 случая. Единствено резултатът за Компания 2 не съвпада с прогнозата. От представените в таблицата случаи, Компания 2 и Компания 6 ще бъдат разгледани в детайли, като се вземе предвид ситуацията и от гледната точка на собствениците на компаниите.

Според **С.Ц., собственик на Компания 6, работеща в сферата на образованието**, състоянието на компанията към сегашния момент (2015 година) спрямо ситуацията преди две години (2013 година) е без промяна – същия брой служители, приблизително същите обороти и без видим растеж. Това съвпада с изготвената от I3SP прогноза на база на данните през 2013 година, а именно че компанията ще продължи дейността си но няма да отчете растеж.

Според модела за прогноза на I3SP, **ограничаващи фактори за растежа на Компания 6** са били липсата на партньорства с конкуренти в комбинация с липсата на концентрация на клиенти на пазара. Такива партньорства за съжаление не са били

изградени през последните две години, но С.Ц. отчита нуждата от изграждане на повече партньорства с други образователни институции.

Ситуацията с Компания 6 към момента (2015 година) се е променила – тя вече не развива иновативен продукт, а участва на пазара в паралелна конкуренция с други компании. Компанията разполага с повече нематериални активи - интелектуална собственост. Факторите за успех, според модела, в този случай са различни и прогнозата е, че Компания 6 ще увеличи мащаба си.

Според **И.М., собственик на Компания 2, занимаваща се с уеб дизайн**, към сегашния момент (2015 година) компанията е увеличила своя мащаб, спрямо ситуацията преди две години (2013 година) – значително по-голям брой служители по-голям оборот. Този резултат не съвпада с изготвената от I3SP прогноза на база на данните през 2013 година, а именно че компанията ще продължи дейността си но няма да отчете растеж.

Според моделът, **ограничаващ фактор за растежа на Компания 2** е била липсата на ясно конкурентно предимство. И.М. осъзнава този факт, отчасти с помощта на прогнозата за успех на компанията, изготвена през 2013, и споделя следното: „През последните две години компанията е дефинирала по-ясно конкурентното си предимство“. Вероятно е моделът за предсказване на успеха, чрез идентифициране на ограничаващите фактори за растеж, да е помогнал в развитието на компанията.

Ситуацията с Компания 2 към момента се е променила и при изготвения нов анализ през 2015 година И.М. посочва, че компанията притежава ясно конкурентно предимство. Прогнозата на I3SP вече е за растеж на компанията.

Като цяло потребителите изразиха задоволство от получените анализи и проявиха интерес към тяхното тълкуване и към прилагане на получените знания.

5.6 Резултати, изводи и заключение

Разгледани бяха изискванията към софтуера за предсказване на успеха на стартиращи компании I3SP, описан беше процесът на проектиране и разработка на софтуера, както и на тестване, апробиране и валидиране на вече завършения продукт. Получените резултати показаха, че продуктът е полезен за целевата група и в голям процент от случаите дава правилни прогнози за успеха на компаниите.

В I3SP е предвидена възможността за актуализация на модела при натрупване на повече данни. Данните от всички потребители на продукта периодично (през няколко месеца) се анализират отново със софтуер за извличане на знания от данни и се генерират нови модели за предсказване на успеха (във вид на текстов файл). Този файл се качва в I3SP и софтуерът вече може да прилага новия модел. Това дава възможност за избор от няколко модела, както и за създаване на специализирани модели (например за технологични стартиращи компании).

Възможно развитие на I3SP е разработката на бъдещи версии, които да включват по-кратка анкета само с ключовите въпроси от моделите за предсказване на успеха, тъй като част от потребителите на продукта изразиха недоволство, че попълването на анкетата отнема прекалено дълго време (около 15 минути). Недостатък на този вариант би бил, че попълването на по-кратка анкета не може да се използва за развитие на модела.

Предстои популяризиране на софтуера – процес, който вече е започнал с изготвяне на уеб сайт за продукта - <http://byankov.com/prediction/>, както и публикации в научни списания, в уеб сайта на автора <http://byankov.com/> и в социални мрежи. Включването

на повече предприемачи в проекта ще позволи по-нататъшно развитие на моделите за предсказване на успеха и на софтуерния продукт.

ЗАКЛЮЧЕНИЕ

В настоящата дисертация беше **проучено, изследвано и анализирано съвременното състояние** у нас и в света в източниците на научна информация в областта на предсказване успеха на стартиращи компании. Беше извършен подробен литературен обзор на проблемната област и на вече известните методи, подходи и модели за решаване на задачата за предсказване на успеха на стартиращи компании. Бяха проучени, изследвани и анализирани факторите за успех на стартиращи компании в света и България. Бяха разгледани и сравнени софтуерни продукти за предсказване на техния успех.

Беше предложен **модел на процеса на създаване на компания**. Беше създаден теоретичен **модел за предсказване на успеха на стартиращи компании** от България. Моделът беше адаптиран за средата посредством провеждане на **качествено изследване** с предприемачи и собственици на български компании. Моделът съдържа следните основни категории за определяне на успеха на компанията: предприемачески екип, бизнес стратегия, структура на индустрията, ресурси. Предложеният модел беше разработен в детайли с подкатегории и фактори за успех на стартиращите компании.

На база на предложения модел за предсказване на успеха на стартиращи компании, беше проведен **експеримент**, включващ **количествено изследване** на стартиращи компании от България. Чрез **факторен анализ** на данните от количественото изследване бяха валидирани категориите и подкатегиите на факторите от предложения модел за предсказване на успеха. Данните бяха анализирани с два софтуерни продукта за **извличане на знания от данни** – IBM SPSS Modeler и Weka. Синтезирани бяха **различни модели за предсказване на успеха**. След анализ на моделите бяха избрани тези, които предлагат най-висока точност на прогнозата за успех на компанията. Основният фактор за успеха на стартиращите компании се оказа наличието на ясно конкурентно предимство. Бяха създадени и модели, специализирани в предсказването на успеха на **технологични стартиращи компании**. Най-добрите получени модели са от типа класификационно дърво. При проведените **анализи с кръстосана валидация**, синтезираните модели предсказаха с висока точност резултатите за успех на компаниите от анализирания тестов набор от данни.

На база изготвените модели за предсказване на успеха беше реализиран прототип на **информационна система за предсказване на успеха** на стартиращи компании (I3SP).

ОСНОВНИ ПРИНОСИ НА ДИСЕРТАЦИОННИЯ ТРУД

Основните приноси дисертационния труд са описани в Таблица 16.

Таблица 16 Основни приноси на дисертационния труд

Принос	Публикации
1. Проучване, изследване и анализ на съвременното състояние у нас и в света в библиотеки, научни бази данни, научни трудове и интернет на моделите за предсказване успеха на стартиращи компании, факторите за успех на стартиращи компании, софтуерните продукти за извличане на знания от данни и тяхното приложение при класификационни модели за предсказване на успеха.	<ul style="list-style-type: none"> Yankov, B.: Overview of Success Prediction Models for New Ventures, International Conference Automatics and Informatics'12, ISSN 1313-1850, pp. 13-16 (2012)
2. Предложение за теоретичен модел за предсказване на успеха на стартиращи компании от България, който е адаптиран за средата и за българските предприемачи посредством качествено изследване.	<ul style="list-style-type: none"> Yankov, B.: A Model for Predicting the Success of New Ventures, Vth International Scientific Conference e-Governance, ISSN 1313-8774, pp. 128-135 (2013) Yankov, B.: Research of Success Factors for Start-up Companies, Doctoral Conference in Mathematics, Informatics and Education, MIE 2013, ISBN 978-954-07-3596-2, pp. 119-124 (2013)
3. Валидиран е предложеният от автора теоретичен модел чрез количествено изследване и чрез софтуерни продукти за извличане на знания от данни. Създадени са модели за предсказване на успеха, адаптирани за компании от България, чрез прилагане на количествени методи, както и модел за предсказване на успеха на технологични компании.	<ul style="list-style-type: none"> Yankov, B., Haralampiev, K., Ruskov, P.: Start-up Companies Predictive Models Analysis, Vanguard Scientific Instruments in Management '2013 (VSIM:13), ISSN 1314-0582, pp. 275-285 (2013) Yankov, B.: Synthesis of Predictive Models for Start-up Companies, International Conference Automatics and Informatics'13, ISSN 1313-1850, (2013) – приета за публикуване Yankov, B., Ruskov, P., Haralampiev, K.: Models and Tools for Technology Start-Up Companies Success Analysis, Journal Economic Alternatives 2014/3, ISSN 1312-7462, pp. 15-24 (2014) Янков, Б.: Сравнение на класификационни модели за стартиращи компании Vanguard Scientific Instruments in Management '2014 (VSIM:14), ISSN 1314-0582, (2014) – приета за публикуване

<p>4. Проектиран и реализиран е прототип на информационна система за предсказване на успеха на стартиращи компании, която позволява бъдещо надграждане и подобряване на алгоритъма на база събраните данни.</p>	<ul style="list-style-type: none"> • Yankov, B., Vitanov, N.: Information System for Forecasting the Success of Bulgarian Start-up Companies, Doctoral Conference in Mathematics, Informatics and Education, MIE 2014, ISBN ISBN 978-954-07-3759-1, pp. 95-102 (2014) • Yankov, B., Ruskov, P., Haralampiev, K.: Models and Tools for Technology Start-Up Companies Success Analysis, Journal Economic Alternatives 2014/3, ISSN 1312–7462, pp. 15-24 (2014)
---	--

ПУБЛИКАЦИИ, СВЪРЗАНИ С ТЕМАТА НА ДИСЕРТАЦИЯТА

Авторът има следните публикации по темата на дисертацията:

1. Yankov, B.: Overview of Success Prediction Models for New Ventures, International Conference Automatics and Informatics '12, ISSN 1313-1850, pp. 13-16 (2012)
2. Yankov, B.: A Model for Predicting the Success of New Ventures, Vth International Scientific Conference e-Governance, ISSN 1313-8774, pp. 128-135 (2013)
3. Yankov, B.: Research of Success Factors for Start-up Companies, Doctoral Conference in Mathematics, Informatics and Education, MIE 2013, ISBN 978-954-07-3596-2, pp. 119-124 (2013)
4. Yankov, B., Haralampiev, K., Ruskov, P.: Start-up Companies Predictive Models Analysis, Vanguard Scientific Instruments in Management '2013 (VSIM:13), ISSN 1314-0582, pp. 275-285 (2013)
5. Yankov, B.: Synthesis of Predictive Models for Start-up Companies, International Conference Automatics and Informatics '13, ISSN 1313-1850, (2013) – приета за публикуване
6. Yankov, B., Ruskov, P., Haralampiev, K.: Models and Tools for Technology Start-Up Companies Success Analysis, Journal Economic Alternatives 2014/3, ISSN 1312–7462, pp. 15-24 (2014)
7. Yankov, B., Vitanov, N.: Information System for Forecasting the Success of Bulgarian Start-up Companies, Doctoral Conference in Mathematics, Informatics and Education, MIE 2014, ISBN ISBN 978-954-07-3759-1, pp. 95-102 (2014)
8. Янков, Б.: Сравнение на класификационни модели за стартиращи компании Vanguard Scientific Instruments in Management '2014 (VSIM:14), ISSN 1314-0582, (2014) – приета за публикуване

ИЗПОЛЗВАНА ЛИТЕРАТУРА

- Abbas, A. (2008). *An Assessment Methodology for Predicting the Success of Technological Enterprises*.
- Achour, M. B. (13 11 2014 г.). *PHP Manual*. Извлечено от PHP: <http://php.net/manual/en/>
- AppSplit - Crowdsourcing for Apps*. (14 07 2014 г.). Извлечено от <http://appspl.it.com>
- Astebro, T. (2000). Key Success Factors for Technological Entrepreneurs' R&D Projects. *IEEE Transactions on Engineering Management*, Vol. 51, No. 3, стр. 314-321.
- Avramov, J. (2012). Some Problems in the Operation of the Management of Mis Funding for Modernization and Innovation for Industry Under OP "Competitiveness". *IV International Science Conference „E-governance”*, (стр. 227-238).
- Bailetti, T. (2012). Technology Entrepreneurship: Overview, Definition, and Distinctive Aspects. *Technology Innovation Management Review*.
- Baron, R. A. (2005). *Entrepreneurship: A process Perspective, First Edition*. Mason. OH: Thomas corp.
- Biggadike, R. C. (1976). Corporate diversification: entry, strategy, and performance. *Cambridge, MA: Harvard, University Press*.
- Blank, S. (2005). *The Four Steps to the Epiphany* Steve Blank. K&S Ranch Press.
- Blank, S. (22 11 2014 г.). Извлечено от Steve Blank - Entrepreneurship and Conservation: <http://steveblank.com/>
- Blueprint New Business Startup Tester*. (15 07 2014 г.). Извлечено от Blueprint Business Plans and Consultancy Services: <http://q1.prototype.thinkblueprint.com.au/new-business-success-calculator/>
- Bouckaert, R. F. (2013). *WEKA Manual for Version 3-6-10*. Hamilton, New Zealand : University of Waikato.
- Briggs, I. M. (1995). *Gifts Differing: Understanding Personality Type*. Davies-Black Publishing.
- Brockhaus, R. (1982). The Psychology of the Entrepreneur. От D. S. Kent, *Encyclopedia of Entrepreneurship In C.A.* (стр. 39-56).
- Brüderl, J. P. (April 1992 г.). Survival Chances of Newly Founded Organizations. *American Sociological Review*, Vol. 57, No. 2, стр. 227–242.
- Carland, J. C. (1996). The Theoretical Bases and Dimensionality of the Carland Entrepreneurship Index. *Proceedings of the RISE'96 Conference* (стр. 1-24). University of Jyväskylä, Finland.

- Carland, J. C. (Oct 2000 r.). A new venture creation model. *Journal of Business and Entrepreneurship*;, стр. 29.
- Chandler, G. &. (1994). Resource-based capabilities, venture strategies, and venture performance. *Journal of Business Venturing*, стр. 331–349.
- Chandler, G. H. (1994). Resource-based capabilities, venture strategies, and venture performance. *Journal of Business Venturing*, 9, стр. 331-349.
- Chapman P., C. J. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*.
- Chrisman, J. B. (1998). The Determinants of New Venture Performance: An Extended Model. *Entrepreneurship, theory and practice : ET & P Vol. 23.*, стр. 5-30.
- Compass - Benchmark Your Business Metrics From 30+ Data Sources*. (15 07 2014 r.).
Извлечено от <https://www.compass.co>
- Davis, A. M. (2005). The Rise and Fall of a Software Startup. *Journal of Information Technology Case and Application Research, Vol. 7, No. 2*, стр. 31-48.
- Ebben, J. J. (2005). Efficiency, flexibility, or both? Evidence linking strategy to performance in small firms. *Strategic Management Journal*, 26, стр. 1249–1259.
- Elsevier. (20 11 2014 r.). Извлечено от ScienceDirect.com - a leading full-text scientific database: <http://www.sciencedirect.com/>
- Emerald. (20 11 2014 r.). Извлечено от Emerald Insight: <http://emeraldinsight.com/>
- Ensley, M. (1999). *Entrepreneurial teams as determinants of new venture performance*. New York: Garland Publishing Inc.
- EquityNet - The Leading Equity Crowdfunding Platform*. (14 07 2014 r.). Извлечено от <https://www.equitynet.com/>
- EuropeanCommision. (2014). HORIZON 2020, The New EU Framework Programme for Research and Innovation 2014-2020.
- Eusostat, European Commission. (2009). *Business Demography: employment and survival*.
- Florida, U. o. (9 November 2007 r.). National Colligate Invention and Innovators Alliance (NCIIA) Participants Guide. *Invention to Venture Workshop in Technology Entrepreneurship Workshop*. NCIIA Publishing.
- Fundly - Online Fundraising Websites to Raise Money for Anything*. (14 07 2014 r.).
Извлечено от <https://fundly.com>
- Gartner, W. B. (1988). "Who is an Entrepreneur?" Is the Wrong Question. *American Journal of Small Business, Spring, 12(4)*, стр. 11-32.

