
**ИВЕЛИНА
БАЛАБАНОВА**

**ГЕОРГИ
ГЕОРГИЕВ**



**МОДЕЛИРАНЕ И ИЗСЛЕДВАНЕ НА ПРОЦЕСИ
В БИЗНЕСА И ИНДУСТРИЯТА С ИЗКУСТВЕН
ИНТЕЛЕКТ И МАШИННО ОБУЧЕНИЕ**

ИНТЕЛИГЕНТНИ БИЗНЕС РЕШЕНИЯ

Методическо ръководство
за лабораторни упражнения



Габрово,
2022

ИВЕЛИНА БАЛАБАНОВА

ГЕРГИ ГЕОРГИЕВ

**МОДЕЛИРАНЕ И ИЗСЛЕДВАНЕ НА ПРОЦЕСИ
В БИЗНЕСА И ИНДУСТРИЯТА
С ИЗКУСТВЕН ИНТЕЛЕКТ И МАШИННО ОБУЧЕНИЕ**

**МЕТОДИЧЕСКО РЪКОВОДСТВО
ЗА ЛАБОРАТОРНИ УПРАЖНЕНИЯ**

**ГАБРОВО
2022**

© доц. д-р инж. Ивелина Стефанова Балабанова, 2022
© гл. ас. д-р инж. Георги Иванов Георгиев, 2022

**МОДЕЛИРАНЕ И ИЗСЛЕДВАНЕ НА ПРОЦЕСИ В БИЗНЕСА И
ИНДУСТРИЯТА С ИЗКУСТВЕН ИНТЕЛЕКТ И МАШИННО
ОБУЧЕНИЕ**

**МЕТОДИЧЕСКО РЪКОВОДСТВО ЗА ЛАБОРАТОРНИ
УПРАЖНЕНИЯ**

Рецензент: доц. д-р инж. Боян Димитров Карапенев

За контакти:

Габрово, 5300, ул. „Хаджи Димитър“ №4
Технически университет – Габрово
Катедра „Комуникационна техника и технологии“

© Университетско издателство „Васил Априлов“ – Габрово, 2022

ISBN:

ПРЕДГОВОР

Настоящото методично ръководство за лабораторни упражнения дава насоки на студентите за основни моменти, ключови аспекти и функционални особености при моделирането на основни административни, маркетингови и информационни бизнес процеси в съвременните икономически и индустриално ориентирани фирми, предприятия и центрове. Разгледани са различни примери, свързани с концептуални модели за обработка на заявки, регистри с клиенти, изготвяне на справки и отчети и други като част от основните задачи на всеки административен отдел.

Бизнес процесите са свързани с пренос, обработка и различни типове манипулации при опериране с масиви от данни в различни формати от конкретни информационни потоци и тяхно съхранение в свързано базирани бази от данни с приоритет на достъпа. Студентите имат възможност не само да придобият знания относно моделиране на целеви бизнес процеси в средата в Bizagi Modeler, но и сами да проектират и администратират бази данни в Microsoft SQL Server, MySQL и т.н. относно организация на обработваните масиви.

Базисен аспект, който се поставя в ръководството, се свързва с изучаването на основните понятия, съвременните методи, алгоритмите, подходите при обработка и извличането на знания от данни или така наречените Data Mining процеси, както и тяхна модификация и адаптиране спрямо спецификата на конкретните задачи. Поставени са основите на изкуствения интелект, машинното обучение и приложната статистика. Основните програмни инструменти, с които обучаемите ще имат възможност да синтезират модели за идентификация и прогнозен анализ, установяват параметрични тенденции и зависимости са MATLAB и STATISTICA. Предвидените изследвания включват широк набор от апарати, между които са:

- ✚ Изкуствените невронни мрежи с обратно разпространение на грешката;*
- ✚ Адаптивните невронно-размити интерфейсни системи;*
- ✚ метод k – най-близки съседи;*
- ✚ метод Дърво на решенията за класификационни и регресионни задачи;*
- ✚ Дискриминантен анализ и Алгоритъм на Naïve Bayes;*
- ✚ Клъстерен, Корелационен и Регресионен анализ.*

Представени са различни примерни казуси при оперирането с икономически данни, които могат да бъдат заменени с потребителски информационни потоци от различни сфери на техниката и индустрията. Това позволява лабораторните упражнения да бъдат използвани в широк кръг от дисциплини, предназначени или свързани с изкуствения интелект и машинното обучение. Поради тази причина ръководството е приложимо за образователни цели от различни специалности като „Дигитална администрация“, „Комуникационни технологии и киберсигурност“, както и техни близки аналози.

Съдържанието на ръководството е оформено в десет упражнения, разделени в три модула, както следва:

- ✚ Модул №1: Концептуално моделиране на административни маркетингови бизнес процеси;*

- ✚ Модул №2: Клъстерен анализ, подходи за разпознаване и класификация във връзка с данни по икономически региони;
- ✚ Модул №3: Прогнозен анализ на финансови данни чрез изкуствен интелект, приложна статистика и машинно обучение.

Авторите изказват своите най-сърдечни благодарности на доц. д-р инж. Боян Димитров Карапенов за направените препоръки и дадени съвети при съставяне на ръководството!

МОДУЛ I

КОНЦЕПТУАЛНО МОДЕЛИРАНЕ НА АДМИНИСТРАТИВНИ МАРКЕТИНГОВИ БИЗНЕС ПРОЦЕСИ

Лабораторно упражнение №1

БИЗНЕС ПРОЦЕСИ ЗА РЕГИСТРАЦИЯ НА КЛИЕНТИ, ОБРАБОТКА НА КЛИЕНТСКИ ПОРЪЧКИ И ЗАЯВЛЕНИЯ

1.1. Цел на упражнението

Да се запознаят студентите с принципите, последователността от процедури и взаимовръзки, типовете информационни потоци при проектирането на модели на бизнес процеси относно обслужването и поддържането на регистрацията на клиенти, постъпили поръчки и заявления от тяхна страна.

1.2. Теоретична постановка

Концепция за моделиране на бизнес процеси

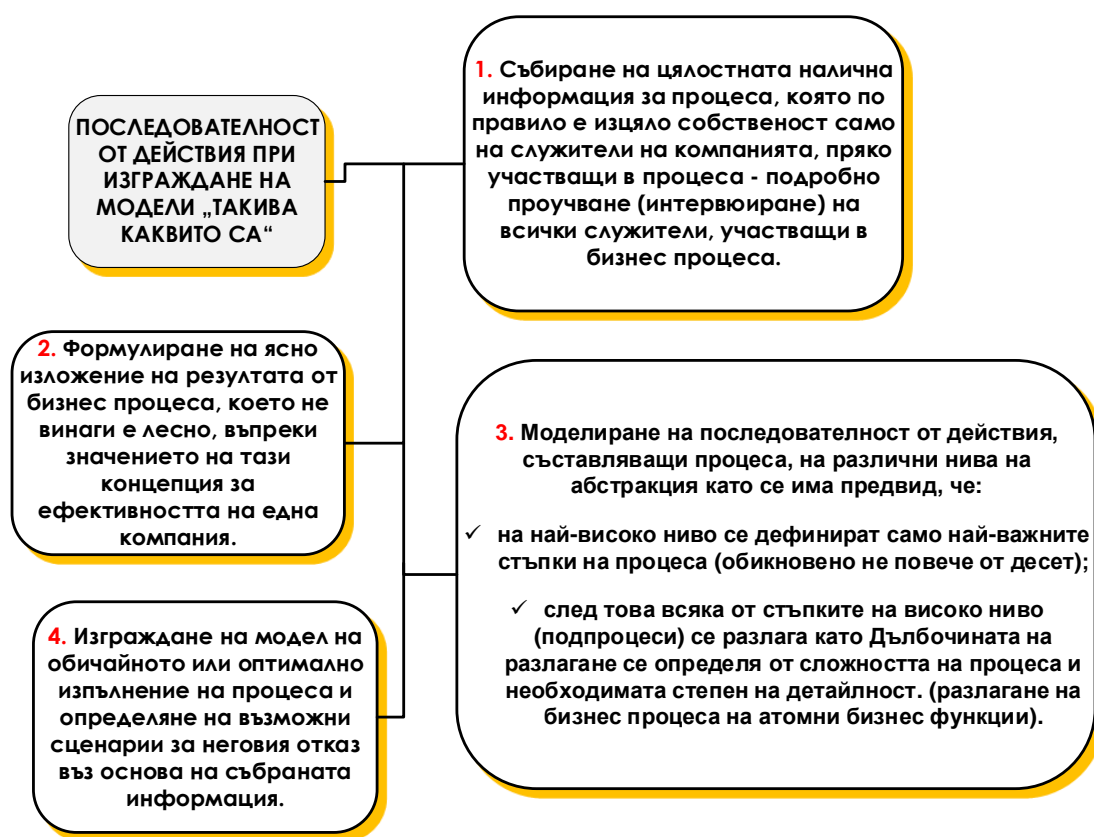
Спецификата на моделиране на бизнес процеси в дадена компания се определя главно от нейния предмет на дейност. Независимо от това процедурите по моделиране трябва да бъдат насочени към решаване на общ за всички видове компании набор от задачи:

- ✚ Прецизно определяне на потенциалния резултат от бизнес процеса и оценяване на значението му за бизнеса;
- ✚ Регламентиране на набора от дейности, които съставляват бизнес процеса. Ясното определение на набора от задачи и действия, които трябва да бъдат изпълнени е изключително важно за подробно разбиране на процеса;
- ✚ Определяне на реда, в който се извършват необходимите действия. Действията в рамките на един бизнес процес могат да се извършват както последователно, така и паралелно. Очевидно е, че паралелното изпълнение, ако е възможно, може да намали общото време за изпълнение на процеса и следователно да повиши неговата ефективност;
- ✚ Дефиниране на зоните на отговорност: Идентификация и следене кой служител или звено на компанията е отговорен за изпълнението на дадено действие или процес като цяло;
- ✚ Определяне на ресурсите, консумирани от бизнес процеса. Знаейки точно кой използва кои ресурси и за какви операции, ефективността на използването на ресурсите може да бъде увеличена чрез планиране и оптимизация;
- ✚ Разбиране на същността на взаимодействията между служителите, участващи в процеса, и фирмените отдели, оценяване и повишаване на ефективността на комуникацията между тях;
- ✚ Изучаване на движението на документите по време на процеса. Бизнес процесите оперират и г различни документи (на хартиен или електронен носител). Важно е да се уточни къде отиват документите или информационните потоци и да се

определи дали тяхното движение е оптимално и дали всички те наистина са необходими;

- ✚ Определяне на потенциалните тесни места и възможностите за подобряване на процесите, които ще бъдат използвани по-късно за оптимизиране на процеса;
- ✚ По-ефективно прилагане на стандартите за качество, като ISO 9000, и успешно преминаване на сертифициране;
- ✚ Използване на моделите на бизнес процесите като ръководство за нови служители;
- ✚ Ефективно автоматизиране на бизнес процесите като цяло или техните отделни стъпки, включително автоматизация на взаимодействието с външната среда - клиенти, доставчици, партньори;
- ✚ Анализиране на съвкупността от бизнес процесите на компанията и на нейна база описване на дейностите на предприятието като цяло [1].

Основната задача при моделирането на бизнес процесите на компанията е описание на съществуващите в него процеси с цел изграждане на техните модели „такива, каквито са“. За целта се преминава през последователността от действия, представена на фиг. 1.1.



Фиг. 1.1. Последователност при изграждане на модели „такива каквито са“

Необходимо е да се идентифицират събития, които могат да прекъснат процеса. В случай на прекъсване, може да е необходимо правилно да бъдат „върнати“

(компенсиращи) онези стъпки на процеса, които вече са завършени. За целта трябва да се определи логиката на компенсиращите действия за всяко прекъсващо събитие.

Важна част от изграждането на модел на бизнес процес е проучването на аспекти на неговата ефективност. Необходимо е да се разработи система от показатели и използват приети такива, за да се оцени ефективността на процеса - освен КРІ (ключов показател за ефективност, може да се наложат и допълнителни индикатори, характеризиращи разглеждания процес [2].

В процедурите на моделиране на бизнес процеси могат да бъдат получени различни графично интерпретирани резултати, обобщени на фиг. 1.2.



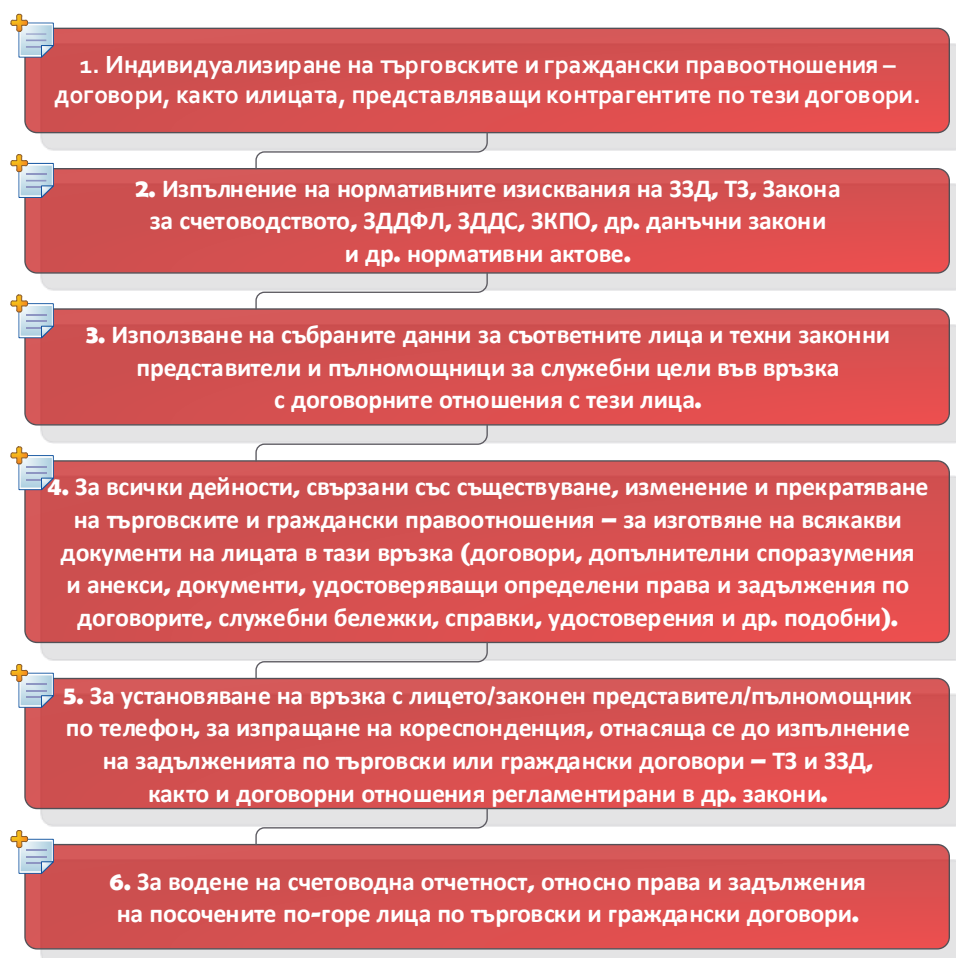
Фиг. 1.2. Резултати при процедурите на моделиране на бизнес процеси

Бизнес процес „Поддържане на регистър с клиенти с всички необходими за продажбите данни – име, адрес, идентификационен номер, кредитно досие и други“

За обработването на лични данни се смята всяко действие или съвкупност от действия, които могат да се извършват по отношение на личните данни с автоматични

или други средства, като събиране, записване, организиране, съхранение, адаптиране или изменение, възстановяване, консултиране, употреба, разкриване или предаване, разпространяване, актуализиране или комбиниране, блокиране, заличаване или унищожаване на данните. Процедурата се състои и в осигуряване на достъп до определена информация само за лица, чиито служебни задължения или конкретно възложени задачи налагат такъв достъп. Лични данни са всяка информация, отнасяща се до физическо лице, което е идентифицирано или може да бъде идентифицирано пряко или непряко чрез идентификационен номер или чрез един или повече специфични признаци [3, 4].

Регистърът, набиращ и съхраняващ лични данни на клиенти и доставчици, може да включва информация с оглед на различни цели/казуси, каквито са включените на фиг. 1.3.

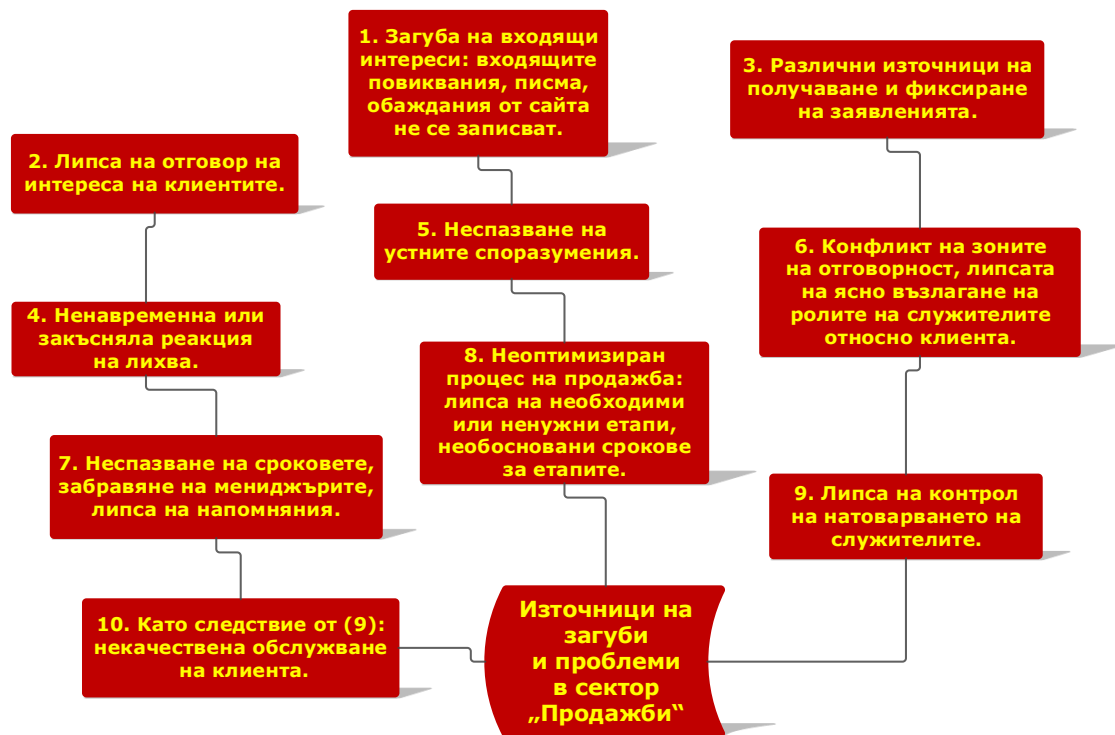


Фиг. 1.3. Информация в регистър на клиенти и доставчици

Бизнес процес „Обработка на клиентски поръчки/заявления“

Правилното използване на бизнес процесите в системи за управление на взаимоотношенията с клиенти (CRM) позволява внедряването на концепцията за управление на процесите на компанията. По този начин могат да бъдат изградени

гъвкави информационни системи, които бързо могат да се адаптират към характеристиките на организацията.



Фиг. 1.4. Источници на загуби в сектор „Продажби“

Благодарение на използването на BPM (Управление на бизнес процеси) технологиите е възможно значително да се:

- ✚ намалят загубите;
- ✚ оптимизират бизнес процесите;
- ✚ подобри цялостната системна ефективност и управление.

Най-често срещаните източници на загуби и проблеми в отдела по продажби са онагледени на фиг. 1.4. Посочените проблеми могат да бъдат значително ограничени чрез изграждане на ефективен бизнес модел „Продажби“, насочен към минимизиране на загубите. В таблица 1.1 са посочени възможните резултати от използването на BPM технологията в CRM системите и ефектите след внедряване на CRM [5, 6].

**Таблица 1.1. Възможности за внедряване на BPM
в CRM и ефекти след внедряване на CRM**

Възможности за използване на BPM технологията в CRM	Ефект след внедряването на CRM системата
1. Автоматична регистрация на входящи повиквания: писма от различни пощенски кутии, обаждания	<ul style="list-style-type: none"> • нито едно искане няма да бъде загубено или забравено; • всички източници на приложения се натрупват в CRM
2. За всяко входящо обаждане стартира бизнес процес, който формира задачи за отговорни лица	<ul style="list-style-type: none"> • гаранция за отговор на обжалването, на служителя е възложена непосредствената задача в системата
3. Ясно регулиране на етапите на бизнес процеса в системата с описание на задачите на всеки етап - методологията на „Пазар на продажбите”	<ul style="list-style-type: none"> • има ясна задача за всяка стъпка; • за всеки клиент има разбиране на какъв етап на продажба се намира
4. Всеки етап от бизнес процеса се регулира във времето, т.е. служителите получават задачи със срокове	<ul style="list-style-type: none"> • осигуряване на навременна обратна връзка с клиентите; • повишена лоялност поради ефективността на добив; • намаляване на цикъла на продажбите
5. След всеки контакт с клиента, мениджърът трябва да запише събитие в системата с информация за резултатите от разговора, в противен случай бизнес процесът няма да продължи	<ul style="list-style-type: none"> • гарантира спазването на устни споразумения с клиента; • независимост от напускането на мениджъра (цялата информация се съхранява в системата)
6. За всеки етап се назначава отговорен служител, който може да варира в зависимост от различните условия на транзакцията	<ul style="list-style-type: none"> • предотвратява се конфликт на отговорност
7. Наличието на система от напомняния за задачи и събития	<ul style="list-style-type: none"> • неспазване на сроковете поради човешки фактори
8. Способността да се контролира натоварването на мениджърите и разпределението на задачите (планиране)	<ul style="list-style-type: none"> • осигуряване на равномерно натоварване на мениджърите и в резултат на това висококачествено обслужване на клиентите
9. Наличие на инструменти за анализ на бизнес процеси (фуния за продажби, анализ на причините за неуспеха на транзакциите и др.)	<ul style="list-style-type: none"> • CRM + BPM предоставя инструменти за анализ и вземане на ефективни управленски решения

Бизнес процес „Бизнес моделиране на основата на BPMN стандарт“

“Бизнес моделирането“ (моделирането на бизнес процеси) е широко понятие, отнасящо се до анализа, дизайна и изпълнението на бизнес процеси, като се използват езици за програмиране (BPMN - Business Process Modeling Notation, BPEL - Business Process Execution Language, UML - Unified Modeling Language, WS-CDL - Web Services Choreography Description Language) и други индустриални стандарти. Бизнес анализаторите и мениджърите използват бизнес моделите, за представяне на настоящото



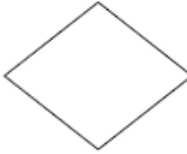



и прогнозиране на бъдещото състояние на своите предприятия. Като сравняват високо прецизно разработени модели на “as is” и “to be” предприятия, специалистите по моделиране могат да установят специфичните трансформации, които могат да имат за резултат количествено измерими подобрения в техните предприятия” [7-9].

Нотацията BPMN е предназначена за графично представяне на работните процеси с различно ниво на детайлност. Основните обекти в BPMN нотацията са:

- ✚ Елементи на работния поток;
- ✚ Данни;
- ✚ Свързващи елементи;
- ✚ Зони на отговорност;
- ✚ Артефакти – представлява група или анотация.

В таблица 1.2 и таблица 1.3 са дадени детайлно описани основните типове обекти и моделирането на разклонения в BPMN стандарт [10].

Таблица 1.2. Основни типове обекти в BPMN нотация

Обект	Описание	Графичен символ
Събитие	Представява причина (тригер) или въздействие (резултат) при изпълнението на процеса. Може да бъде начално, междинно или крайно за процеса.	
Действие, дейност	Представява дейностите, стъпките, задачите, които се изпълняват в процеса.	
Условно разклонение	Използва се за разделяне или сливане на клонове в процеса	
Поток на дейностите	Служи за изобразяване на последователността, в която се изпълняват дейностите от процеса	
Поток от информация	Използва се за изобразяване обмена на информация между двама участници в процеса	
Асоциация	Използва се за изобразяване на връзки между артефакти и елементи на работния поток.	




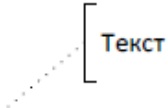


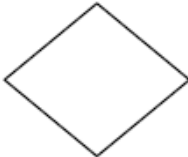



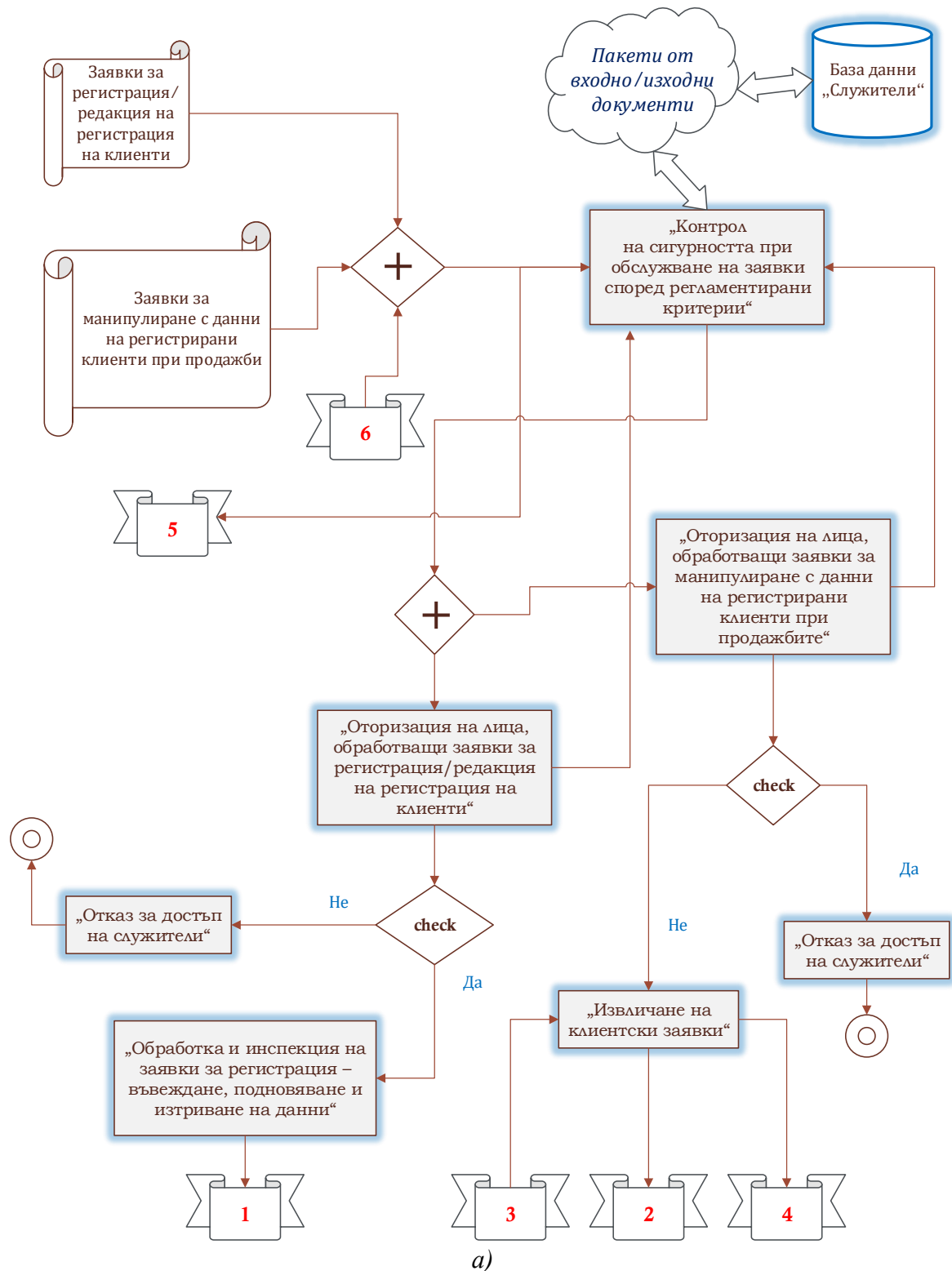
Пул	Използва се за изобразяване на участник в процеса – експерт, организационно звено, клиент	
Данни	Използва се за изобразяване на входяща / изходяща информация или данни в процеса	
Група	Използва се за групиране на графични елементи, които принадлежат към обща категория.	
Анотация	Използва се за въвеждане на допълнителна текстова информация в процеса	
Подпроцес	Използва се за изобразяване на връзка към процес (подпроцес), моделиране на отделна диаграма	