- GiveForward - The Number 1 Way to Raise Money for a Loved One.* (14 07 2014 г.).
Извлечено от <http://www.giveforward.com/>
- GoFundMe - Number One for Crowdfunding and Fundraising Websites.* (14 07 2014 г.).
Извлечено от <http://www.gofundme.com/>
- Google Inc. (20 11 2014 г.). Извлечено от Google Scholar: <http://scholar.google.bg/>
- Google Inc. (12 08 2014 г.). *Google APIs Explorer*. Извлечено от Google Developers:
<https://developers.google.com/apis-explorer/#p/>
- Google Inc. (12 08 2014 г.). *Google Forms*. Извлечено от Google Документи:
<http://www.google.com/google-d-s/createforms.html>
- Google Inc. (15 11 2014 г.). *Google Sheets API version 3.0*. Извлечено от Google Developers: <https://developers.google.com/google-apps/spreadsheets/>
- Google Inc. (12 08 2014 г.). *Google Spreadsheets API*. Извлечено от Google Developers:
<https://developers.google.com/google-apps/spreadsheets/>
- Google Inc. (15 11 2014 г.). *Google Visualization API Reference*. Извлечено от Google Developers: <https://developers.google.com/chart/interactive/docs/reference>
- Google Inc. (н.д.). *Google Тенденции*. Изтеглено на 25 05 2014 г. от
<http://www.google.bg/trends/explore#q=entrepreneurship%2C%20%2Fm%2F02nwq&cmpt=q>
- Grabenwarter, U. (2005). *Exposed to the J-curve: Understanding and Managing Private Equity Fund Investments*. Euromoney Investors Plc.
- Gruber, M. (2004). Marketing in new ventures: theory and empirical evidence. *Schmalenbach Business Review*, Vol. 56 No. 2, стр. 164-99.
- Hall, M. F. (Volume 11, Issue 1 2009 г.). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*.
- HP Vertica*. (11 01 2015 г.). Извлечено от Real-Time Analytics Platform:
<http://www.vertica.com/>
- IBM. (2010). *Predictive Modeling with IBM SPSS Modeler - Student Guide*. IBM Corp.
- IBM. (22 11 2014 г.). *IBM SPSS Statistics*. Извлечено от <http://www-01.ibm.com/software/analytics/spss/products/statistics/>
- IBM Corporation. (2012). *IBM SPSS Modeler 15 User's Guide*. IBM Corporation.
- IBM SPSS Statistics*. (01 11 2014 г.). Извлечено от IBM: <http://www-01.ibm.com/software/analytics/spss/products/statistics/>

- Indiegogo - Crowdfunding for What Matters to You.* (14 7 2014 г.). Извлечено от <https://www.indiegogo.com/>
- Jelinek, M. (1996). *International Journal of Technology Management*. 799-813.
- Jung, C. G. (1971). *Psychological Types - Collected Works of C.G. Jung, Volume 6*. Princeton University Press.
- Kaiser, H. R. (1974). Little jiffy, mark iv. *Educational and Psychological Measurement*, 111-117.
- Keeley, R. R. (1989). Determinants of new venture success before 1982 and after a preliminary look at two eras. *IESE DI-173-E*.
- KickStarter - The World's Largest Funding Platform for Creative Projects.* (14 07 2014 г.). Извлечено от <https://www.kickstarter.com>
- Lawrence, P. S. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*.
- Learning, P. f. (1993). *Quinlan, J.* San Mateo CA: Morgan Kaufmann.
- Likert, P. (1932). *A technique for the measurement of attitudes*. New York: The Science Press.
- Luo, Y. (1999). Environment-strategy-performance relations in small businesses in China: A case of township and village enterprises in Southern China. *Journal of Small Business Management*, стр. 37-52.
- Lussier, R. N. (1995). There Are Few Differences Between Successful and Failed Small Businesses. *Journal of Small Business Strategy*, Vol. 6, No. 1.
- Manev, I. M. (6 2012 г.). Social Capital and Strategy Effectiveness: An Empirical Study of Entrepreneurial Ventures in a Transition Economy. *Конкурентные стратегии и тактики*, стр. 57-70.
- McDougall. (1987). An Analysis of Strategy, Entry Barriers, and Origin as Factors Explaining New Venture Performance. *Unpublished Doctoral Dissertation*. University of South Carolina.
- McGill, U. (2014). *McGill University*. Извлечено от McGill University: <http://www.mcgill.ca/>
- Microsoft. (13 11 2014 г.). *ASP.NET*. Извлечено от <http://www.asp.net/>
- Microsoft. (13 01 2015 г.). *Explore SQL Server 2012-2014*. Извлечено от <http://www.microsoft.com/en-us/server-cloud/products/sql-server/>
- Microsoft Excel - Софтуер за електронни таблици.* (01 11 2014 г.). Извлечено от Microsoft Office: <http://office.microsoft.com/bg-bg/excel/>

- Moore, G. A. (1991). *Crossing the chasm: Marketing and selling technology products to mainstream customers*. New York, N.Y.: HarperBusiness.
- NCIIA. (2007). *National Colligate Invention and Innovators Alliance (NCIIA) Participants Guide*. University of Central Florida, NCIIA Publishing .
- Nelson, R. (1999). *The Handy History Answer Book*. Visible Ink Press.
- Nicholas, S. A. (2003). Engineering Entrepreneurship: does entrepreneurship have a role in engineering education? *Antennas and Propagation Magazine, IEEE*.
- OAuth community. (15 11 2014 г.). Извлечено от OAuth Community Site: <http://oauth.net/>
- Oracle Corp. (12 11 2014 г.). Извлечено от Java: <https://www.java.com/en/>
- Park, M. L. (2000). A New Venture Performance Model in the Korean Information and Telecommunications Industry. *ETRI Journal, Volume 22, Number 4*, 51-65.
- Peng, M. L. (2000). Managerial ties and firm performance in a transition economy: The nature of a micro-macro link. *Academy of Management Journal*, 43 (3), стр. 486–501.
- Peng, M. W. (2001). How entrepreneurs create wealth in transition economies. *Academy of Management Executive* 15 (1), стр. 95-108.
- Poindexter, J. B. (1975). *The Efficiency of financial Markets: The Venture Capital Case*. New York University, Graduate School of Business Administration.
- ProQuest. (20 11 2014 г.). *UMI Dissertarion Express*. Извлечено от <http://dissexpress.umi.com/dxweb/search.html>
- Ries, E. (2011). *The Lean Startup*. Crown Business.
- Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business Publishing.
- Ries, E. (22 11 2014 г.). Извлечено от The Lean startup - The Movement That is Transforming How New Products Are Built and Launched: <http://theleanstartup.com/>
- Robertson, M. C. (2003). Barriers to start-up and their effect on aspirant entrepreneurs. *Education + Training, Vol. 45 Iss: 6*, стр. 308–316.
- RocketHub - Crowdfunding. (14 07 2014 г.). Извлечено от <http://www.rockethub.com/>
- RuleQuest. (11 12 2014 г.). *Data Mining Tools See5 and C5.0*. Извлечено от RuleQuest Research Data Mining Tools: <http://www.rulequest.com/see5-info.html>
- Ruskov, P. H. (2012). Online Investigation of SMEs Competitive Advantage. *MEB 2012, 10th International Conference on Management, Enterprise and Benchmarking*, (стр. 143-159).

- Sandberg, W. R. (1986). *New venture performance: The role of strategy and industry structure*. Lexington Books.
- Sell a Band - Where Fans Invest in Music*. (14 07 2014 г.). Извлечено от <https://www.sellaband.com/>
- Shepherd, D. (May 1999 г.). Venture Capitalists' Assessment of New Venture Survival. *Management Science*, Vol. 45, No. 5, стр. 621-632.
- Shepherd, D. E. (September–November 2000 г.). New venture strategy and profitability: A venture capitalist's assessment. *Journal of Business Venturing*, Volume 15, Issues 5–6, стр. 449–467.
- Shirokova, G. S. (2010). Factors of new venture performance in Russia. *Management, Research Review*, Vol. 33 Iss: 5, стр. 484–498.
- Small Business Services (SBS), Small Firms: Big Business!* (2002). London.
- Startup Risk Calculator*. (15 07 2014 г.). Извлечено от EquityNet - The Leading Equity Crowdfunding Platform: <https://www.equitynet.com/crowdfunding-tools/startup-risk-calculator.aspx>
- Surveys, W. T.-W. (18 07 2014 г.). *Usage Statistics and Market Share for Operating Systems for Websites, July 2014*. Извлечено от http://w3techs.com/technologies/overview/operating_system/all
- Teal, E. a. (2003). The determinants of new venture success: strategy, industry structure, and the founding entrepreneurial team. *The Journal of Private Equity*, Vol. 6 No. 4, 38-51.
- Teal, E. J. (1998). The determinants of new venture success: Strategy, industry structure, and the founding entrepreneurial team. *Unpublished doctoral dissertation*. Athens, Greece: The University of Georgia.
- The R Foundation. (11 01 2015 г.). Извлечено от The R Project for Statistical Computing: <http://www.r-project.org/>
- Tranfield, D. D. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management* 14(3), стр. 207-222.
- Tyebjee, T. B. (1984). A Model of Venture Capitalist Investment Activity. *Management Science*, Vol. 30, No. 9, стр. 1051-1066.
- Waikato, M. L. (11 09 2014 г.). *Weka 3: Data Mining Software in Java*. Извлечено от Weka: <http://www.cs.waikato.ac.nz/ml/weka/index.html>
- Wells, W. A. (1974). *Venture Capital Decision Making, unpublished doctoral dissertation*. Carnegie-Mellon University.

Yankov, B. (2012). Overview of Success Prediction Models for New Ventures. *International Conference Automatics and Informatics '12, ISSN 1313-1850*, (стр. 13-16). Sofia.

Yankov, B. (2013). A Model for Predicting the Success of New Ventures. *Vth International Scientific Conference e-Governance, ISSN 1313-8774*, , (стр. 128-135). Sozopol.

БАН, и. д. (2012). *Речник на българския език. Т. I–XIV*. София.

Национален Статистически Институт - Република България. (2014). Извлечено от Национален Статистически Институт - Република България: <http://www.nsi.bg/bg>

НСИ, Национален Статистически Институт - Република България. (2013). *Демография на предприятията към 31.12.2012 година*.

Стойчева, М. Е. (2013). Практическо обучение в областта на предприемачеството. *VII Национална конференция Образованието в информационното общество*, (стр. 207-216). Пловдив.

Харалампиев, К. (2012). *IBM SPSS – Статистически решения на приложни изследователски задачи. (Второ преработено и допълнено издание)*.

ДЕКЛАРАЦИЯ ЗА ОРИГИНАЛНОСТ

Декларирам, че представената във връзка с провеждането на процедура за придобиване на образователната и научна степен „доктор” в Софийски университет “Св. Климент Охридски“ дисертация на тема: “ Методология за предсказване на успеха за технологични стартиращи компании в България“ е мой труд и в нейното разработване не са ползвани чужди публикации и разработки в нарушение на авторските им права.

Цитиранията на всички източници на информация, текст, илюстрации, таблици, изображения и други са обозначени според стандартите.

Резултатите и приносите на проведеното дисертационно изследване са оригинални и не са заимствани от изследвания и публикации, в които нямам участие.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1: РЕЧНИК

Термин / съкращение	Дефиниция
AJAX (Asynchronous JavaScript And XML)	Група техники в уеб програмирането, позволяващи създаването на асинхронни уеб приложения, чрез използване на езиците JavaScript и XML.
BI (Business Intelligence)	Набор от технологии и инструменти за трансформиране на неструктурирани данни в смислена и полезна информация за бизнес анализ. Целта на BI е да позволи лесна интерпретация на големи обеми от данни, благодарение на което да бъдат идентифицирани и създадени нови възможности и бизнес стратегии.
CLR (Common Language Runtime)	Виртуална машина, компонент на Microsoft .NET Framework, която управлява изпълнението на .NET програми.
CLI (Common Language Infrastructure)	Спецификация, разработена от Microsoft, която описва изпълнимия код и средата и се прилага в Microsoft .NET Framework.
CRISP-DM (Cross Industry Standard Process for Data Mining)	Методология, която описва стандарт на процеса на извличане на данни (data mining) и е създадена от консорциум от компании като проект на Европейския Съюз през 1997 година. CRISP-DM е описана като йерархичен модел на процес, съдържащ групи от задачи, описани на четири нива на абстракция (от общо към частно): фази, общи задачи, специализирани задачи и инстанции на процеси.
CSV (Comma-separated Values)	В превод „стойности, разделени със запетая“, файлов формат за съхранение на таблични данни под формата на текст, дефиниран в препоръка RFC4180.
CSS (Cascading Style Sheets)	Език за описание на стилове, специфициран и поддържан от W3C, чрез който се описва представянето на документ, написан на език за маркиране (най-често HTML, XHTML, XML).
DB (database)	База от данни – структуриран набор от данни, съхранявани на компютър
DBMS (database management system)	Система за управление на база данни и софтуерно приложение, което извлича и анализира данни от база данни, като взаимодейства с потребителя и с други приложения.
DM (Data Mining)	В превод „извличане на знания от данни“, процес на откриване на смислени корелации, зависимости, повтарящи се образци, тенденции и аномалии в големи масиви от данни, чрез използване на техники и алгоритми от областта на машинното обучение, разпознаването на образи, статистиката, невронните мрежи и визуализацията на данни.

DOM (Document Object Model)	Обектен модел, използван за представяне на XML-базирани документи, използван също в HTML и XHTML.
GPL (GNU General Public License)	<p>В превод „Общ публичен лиценз на ГНУ“ е лиценз, издаден от Фондацията за свободен софтуер, с цел той да бъде използван за лицензирането на софтуер като „свободен“.</p> <p>GPL гарантира на потребителите на компютърни програми следните права:</p> <ul style="list-style-type: none"> • Свободата да ползват програмата за каквато и да е цел, • Свободата да изучават как работи програмата и да я променят, • Свободата да разпространяват копия, • Свободата да подобряват програмата и да дават на обществото достъп до подобренията.
HTML (HyperText Markup Language)	В превод „език за маркиране на хипертекст“, основният маркиращ език за описание и дизайн на уеб страници, стандартизиран и поддържан от W3C.
HTTP (HyperText Transfer Protocol)	В превод „протокол за трансфер на хипертекст“, мрежов протокол, от приложния слой на OSI модела, за пренос на информация в компютърни мрежи. Най-новата версия на протокола е описана в RFC 2616.
IBM SPSS Modeler	Софтуерна платформа за извличане на знания от данни и създаване на модели за предсказване на успеха, подпомагащи вземането на решения от индивиди, групи, системи и компании.
I3SP (Information System for Start-ups Success Prediction)	В превод „информационна система за предсказване успеха на стартиращи компании“, софтуерна информационна система, разработена към настоящата дисертация, която служи за прогнозиране на вероятността за успех на стартиращи компании и визуализация на класификационно дърво с факторите за успех.
Java	Обектно-ориентиран език за програмиране. Кодът, написан на Java се компилира в специфичен за езика байт код. Програмите, написани на Java се изпълняват във виртуална машина (Java Virtual Machine).
JavaScript	Език за програмиране, който се интерпретира и изпълнява от повечето уеб браузъри. Поддържа обектно-ориентиран и процедурен стил на програмиране.
JSON (JavaScript Object Notation)	Текстово базиран отворен стандарт, създаден за обмен на обекти от данни, състоящи се от двойки ключ-стойност. JSON произлиза от JavaScript, но е езиково независим.
JSP (Java Server Pages)	В превод „Сървърни страници на Java“, технология за разработка на уеб сървърни приложения, използваща езика за програмиране Java.
Microsoft .NET Framework	Платформа за разработка на приложения, създадена от Microsoft, която предоставя програмен модел, библиотека от класове и среда за изпълнение на програмен код.

MVC (Model-View-Controller)	В превод „модел-изглед-контролер“, шаблон за дизайн на софтуерна архитектура, използван при реализацията на потребителски интерфейси. MVC разделя софтуерното приложение на три взаимосвързани части, разделяйки вътрешното представяне на информацията от информацията, представена на потребителя.
NCIIA (The National Collegiate Inventors and Innovators Alliance)	В превод „Национален университетски съюз на изобретателите и иноваторите“, организация в САЩ, която обучава предприемачи и финансира студентски компании.
NVP (New Venture Performance)	В превод „производителност на стартираща компания“, крайният резултат от процесът на създаване и организиране на нов бизнес, който разработва, създава и маркеттира продукти и услуги, за да задоволи пазарни потребности с цел печалба и растеж.
oAuth	Отворен протокол, който позволява сигурна оторизация за уеб, мобилни и десктоп приложения. Протоколът предоставя стандартна рамка, посредством която приложения на трети страни могат да получат ограничен достъп до HTTP услуга.
OS (Operating System)	В превод „операционна система (ОС)“, основна част от компютърния системен софтуер, която управлява и координира работата на процесора и устройствата в компютърната система, както и обслужва работата на приложния софтуер.
OSI (Open Systems Interconnection) model	В превод „отворен модел за свързване на системи“, теоретичен модел, описващ принципния начин на комуникация и строежа на компютърните мрежи. Като главна градивна единица са използвани седем на брой слоеве, като всеки слой предоставя интерфейс и услуги към по-горния слой, като в същото време получава услуги от слоя под него.
PERL	Скриптов многоплатформен програмен език с общо предназначение, който се разпространява под лиценза GNU.
RIA (Rich Internet Application)	В превод „интернет приложения с богата функционалност“, уеб приложения, които притежават много от характеристиките на десктоп приложения.
SME (Small and medium-sized enterprises)	В превод „малки и средно големи компании“. SMEs са компании, чийто брой на персонала попада в определени ограничения (до 250 служители за Европейския Съюз), определящи ги като микро, малки и средно големи компании.
W3C (World Wide Web Consortium)	В превод „световен интернет консорциум“, основната световна организация, стандартизираща интернет, познат още като world wide web, www и w3.
Weka	Набор от алгоритми за машинно учене които служат за решаване на проблеми свързани с извличане на знания от данни. Weka е написана на програмния език Java и алгоритмите могат да бъдат извикани директно или от Java код.

xHTML (eXtensible HyperText Markup Language)	В превод „разширен език за маркиране на хипертекст“, език от семейството на XML-базираните езици, който наследява HTML.
XML (eXtensible Markup Language)	В превод „разширен език за маркиране“, език дефиниращ правила за създаване на специализирани маркиращи езици, както и синтаксисът, на който тези езици трябва да се подчиняват. XML указва как да бъде структуриран един документ чрез маркиране на съдържането с етикети, което го прави метаязык.

ПРИЛОЖЕНИЕ 2: РЕЗУЛТАТИ ОТ АНАЛИЗИ СЪС СОФТУЕРНИ ПРИЛОЖЕНИЯ ЗА ИЗВЛИЧАНЕ НА ЗНАНИЯ ОТ ДАННИ

Резултати от факторен анализ със софтуерния продукт IBM SPSS Statistics

Представени са подробно изходните данни от факторния анализ, изготвен чрез IBM SPSS Statistics.

Резюме на факторния анализ:

Notes		
Output Created		15-Nov-2014 11:23:36
Comments		
Input	Data	survey data spss 12 fa.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	142
Missing Value Handling	Definition of Missing	MISSING=EXCLUDE: User-defined missing values are treated as missing.
	Cases Used	MEAN SUBSTITUTION: For each variable used, missing values are replaced with the variable mean.
Syntax		<p>FACTOR</p> <p>/VARIABLES</p> <p>Личн_Могадасесправямвработатасиб</p> <p>Личн_Азсъмверенвсебеси</p> <p>Личн_Азсъмпредприемчивинициативе</p> <p>Личн_Азсъмамбициозен</p> <p>Личн_Склоненсъмдапоемамрискове</p> <p>Опит_Имампредприемаческиуменияум</p> <p>Опит_Имамуменияпомениджмънт</p> <p>Опит_Имамуменияпомаркетинг</p> <p>Опит_ИмамуменияпочовешкиресурсиН</p> <p>Опит_Имамуменияпоинвестиране</p> <p>Опит_Имампридишенопитвстартиране</p> <p>Опит_Имамопитнаподобнапозицияспр</p> <p>Опит_Имамопитвсщатасферанаиндус</p> <p>Опит_Имамуправленскиопит</p> <p>Екип_Фирматаразполагасъсзавършен</p> <p>Екип_Екипътпритежавазнания</p> <p>Екип_Екипътпритежаваумения</p> <p>Екип_Екипътпритежаваположителнан</p> <p>План_Планирамобхващайкиширокаобл</p> <p>План_Планирамзадълбочено</p> <p>План_Правяформаленписменстратегии</p> <p>План_Планирамчесто</p> <p>План_Планирампофункционалниоблас</p> <p>План_Припланиранеторазглеждаммно</p> <p>ЦелиЦелитекоитопоставямпредбизне</p> <p>Целевиятпазарендялнафирматае</p> <p>Страт_Успявамдасепридържамкъмпър</p> <p>Страт_Стратегиятакоятосъмизбрале</p> <p>Страт_Набелязанатастратегияеподр</p> <p>Страт_Имамяснастратегиязабизнеса</p> <p>Бизнесътминавлизанапазаранаследн</p> <p>КонкурентнопредимствоМоятбизнесп</p> <p>СтратегическипартньорстваФирмата</p> <p>Партньор_клиенти</p> <p>Партньор_доставчици</p> <p>Партньор_други</p>

	Индустр_Повечетопечалба
	Индустр_ИндустриятаБВП
	Индустр_Виндустриятаърсене
	Индустр_Индустриятарастеж
	Конкур_Конкурентитевиндустрията
	Конкур_Компаниятаеизложенаанаконк
	МАктив_складовиналичности
	МАктив_недвижимиимоти
	МАктив_паривброй
	НМАктив_разпознаваембренд
	НМАктив_софтуернисистемизау
	НМАктив_изграденабизнесрепут
	НМАктив_лицензи
	НМАктив_изграденибизнесконта
	НМАктив_канализадистрибу
	НМАктив_пазаринатруда
	НМАктив_капиталовипазари
	НМАктив_материалиидостав
	/MISSING MEANSUB
	/ANALYSIS
	Личн_Могадасесправямвработатасиб
	Личн_Азсъмугверенвсебеси
	Личн_Азсъмпредприемчивинициативе
	Личн_Азсъмамбициозен
	Личн_Склоненсъмдапоемамрискове
	Опит_Имампредприемаческиуменияум
	Опит_Имамуменияпомениджмънт
	Опит_Имамуменияпомаркетинг
	Опит_ИмамуменияпочовешкиресурсиН
	Опит_Имамуменияпоинвестиране
	Опит_Имампредишенопитвстартиране
	Опит_Имамопитнаподобнапозицияспр
	Опит_Имамопитвсъщатасферанаиндус
	Опит_Имамуправленскиопит
	Екип_Фирматаразполагасъсзавършен
	Екип_Екипътпритежавазнания
	Екип_Екипътпритежаваумения
	Екип_Екипътпритежаваположителнан
	План_Планирамобхващайкиширокаобл
	План_Планирамадълбочено
	План_Правяформаленписменстратегии
	План_Планирамачесто
	План_Планирампотенциалниоблас
	План_Припланиранеторазглеждамно
	ЦелиЦелитекоитопоставямпредбизне
	Целевиятпазарендялнафирматае
	Страт_Успявамдасепридържахкъмпър
	Страт_Стратегиятакоятохъмизбрале
	Страт_Набелязанатастратегияеподр
	Страт_Имамянастратегиязабизнеса
	Бизнесътминавлизнапазараанаследн
	КонкурентнопредимствоМоятбизнесп
	СтратегическипартньорстваФирмата
	Партньор_клиенти
	Партньор_доставчици
	Партньор_други
	Индустр_Повечетопечалба
	Индустр_ИндустриятаБВП
	Индустр_Виндустриятаърсене
	Индустр_Индустриятарастеж
	Конкур_Конкурентитевиндустрията
	Конкур_Компаниятаеизложенаанаконк
	МАктив_складовиналичности
	МАктив_недвижимиимоти
	МАктив_паривброй
	НМАктив_разпознаваембренд

		HMAктив_софтуернисистемизау HMAктив_изграденабизнесрепут HMAктив_лицензи HMAктив_изграденибизнесконта HMAктив_канализадистрибу HMAктив_пазаринатруда HMAктив_капиталовипазари HMAктив_материалиидостав /PRINT INITIAL KMO AIC EXTRACTION ROTATION /FORMAT BLANK(.5) /PLOT EIGEN /CRITERIA FACTORS(11) ITERATE(25) /EXTRACTION PC /CRITERIA ITERATE(25) /ROTATION VARIMAX /METHOD=CORRELATION.
Resources	Processor Time	00 00:00:03.500
	Elapsed Time	00 00:00:02.782
	Maximum Memory	328872 (321.164K) bytes
	Required	

Списък с променливи, изключени от факторния анализ (изписани са пълните етикети на променливите):

- Личност и ценности [Това, което ми се случва в живота зависи от мен самия.]
- Личност и ценности [Не се притеснявам да вземам решения без да разполагам с необходимата информация.]
- Умения и опит [Имам технически умения (за технологии, свързани с моя бизнес).]
- Какво образование сте завършили?
- В коя област е завършеното от Вас образование?
- Каква е Вашата възрастова група?
- Предприемачи / собственици на бизнес ли са Вашите родители?
- Планиране [Често правя анализ на конкуренцията.]
- Целевата годишна печалба на фирмата е (в лв)
- Начин на навлизане на пазара
- Продуктът, който предлагам, е възприет от пазара в следната степен:
- Бизнесът ми има следното основно конкурентно предимство:
- Стратегически партньорства [Фирмата участва в партньорства с правителството.]
- Характеристики на индустрията, в която фирмата оперира [Има бариери (пречки) за навлизане на нови фирми в индустрията.]
- Характеристики на индустрията, в която фирмата оперира [Има бариери (пречки) за излизане от бизнеса.]
- Характеристики на индустрията, в която фирмата оперира [В индустрията има много фалити на фирми.]
- Характеристики на индустрията, в която фирмата оперира [Индустриалният сектор, в който фирмата оперира, е голям в сравнение с останалите сектори.]
- Характеристики на индустрията, в която фирмата оперира [В индустрията има криза.]
- Характеристики на индустрията, в която фирмата оперира [Индустрията е предсказуема.]
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Иновации
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Маркетинг и реклама
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Екип
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Знания
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Приспособимост на бизнеса
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Стратегия и планиране
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Външна среда
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Обслужване на клиентите
- Смятам, че следните фактори са ключови за успеха за моя бизнес:Цена
- Фирмата работи в следния индустриален сектор:

- Индустрията, в която работи фирмата, се намира в стадий на:
- Конкурентност на индустрията [В индустрията има силна конкуренция.]
- Конкурентност на индустрията [В индустрията присъстват и малки фирми.]
- Характеристики на клиентите [В индустрията има концентрация на клиенти.]
- Характеристики на клиентите [Клиентите са различни типове.]
- Материални активи [Компанията притежава оборудване / машини.]
- Нематериални активи [Компанията притежава патенти / търговски марки.]
- Нематериални активи [Компанията притежава разработени собствени продукти.]
- Със собствени средства на основателите - Основният източник, използван за първоначално финансиране на фирмата е:
- Средства от семейство / приятели - Основният източник, използван за първоначално финансиране на фирмата е:
- Стипендия / фонд за малък бизнес - Основният източник, използван за първоначално финансиране на фирмата е:
- Заем / кредит - Основният източник, използван за първоначално финансиране на фирмата е:
- Бизнес инкубатор - Основният източник, използван за първоначално финансиране на фирмата е:
- Индивидуален инвеститор (бизнес ангел) - Основният източник, използван за първоначално финансиране на фирмата е:
- Инвестиционен фонд за рисков капитал (VC) - Основният източник, използван за първоначално финансиране на фирмата е:
- Бартер - Основният източник, използван за първоначално финансиране на фирмата е:
- Партньорство - Основният източник, използван за първоначално финансиране на фирмата е:
- Индивидуален инвеститор (бизнес ангел) - Основният източник, използван за първоначално финансиране на фирмата е:

Резултати от факторния анализ:

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.697
Bartlett's Test of Sphericity	Approx. Chi-Square	3154.260
	df	1431
	Sig.	.000

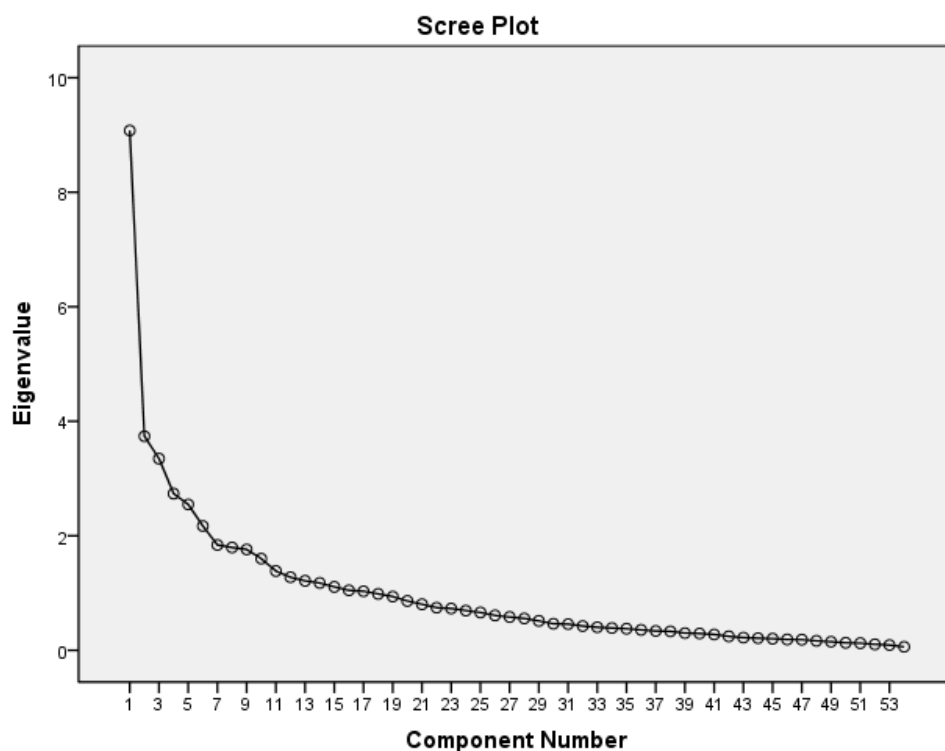
Communalities		
	Initial	Extraction
Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]	1.000	.237
Личност и ценности [Аз съм уверен в себе си.]	1.000	.649
Личност и ценности [Аз съм предприемчив, инициативен.]	1.000	.699
Личност и ценности [Аз съм амбициозен.]	1.000	.645
Личност и ценности [Склонен съм да поемам рискове.]	1.000	.569
Умения и опит [Имам предприемачески умения (умения в стартирането на бизнес).]	1.000	.657
Умения и опит [Имам умения по мениджмънт.]	1.000	.733
Умения и опит [Имам умения по маркетинг.]	1.000	.743
Умения и опит [Имам умения по човешки ресурси (HR).]	1.000	.677
Умения и опит [Имам умения по инвестиране.]	1.000	.612
Умения и опит [Имам предишен опит в стартирането на бизнес.]	1.000	.659
Умения и опит [Имам опит на подобна позиция (спрямо заеманата в момента).]	1.000	.759
Умения и опит [Имам опит в същата сфера на индустрията.]	1.000	.663
Умения и опит [Имам управленски опит.]	1.000	.656
Работа в екип [Фирмата разполага със завършен екип.]	1.000	.650
Работа в екип [Екипът притежава необходимите знания.]	1.000	.853
Работа в екип [Екипът притежава необходимите умения.]	1.000	.830
Работа в екип [Екипът притежава необходимата положителна нагласа.]	1.000	.699
Планиране [Планирам, обхващайки широка област.]	1.000	.492
Планиране [Планирам задълбочено.]	1.000	.632