Таблица 1.3. Разклонения в BPMN нотация

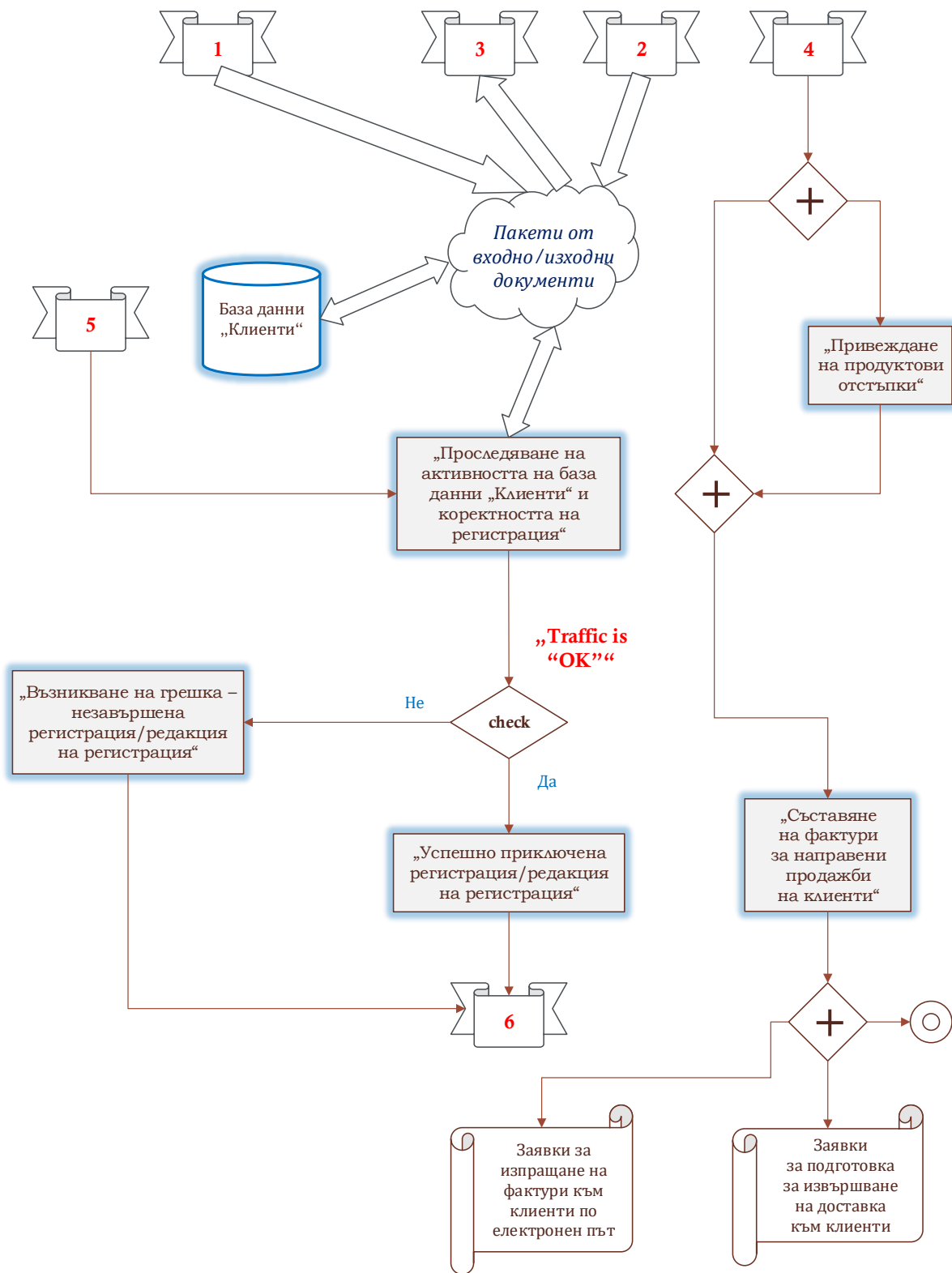
Обект	Описание	Графичен символ
Изключващо условно разклонение	Клоновете на процеса са алтернативни и може да се изпълни само един от тях	 
Включващо условно разклонение	Може да се изпълни един или повече от клоновете в процеса	
Паралелно разклонение	Изпълняват се всички клонове в процеса без да се изпълнява проверка за изпълнени условия	
Разклонение, базирано на събитие	На база настъпило събитие се определя кой клон от процеса трябва да се изпълнява.	

1.3. Задачи за изпълнение

- ❖ Въз основа на структурната диаграма от фиг. 1.5 да се моделира в програмна среда **Bizagi Modeler** бизнес процесът „Поддържане на регистър с клиенти с всички необходими за продажбите данни – име, адрес, идентификационен номер, кредитно досие и други“;



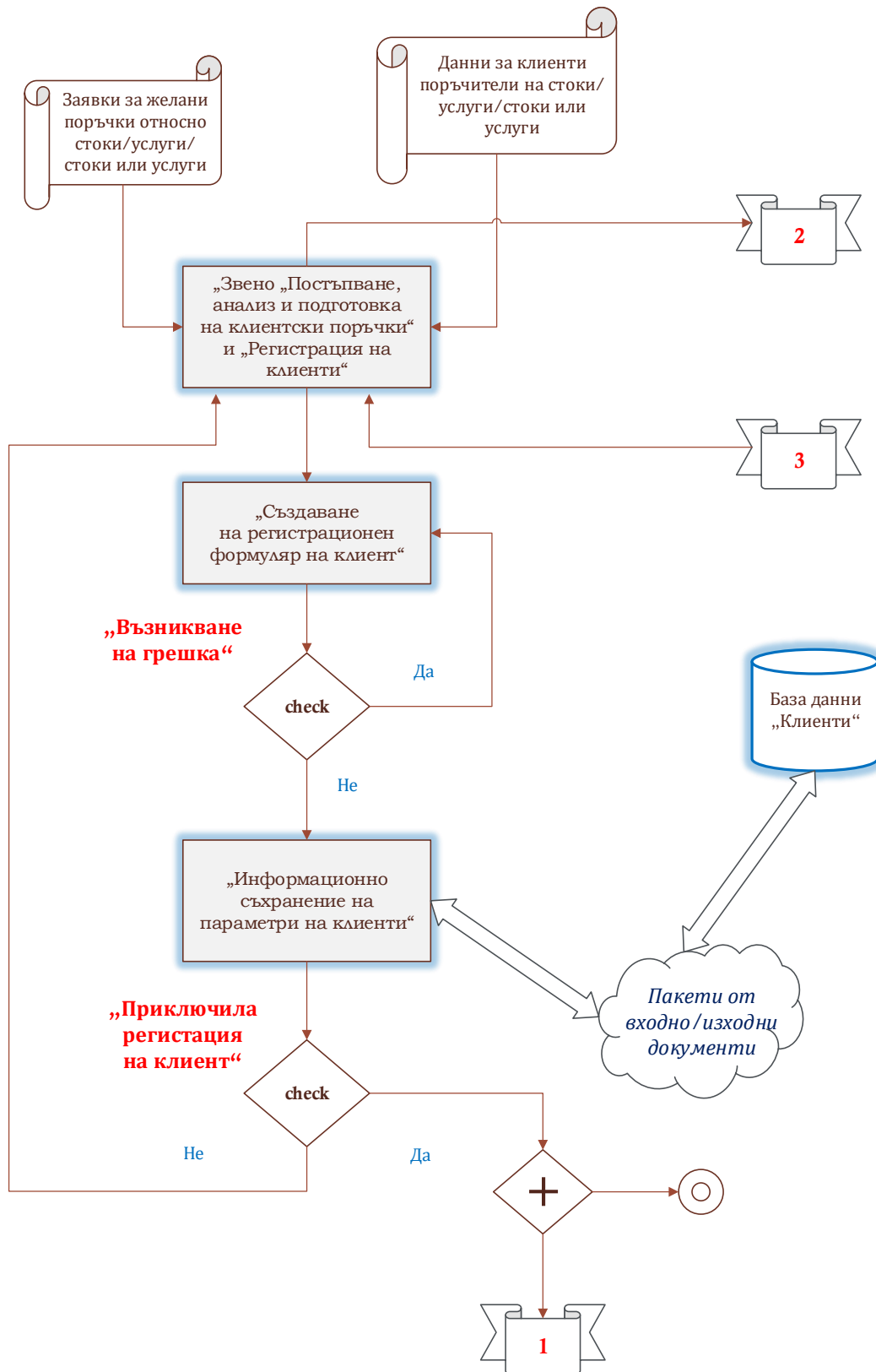
a)



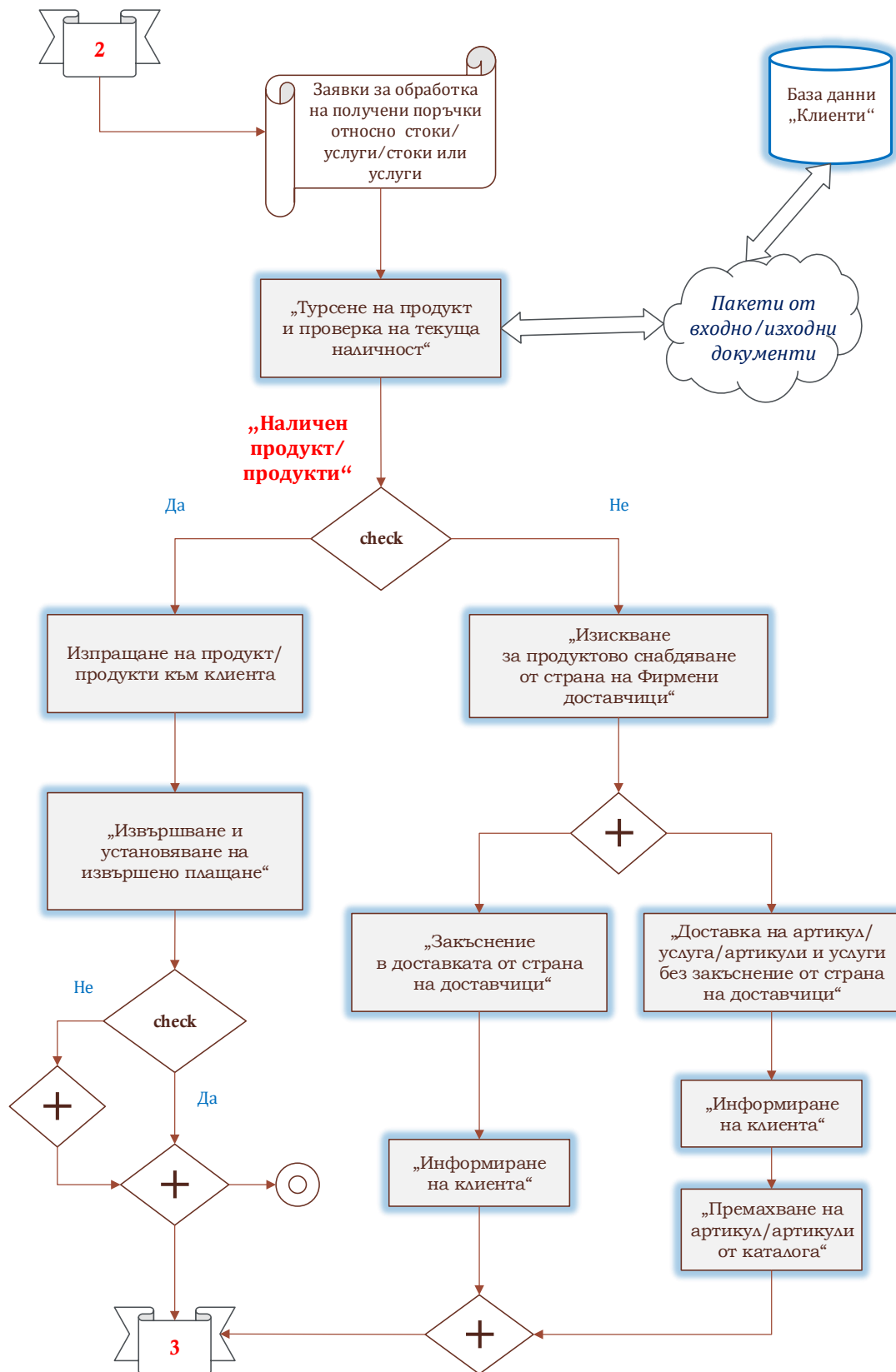
б)

Фиг. 1.5. Информационна диаграма на бизнес процеса „Поддържане на регистър с клиенти с всички необходими за продажбите данни – име, адрес, идентификационен номер, кредитно досие и други“ – модул №1 а) и модул №2 б)

- ❖ Да се проектира модел съобразно изложената функционална диаграма на бизнес процеса „Обработка на клиентски поръчки/заявления“ на фиг. 1.6, използвайки възможностите на продукта [Bizagi Modeler](#);



a)



б)

Фиг. 1.6. Информационна диаграма на бизнес процеса „Обработка на клиентски поръчки/заявления“ – модул №1 а) и модул №2 б)

- ❖ Да се моделират разглежданите бизнес процеси посредством средата **PowerDesigner** и съпоставят наличните програмни ресурси между използваните продукти;
- ❖ Да се анализира функционалността на разглежданите примерни концептуални диаграми на бизнес процеси и проектират модели на оптимални архитектури с включване на допълнителни персонални потребителски роли и бази данни за обмен на входно-изходни документи.

1.4. Контролни въпроси

1. Какъв набор от задачи обхваща концепцията за моделиране на бизнес процеси?
2. В какво се състои същността на моделите “такива, каквито са“?
3. С какви типове данни се свързва създаването и поддържането на „регистри с информация за клиенти“?
4. Кои са потенциалните източници на загуби при реализация на дейности по продажби на артикули, продукти и услуги на клиенти?
5. По какви програмни стандарти могат да бъдат моделирани, изследвани и анализирани бизнес процеси по икономически и индустриални области?

Лабораторно упражнение №2

БИЗНЕС ПРОЦЕСИ ЗА ЕКСПЕДИЦИЯ И ФАКТУРИРАНЕ, УПРАВЛЕНИЕ НА ДАНЪЦИТЕ И КРЕДИТЕН КОНТРОЛ, ПРОСЛЕДЯВАНЕ НА ВЗЕМАНИЯ И БЛОКИРАНЕ НА НЕКОРЕКТНИ ПЛАТЦИ

2.1. Цел на упражнението

Да се запознаят студентите с основните и спомагателни дейности и съпътстващи процедури при проектиране на модели на бизнес процеси във връзка с процесите експедиция и фактуриране, мониторинг и контрол на данъците и отпускането на кредити, проследяване на вземания и блокиране на некоректни платци.

2.2. Теоретична постановка

Бизнес функция „Експедиция на стоки“

Експедицията на стоки се осъществява в три направления:

- ✚ подготовка на поръчките - определяне на мястото, от което да се вземе стоката, нужното количество и т.н.;
- ✚ комисиониране - стоките се извеждат от сектора за съхранение и се придвижват към сектора за експедиция;
- ✚ изнасяне на стоките - в зависимост от препоръките за опаковане и транспортиране на дадената стока, тя се фолира, опакова в различни средства за по-лесното и пренасяне, било то кашони, палети и др.

При изготвянето на поръчката всяка стока се описва, за да бъде изготвена нужната документация свързана с експедирането на стоките. След като поръчката е изготвена и документацията за нея е оформена, стоката е готова да напускане на склада.

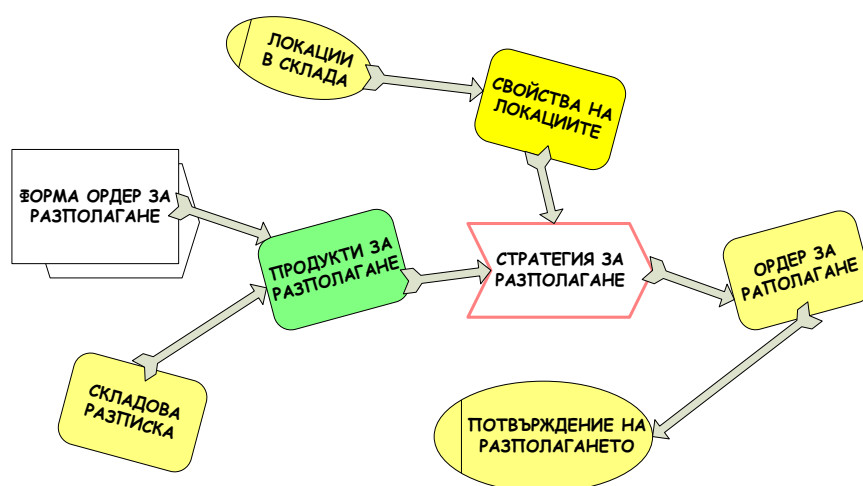
Цялостната организация на придвижването на стоките до склада, вътре в него и до напускането може да бъде базирана на различни софтуерни решения. Ето някои примери за внедрени решения, използвани от компании в България [11, 12]:

- ✚ SmartStock.WMS - интегрирано решение за управление на складове в реално време (High Performance Real Time Warehouse Management Solution) на база на Wi-Fi безжична мрежа. В основата на системата е заложено уникално идентифициране на всички стокони единици и места в склада, чрез използването на баркод технология. WMS оптимизира изпълнението на всички операции в склада - приемане на стока, поставяне, вътрешно преместване, пикиране/комисиониране, експедиция. Решението предлага широк набор от инструменти за управление на складовите процеси - следене на складовите наличности, управление на процесите за изпълнение на поръчката, приемане/експедиция на стоката, система за оптимално разпределение на складовите площи, инвентаризация, отчети по логистични показатели, различни видове комуникация. Системата създава електронни документи, отразяващи товаропотока, а това изключва грешки, свързани с загуба на документи или тяхна повреда. SmartStock.WMS оптимизира и улеснява всички складови операции, осигурявайки максимална ефективност на работа. Системата управлява

действията и следи за тяхното точно и коректно изпълнение. Последователният контрол на всички складови операции отстранява риска от човешка грешка;

✚ BORA BS Material Management – софтуерът разполага с три основни модула, както следва:

- ✓ „Склад“ и „Инвентаризация“ – Inventory Management - осигурява контролиране на материалните активи по веригата на доставка, обслужване на производството и продажбата на клиенти. Това отразява “външно” за склада движение на стоките и статичната картина за наличностите на стоки като материални активи и тяхното партидно отчитане. Контролът на последните се извършва чрез модул „Инвентаризация“;
- ✓ “Управление и мониторинг на склада” – Warehouse management - извършва контрола на вътрешното разположение и движение по клетки — локации в склада;
- ✓ “Пласмент и експедиция” – Shipment Management - организира процесите на контрол на материалите за експедиция след напускането им от склада.

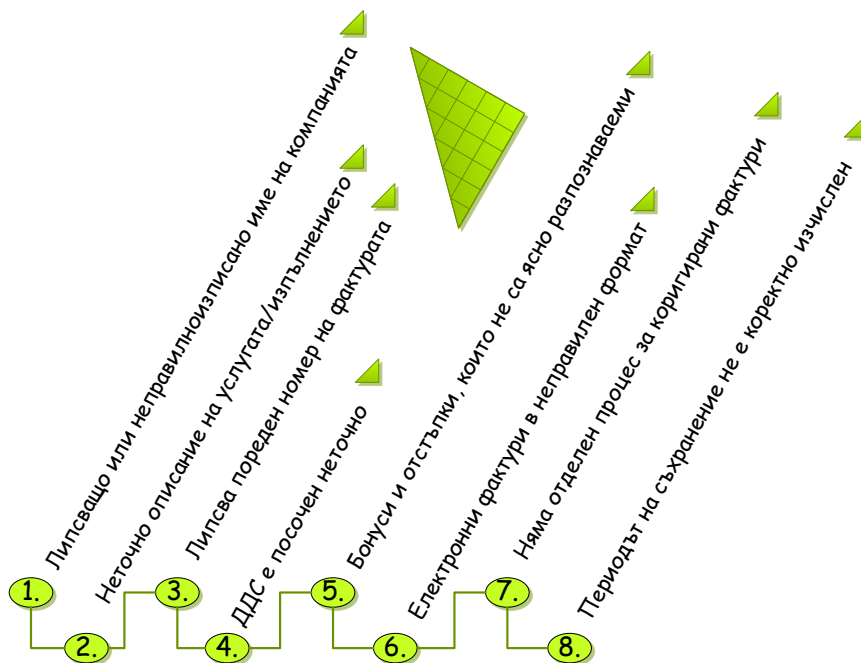


Фиг. 2.1. Стратегия за експедиция на стоки в склада по партиди

Една от възможните стратегии за разполагане на стоки (експедиция на стоки) в склада се свързва с извършване на процеса по партиди, илюстрирано на фиг. 2.1. Стратегията се изпълнява въз основа на приключен складов документ, състоянието на локациите в склада и правилата (свойствата) на локациите.

Бизнес функция „Фактуриране на закупени стоки“

Една от големите ползи, с които разполагат компаниите в резултат от внедряване на системите за управление на информацията, е автоматизацията на плащанията и обработката на фактури, базирана на M-Files (решение за управление на информацията - EIM).



Фиг. 2.2. Допускани грешки в процеса на фактуриране

1. Наименование	2. Номер	3. Дата на издаване
4. Име на доставчик и получател	5. Адрес на доставчик и получател	
5. Единен идентификационен код на доставчик и получател		
7. Предмет и стойностно изражение на съответната операция		
8. / 9. Регистрационен номер на доставчика / получателя		
10. Дата на доставката или дата, на която е получено плащането		
11. Данъчна основа на доставката	12. Данъчна ставка	
13. Основание за прилагане на друга ставка различна от 20%		
14. Размер на ДДС	15. Сума за плащане	
16. При доставка ново превозно средство, обстоятелства, определящи го като ново		

Фиг. 2.3. Задължителни реквизити за изрядна фактура според Закона за ДДС

Подходящ пример за това е компанията American Solutions for Business (ASB) - бързоразвиваща се фирма за печат и дистрибуция на промоционални продукти със

стотици продажби и хиляди доставчици в САЩ. Използвайки M-Files, ASB е в състояние да предостави достъп до фактури на упълномощени лица, директно от други бизнес приложения, като например PeopleSoft. По този начин те успяват да:

- ✚ оптимизират времето за обработка на фактурите на доставчика,
- ✚ изпълнят по-бързо заявките;
- ✚ рационализират движението на документи в цялата организация.

ASB изчислява, че са намалили времето за обработка на фактурите с три дни и могат да се похвалят с възможност да контролират целия процес с голяма точност.

Фигура 2.2 показва основните грешки, които се допускат в процеса на фактурирането и могат да бъдат доведат до забавяне на плащането и т.н., а фиг. 2.3 онагледява задължителните присъстващи реквизити за изрядност при изготвяне на фактури според Закона за ДДС [13].

Управление на данъците и кредитен контрол

По отношение на бизнеса индивидуалните и корпоративните данъчни ставки се отнасят до:

- ✚ Директно данъчно облагане, включително данъчни компенсации, корпоративно данъчно планиране и реорганизиране на бизнеса;
- ✚ Индиректно данъчно облагане, включително ДДС, ДДФЛ (Данък върху доходите на физически лица) и Социално осигуряване;
- ✚ Международно данъчно облагане;
- ✚ Работна заплата и планиране на печалбата;
- ✚ Планиране на транзакциите – сливания, продажби, реорганизации.

Основните техники за редуциране на кредитния риск могат да бъдат формулирани по следния начин:

- ✚ Необезпечени експозиции — Балансова стойност;
- ✚ Обезпечени експозиции — Балансова стойност;
- ✚ Експозиции, гарантирани с обезпечение;
- ✚ Експозиции, гарантирани с финансови гаранции;
- ✚ Експозиции, гарантирани с кредитни деривати.

Фигура 2.4 обобщава основните експозиции, имащи отношение към кредитния риск и ефекта от неговата редукция.

Отделите “Управление на риска” осъществяват следните функции в банковите институции:

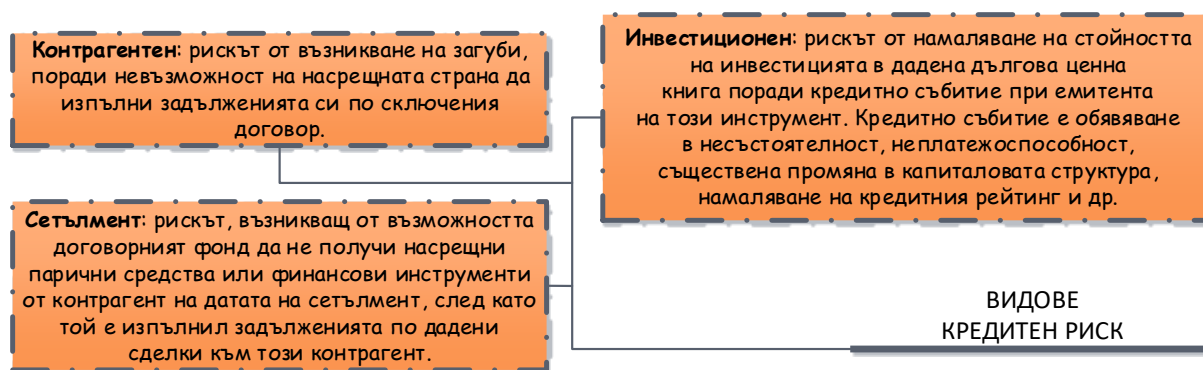
- ✚ Разработват и прилагат системата за управление на риска на всяка колективна инвестиционна схема;
- ✚ Изпълняват правилата и процедурите по управление на риска;
- ✚ Гарантират съответствието с одобрената вътрешна система за ограничаване на риска на колективната инвестиционна схема, включително с нормативно

определените лимити за стойността на общата рискова експозиция и риска на насрещната страна съгласно ЗДКИСДПКИ;

- ✚ Консултират управителния орган на управляващото дружество относно определянето на рисковия профил на всяка колективна инвестиционна схема;
- ✚ Докладват редовно пред управителния орган на управляващото дружество и на лицата, осъществяващи надзорни функции, когато е приложимо, относно:
 - ✓ съответствието между текущото ниво на риск, на което е изложена всяка управлявана от него колективна инвестиционна схема и одобрените рискови профили на тази схема;
 - ✓ съответствието на всяка колективна инвестиционна схема с вътрешната ѝ система за ограничаване на риска;
 - ✓ адекватността и ефективността на процеса за управление на риска и по-специално показващ дали са предприети подходящи коригиращи мерки в случаите, когато са констатирани недостатъци;
- ✚ Докладват редовно пред УС на УД, представяйки текущото ниво на риск, на което е изложена всяка колективна инвестиционна схема и за текущите или предвиджани нарушения на ограниченията, като по този начин се осигурява предприемането на навременни и подходящи действия.



Фиг. 2.4. Класове експозиции, имащи отношение към кредитния риск и ефекта от неговата редукция



Фиг. 2.5. Видове кредитен риск

Възможността от намаляване на стойността на експозицията в един финансов инструмент при неочаквани събития от кредитен характер, свързани с емитентите на финансови инструменти, с насрещната страна по борсови и извънборсови сделки, както и държавите, в които те извършват дейност [14, 15]. Видовете кредитен риск са дадени на фиг. 2.5.

Бизнес процеси „Проследяване на вземанията от клиенти и блокиране на поръчки на некоректни платци“ и „Блокиране на поръчки на некоректни платци“

Управлението на вземанията от клиентите е изключително съществена част, свързана с управлението на ликвидността и паричните потоци на фирмите. Добрата ликвидност е приоритет и задължително изискване за всички собственици и мениджъри на фирми. Високо ликвидната компания има предимството да се възползва от различните бизнес възможности при тяхно потенциално възникване. Управлението на ликвидността е една от най-важните отговорности на финансовите мениджъри и се следи с повишено внимание от висшия мениджмънт на всяка една компания [15].

Управлението на този процес включва не само напомняне на клиентите да си платят задължението, но също така и идентифициране на причините, поради които се стига до лоши вземания и неплащане от задълженията [16]. Доброто управление на вземанията е многообхватен процес, включващ определени процедури (фиг. 2.6) и преминаващ през регламентирани етапи, на които се решават целеви въпроси (фиг. 2.7).

В теорията и практиката са известни няколко подхода, свързани с решаването на проблема с трудносъбираемите вземания, представени на фиг. 2.8. На фиг. 2.9 е показан алгоритъм на модел за обезценка на финансови активи в Международни стандарти за финансово отчитане МСФО 9.

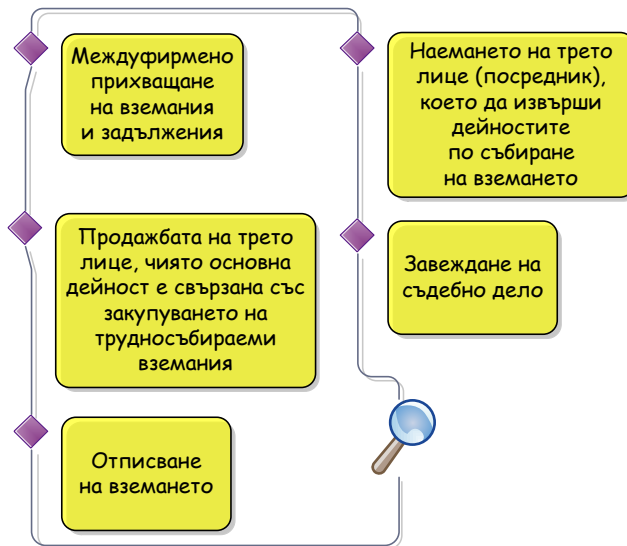
На първо място, подходът решава проблема със закъснялото признаване на възникнали, но неотчетени кредитни загуби. Изискването за признаване на коректив за обезценка за всички финансови активи в размер, равен поне на очакваните кредитни загуби за 12 месеца, намалява ефекта от систематичното надценяване на приходите от лихви. Въпреки че липсва каквато и да е теоретична обосновка за избрания период от 12 месеца, по този начин се прави опит за разпределение на стойността на първоначално очакваните кредитни загуби за срока на инструмента [17].