Планиране [Правя формален (писмен) стратегически план.]	1.000	.546
Планиране [Планирам често.]	1.000	.481
Планиране [Планирам по функционални области / отдели.]	1.000	.506
Планиране [При планирането разглеждам множество алтернативи.]	1.000	.513
Цели [Целите, които поставям пред бизнеса си, са амбициозни.]	1.000	.475
Целевият пазарен дял на фирмата е:	1.000	.281
Стратегическа посока [Успявам да се придържам към първоначално избраната стратегия.]	1.000	.492
Стратегическа посока [Стратегията, която съм избрал, е агресивна.]	1.000	.521
Стратегическа посока [Набелязаната стратегия е подробна.]	1.000	.718
Стратегическа посока [Имам ясна стратегия за бизнеса си.]	1.000	.671
Бизнесът ми навлиза на пазара на следния етап:	1.000	.546
Конкурентно предимство [Моят бизнес притежава ясно конкурентно предимство.]	1.000	.547
Стратегически партньорства [Фирмата участва в партньорства с конкуренти.]	1.000	.386
Стратегически партньорства [Фирмата участва в партньорства с клиенти.]	1.000	.675
Стратегически партньорства [Фирмата участва в партньорства с доставчици.]	1.000	.652
Стратегически партньорства [Фирмата участва в партньорства с други заинтересовани страни.]	1.000	.608
Характеристики на индустрията, в която фирмата оперира [Повечето компании в индустрията имат добра печалба.]	1.000	.564
Характеристики на индустрията, в която фирмата оперира [Индустрията, в която фирмата оперира, генерира голяма част от brutния вътрешен продукт (БВП).]	1.000	.489
Характеристики на индустрията, в която фирмата оперира [В индустрията има стабилно търсене.]	1.000	.572
Характеристики на индустрията, в която фирмата оперира [Индустрията е в период на растеж.]	1.000	.606
Конкурентност на индустрията [Конкурентите в индустрията са агресивни.]	1.000	.750
Конкурентност на индустрията [Компанията е изложена на конкурентни атаки.]	1.000	.661
Материални активи [Компанията притежава складови наличности.]	1.000	.616
Материални активи [Компанията притежава недвижими имоти.]	1.000	.542
Материални активи [Компанията притежава пари в брой.]	1.000	.409
Нематериални активи [Компанията притежава разпознаваем бранд.]	1.000	.694
Нематериални активи [Компанията притежава софтуерни системи за управление на бизнес процесите (ERP, CRM).]	1.000	.527
Нематериални активи [Компанията притежава изградена бизнес репутация.]	1.000	.622
Нематериални активи [Компанията притежава лицензи.]	1.000	.511
Нематериални активи [Компанията притежава изградени бизнес контакти (социална мрежа).]	1.000	.568
Нематериални активи [Компанията има достъп до канали за дистрибуция.]	1.000	.584
Нематериални активи [Компанията има достъп до пазари на труда.]	1.000	.549
Нематериални активи [Компанията има достъп до капиталови пазари.]	1.000	.433
Нематериални активи [Компанията има достъп до материали и доставчици.]	1.000	.573
Extraction Method: Principal Component Analysis.		

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.078	16.812	16.812	9.078	16.812	16.812	4.677	8.661	8.661
2	3.739	6.924	23.735	3.739	6.924	23.735	3.621	6.705	15.367
3	3.347	6.199	29.934	3.347	6.199	29.934	3.304	6.118	21.485
4	2.736	5.067	35.001	2.736	5.067	35.001	3.110	5.760	27.245
5	2.548	4.718	39.719	2.548	4.718	39.719	2.792	5.170	32.415
6	2.171	4.020	43.738	2.171	4.020	43.738	2.747	5.087	37.502
7	1.839	3.406	47.144	1.839	3.406	47.144	2.594	4.804	42.306
8	1.795	3.324	50.468	1.795	3.324	50.468	2.481	4.595	46.901
9	1.762	3.262	53.730	1.762	3.262	53.730	2.392	4.430	51.331
10	1.600	2.963	56.694	1.600	2.963	56.694	2.273	4.208	55.540
11	1.384	2.563	59.256	1.384	2.563	59.256	2.007	3.717	59.256
12	1.275	2.362	61.618						

13	1.213	2.247	63.865						
14	1.176	2.178	66.043						
15	1.108	2.051	68.094						
16	1.047	1.939	70.033						
17	1.031	1.910	71.943						
18	.985	1.823	73.766						
19	.938	1.736	75.503						
20	.862	1.596	77.098						
21	.803	1.487	78.585						
22	.745	1.379	79.964						
23	.729	1.351	81.315						
24	.694	1.284	82.599						
25	.660	1.222	83.822						
26	.609	1.128	84.949						
27	.582	1.077	86.027						
28	.557	1.031	87.058						
29	.513	.950	88.007						
30	.463	.857	88.864						
31	.456	.845	89.710						
32	.423	.784	90.493						
33	.403	.747	91.240						
34	.388	.718	91.958						
35	.376	.697	92.655						
36	.356	.660	93.315						
37	.337	.623	93.938						
38	.331	.612	94.551						
39	.301	.557	95.108						
40	.292	.541	95.649						
41	.275	.510	96.159						
42	.243	.450	96.608						
43	.220	.408	97.016						
44	.211	.392	97.408						
45	.202	.375	97.783						
46	.189	.350	98.133						
47	.185	.342	98.475						
48	.164	.304	98.779						
49	.149	.277	99.056						
50	.130	.241	99.297						
51	.124	.229	99.526						
52	.104	.192	99.719						
53	.092	.171	99.890						
54	.060	.110	100.000						

Extraction Method: Principal Component Analysis.



Component Matrix ^a											
	Component										
	1	2	3	4	5	6	7	8	9	10	11
Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]											
Личност и ценности [Аз съм уверен в себе си.]											
Личност и ценности [Аз съм предприемчив, инициативен.]											
Личност и ценности [Аз съм амбициозен.]											
Личност и ценности [Склонен съм да поемам рискове.]											
Умения и опит [Имам предприемачески умения (умения в стартирането на бизнес).]	.576										
Умения и опит [Имам умения по мениджмънт.]	.534										
Умения и опит [Имам умения по маркетинг.]					.512						
Умения и опит [Имам умения по човешки ресурси (HR).]											
Умения и опит [Имам умения по инвестиране.]	.619										
Умения и опит [Имам предишен опит в стартирането на бизнес.]											
Умения и опит [Имам опит на подобна позиция (спрямо заеманата в момента).]					.557						
Умения и опит [Имам опит в същата сфера на индустрията.]					.594						
Умения и опит [Имам управленски опит.]											
Работа в екип [Фирмата разполага със завършен екип.]				.539							
Работа в екип [Екипът притежава необходимите знания.]											
Работа в екип [Екипът притежава необходимите умения.]											
Работа в екип [Екипът притежава необходимата положителна нагласа.]											
Планиране [Планирам, обхващайки широка област.]											
Планиране [Планирам задълбочено.]	.531										
Планиране [Правя формален (писмен) стратегически план.]											

Планиране [Планирам често.]												
Планиране [Планирам по функционални области / отдели.]	.541											
Планиране [При планирането разглеждам множество алтернативи.]	.509											
Цели [Целите, които поставям пред бизнеса си, са амбициозни.]	.514											
Целевият пазарен дял на фирмата е:												
Стратегическа посока [Успявам да се придържам към първоначално избраната стратегия.]												
Стратегическа посока [Стратегията, която съм избрал, е агресивна.]												
Стратегическа посока [Набелязаната стратегия е подробна.]	.659											
Стратегическа посока [Имам ясна стратегия за бизнеса си.]	.656											
Бизнесът ми навлиза на пазара на следния етап:												
Конкурентно предимство [Моят бизнес притежава ясно конкурентно предимство.]												
Стратегически партньорства [Фирмата участва в партньорства с конкуренти.]												
Стратегически партньорства [Фирмата участва в партньорства с клиенти.]												
Стратегически партньорства [Фирмата участва в партньорства с доставчици.]								.616				
Стратегически партньорства [Фирмата участва в партньорства с други заинтересовани страни.]												
Характеристики на индустрията, в която фирмата оперира [Повечето компании в индустрията имат добра печалба.]												
Характеристики на индустрията, в която фирмата оперира [Индустрията, в която фирмата оперира, генерира голяма част от brutния вътрешен продукт (БВП).]												
Характеристики на индустрията, в която фирмата оперира [В индустрията има стабилно търсене.]												
Характеристики на индустрията, в която фирмата оперира [Индустрията е в период на растеж.]												
Конкурентност на индустрията [Конкурентите в индустрията са агресивни.]								.506				
Конкурентност на индустрията [Компанията е изложена на конкурентни атаки.]												
Материални активи [Компанията притежава складови наличности.]												
Материални активи [Компанията притежава недвижими имоти.]												
Материални активи [Компанията притежава пари в брой.]												
Нематериални активи [Компанията притежава разпознаваем бранд.]												
Нематериални активи [Компанията притежава софтуерни системи за управление на бизнес процесите (ERP, CRM).]												
Нематериални активи [Компанията притежава изградена бизнес репутация.]												
Нематериални активи [Компанията притежава лицензи.]												
Нематериални активи [Компанията притежава изградени бизнес контакти (социална мрежа).]	.659											
Нематериални активи [Компанията има достъп до канали за дистрибуция.]												
Нематериални активи [Компанията има достъп до пазари на труда.]												
Нематериални активи [Компанията има достъп до капиталови пазари.]												
Нематериални активи [Компанията има достъп до материали и доставчици.]												
Extraction Method: Principal Component Analysis.												
a. 11 components extracted.												

Rotated Component Matrix ^a												
	Component											
	1	2	3	4	5	6	7	8	9	10	11	
Личност и ценности [Мога да се справям в работата си без нужда от съдействие.]												
Личност и ценности [Аз съм уверен в себе си.]		.754										
Личност и ценности [Аз съм предприемчив, инициативен.]		.761										
Личност и ценности [Аз съм амбициозен.]		.683										

Личност и ценности [Склонен съм да поемам рискове.]	.643										
Умения и опит [Имам предприемачески умения (умения в стартирането на бизнес).]											
Умения и опит [Имам умения по мениджмънт.]				.775							
Умения и опит [Имам умения по маркетинг.]				.784							
Умения и опит [Имам умения по човешки ресурси (HR).]				.777							
Умения и опит [Имам умения по инвестиране.]				.546							
Умения и опит [Имам предишен опит в стартирането на бизнес.]				.645							
Умения и опит [Имам опит на подобна позиция (спрямо заеманата в момента).]				.782							
Умения и опит [Имам опит в същата сфера на индустрията.]				.748							
Умения и опит [Имам управленски опит.]				.695							
Работа в екип [Фирмата разполага със завършен екип.]			.743								
Работа в екип [Екипът притежава необходимите знания.]			.883								
Работа в екип [Екипът притежава необходимите умения.]			.886								
Работа в екип [Екипът притежава необходимата положителна нагласа.]			.775								
Планиране [Планирам, обхващайки широка област.]	.635										
Планиране [Планирам задълбочено.]	.768										
Планиране [Правя формален (писмен) стратегически план.]	.653										
Планиране [Планирам често.]	.633										
Планиране [Планирам по функционални области / отдели.]	.659										
Планиране [При планирането разглеждам множество алтернативи.]	.659										
Цели [Целите, които поставям пред бизнеса си, са амбициозни.]											
Целевият пазарен дял на фирмата е:											
Стратегическа посока [Успявам да се придържам към първоначално избраната стратегия.]											
Стратегическа посока [Стратегията, която съм избрал, е агресивна.]											
Стратегическа посока [Набелязаната стратегия е подробна.]	.704										
Стратегическа посока [Имам ясна стратегия за бизнеса си.]	.653										
Бизнесът ми навлиза на пазара на следния етап:											
											.515
Конкурентно предимство [Моят бизнес притежава ясно конкурентно предимство.]											
Стратегически партньорства [Фирмата участва в партньорства с конкуренти.]											
Стратегически партньорства [Фирмата участва в партньорства с клиенти.]					.689						
Стратегически партньорства [Фирмата участва в партньорства с доставчици.]					.728						
Стратегически партньорства [Фирмата участва в партньорства с други заинтересовани страни.]					.699						
Характеристики на индустрията, в която фирмата оперира [Повечето компании в индустрията имат добра печалба.]								.684			
Характеристики на индустрията, в която фирмата оперира [Индустрията, в която фирмата оперира, генерира голяма част от brutния вътрешен продукт (БВП).]								.609			
Характеристики на индустрията, в която фирмата оперира [В индустрията има стабилно търсене.]								.634			
Характеристики на индустрията, в която фирмата оперира [Индустрията е в период на растеж.]								.704			
Конкурентност на индустрията [Конкурентите в индустрията са агресивни.]											.821
Конкурентност на индустрията [Компанията е изложена на конкурентни атаки.]											.700
Материални активи [Компанията притежава складови наличности.]								.770			
Материални активи [Компанията притежава недвижими имоти.]								.685			
Материални активи [Компанията притежава пари в брой.]								.504			

Нематериални активи [Компанията притежава разпознаваем бранд.]										.736				
Нематериални активи [Компанията притежава софтуерни системи за управление на бизнес процесите (ERP, CRM).]										.647				
Нематериални активи [Компанията притежава изградена бизнес репутация.]										.619				
Нематериални активи [Компанията притежава лицензи.]										.636				
Нематериални активи [Компанията притежава изградени бизнес контакти (социална мрежа).]														
Нематериални активи [Компанията има достъп до канали за дистрибуция.]														
Нематериални активи [Компанията има достъп до пазари на труда.]													.536	
Нематериални активи [Компанията има достъп до капиталови пазари.]														
Нематериални активи [Компанията има достъп до материали и доставки.]										.539				
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 10 iterations.														

Component Transformation Matrix											
Component	1	2	3	4	5	6	7	8	9	10	11
1	.557	.392	.295	.345	.246	.306	.257	.140	.087	.282	.054
2	-.262	-.482	-.157	.080	.276	.291	.408	.470	-.058	.046	.338
3	-.345	.198	.522	-.355	-.384	.118	.160	.284	.410	.083	.023
4	-.231	.302	-.601	.012	.200	.159	-.109	-.014	.586	.249	-.113
5	.568	-.369	-.116	-.620	-.022	.288	-.036	-.063	.229	-.020	-.049
6	-.267	-.071	.393	-.098	.656	.191	.019	-.420	.128	-.256	-.185
7	.029	.554	-.189	-.413	.233	.004	.056	.315	-.295	-.473	.136
8	-.085	.032	.056	.183	-.181	.644	-.605	-.021	-.087	-.172	.322
9	.106	-.029	.146	-.074	.262	-.466	-.291	-.002	.290	.089	.706
10	.034	-.143	.162	.003	.274	-.136	.517	.607	-.029	.132	-.452
11	.185	-.117	-.020	.381	-.123	-.110	.080	.176	.479	-.710	-.089
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.											

Резултати от синтезиране на модели за предсказване на успеха със софтуерния продукт IBM SPSS Modeler

Модел за предсказване на успеха на стартиращи компании

Подробни резултати от синтезиране на модел за предсказване на успеха на стартиращи компании от България (от целия набор от данни), чрез алгоритъм C5.0 с прилагане на кръстосана валидация.