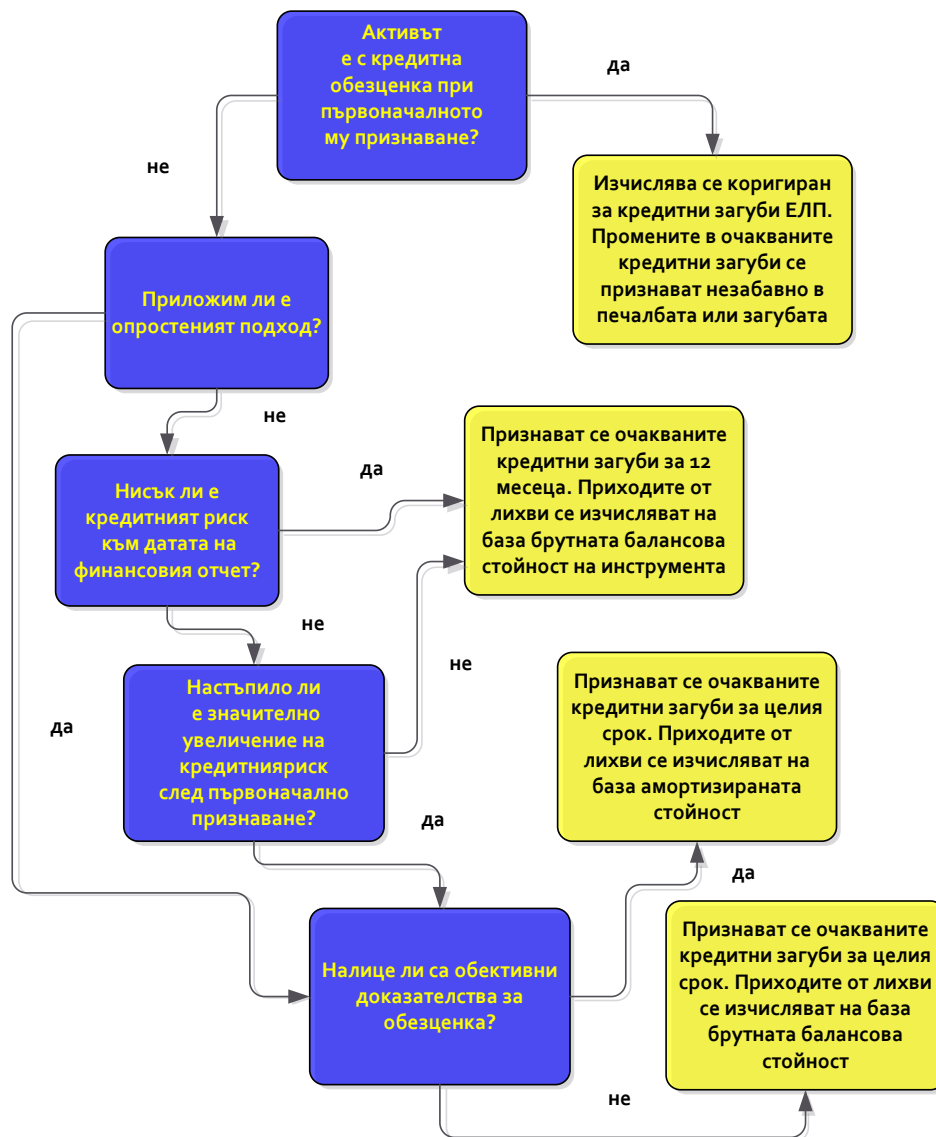
Фиг. 2.6. Процедури по отношение на управление на вземанията



Фиг. 2.7. Етапи при управление на вземанията



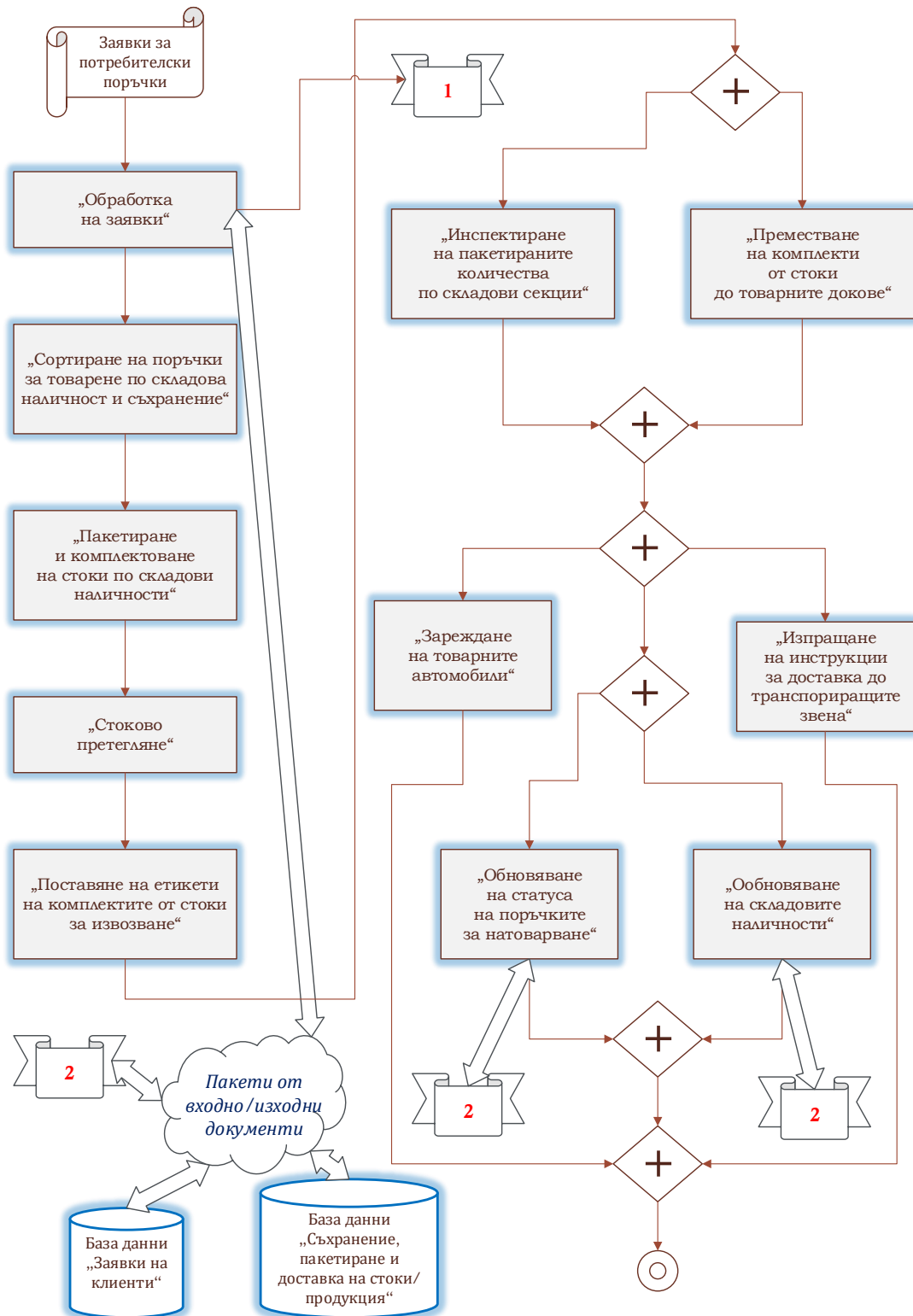
Фиг. 2.8. Подходи при трудносъбираеми вземания



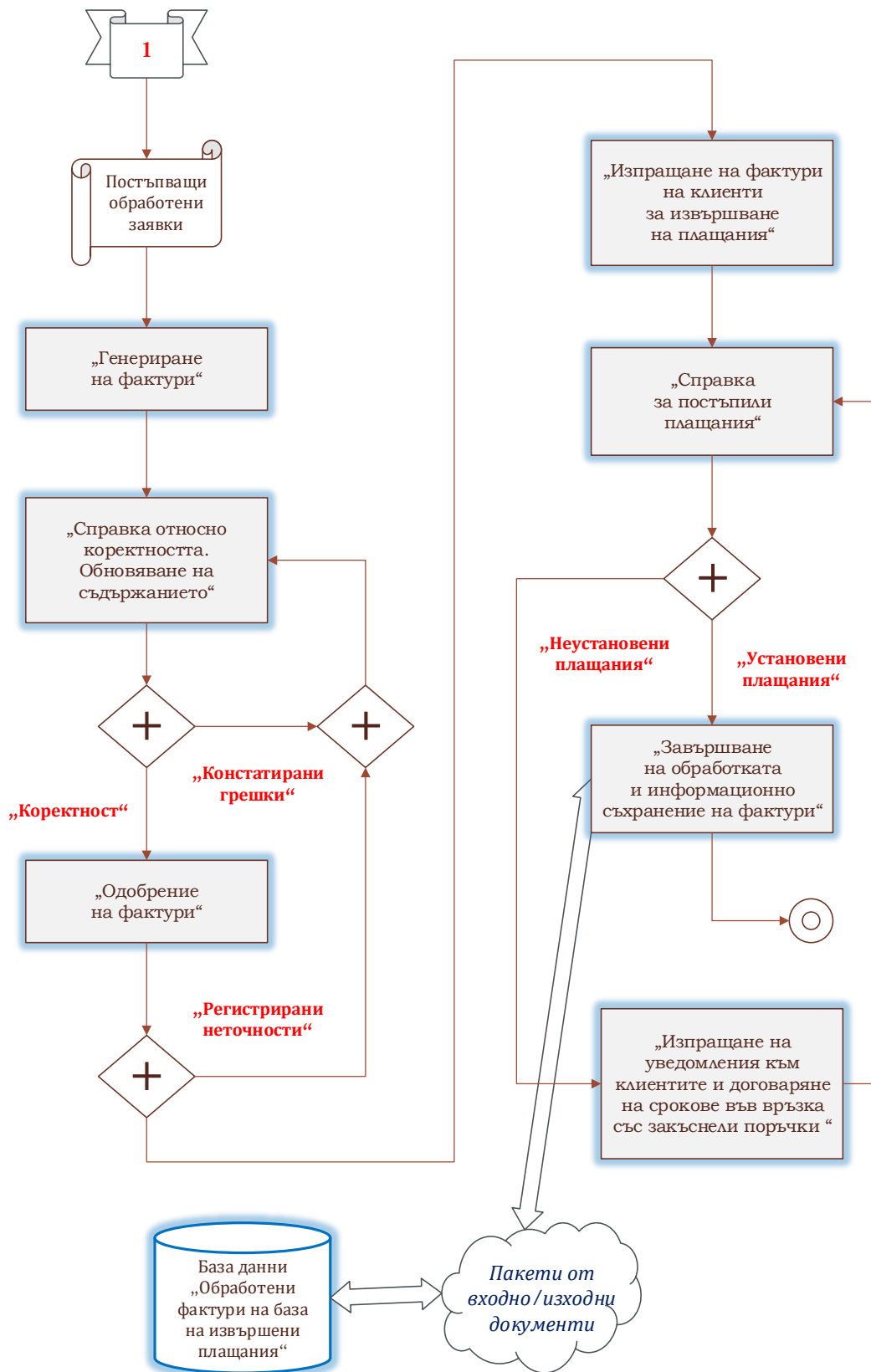
Фиг. 2.9. Модел за обезценка на финансови активи в МСФО 9

2.3. Задачи за изпълнение

- ❖ Да се построи модел на бизнес процес „Експедиция и фактуриране“ като за база бъде използвана функционалната диаграма от фиг. 2.10 чрез продукта Bizagi Modeler;



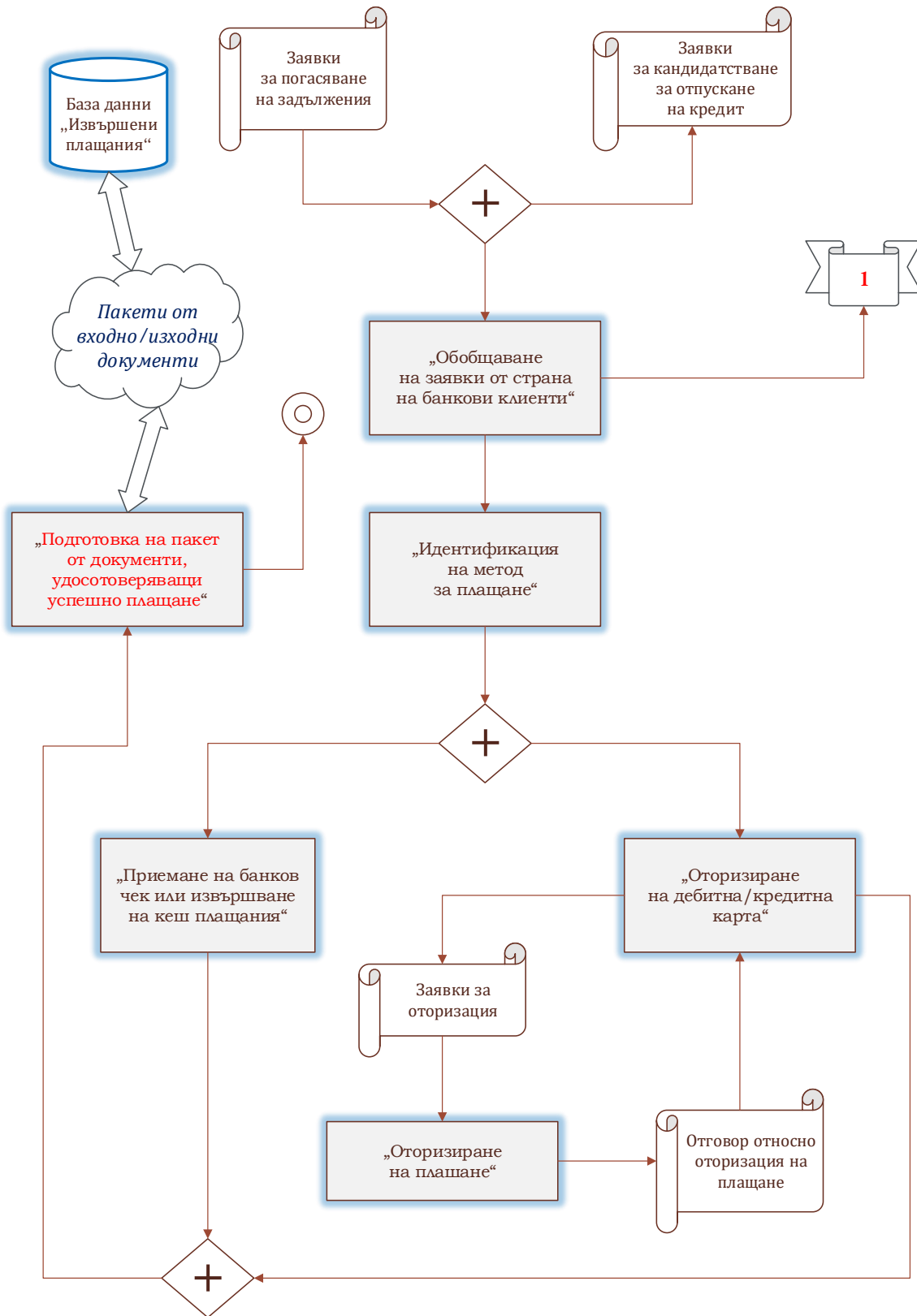
a)