Analysis

Tree depth: 7

Cross Validation

Mean: 60,3

Standard Error: 5,3

Analysis of survey data spss 12.sav (2014-9-5 8:31:31)

Number of records: 117

Analysis Accuracy: 83,761%

Fields

Target

Фирматаеуспешна

Inputs

КонкурентнопредимствоМоятбизнесп

ФУспех_Външнасреда

Опит_Имамопитнаподобнапозицияспр

НМАктив_изграденабизнесрепут

Начиннанавлизаненапазара

Индустр_Повечетопечалба

Партньор_други

Клиент_Виндустриятаимаконцентрац

НМАктив_разпознаваембренд

Build Settings

Use partitioned data: true

Partition: Partition

Calculate predictor importance: true

Calculate raw propensity scores: false

Calculate adjusted propensity scores: false

Use weight: false

Output type: Decision tree

Group symbolics: false

Use boosting: false

Cross-validate: true

Number of folds: 10

Mode: Simple

Favor: Accuracy

Expected noise (%): 0

Use misclassification costs: false

Training Summary

Algorithm: C5

Model type: Classification

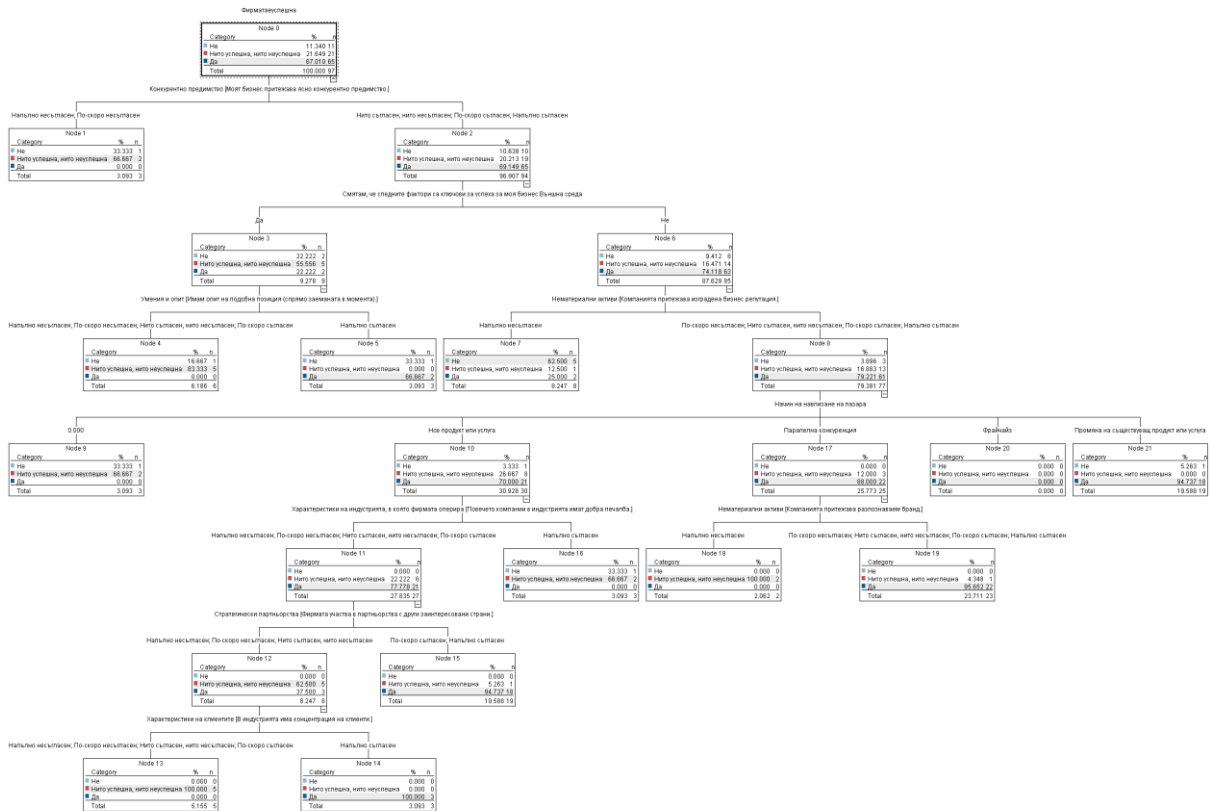
Stream: Z:\Boyan-2014-08-15.str

User: Gunshot

Date built: 14-9-5 8:30

Application: IBM® SPSS® Modeler 14.2

Elapsed time for model build: 0 hours, 0 mins, 1 secs



КонкурентнопредмствоМоятбизнессп in [1 2] [Mode: 2] => 2,0 (3; 0,667)

КонкурентнопредмствоМоятбизнессп in [3 4 5] [Mode: 3] (94)

ФУспех_Външнасреда = 1 [Mode: 2] (9)

Опит_Имамопитнаподобнапозицияспр in [1 2 3 4] [Mode: 2] => 2,0 (6; 0,833)

Опит_Имамопитнаподобнапозицияспр in [5] [Mode: 3] => 3,0 (3; 0,667)

ФУспех_Външнасреда = 2 [Mode: 3] (85)

НМАктив_изграденабизнесрепут in [1] [Mode: 1] => 1,0 (8; 0,625)

НМАктив_изграденабизнесрепут in [2 3 4 5] [Mode: 3] (77)

Начиннанавлизаненапазара = 0 [Mode: 2] => 2,0 (3; 0,667)

Начиннанавлизаненапазара = 1 [Mode: 3] (30)

Индустр_Повечетопечалба in [1 2 3 4] [Mode: 3] (27)

Партньор_други in [1 2 3] [Mode: 2] (8)

Клиент_Виндустриятаимаконцентрац in [1 2 3
 4] [Mode: 2] => 2,0 (5; 1,0)
 Клиент_Виндустриятаимаконцентрац in [5] [
 Mode: 3] => 3,0 (3; 1,0)
 Партньор_други in [4 5] [Mode: 3] => 3,0 (19;
 0,947)
 Индустр_Повечетопечалба in [5] [Mode: 2] => 2,0 (3;
 0,667)
 Начиннанавлизаненапазара = 2 [Mode: 3] (25)
 НМАктив_разпознаваембренд in [1] [Mode: 2] => 2,0 (2;
 1,0)
 НМАктив_разпознаваембренд in [2 3 4 5] [Mode: 3] => 3,0
 (23; 0,957)
 Начиннанавлизаненапазара = 3 [Mode: 3] => 3,0 (0)
 Начиннанавлизаненапазара = 4 [Mode: 3] => 3,0 (19; 0,947)

Модел за предсказване на успеха на технологични стартиращи компании

Подробни резултати от синтезиране на модел за предсказване на успеха на
технологични стартиращи компании от България (от данните, касаещи само
 технологични компании), чрез алгоритъм C5.0 с прилагане на кръстосана валидация.

Analysis

Tree depth: 7

Cross Validation

Mean: 60,0

Standard Error: 6,2

Analysis of survey data spss 12 tech.sav (2014-11-16 10:01:41)

Number of records: 72

Analysis Accuracy: 87,5%

Fields

Target

Фирматаеуспешна

Inputs

Индустр_Повечетопечалба

ФУспех_Цена

Опит_Имауменияпомениджмънт

ФУспех_Маркетингирек

Страт_Стратегиятакоятосъмизбрале

Конкур_Виндустриятаимасилнаконку

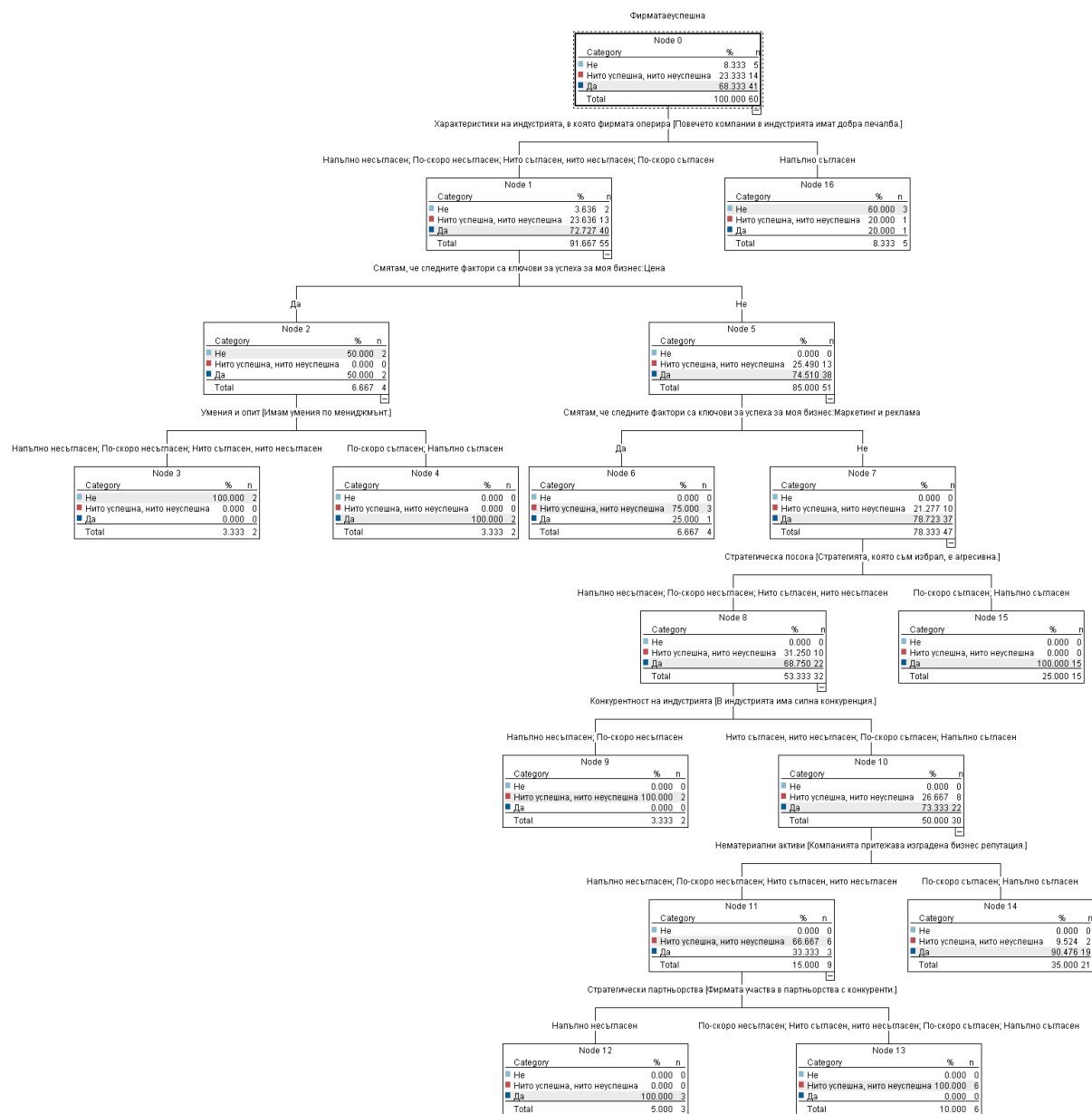
НМАктив_изграденабизнесрепут
СтратегическипартньорстваФирмата

Build Settings

Use partitioned data: true
Partition: Partition
Calculate predictor importance: true
Calculate raw propensity scores: false
Calculate adjusted propensity scores: false
Use weight: false
Output type: Decision tree
Group symbolics: false
Use boosting: false
Cross-validate: true
Number of folds: 10
Mode: Simple
Favor: Accuracy
Expected noise (%): 0
Use misclassification costs: false

Training Summary

Algorithm: C5
Model type: Classification
Stream: F:\Boyan-2014-11-12 tech.str
User: Gunshot
Date built: 14-11-16 10:00
Application: IBM® SPSS® Modeler 14.2
Elapsed time for model build: 0 hours, 0 mins, 33 secs



Индустр_Повечетопечалба in [1 2 3 4] [Mode: 3] (55)

ФУспех_Цена = 1 [Mode: 1] (4)

Опит_Имауменияпомениджмънт in [1 2 3] [Mode: 1] => 1,0 (2; 1,0)

Опит_Имауменияпомениджмънт in [4 5] [Mode: 3] => 3,0 (2; 1,0)

ФУспех_Цена = 2 [Mode: 3] (51)

ФУспех_Маркетингирек = 1 [Mode: 2] => 2,0 (4; 0,75)

ФУспех_Маркетингирек = 2 [Mode: 3] (47)

Страт_Стратегиятакоятосъмизбрал in [1 2 3] [Mode: 3] (32)

Конкур_Виндустриятаимасилнаконку in [1 2] [Mode: 2]

=> 2,0 (2; 1,0)

(30) Конкур_Виндустриятаимасилнаконку in [3 4 5] [Mode: 3]

] (9) НМАктив_изграденабизнесрепут in [1 2 3] [Mode: 2

Mode: 3] => 3,0 (3; 1,0) СтратегическипартньорстваФирмата in [1] [

5] [Mode: 2] => 2,0 (6; 1,0) СтратегическипартньорстваФирмата in [2 3 4

=> 3,0 (21; 0,905) НМАктив_изграденабизнесрепут in [4 5] [Mode: 3]

Страт_Стратегиятакояотосъмизбрале in [4 5] [Mode: 3] => 3,0 (15; 1,0)

Индустр_Повечетопечалба in [5] [Mode: 1] => 1,0 (5; 0,6)

Програмен код на алгоритъма C5.0

Представена е извадка от програмния код на алгоритъна C5.0, използван за синтезиране на модели от типа класификационно дърво. Представеният фрагмент е функцията „int main(int Argc, char *Argv[])“ от програмния код на Linux версията на алгоритъма, която е реализирана на езикът С и се разпространява под отворения лиценз Gnu GPL.

Пълният код е достъпен за изтегляне на уеб сайта на производителя

<http://www.rulequest.com/download.html>

```

/*****
/*
/* Copyright 2010 Rulequest Research Pty Ltd.
/*
/* This file is part of C5.0 GPL Edition, a single-threaded version
/* of C5.0 release 2.07.
/*
/* C5.0 GPL Edition is free software: you can redistribute it and/or
/* modify it under the terms of the GNU General Public License as
/* published by the Free Software Foundation, either version 3 of the
/* License, or (at your option) any later version.
/*
/* C5.0 GPL Edition is distributed in the hope that it will be useful,
/* but WITHOUT ANY WARRANTY; without even the implied warranty of
/* MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU
/* General Public License for more details.
/*
/* You should have received a copy of the GNU General Public License
/* (gpl.txt) along with C5.0 GPL Edition. If not, see
/*
/* <http://www.gnu.org/licenses/>.
/*
*****/

/*****
/*

```

```

/*      Main routine, C5.0                                     */
/*      -----                                                */
/*                                                                 */
/*****
/*****

#include "defs.i"
#include "extern.i"
#include <signal.h>

#include <sys/unistd.h>
#include <sys/time.h>
#include <sys/resource.h>

#define SetFOpt(V)      V = strtod(OptArg, &EndPtr);\
                        if ( ! EndPtr || *EndPtr != '\00' ) break;\
                        ArgOK = true
#define SetIOpt(V)      V = strtol(OptArg, &EndPtr, 10);\
                        if ( ! EndPtr || *EndPtr != '\00' ) break;\
                        ArgOK = true

int main(int Argc, char *Argv[])
/* ---- */
{
    int                o;
    extern String OptArg, Option;
    char               *EndPtr;
    Boolean             FirstTime=true, ArgOK;
    double             StartTime;
    FILE                *F;
    CaseNo              SaveMaxCase;
    Attribute           Att;

    struct rlimit RL;

    /* Make sure there is a largish runtime stack */

    getrlimit(RLIMIT_STACK, &RL);

    RL.rlim_cur = Max(RL.rlim_cur, 20 * 1024 * 1024);

    if ( RL.rlim_max > 0 )    /* -1 if unlimited */
    {
        RL.rlim_cur = Min(RL.rlim_max, RL.rlim_cur);
    }

    setrlimit(RLIMIT_STACK, &RL);

    /* Check for output to be saved to a file */

    if ( Argc > 2 && ! strcmp(Argv[Argc-2], "-o") )
    {
        Of = fopen(Argv[Argc-1], "w");
        Argc -= 2;
    }

    if ( ! Of )
    {
        Of = stdout;
    }
}