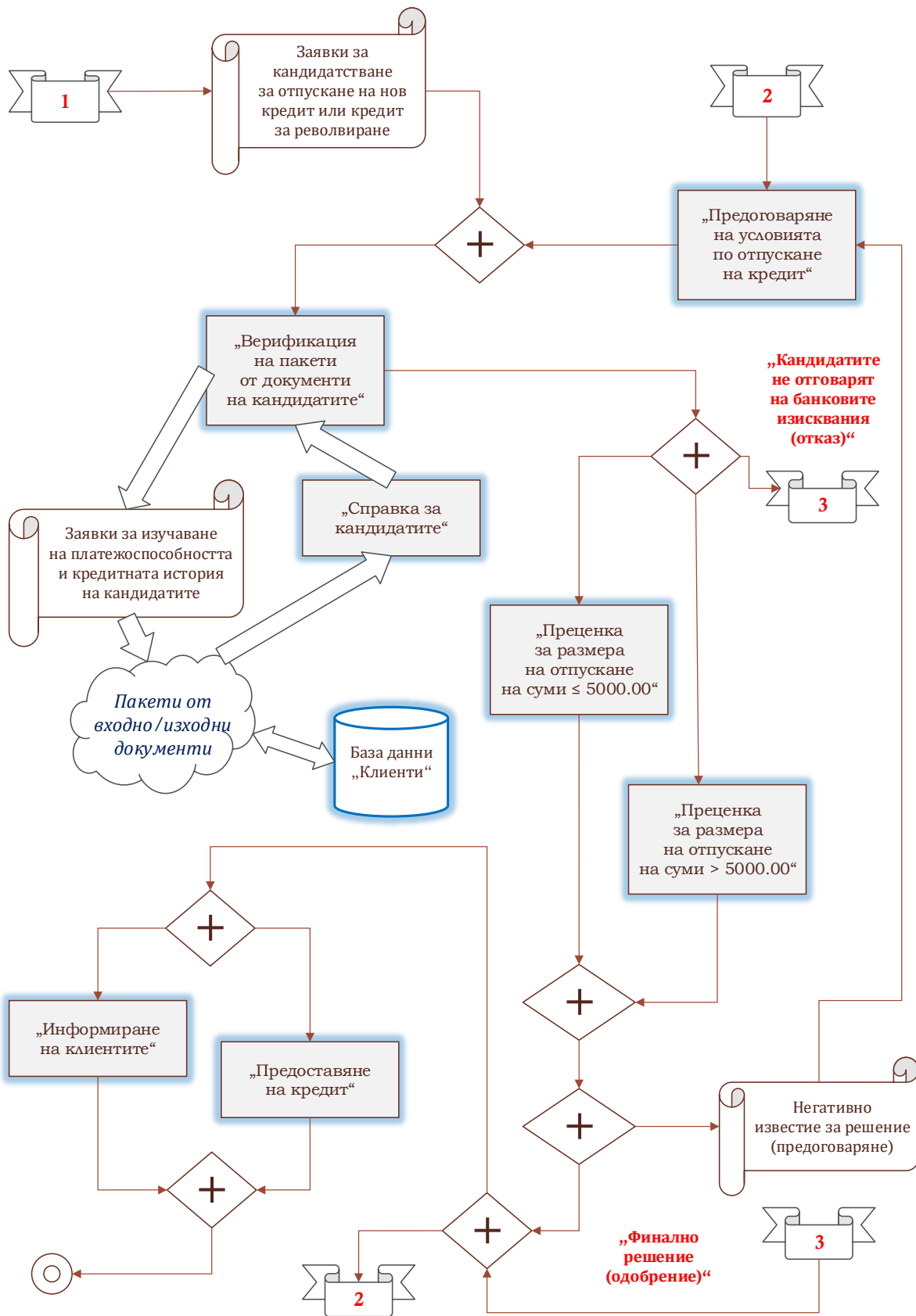
б)

Фиг. 2.10. Адаптирана информационна диаграма на бизнес процес „Експедиция и фактуриране“ – модул №1 а) и модул №2 б)

- ❖ Да се проектира BPMN модел на бизнес процеса „Управление на данъците и кредитен контрол“ на основата на примерната функционална диаграма на фиг. 2.11 в средата [Bizagi Modeler](#);



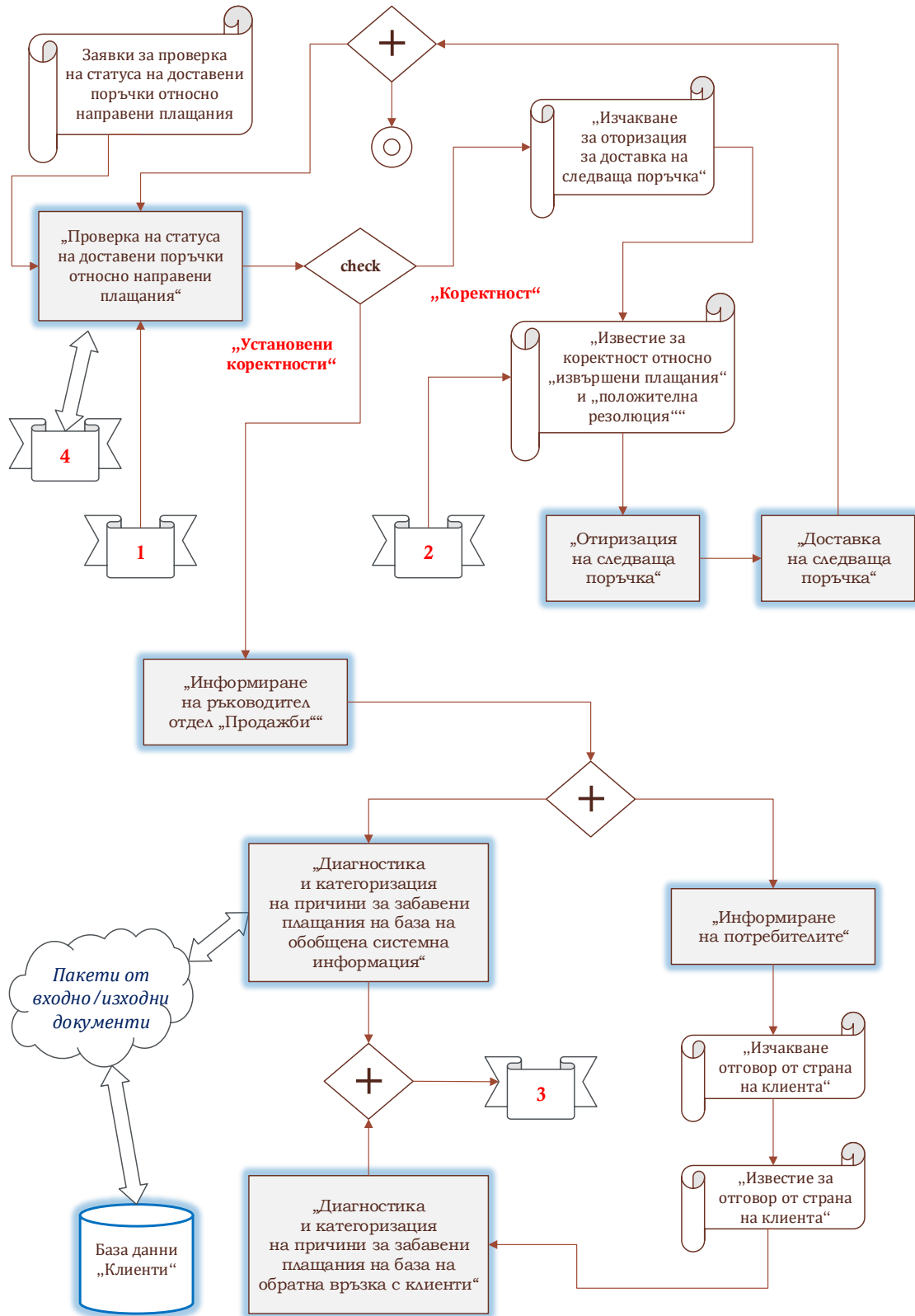
a)



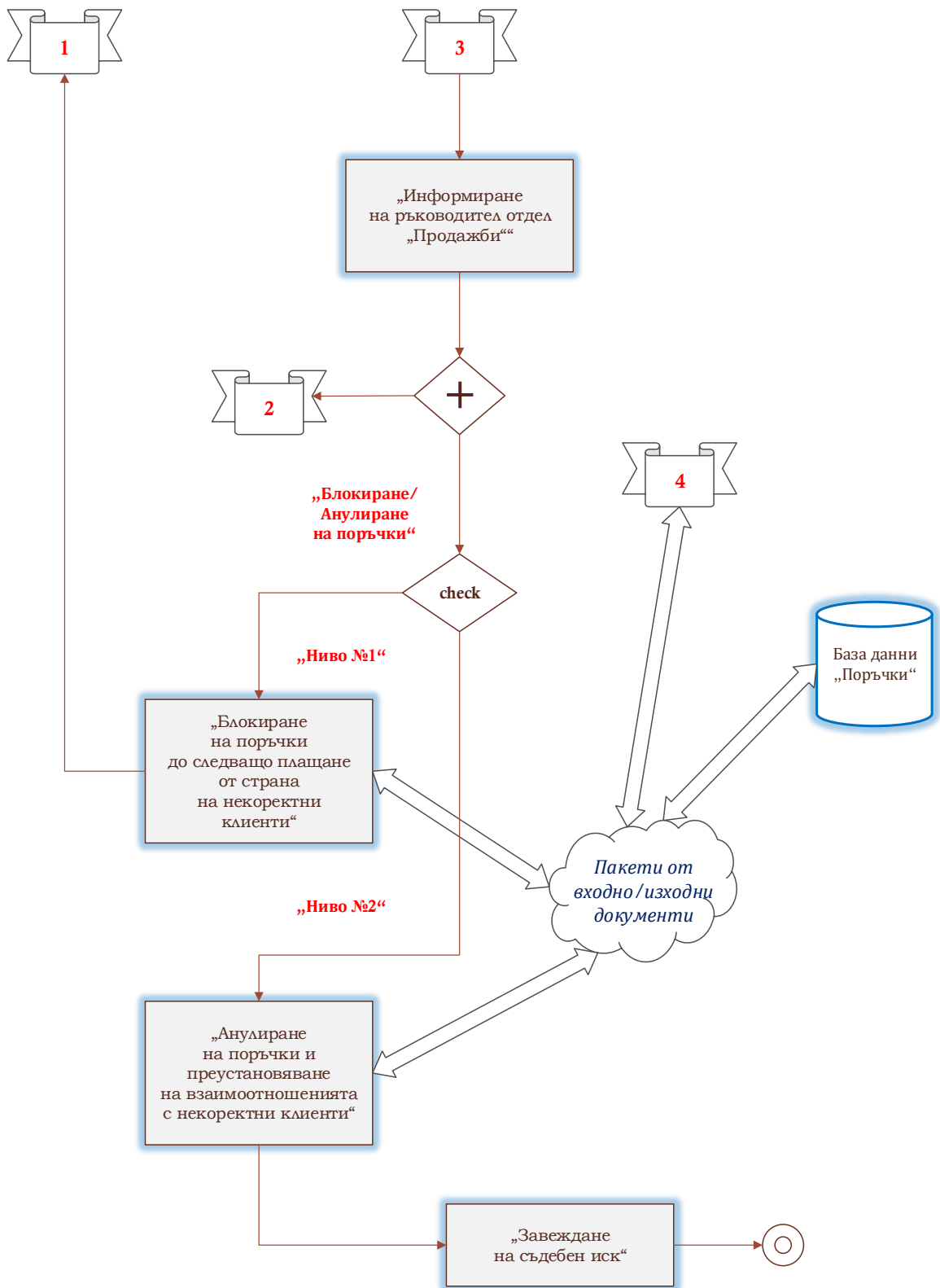
б)

Фиг. 2.11. Адаптирана информационна диаграма на бизнес процеса „Управление на данъците и кредитен контрол“ – модул №1 а) и модул №2 б)

- ❖ Да се моделират като обща бизнес концепция чрез BPMN стандарт процесите „Проследяване на вземанията от клиенти и блокиране на поръчки на некоректни платци“ и „Блокиране на поръчки на некоректни платци“ според представената функционална диаграма на фиг. 2.12 посредством продукта [Bizagi Modeler](#);



a)



б)

Фиг. 2.12. Адаптирана информационна диаграма на бизнес процеси „Проследяване на вземанията от клиенти и блокиране на поръчки на некоректни платци“ и „Блокиране на поръчки на некоректни платци“ – модул №1 а) и модул №2 б)

- ❖ Да се проектират модели на указаните целеви бизнес функции в среда на платформата **PowerDesigner** и съпоставят наличните програмни ресурси между използваните програмни инструменти;
- ❖ Да се изучат, анализират и оценят спецификата и особеностите на дадените примерни концептуални диаграми на бизнес процеси и оптимизират по отношение на включване на допълнителни персонални потребителски роли и бази данни към входно-изходни документи за обмен на данни от различни формати.

2.4. Контролни въпроси

1. На какво се основават стратегиите за експедиция на стоки в складови помещения по секционирание на партиди?
2. Какви са допусканията грешки и необходимите реквизити, определящи изрядността при фактуриране?
3. В какво се състоят прилаганите в бизнес практиката корпоративни и данъчни ставки?
4. Въз основа на каква база се оценява кредитният риск и ефективността от редуциране?
5. Какви критерии са обвързани с отделните етапи, имащи отношение към управление на вземанията?
6. Кое определя значимостта на модела на обезценка на финансови активи МСФО9?

**БИЗНЕС ПРОЦЕСИ ЗА МАРКЕТИНГОВИ КАМПАНИИ,
ОТВОРЕН ИНТЕРФЕЙС ЗА ИМПОРТИРАНЕ НА ДАННИ
И ИЗГОТВЯНЕ НА СПРАВКИ И ОТЧЕТИ**

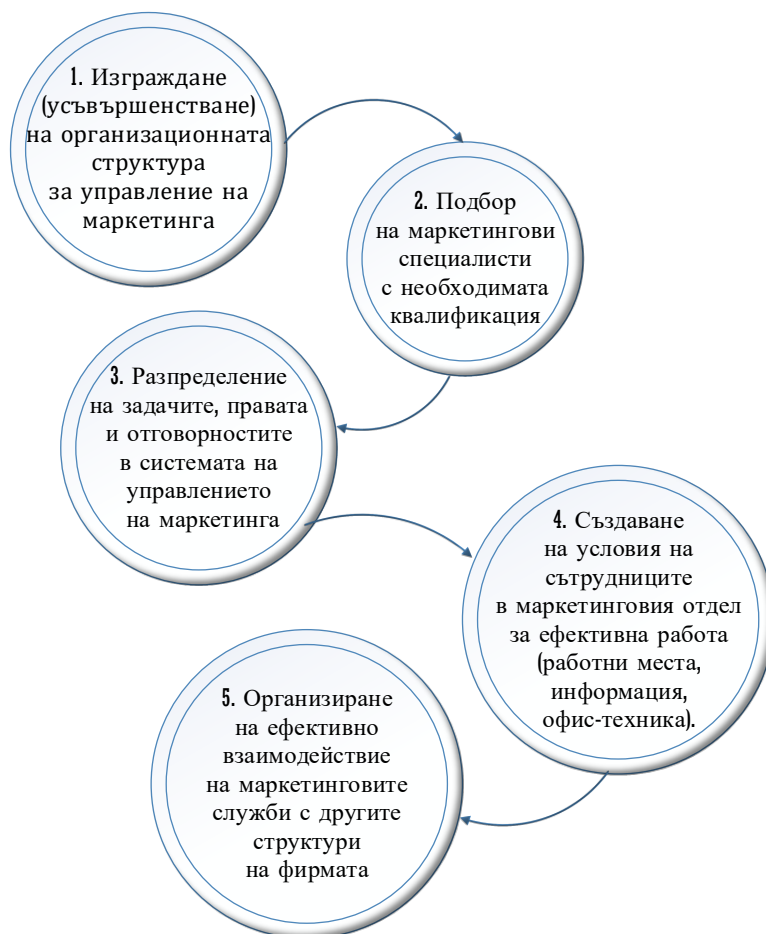
3.1. Цел на упражнението

Да се запознаят студентите със съвкупността от процедурни последователности при проектиране на модели на бизнес процеси относно дейности във връзка с маркетингови кампании, прилагане и опериране с отворен интерфейс за информационен импорт към други системи, както и изготвяне на справки и отчети във връзка с основни бизнес активности.

3.2. Теоретична постановка

Планиране на международния Маркетинг

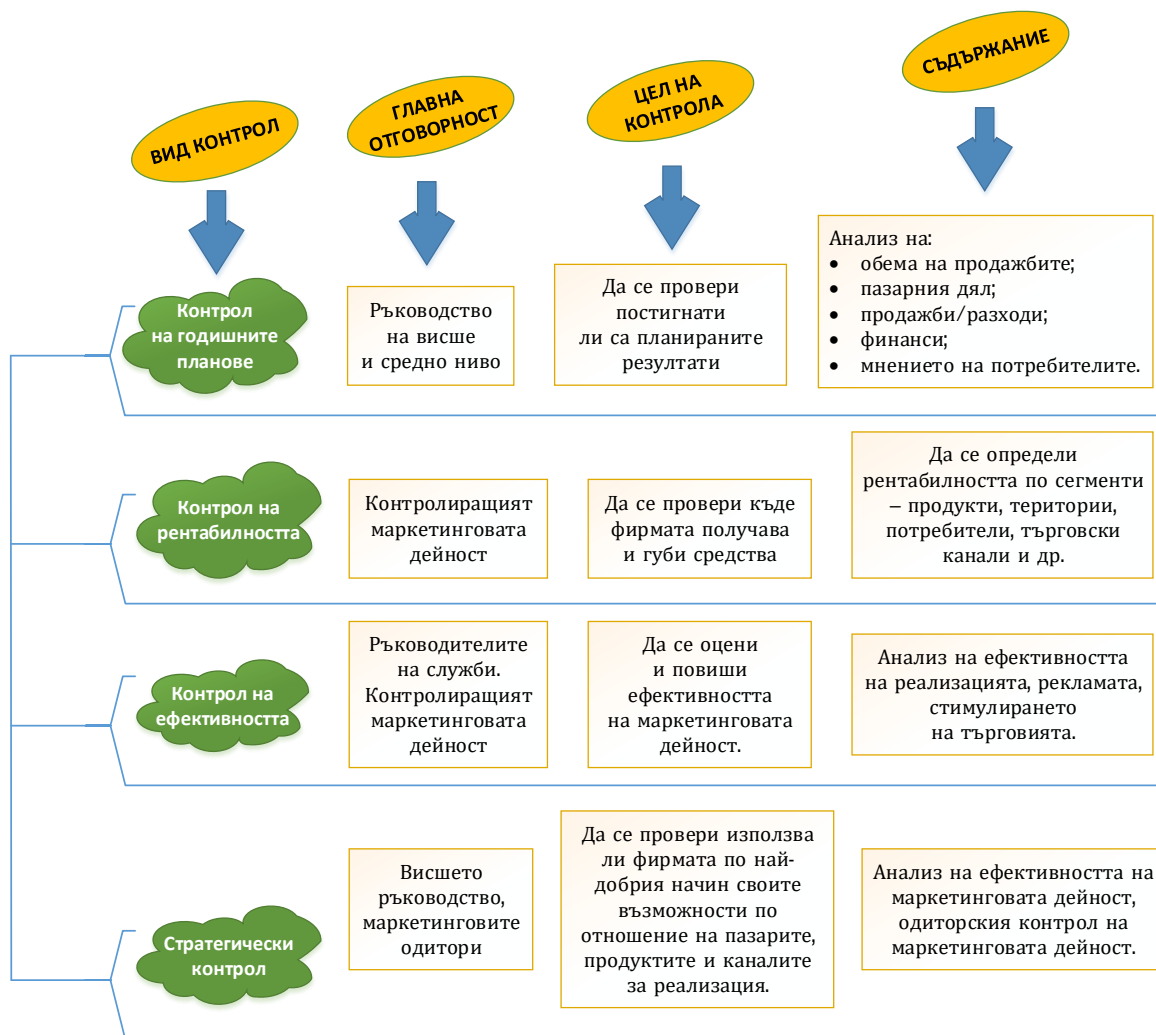
Фирменият мениджър активно трябва да има предвид: целите и методите на планирането; разпределяне на задълженията в областта на планирането във фирмата; структурата и щата на отдела за планиране и изискванията към сътрудниците в отдела за планиране. Фигура 3.1 представя структурата на организиране на международния маркетинг.



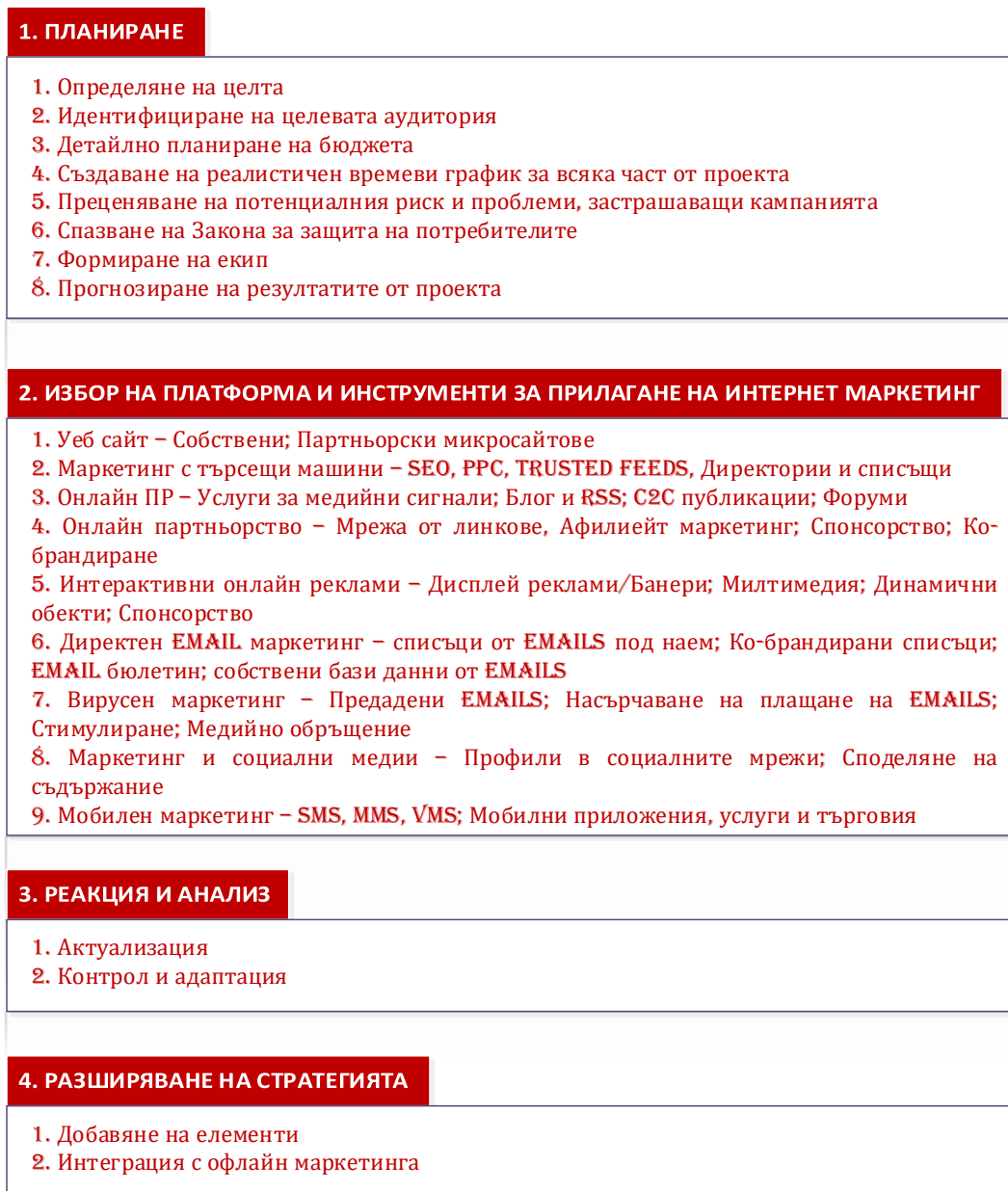
Фиг. 3.1. Организиране на международния маркетинг

Всяка структура за управлението на маркетинга може да се изгражда на основата на един от следните показатели:

- + **Функции** - Функционалната организация на маркетинга е най-проста, обаче нейната ефективност пада с увеличаването на номенклатурата на произвежданите стоки и разширяването на броя на пазарите. Това е обусловено от факта, че във функционалната организация отсъства личността, отговаряща за маркетинга на отделните продукти като цяло, или за дейността на определени пазари;
- + **Географски зони** - Географската организация представлява структура за управление на маркетинга, която се изгражда по отделни региони. Тя позволява на търговците да упражняват своята дейност в пределите на обслужваната територия, добре да познават своите потребители и ефективно да работят с минимални разходи на време и средства за придвижване;
- + **Продукти и потребителски пазари** - Продуктовата организация е структура за управление на маркетинга, в която за реализирането на стратегията и текущите планове на маркетинга, за определен продукт или група продукти, отговаря ръководителят на продукта, при който на подчинение са всички сътрудници, изпълняващи необходимите за дадения продукт маркетингови функции [18, 19].



Фиг. 3.2. Видове маркетингов контрол и основни дейности



Фиг. 3.3. Стратегическо маркетингово планиране в Интернет среда

Контрол в управлението на международния Маркетинг

Контролът на международния маркетинг е процес в основата, на който се извършва измерване и оценяване на резултатите от реализирането на стратегиите и плановете на маркетинга, изпълненията на коригиращите действия, осигуряващи постигането на маркетинговите цели [20].

Осъществяването на контролните функции се базира на определени нормативи, стандарти, в които е отразено очакваното ниво на оценяваните характеристики. Контролът дава възможност на ръководителите на различни равнища на управление да установяват - до каква степен взетите от тях решения в процеса на съставянето и изпълнението на плановете са правилни и възниква ли необходимост от внасянето на корекции. Маркетинговият контрол се разграничава в четири основни групи (фиг. 3.2).

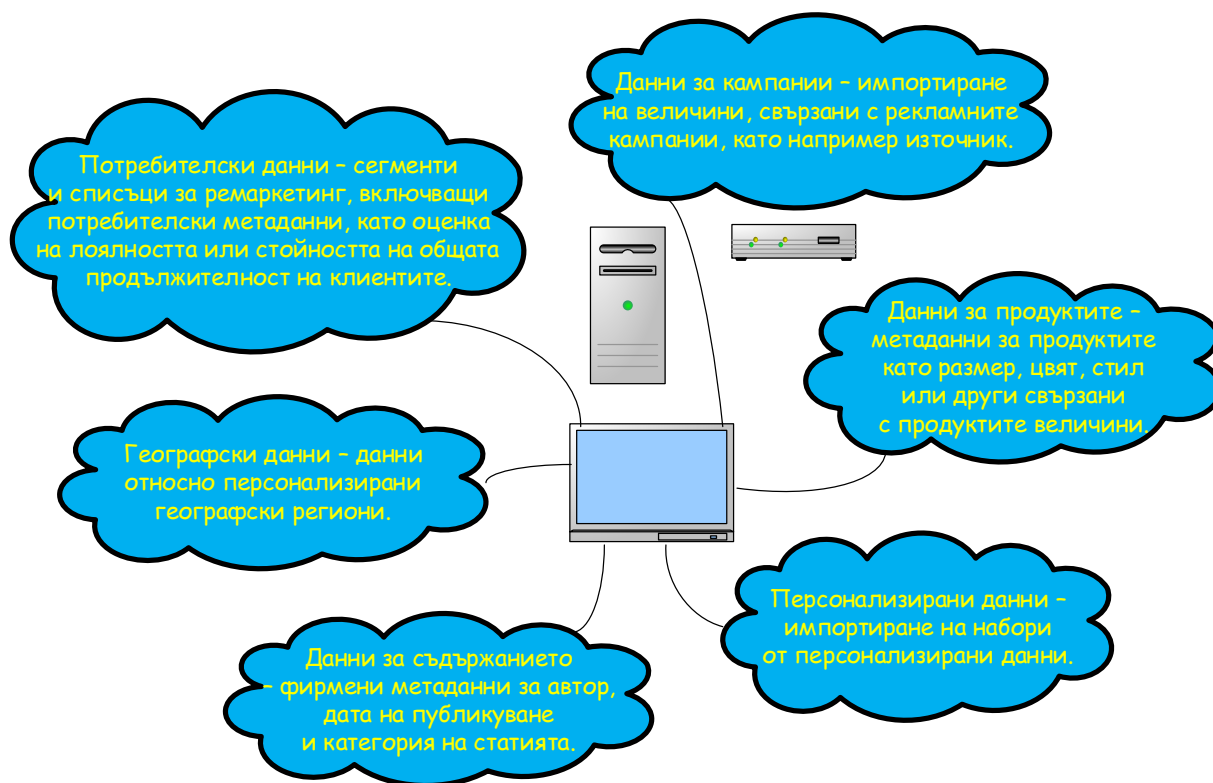
Организация на Интернет маркетингова стратегия

Интернет маркетинговата стратегия отнема значително време за изясняване на концепцията, ефективността и проследяването на онлайн присъствието на дадена организация. Тези непрекъснати процеси изискват бърза мисъл и значителен ресурс.

Основната цел на компаниите, желаещи подобряване на ефективността на онлайн маркетинга, се свързва с адекватно стратегическо планиране, показано на фиг. 3.3. Успешната Интернет маркетингова стратегия не се свързва само с целеполагане, сегментиране и позициониране. Разглежда се като сложен процес на осъществяване на всички компоненти в организацията. В динамична среда с непрекъснати промени, организациите са длъжни да бъдат гъвкави, креативни и адаптивни. Правилното боравене с Интернет инструментите, онагледени по-горе, би допринесло за увеличаване на клиентите и печалбите, а умелото им съчетаване гарантира дългосрочни, стабилни и печеливши резултати за бизнеса.

Бизнес процес „Отворен интерфейс за импорт на данни и връзка с други системи“

Една от съвременните технологични услуги на Google, която подпомага компаниите в управлението на бизнеса, е Google анализ. Google анализ дава възможност за обединяване, обработка и анализиране на данни от хранилища на използвани инструменти за управлението на бизнеса си, с които фирмите разполагат, и тези събирани от Google анализ – например: Данни за управление на връзките с клиенти, тези от електронна търговия и от *Google Анализ*.