```

```

}

KRInit = time(0) & 07777;

PrintHeader("");

/* Process options */

while ( (o = ProcessOption(Argc, Argv, "f+bpv+t+sm+c+S+I+ru+egX+wh")) )
{
    if ( FirstTime )
    {
        fprintf(Of, T_OptHeader);
        FirstTime = false;
    }

    ArgOK = false;

    switch (o)
    {
        case 'f':    FileStem = OptArg;
                     fprintf(Of, T_OptApplication, FileStem);
                     ArgOK = true;
                     break;
        case 'b':    BOOST = true;
                     fprintf(Of, T_OptBoost);
                     if ( TRIALS == 1 ) TRIALS = 10;
                     ArgOK = true;
                     break;
        case 'p':    PROBTHRESH = true;
                     fprintf(Of, T_OptProbThresh);
                     ArgOK = true;
                     break;
#ifdef VerbOpt
        case 'v':    SetIOpt(VERBOSITY);
                     fprintf(Of, "\tVerbosity level %d\n", VERBOSITY);
                     ArgOK = true;
                     break;
#endif
        case 't':    SetIOpt(TRIALS);
                     fprintf(Of, T_OptTrials, TRIALS);
                     Check(TRIALS, 3, 1000);
                     BOOST = true;
                     break;
        case 's':    SUBSET = true;
                     fprintf(Of, T_OptSubsets);
                     ArgOK = true;
                     break;
        case 'm':    SetFOpt(MINITEMS);
                     fprintf(Of, T_OptMinCases, MINITEMS);
                     Check(MINITEMS, 1, 1000000);
                     break;
        case 'c':    SetFOpt(CF);
                     fprintf(Of, T_OptCF, CF);
                     Check(CF, 0, 100);
                     CF /= 100;
                     break;
        case 'r':    RULES = true;
                     fprintf(Of, T_OptRules);
                     ArgOK = true;
                     break;
    }
}

```



```

case 'S':   SetFOpt(SAMPLE);
            fprintf(Of, T_OptSampling, SAMPLE);
            Check(SAMPLE, 0.1, 99.9);
            SAMPLE /= 100;
            break;
case 'I':   SetIOpt(KRInit);
            fprintf(Of, T_OptSeed, KRInit);
            KRInit = KRInit & 07777;
            break;
case 'u':   SetIOpt(UTILITY);
            fprintf(Of, T_OptUtility, UTILITY);
            Check(UTILITY, 2, 10000);
            RULES = true;
            break;
case 'e':   NOCOSTS = true;
            fprintf(Of, T_OptNoCosts);
            ArgOK = true;
            break;
case 'w':   WINNOW = true;
            fprintf(Of, T_OptWinnow);
            ArgOK = true;
            break;
case 'g':   GLOBAL = false;
            fprintf(Of, T_OptNoGlobal);
            ArgOK = true;
            break;
case 'X':   SetIOpt(FOLDS);
            fprintf(Of, T_OptXval, FOLDS);
            Check(FOLDS, 2, 1000);
            XVAL = true;
            break;
    }

    if ( ! ArgOK )
    {
        if ( o != 'h' )
        {
            fprintf(Of, T_UnregnizedOpt,
                    Option,
                    ( ! OptArg || OptArg == Option+2 ? "" : OptArg ));
            fprintf(Of, T_SummaryOpts);
        }
        fprintf(Of, T_ListOpts);
        Goodbye(1);
    }
}

if ( UTILITY && BOOST )
{
    fprintf(Of, T_UBWarn);
}

StartTime = ExecTime();

/* Get information on training data */

if ( ! (F = GetFile(".names", "r")) ) Error(NOFILE, "", "");
GetNames(F);

if ( ClassAtt )
{

```

```

    fprintf(Of, T_ClassVar, AttName[ClassAtt]);
}

NotifyStage(READDATA);
Progress(-1.0);

/* Allocate space for SomeMiss[] and SomeNA[] */

SomeMiss = AllocZero(MaxAtt+1, Boolean);
SomeNA    = AllocZero(MaxAtt+1, Boolean);

/* Read data file */

if ( ! (F = GetFile(".data", "r")) ) Error(NOFILE, "", "");
GetData(F, true, false);
fprintf(Of, TX_ReadData(MaxCase+1, MaxAtt, FileStem));

if ( XVAL && (F = GetFile(".test", "r")) )
{
    SaveMaxCase = MaxCase;
    GetData(F, false, false);
    fprintf(Of, TX_ReadTest(MaxCase-SaveMaxCase, FileStem));
}

/* Check whether case weight attribute appears */

if ( CWtAtt )
{
    fprintf(Of, T_CWtAtt);
}

if ( ! NOCOSTS && (F = GetFile(".costs", "r")) )
{
    GetMCosts(F);
    if ( MCost )
    {
        fprintf(Of, T_ReadCosts, FileStem);
    }
}

/* Note any attribute exclusions/inclusions */

if ( AttExIn )
{
    fprintf(Of, "%s", ( AttExIn == -1 ? T_AttributesOut : T_AttributesIn
));

    ForEach(Att, 1, MaxAtt)
    {
        if ( Att != ClassAtt &&
            Att != CWtAtt &&
            ( StatBit(Att, SKIP) > 0 ) == ( AttExIn == -1 ) )
        {
            fprintf(Of, "    %s\n", AttName[Att]);
        }
    }
}

/* Build decision trees */

if ( ! BOOST )

```

```

{
    TRIALS = 1;
}

InitialiseTreeData();
if ( RULES )
{
    RuleSet = AllocZero(TRIALS+1, CRuleSet);
}

if ( WINNOW )
{
    NotifyStage(WINNOWATTS);
    Progress(-MaxAtt);
    WinnowAtts();
}

if ( XVAL )
{
    CrossVal();
}
else
{
    ConstructClassifiers();

    /* Evaluation */

    fprintf(Of, T_EvalTrain, MaxCase+1);

    NotifyStage(EVALTRAIN);
    Progress(-TRIALS * (MaxCase+1.0));

    Evaluate(CMINFO | USAGEINFO);

    if ( (F = GetFile(( SAMPLE ? ".data" : ".test" ), "r")) )
    {
        NotifyStage(READTEST);
        fprintf(Of, "\n");

        FreeData();
        GetData(F, false, false);

        fprintf(Of, T_EvalTest, MaxCase+1);

        NotifyStage(EVALTEST);
        Progress(-TRIALS * (MaxCase+1.0));

        Evaluate(CMINFO);
    }
}

fprintf(Of, T_Time, ExecTime() - StartTime);

#ifdef VerbOpt
    Cleanup();
#endif

return 0;
}

```

Резултати от синтезиране на модели за предсказване на успеха със софтуерния продукт Weka

Подробни резултати от синтезиране на модел за предсказване на успеха на **стартиращи компании от България** (от целия набор от данни), чрез алгоритъм J48 с прилагане на кръстосана валидация.

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: survey to spss 12 for weka-weka.filters.unsupervised.attribute.Remove-R102,104-107

Instances: 117

Attributes: 103

[list of attributes omitted]

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

КонкурентнопредимствоМоятбизнесп <= 2: 2.0 (6.0/1.0)

КонкурентнопредимствоМоятбизнесп > 2

| ФУспех_Цена <= 1

| | Опит_Имауменияпомениджмънт <= 3: 1.0 (2.0)

| | Опит_Имауменияпомениджмънт > 3: 3.0 (2.0)

| ФУспех_Цена > 1

| | ФУспех_Външнасреда <= 1

| | | Опит_Имаопитнаподобнапозицияспр <= 4: 2.0 (5.0)

| | | Опит_Имаопитнаподобнапозицияспр > 4: 3.0 (3.0/1.0)

| | ФУспех_Външнасреда > 1

| | | НМАктив_изграденабизнесрепут <= 1

| | | | Продуктъткойтопредлагамевъзприет <= 2: 3.0 (3.03/1.03)

| | | | Продуктъткойтопредлагамевъзприет > 2: 1.0 (4.04/0.04)

| | | НМАктив_изграденабизнесрепут > 1

| | | | НМАктив_разпознаваембренд <= 2

```

| | | | | РаботасединосновенклиентОсновния <= 1: 3.0 (2.0)
| | | | | РаботасединосновенклиентОсновния > 1
| | | | | Личн_Несепритеснявамдавамреша <= 3: 2.0 (9.15/1.0)
| | | | | Личн_Несепритеснявамдавамреша > 3: 3.0 (4.0/1.0)
| | | | НМАктив_разпознаваембренд > 2: 3.0 (76.78/13.78)

```

Number of Leaves : 11

Size of the tree : 21

Time taken to build model: 0.25 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	78	66.6667 %
Incorrectly Classified Instances	39	33.3333 %
Kappa statistic	0.2537	
Mean absolute error	0.2018	
Root mean squared error	0.3835	
Relative absolute error	77.7491 %	
Root relative squared error	107.3047 %	
Total Number of Instances	117	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.323	0.116	0.5	0.323	0.392	0.551	2.0
	0.893	0.595	0.728	0.893	0.802	0.68	3.0
	0.091	0.038	0.2	0.091	0.125	0.265	1.0
	0	0	0	0	?		
Weighted Avg.	0.667	0.416	0.618	0.667	0.63	0.607	

=== Confusion Matrix ===

```
a b c d <-- classified as  
10 18 3 0 | a = 2.0  
7 67 1 0 | b = 3.0  
3 7 1 0 | c = 1.0  
0 0 0 0 | d =
```

ПРИЛОЖЕНИЕ 3: ПРОГРАМЕН КОД НА ИНФОРМАЦИОННАТА СИСТЕМА I3SP

В приложението са представени фрагменти от програмния код на информационната система „I3SP“, разработена към дисертацията. Представен е кода на основните класове, реализиращи бизнес логиката и на някои от фронтенд шаблоните.

Програмен код на клас „GoogleService“:

```
class GoogleService {
    public $client;

    public function __construct() {
        $this->client = new Google_Client();
        $this->client->setApplicationName(APP_NAME);
        $this->client->setClientId(CLIENT_ID);
        $this->client->setClientSecret(CLIENT_SECRET);
        $this->client->setRedirectUri(REDIRECT_URI);
        $this->client->setAccessType('online');
        $this->client->setScopes(array(
            'https://www.googleapis.com/auth/drive',
            'https://www.googleapis.com/auth/drive.file',
            'https://www.googleapis.com/auth/userinfo.email',
            'https://www.googleapis.com/auth/userinfo.profile',
            'https://spreadsheets.google.com/feeds'));
    }

    //get logged user info
    function getLoggedInUserInfo()
    {
        $token_decoded = json_decode($this->client->getAccessToken(), true);
        $query =
        'https://www.googleapis.com/oauth2/v1/userinfo?access_token=' .
        $token_decoded["access_token"];
        $user_info = file_get_contents($query);
        /*
         * $json should be in format
         { "id": "xx",
           "name": "xx",
           "given_name": "xx",
           "family_name": "xx",
           "link": "xx",
           "picture": "xx",
           "gender": "xx",
           "locale": "xx",
           "email" : "xx" }
         */
        return json_decode($user_info, false);
    }

    //get access token json decoded
    function getAccessTokenDecoded()
    {
        $token_decoded = json_decode($this->client->getAccessToken(), true);
        return $token_decoded["access_token"];
    }

    //check for access user token
    function isAccessTokenValid()
    {

```

```

        $token_decoded = json_decode($this->client->getAccessToken(), true);
        $query =
'https://www.googleapis.com/oauth2/v1/tokeninfo?access_token=' .
$token_decoded["access_token"];
        $ch = curl_init($query);
        curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
        curl_setopt($ch, CURLOPT_SSL_VERIFYPEER, false);

        $json = '';
        if( ($json = curl_exec($ch) ) === false)
        {
            return false;
        }
        else
        {
            $json = json_decode($json, true);
            if(!isset($json) || array_key_exists('error', $json))
            {
                return false;
            }
            /*
             * $json should be in format
            {
                "audience": "8819981768.apps.googleusercontent.com",
                "user_id": "123456789",
                "scope": "profile email",
                "expires_in": 436
            }
            */
            return true;
        }
    }

    //get survey data row by file id and row_index
    function getSurveyDataIndex($file_id, $row_index)
    {
        $json_token = json_decode($this->client->getAccessToken());
        $curlParams = array (
            CURLOPT_RETURNTRANSFER => true,
            CURLOPT_FOLLOWLOCATION => 0,
            CURLOPT_FAILONERROR => false,
            CURLOPT_SSL_VERIFYPEER => false,
            CURLOPT_VERBOSE => false,
        );

        $ch = curl_init();
        curl_setopt_array($ch, $curlParams);
        $getURL =
'https://spreadsheets.google.com/tq?tqx=out:csv&key='.$file_id.'&tq=select+
*+limit+1+offset+' . $row_index;
        curl_setopt($ch, CURLOPT_URL, $getURL);

        $headers = array();
        $headers[] = "Authorization: Bearer " . $json_token->access_token;
        curl_setopt($ch, CURLOPT_HTTPHEADER, $headers);
        curl_setopt($ch, CURLOPT_CUSTOMREQUEST, 'GET');
        curl_setopt($ch, CURLOPT_USERAGENT, 'PHP Prediction App');
        curl_setopt($ch, CURLOPT_AUTOREFERER, true);
        $ret = curl_exec($ch);
        $info = curl_getinfo($ch);
        curl_close($ch);
    }

```



```

        if($info['http_code'] != '200'
            || str_replace(array('\'', '\"'), '', trim($ret)) == '' )
            return false;
        else
        {
            //return string with content
            return $ret;
        }
    }

    //get survey data row by file id and timespamp
    function getSurveyDataByDate($file_id,$date)
    {
        //get access token
        $json_token = json_decode($this->client->getAccessToken());

        $curlParams = array (
            CURLOPT_RETURNTRANSFER => true,
            CURLOPT_FOLLOWLOCATION => 0,
            CURLOPT_FAILONERROR => false,
            CURLOPT_SSL_VERIFYPEER => false,
            CURLOPT_VERBOSE => false,
        );

        $ch = curl_init();
        curl_setopt_array($ch, $curlParams);

        //set google spreadsheet query
        $query = urlencode("select * where A = datetime '$date'");

        $getUrl =
        "https://spreadsheets.google.com/tq?tqx=out:csv&key=$file_id&tq=$query";

        curl_setopt($ch, CURLOPT_URL, $getUrl);

        //create request to google API
        $headers = array();
        $headers[] = "Authorization:Bearer " . $json_token->access_token;
        curl_setopt($ch, CURLOPT_HTTPHEADER, $headers);
        curl_setopt($ch, CURLOPT_CUSTOMREQUEST, 'GET');
        curl_setopt($ch, CURLOPT_USERAGENT, 'PHP Prediction App');
        curl_setopt($ch, CURLOPT_AUTOREFERER, true);
        $ret = curl_exec($ch);
        $info = curl_getinfo($ch);
        curl_close($ch);

        if($info['http_code'] != '200'
            || str_replace(array('\'', '\"'), '', trim($ret)) == '' )
            return false;
        else
        {
            //return string with content
            return $ret;
        }
    }

    //get whole file from google drive by file id
    function getSurveyDataWhole($fileId)
    {

```

```

        $driveService = new Google_DriveService($this->client);
        // Retrieve metadata for the file specified by $fileId.
        $file = $driveService->files->get($fileId);

        // Get the contents of the file.
        // $request = new
        Google_HttpRequest($file['exportLinks']['application/vnd.openxmlformats-
        officedocument.spreadsheetml.sheet']);
        $response = $this->client->getIo()->authenticatedRequest($request);
        $Data = $response->getResponseBody();

        return $Data;
    }
}

```

Програмен код на клас „AnalyseService“:

```

class AnalyseService {
    /* expects array with id , name */
    private $surveyFile;
    private $rowIndex;
    private $sql;

    function __construct($sql,$surveyFile) {
        $this->surveyFile = $surveyFile;
        $this->sql = $sql;
    }

    //display the survey contents page
    function displaySurveyContents()
    {
        global $smarty;
        global $googleService;

        $smarty->assign('PUBLIC_API_KEY', PUBLIC_API_KEY);
        $smarty->assign('CLIENT_ID_JSAPI', CLIENT_ID_JSAPI);
        $smarty->assign('APP_URL', APP_URL);
        $smarty->assign('row_index', 1);
        $smarty->assign('access_token', $googleService-
        >getAccessTokenDecoded());

        $smarty->display('survey_data.tpl');
    }

    //display the analyse history page
    function displayAnalysesHistory()
    {
        global $smarty;
        $user_id = $_SESSION["user"]["id"];

        //get all save analyses for this user from the DB
        $this->sql->query("SELECT a.*, p.name as prediction_name FROM
        analyses a LEFT JOIN prediction_trees p on p.id = a.prediction_tree_id
        WHERE a.user_id=$user_id", SQL_ALL);
        $smarty->assign('analyses', $this->sql->record);
        $smarty->display('analyses_history.tpl');
    }
}

```

```

//saves an analyse to the DB
private function AnalyseSave()
{
    global $googleService;
    if(isset($_POST["analyse_save"]) && isset($_POST["row_index"]) &&
isset($_POST["file_id"]))
    {
        //prepare all parameters for the DB
        $row_index = $this->sql->db-
>real_escape_string($_POST["row_index"]);
        $percent_yes = $this->sql->db-
>real_escape_string($_POST["percent_yes"]);
        $percent_no = $this->sql->db-
>real_escape_string($_POST["percent_no"]);
        $percent_yesno = $this->sql->db-
>real_escape_string($_POST["percent_yesno"]);
        $count = $this->sql->db->real_escape_string($_POST["count"]);
        $date = date("Y-m-d H:i:s");
        $note = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["note"]));
        $label = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["label"]));
        $user_id = $_SESSION["user"]["id"];
        $gfileid = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["file_id"]));
        $gfilename = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["file_name"]));
        $prediction_file_id = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["prediction_file_id"]));
        $row_date = $this->sql->db-
>real_escape_string(htmlspecialchars($_POST["row_date"]));

        $this->sql->query("SELECT * FROM analyses WHERE
user_id=$user_id and google_file_id='$gfileid' and row_index=$row_index
and prediction_tree_id=$prediction_file_id", SQL_INIT);
        $result = $this->sql->record;
        //decide whether to insert or update the record
        if(isset($result) && $result > 0)
        {
            $query = "UPDATE analyses SET label='$label', note='$note',
creation_date='$date', result_yes=$percent_yes, result_no=$percent_no,
result_yesno=$percent_yesno, count=$count ";
            if(isset($row_date) and date_parse($row_date))
            {
                {
                    $parsedDate = date_parse($row_date);
                    $parsedYear = $parsedDate["year"];
                    if($parsedYear > 2000) // check for valid date year
                    {
                        $query = $query . ", row_date='$row_date' ";
                    }
                }
                $query = $query . " WHERE user_id=$user_id and
google_file_id='$gfileid' and row_index=$row_index and
prediction_tree_id=$prediction_file_id";
            }
            else
            {
                $query = "INSERT INTO analyses (prediction_tree_id,
google_file_id, google_file_name, row_index, label, note, user_id,
creation_date, result_yes, result_no, result_yesno, count, row_date) "

```

```

"VALUES($prediction_file_id,'$gfileid','$gfilename',$row_index,'$label','$note',
$user_id,'$date',$percent_yes,$percent_no,$percent_yesno,$count,'$row_date');"

    }
    $this->sql->query($query,SQL_NONE);
}

//dispay the analyse page
function displayAnalyse()
{
    if(isset($_POST["analyse_save"]))
    {
        //if save button clicked -> store the analyse into DB
        $this->AnalyseSave();
        //redirect to history page
        redirect(APP_URL.'?p=analyses_history');
        exit();
    }

    global $smarty;
    global $googleService;

    $row_index = null;
    $row_date = null;

    //perform various checks for the input parameters for this page
    if(isset($_GET["row_index"]) )
    {
        if(is_numeric($_GET["row_index"]) && $_GET["row_index"] >= 0)
        {
            if(isset($_GET["dbdate"]) && strlen($_GET["dbdate"]) > 0)
            {
                $row_date = $_GET["dbdate"];
            }
            $row_index = round($_GET["row_index"]);
            $user_id = $_SESSION["user"]["id"];

            $prediction_file_path =
$_SESSION['prediction_file']["file_path"];
            $prediction_file_id = $_SESSION['prediction_file']["id"];
            if(isset($_GET["prediction_file_id"]) &&
trim($_GET["prediction_file_id"]) != '')
            {
                $prediction_file_id = $this->sql->db-
>real_escape_string(trim($_GET["prediction_file_id"]));
                //get selected file
                $this->sql->query("SELECT * FROM prediction_trees WHERE
id=$prediction_file_id", SQL_INIT);
                if(isset($this->sql->record) && !empty($this->sql-
>record))
                {
                    $res = $this->sql->record;
                    $prediction_file_path = $res["file_path"];
                }
            }
        }
    }
}

```

```

        $this->sql->query("SELECT * FROM prediction_trees",
SQL_ALL);
        $predictions = $this->sql->record;

        $file_id = '';
        $file_name = '';
        if(!isset($_GET["from_feed"]) && isset($_GET["file_id"]))
&& trim($_GET["file_id"]) != '')
        {
            $file_id = trim($_GET["file_id"]);
            $this->sql->query("SELECT * FROM analyses WHERE
user_id=$user_id and google_file_id='$file_id' and row_index=$row_index",
SQL_INIT);
            $result = $this->sql->record;
            if(isset($result))
            {
                $file_name = $result["google_file_name"];
            }
        }
        else if(isset($this->surveyFile))
        {
            $file_id = $this->surveyFile["id"];
            $file_name = $this->surveyFile["name"];
        }

        // $googleService->getSurveyDataWhole($file_id);
        // get data from google drive about the selected row by date
or index
        if(isset($row_date))
        {
            $survey_string = $googleService-
>getSurveyDataByDate($file_id,$row_date);
        }
        else
        {
            $survey_string = $googleService-
>getSurveyDataIndex($file_id,$row_index);
        }

        if($survey_string)
        {
            //convert the csv string to assoc array NO TRANSLATIONS
            $survey_data =
FileService::convertCSVStringToAssocArray($survey_string);

            //load dictionary data into array
            $dictionary = FileService::convertDictionaryToArray();

            //load the prediction tree into array and use the
offset to determine the tree levels
            $prediction_tree_array =
FileService::convertPredictionTreeToArrayWithTranslation($prediction_file_p
ath,$dictionary);
            //var_dump($prediction_tree_array);

            //quick Survey data validation - only check if core
survey data is there / not whole file

            if(FileService::validateSurveyData($survey_data,$prediction_tree_array))
            {

```

```

//see the procedure declaration for more info
$dictionary_array =
FileService::getDictionaryTranslation($survey_data,$dictionary);
//var_dump($dictionary_array);

//analyse and get results
$analyseResult = $this-
>Analyse($prediction_tree_array,$dictionary_array);

//generate json for the three visualisation
$prediction_tree_json =
AnalyseService::convertPredictionTreeToJson($smarty,$prediction_tree_array,
$dictionary_array);

//check if that file / row has not been previously
analysed

$label = '';
$note = '';
$this->sql->query("SELECT * FROM analyses WHERE
user_id=$user_id and google_file_id='$file_id' and row_index=$row_index and
prediction_tree_id = $prediction_file_id", SQL_INIT);
$result = $this->sql->record;
if(isset($result))
{
    $label = $result["label"];
    $note = $result["note"];
}

//prepare the JSON that stores the info for the
tree visualization
$jsonTree = json_encode( $prediction_tree_json);
$jsonTreePathIndexs = json_encode($analyseResult-
>path_indexes_array);

$smarty->assign('count', round($analyseResult-
>count));
$smarty->assign('percent_yes', $analyseResult-
>percent_yes);
$smarty->assign('percent_yesno', $analyseResult-
>percent_yesno);
$smarty->assign('percent_no', $analyseResult-
>percent_no);
$smarty->assign('warning', $analyseResult-
>warning);
$smarty->assign('jsonTree', $jsonTree);
$smarty->assign('jsonTreePathIndexs',
$jsonTreePathIndexs);
$smarty->assign('survey_raw_data', $survey_data);

$smarty->assign('predictions', $predictions);
$smarty->assign('row_index', $row_index);
$smarty->assign('row_date', $row_date);
$smarty->assign('prediction_file_id',
$prediction_file_id);
$smarty->assign('file_id', $file_id);
$smarty->assign('file_name', $file_name);

$smarty->assign('label', $label);
$smarty->assign('note', $note);

```

```

    }
    else {
        $smarty->assign('notification', 10);
        $smarty->assign('additional_notes', 'Избрания файл
за анализ има липса на ключови данни или не е валиден анкетен файл !
Анализа не може да бъде извършен. Моля изберете друг файл от секция
настройки.' );
    }
}
else {
    $smarty->assign('notification', 10);
    $smarty->assign('additional_notes', 'Възникна грешка
при опита за достъп до избрания ред от анкетния файл! Моля опитайте отново
или въведете друг по-малък ред/индекс' );
}
}
else
{
    $smarty->assign('notification', 10);
    $smarty->assign('additional_notes', 'Въведен невалиден
ред/индекс на анкетен файл!' );
}
}

$smarty->assign('PUBLIC_API_KEY', PUBLIC_API_KEY);
$smarty->assign('CLIENT_ID_JSAPI', CLIENT_ID_JSAPI);
$smarty->assign('APP_URL', APP_URL);
$smarty->assign('access_token', $googleService-
>getAccessTokenDecoded());

$smarty->display('survey_data_analyse.tpl');
}

//perform the prediction tree analyse
private function Analyse($prediction_tree_array,$dictionary_array)
{
    //create the result object
    $result = new AnalyseResult();

    //print_r($prediction_tree_array);
    $current_level = 0;

    //store the steps of the algorithm
    $prediction_path_array = array();

    for($i=0; $i < sizeof($prediction_tree_array); $i++)
    {
        if($current_level == $prediction_tree_array[$i]["offset"])
        {
            $current_iteration_name =
            $prediction_tree_array[$i]["name"];

            //replace , with . because in the dictionary there are no
values with ,
            $current_iteration_values = str_replace(',', '.',
            $prediction_tree_array[$i]["values"]);

            $pos =
            strpos($current_iteration_values,$dictionary_array[$current_iteration_nam
e]["value"]);
            if(! ($pos === false) )

```

```

    {
        //entering this node
        $current_level++;
        $prediction_path_array[] = $prediction_tree_array[$i];

        //exit Mode reached
        if($prediction_tree_array[$i]["exitMode"])
        {
            //success! - results
            $result->percent_yes =
$prediction_tree_array[$i]['percent_yes'];
            $result->percent_yesno =
$prediction_tree_array[$i]['percent_yesno'];
            $result->percent_no =
$prediction_tree_array[$i]['percent_no'];
            $result->count =
$prediction_tree_array[$i]["Ncount"];
            break;
        }
    }
    else if($dictionary_array[$current_iteration_name]["value"]
== '0.000')
    {
        //insufficient data for the algorithm
        $result->percent_yes =
$prediction_tree_array[$i]['percent_yes'];
        $result->percent_yesno =
$prediction_tree_array[$i]['percent_yesno'];
        $result->percent_no =
$prediction_tree_array[$i]['percent_no'];
        $result->count = $prediction_tree_array[$i]["Ncount"];
        $result->warning = "Показаните резултати са непълни!
Има липса на данни или празни полета в анкетния ред.";
        break;
    }
}

//check if there are results after running the analyse if not give
message
if(!isset($result->percent_yes))
{
    $result->percent_yes = 0;
    $result->percent_yesno = 0;
    $result->percent_no = 0;
    $result->count = 0;
    $result->warning = "Възникна грешка при изпълнение на анализа.
Показаните резултати не са пълни или са неверни!";
}

$prediction_path_indexes = array();
for($i = 0; $i < sizeof($prediction_path_array); $i++)
{
    $prediction_path_indexes[] =
$prediction_path_array[$i]["index"];
}

$result->path_array = $prediction_path_array;
$result->path_indexes_array = $prediction_path_indexes;

return $result;

```



```

    }

    //create the JSON for the prediction tree visualisation
    private static function
    convertPredictionTreeToJson($smarty,$prediction_array,$dictionary_array)
    {
        $visual_tree_json = array();
        $visual_tree_json[] = array('Column1','Column2','Column3');

        if(sizeof($prediction_array) > 0)
        {
            $visual_tree_json[] = array(array( "v" => '0' , "f" => '<div
class="pnode">' . ($prediction_array[0]["question_name"] == '' ?
$prediction_array[0]["name_translated"] :
$prediction_array[0]["question_name"]) . '</div>' ), null, null);
        }
        for($i=0;$i<sizeof($prediction_array);$i++)
        {
            $hint='';
            if($i<sizeof($prediction_array)-1 &&
isset($dictionary_array[$prediction_array[$i+1]["name"]]["hint"]) &&
strlen($dictionary_array[$prediction_array[$i+1]["name"]]["hint"]) > 0)
            {
                $hint = ' <b>' .
$dictionary_array[$prediction_array[$i+1]["name"]]["hint"] . '</b>';
            }
            //this should return Напълно вярно или Да или Рано или
прозивиолентекс

            if($dictionary_array[$prediction_array[$i]["name"]]["value_name"])
            {
                //this will return code Example: 1.000
                $survey_value =
$dictionary_array[$prediction_array[$i]["name"]]["value_name"];
                //we need the value after translation - $prediction_array
holds it under "values_translated" key
                $survey_values = str_replace($survey_value,
'<b>'.$survey_value.'</b>', $prediction_array[$i]["values_translated"]);
            }
            else
            {
                $survey_values =
$prediction_array[$i]["values_translated"];
            }

            if($i<sizeof($prediction_array)-1)
                $next_node_name = ($prediction_array[$i+1]["question_name"]
== '' ? $prediction_array[$i+1]["name_translated"] :
$prediction_array[$i+1]["question_name"] );
            else
                $next_node_name = '';

            $smarty->assign('values_translated', $survey_values);
            $smarty->assign('percent_yes',
$prediction_array[$i]["percent_yes"]);
            $smarty->assign('percent_yesno',
$prediction_array[$i]["percent_yesno"]);
            $smarty->assign('percent_no',
$prediction_array[$i]["percent_no"]);
            $smarty->assign('Ncount', $prediction_array[$i]["Ncount"]);

```

```

        $smarty->assign('current_mode',
$prediction_array[$i]["currentMode"]);
        $smarty->assign('next_node_name',
($prediction_array[$i]["exitMode"]!=''? '' : $next_node_name . $hint));
        $visual_tree_json[] = array(
            array(
                "v" => (string)$prediction_array[$i]["index"] ,

                "f" => $smarty->fetch('prediction_node.tpl')),
            (string)$prediction_array[$i]["parent"], null);
    }

    return $visual_tree_json;
}
}

```

Програмен код на клас „UserService“:

```

class UserService {
    private $sql;
    /* is set DB array for user */
    private $loggedUser;
    private $isFirstLogin;
    private $authUrl;

    public function __construct($sql, $authUrl)
    {
        $this->sql = $sql;
        $this->authUrl = $authUrl;
    }

    //authenticates google user to the system
    public function authenticate($googleUser)
    {
        $email = $this->sql->db->real_escape_string($googleUser->email);
        $name = $this->sql->db->real_escape_string($googleUser->name);
        $id = $this->sql->db->real_escape_string($googleUser->id);

        $firstTimeLogged = false;
        //get this user and check
        //what is its admin_role and status
        $this->sql->query("SELECT * FROM users WHERE email = '$email'",
SQL_INIT);
        $loggedUser = $this->sql->record;

        //insert the google user if not presented into the DB
        if(!$loggedUser)
        {
            $res = $this->sql->query("INSERT INTO
users(google_id,email,name,admin_role,status) values
('$id','$email','$name',2,1)", SQL_NONE);

            if($res)
            {
                $firstTimeLogged = true;
                $this->sql->query("SELECT * FROM users WHERE email =
'$email'", SQL_INIT);
                $loggedUser = $this->sql->record;
            }
        }
    }
}

```

```

        else
        {
            return false;
        }
    }

    //set user profile
    $loggedUser["picture"] = $googleUser->picture;
    $this->loggedUser = $loggedUser;
    if($this->loggedUser["status"] == 2)
    {
        //ok
        $_SESSION["user"] = $this->loggedUser;
        $_SESSION["mode"] = 'normal';
    }
    else if ($this->loggedUser["status"] == 3)
    {
        //suspended
        $_SESSION["mode"] = 'suspend';
    }
    else
    {
        if($firstTimeLogged)
        {
            $_SESSION["pending_name"] = $this->loggedUser["name"];
            $_SESSION["pending_email"] = $this->loggedUser["email"];
            $_SESSION["mode"] = 'pending_first_time';
        }
        else
        {
            $_SESSION["mode"] = 'pending';
        }
    }

    return true;
}

//get user from session if exists
function setup()
{
    //authenticates and populate $this->loggedUser;
    if(isset($_SESSION["user"]))
    {
        $this->loggedUser = $_SESSION["user"];
    }
}

//logout user - remove stored data from session
function logout()
{
    $this->loggedUser=null;
    unset($_SESSION['user']);
    unset($_SESSION['access_token']);
    unset($_SESSION['isFirstLogin']);
    session_destroy();
}

//display default page for either not logged user or admin/normal
logged user
public function displayDefault()
{
    global $smarty;

```

```

if(isset($this->loggedUser))
{
    if($this->loggedUser["status"] == 2)
    {
        //authorised
        if($this->loggedUser["admin_role"] == 1)
        {
            $this->displayAdminDefaultView();
        }
        else
        {
            $this->displayUserDefaultView();
        }
    }
}
else if(isset($_SESSION["mode"]))
{
    if ($_SESSION["mode"] == 'suspend')
    {
        $this->displayUserSuspendedView();
        $_SESSION["mode"] = '';
    }
    else if ($_SESSION["mode"] == 'pending')
    {
        $this->displayUserPendingView();
        $_SESSION["mode"] = '';
    }
    else if($_SESSION["mode"] == 'pending_first_time')
    {
        $this->isFirstLogin = true;
        $this->displayUserPendingView();
        $_SESSION["mode"] = '';
    }
    else
    {
        $this->displayDefaultView();
    }
}
else{
    $this->displayDefaultView();
}
}

//display default for logged users
private function displayDefaultView()
{
    global $smarty;
    if(isset($this->loggedUser))
    {
        if($this->loggedUser["admin_role"] == 1)
        {
            $this->displayAdminDefaultView();
        }
        else
        {
            $this->displayUserDefaultView();
        }
    }
    else
    {

```

```

        $smarty->assign('authUrl', $this->authUrl);
        $smarty->assign('mode', 'public');
        $smarty->display('index_public.tpl');
    }
}

//pending user view
private function displayUserPendingView()
{
    global $smarty;
    //global $smarty;
    $mode = 'pending';
    if($this->isFirstLogin)
    {
        $mode = 'pending_first_time';
        //send email to admin
        $pending_user_name = $_SESSION["pending_name"];
        $pending_user_email = $_SESSION["pending_email"];

        $this->sql->query('SELECT * FROM users WHERE
admin_role=1', SQL_INIT);
        $record = $this->sql->record;
        if(isset($record))
        {
            $smarty->assign('name', $pending_user_name);
            $smarty->assign('email', $pending_user_email);
            $smarty->assign('APP_URL', APP_URL);
            $email_body = $smarty->fetch("email_newuser.tpl");
            $email = new Email('prediction@' .
APP_DOMAIN, $record["email"], 'Нов потребител в ' . APP_NAME, $email_body );
            $email->send();
        }
    }
    $smarty->assign('mode', $mode);
    $smarty->display('index_public.tpl');
}

//suspend user view
private function displayUserSuspendedView()
{
    global $smarty;
    $smarty->assign('mode', 'blocked');
    $smarty->display('index_public.tpl');
}

//display normal logged user view
function displayUserDefaultView()
{
    global $smarty;
    global $googleService;

    //set prediction default algorithm if not any
    if(!isset($_SESSION['prediction_file']) ||
isset($_POST['prediction_id']))
    {
        //if prediction algorithm changed from page
        if(isset($_POST['prediction_id']))
        {

```

```

        $prediction_id = $this->sql->db-
>real_escape_string($_POST['prediction_id']);
        $this->sql->query("SELECT * FROM prediction_trees WHERE
id=$prediction_id", SQL_INIT);
    }
    //if not set any set the default one
    else if(!isset($_SESSION['prediction_file']))
    {
        $this->sql->query("SELECT * FROM prediction_trees WHERE
is_default = 1", SQL_INIT);
    }

    if(isset($this->sql->record))
    {
        $data = $this->sql->record;
        if(!empty($data))
        {
            $prediction = array();
            $prediction["id"] = $data['id'];
            $prediction["name"] = $data['name'];
            $prediction["file_path"] = $data['file_path'];
            $_SESSION['prediction_file'] = $prediction;
        }
    }
}

//set survey file - picked from google
if(isset($_POST["file_id"]) && isset($_POST["file_name"])) )
{
    if( isset($_SESSION["current_file"]) )
    {
        unset($_SESSION["current_file"]);
    }

    $_SESSION["current_file"] = array(
        "id" => $_POST["file_id"],
        "name" => $_POST["file_name"]
    );

    redirect(APP_URL . '?p=survey_data');
}

$this->sql->query("SELECT * FROM prediction_trees", SQL_ALL);
if(isset($this->sql->record))
{
    $smarty->assign('predictions',$this->sql->record);
    // $smarty-
>assign('prediction_selected',$_SESSION['prediction_file']);
}
$smarty->assign('PUBLIC_API_KEY', PUBLIC_API_KEY);
$smarty->assign('CLIENT_ID_JSAPI', CLIENT_ID_JSAPI);
$smarty->assign('APP_URL', APP_URL);
$smarty->assign('access_token', $googleService-
>getAccessTokenDecoded());
$smarty->display('index.tpl');
}

//admin default view
public function displayAdminDefaultView()

```

```

{
    global $smarty;
    if($this->loggedUser["admin_role"] == 1)
    {
        //show page list of users
        $this->sql->query("SELECT * FROM users", SQL_ALL);
        $allUsers = $this->sql->record;

        //var_dump($allUsers);
        $smarty->assign('users', $allUsers);
        $smarty->display("admin_manageusers.tpl");

    }
    else
    {
        $this->displayNotAuthorised();
    }
}

//not authorized view
private function displayNotAuthorised()
{
    global $smarty;
    $smarty->assign('mode', 'notauthorised');
    $smarty->display('index_public.tpl');
}
}

```

Програмен код на фронтенд шаблон „\templates\admin_manageusers.tpl“:

```

{config_load file="user.conf" section="setup"}
{include file="header.tpl" title="Настройки на потребителите в
приложението"}
{if isset($notification)}
{include file="notification.tpl" notification=$notification}
{/if}
<fieldset>
    <legend>Потребители в приложението</legend>
    <br>
<table class="tabledefault">
    <tr>
        <th>Потребител</th>
        <th>Емайл</th>
        <th>Администратор</th>
        <th>Статус</th>
    </tr>
    {foreach name=outer item=user from=$users}
        <tr>
            <td>{$user["name"]}</td>
            <td>{$user["email"]}</td>
            <td>{if $user["admin_role"] == 2}Да{else}Не{endif}</td>
            <td>{if $user["status"] == 2}
                Активен {if $user["admin_role"] == 2}<a
href="{$_SCRIPT_NAME}?p=admin_manageusers&block=&user={$user["id"]}">Забрани
</a>{endif}
                {elseif $user["status"] == 3}

```

```

Забранен (<a
href="{$_SCRIPT_NAME}?p=admin_manageusers&activate=&user={$user["id"]}">Акти
вирай</a>)
{else}
Изчаква одобрение (<a
href="{$_SCRIPT_NAME}?p=admin_manageusers&activate=&user={$user["id"]}">Акти
вирай</a>)
{/if}</td>

</tr>
{/foreach}
</table>
</fieldset>
{include file="footer.tpl"}

```

Програмен код на фронтенд шаблон „\templates\survey_data_analyse.tpl“:

```

{config_load file="user.conf" section="setup"}
{include file="header.tpl" title="Резултати от анкетата"}

{if isset($smarty.session.current_file) || (isset($file_id) && $file_id !=
'')} }

{if isset($row_index) }
<fieldset>
<legend>Прогноза за успех</legend>
{if isset($warning)}<div class="pt20 pb20"><b
style="color:red">{$warning}</b></div>{/if}
<h3>Резултати:</h3>
<br>
<div class="fl pr20">

{include file="chart.tpl" percent_Yes=$percent_yes
percent_No=$percent_no percent_YesNo=$percent_yesno ncount=$count}
</div>
<div class="fl pl20">

<form action="{$_SCRIPT_NAME}?p=survey_data_analyse"
method="post">
<table >
<tr>
<td>Анкетен файл / ред: </td>
<td>{$file_name} / {$row_index}</td>
</tr>
<tr>
<td>Алгоритъм: </td>
<td>
<select name="prediction_id" id="prediction_id">
{foreach name=outer item=prediction_file
from=$predictions}
<option value="{$_prediction_file["id"]}" {if
isset($prediction_file_id) && $prediction_file_id eq
$_prediction_file["id"]}selected="selected"{/if}>{$prediction_file["name"]}<
/option>
{/foreach}
</select>
</td>
</tr>
<tr>
<td>Етикет: </td>

```



```

        <td><input type="text" name="label" id="label"
value="{ $label}"/></td>
    </tr>
    <tr>
        <td colspan="2">
            Бележка: <br>
            <textarea name="note" id="note" rows="4"
cols="45">{ $note}</textarea>
        </td>
    </tr>
    <tr>
        <td colspan="2">
            <input type="hidden" id="row_index"
name="row_index" value="{ $row_index}" />
            <input type="hidden" id="row_date" name="row_date"
value="{ $row_date}" />
            <input type="hidden" id="prediction_file_id"
name="prediction_file_id" value="{ $prediction_file_id}"/>
            <input type="hidden" id="file_id" name="file_id"
value="{ $file_id}" />
            <input type="hidden" id="file_name"
name="file_name" value="{ $file_name}" />
            <input type="hidden" id="count" name="count"
value="{ $count}" />
            <input type="hidden" id="percent_yes"
name="percent_yes" value="{ $percent_yes}" />
            <input type="hidden" id="percent_no"
name="percent_no" value="{ $percent_no}" />
            <input type="hidden" id="percent_yesno"
name="percent_yesno" value="{ $percent_yesno}" />
            <input type="hidden" id="analyse_save"
name="analyse_save"/>

            <button>Запази промените</button>
            <a class="myButton" id="btnRecalc"
href="{ $SCRIPT_NAME}?p=survey_data_analyse&row_index={ $row_index}&file_id={
$file_id}&prediction_file_id=">Преизчисли</a>
            <a class="myButton"
href="{ $SCRIPT_NAME}?p=survey_data_analyse">Анализирай нов ред</a>
        </td>
    </tr>
</table>
</form>
</div>
<div class="clear">
</fieldset>

<div id="pnlChart"></div>
<script type="text/javascript"
src="//www.google.com/jsapi?key={ $PUBLIC_API_KEY}"></script>
<script type="text/javascript">
    {literal}
        google.load('visualization', '1', {packages: ['orgchart']});
    {/literal}
</script>
<script type="text/javascript">
    function drawChart() {
        var dataArr = { $jsonTree};
        var jsonTreePathIndexs = $.parseJSON('{ $jsonTreePathIndexs}');

```

```

var data = google.visualization.arrayToDataTable(dataArr);
//set different style for the path of the algorithm
data.setRowProperty(0, 'style', "border: 3px solid orange");
for(var i=0;i<jsonTreePathIndexs.length;i++)
{
    //console.log(jsonTreePathIndexs[i]);
    if(i != jsonTreePathIndexs.length-1)
        data.setRowProperty(jsonTreePathIndexs[i], 'style',
"border: 3px solid orange");
    else
        data.setRowProperty(jsonTreePathIndexs[i], 'style',
"border: 2px solid red");
}
var chart = new
google.visualization.OrgChart(document.getElementById('pnlChart'));
{literal}
chart.draw(data, {allowHtml:true});
{/literal}
}

$(function(){

    google.setOnLoadCallback(drawChart);

    $('#btnRecalc').click(function(e){
        e.preventDefault();
        location.href = $('#btnRecalc').attr('href') +
$('#prediction_id').val();
    });
});
</script>
</script>

{if isset($survey_raw_data)}
    Всички отговори на изчисления ред за анализ:<br>
    {foreach name=outer key=question item=answer
from=$survey_raw_data}
        <br>{$question} : {$answer}
    {/foreach}
{/if}
{else}
    {if isset($notification)}
        {include file="notification.tpl" notification=$notification
additional_notes=$additional_notes}
    {/if}
    <fieldset>
        <legend>Анализ на ред/индекс</legend>
        <div>
            <form method="get"
action="{$_SCRIPT_NAME}?p=survey_data_analyse">
                <input type="hidden" name="p" id="p"
value="survey_data_analyse"/>
                Анкетен файл: {$smarty.session.current_file.name}
                <br/>
                Избери ред за анализ / индекс: <input type="text"
id="row_index" name="row_index" value="" /><button>Направи
прогноза</button>
            </form>
        </div>
    </fieldset>
{/if}

```

```

{else}
    <fieldset>
    <legend>Анализ на ред/индекс</legend>
    <div class="warningbox">Внимание: Няма избран анкетен файл</div>
    За да видите съдържанието на даден файл и впоследствие да го
    анализирате, моля първо изберете такъв от <a
href="{${SCRIPT_NAME}}?p=config">страницата със настройки</a>.
    </fieldset>
{/if}

{include file="footer.tpl"}

```

Изнесена информация за структурата на базата данни във формат .SQL:

```

CREATE TABLE IF NOT EXISTS `analyses` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `prediction_tree_id` int(11) NOT NULL,
  `google_file_id` varchar(150) NOT NULL,
  `google_file_name` varchar(150) NOT NULL,
  `row_index` int(11) NOT NULL,
  `label` varchar(250) NOT NULL,
  `note` varchar(500) NOT NULL,
  `user_id` int(11) NOT NULL,
  `creation_date` datetime NOT NULL,
  `result_yes` float NOT NULL,
  `result_no` float NOT NULL,
  `result_yesno` float NOT NULL,
  `count` int(11) NOT NULL,
  `row_date` datetime DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 ;

CREATE TABLE IF NOT EXISTS `prediction_trees` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `name` varchar(150) NOT NULL,
  `file_path` varchar(500) NOT NULL,
  `user_id` int(11) NOT NULL,
  `is_default` bit(1) NOT NULL,
  `upload_date` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP,
  PRIMARY KEY (`id`),
  KEY `file_path` (`file_path` (255))
) ENGINE=InnoDB DEFAULT CHARSET=utf8 ;

CREATE TABLE IF NOT EXISTS `users` (
  `id` int(11) NOT NULL AUTO_INCREMENT,
  `google_id` varchar(100) NOT NULL,
  `email` varchar(254) NOT NULL,
  `name` varchar(150) NOT NULL,
  `admin_role` int(11) NOT NULL COMMENT '1-admin;2-user',
  `status` int(11) NOT NULL COMMENT '1 pending; 2-authorized; 3-blocked',
  PRIMARY KEY (`id`),
  UNIQUE KEY `google_id` (`google_id`,`email`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8 ;

```