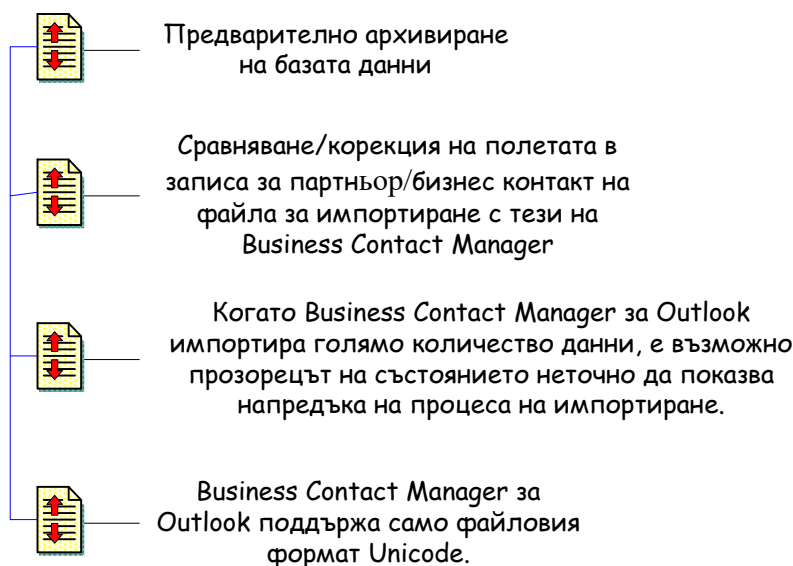


Фиг. 3.4. Типове разширени данни

Потребителите могат да качват и интегрират директно информация с профила си в Google Анализ на три различни етапа по пътя на събирането и обработването им, съответно:

- ✚ Импортиране на данни за посещенията – осигурява се алтернатива на използването на проследяващ код, приложен програмен интерфейс (API) за събиране, комплекти за разработване на софтуер (SDK) за мобилни приложения или Measurement Protocol. Такива данни могат да бъдат Данни за възстановяване на средства (синхронизиране на вътрешното отчитане на електронната търговия с Google анализ);
- ✚ Импортиране на разширени данни, онагледени на фиг. 3.4:
- ✚ Импортиране на обобщени данни – възможност за обединяване на качените показатели. Импортираните обобщени данни се прилагат към избраните изгледи на собствено за отчитане, след като завършат всички процеси по обработване и обобщаване на събраните данни. Поддържан тип данни за качване - Данни за разходите (данни за рекламни мрежи на трети страни (извън Google), като кликания, разходи и импресии с цел получаване на по-пълна представа на разходите за реклама) [21, 22].

Business Contact Manager за Outlook способства импортирането на бизнес записи, съхранявани в други програми като Excel или фирмената папка с контакти на Outlook. Преди процесите по импортиране на данни трябва да бъдат съобразени определени особености, посочени на фиг. 3.5.



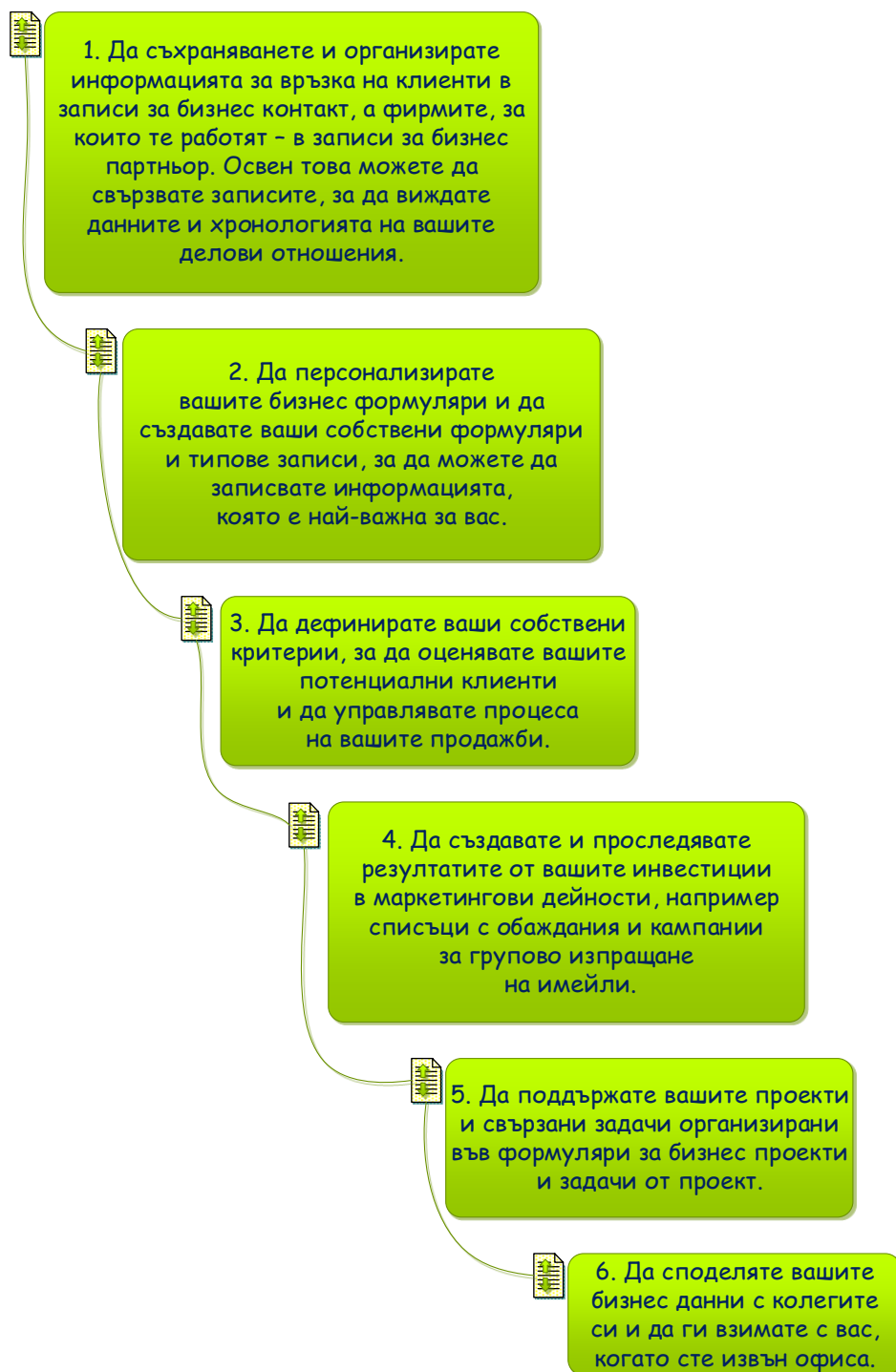
Фиг. 3.5. *Особености при импортиране на бизнес данни в Business Contact Manager за Outlook*

Клиентските типове файлове, които могат да бъдат импортирани към услугата Business Contact Manager за Outlook, са дадени в таблица 3.1.

Таблица 3.1. Импортирани типове файлове в Business Contact Manager за Outlook

Тип на файла	Описание
Данни на Business Contact Manager за Outlook (.bcm)	Всички данни в Business Contact Manager за Outlook, включително партньори, бизнес контакти, възможности, бизнес проекти, задачи от проекти и тяхната хронология на комуникация. Това включва също персонализации, маркетингови дейности и списъци на продуктите и услугите.
Персонализации на Business Contact Manager (.bcmx)	Включва оформлени на формуляри по избор, страници и изгледи, отчети по избор и шаблони за проект. <u>Забележка:</u> Само собственикът на базата данни може да импортира персонализации.
Стойности, разделени със запетая (.csv). Списъците на продуктите и услугите обикновено са в .csv файлов формат.	Включва данни във файловия формат, съдържащ стойности, разделени със запетая (.csv). Данните от почти всяка програма могат да бъдат записани във файловия формат .csv, така че дори ако файловият формат за дадено приложение не е в списъка, може да се запишат данните от приложението във файловия формат .csv и след това данните да бъдат импортирани в Business Contact Manager за Outlook.
Работна книга на Excel (.xls, .xlsx)	Включва файлове с данни във формат на работна книга на Microsoft Excel.
База данни на Access (.mdb, .accdb)	Включва данни във файловите формати на Microsoft Access.
Папка с контакти на Outlook (.pst)	Включва контакти във файловия формат на Microsoft Outlook (.pst). Ако вашата база данни за Business Contact Manager е споделена, всеки потребител може да иска да импортира своите контакти на Outlook.
ACT! (.dbf, .pad)	Отнася се за данни, които са записани във файловия формат на ACT!. Business Contact Manager за Outlook включва инструмент за конвертиране на данни, който позволява да се конвертират данните на ACT! във формата .bcm, преди да ги импортирате. <u>Забележки:</u> Ако са избрани файловите формати на ACT! (.dbf, .pad) или QuickBooks (. QIF), се извършва избор на „Преобразуване на данни“ за конвертиране на данните във файлов формат на Business Contact Manager (.bcm). Този инструмент помага за импортиране на ъказаните типове файлове. Приложението ACT! трябва да бъде инсталирано на съответен потребителски компютър, за да може да се извърши импортиране на данни, които са във файлов формат на ACT!.
QuickBooks (. QIF)	Може да се импортира всяка версия на файл на QuickBooks в .qif формат. Business Contact Manager за Outlook включва инструмент за конвертиране на данни. Това позволява процедурата по конвертиране на данните на QuickBooks във формата .bcm, преди да бъдат импортирани.
Данни на Business Contact Manager за Outlook (.bcm)	Всички данни в Business Contact Manager за Outlook, включително партньори, бизнес контакти, възможности, бизнес проекти, задачи от проекти и тяхната хронология на комуникация. Това включва също персонализации, маркетингови дейности и списъци на продуктите и услугите.

По-новите версии на отворения интерфейс притежават набор от възможности за потребителите, описани на фиг. 3.6 [23, 24].



Фиг. 3.6. Възможности на Business Contact Manager Outlook

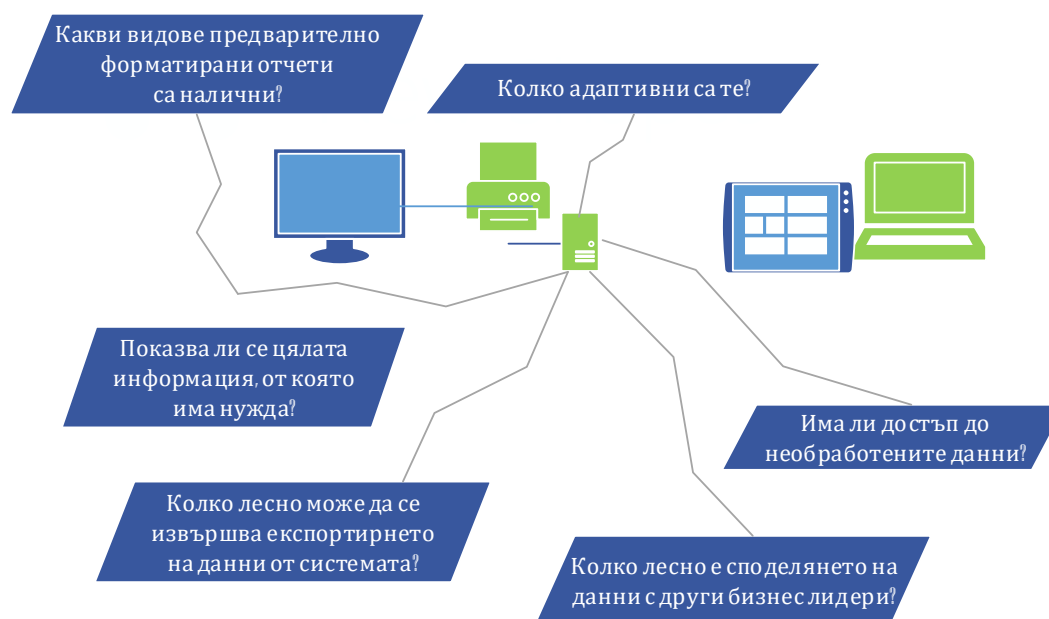
Бизнес процес „Справки и отчети, свързани с извършени бизнес активности“

Изготвянето на отчети и прогнозни анализи е ефективно решение при управлението на бизнес процеси, способстващо за:

- ✚ тяхното подобряване и повишаване на производителността;
- ✚ установяването на тенденциите в движението на пазара;

- ✚ идентифицирането на очакванията на клиентите;
- ✚ увеличаване на общия дял на продажбите.

Необходимо е данните да бъдат обобщавани и представяни в подходящи формати, които да улесняват мениджърите при тяхно манипулиране. Във връзка с оценяване на системата за управление на бизнес процеси по отношение на отчетността трябва да се даде отговор на определени въпроси, схематично дефинирани на фиг. 3.7.



Фиг. 3.7. Оценяване на системата за управление на бизнес процеси относно отчетността

Съществуват различни типове маркетингови отчети като независимо дали са предназначени за маркетинговия отдел, СЕО или клиенти, съдържат:

- ✚ Заглавие;
- ✚ Отчетен период;
- ✚ Обобщение.

Относно маркетинговото отчитане от съществено значение е:

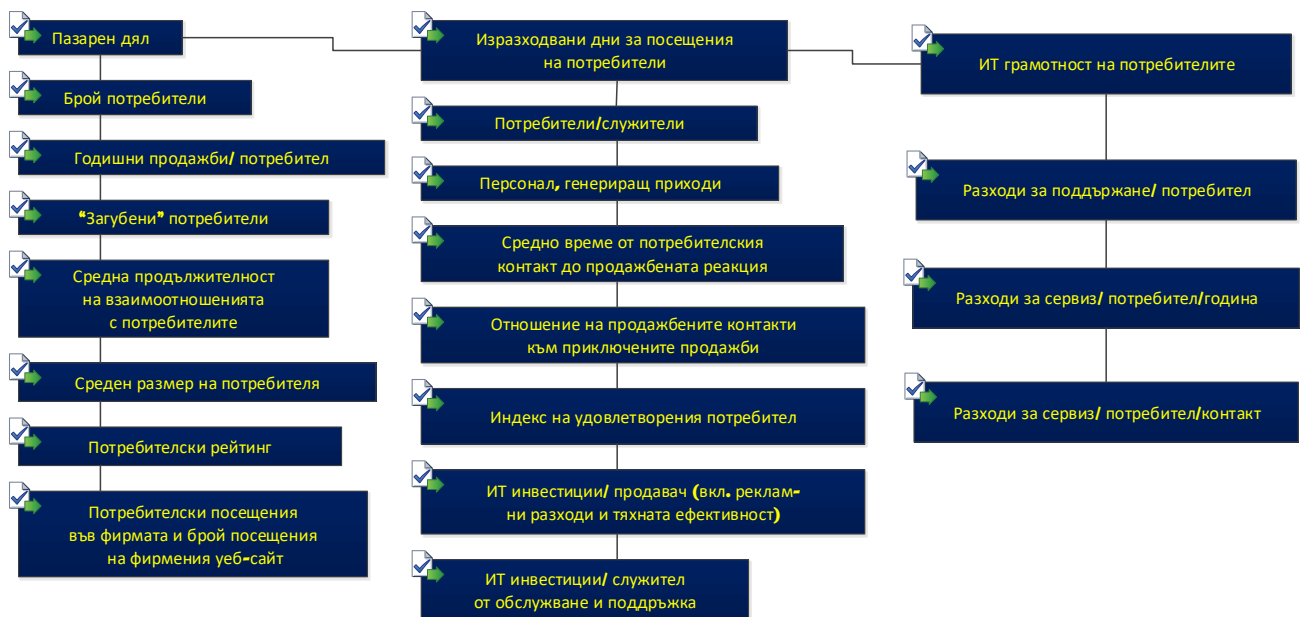
- ✚ Данните да бъдат графично изобразявани, така че да бъде възможна едновременната направа на множествени заключения или вземането на множество от решения;
- ✚ Наличието на ясно дефинирана цел, описания и допълнителни пояснения в записката на докладите/отчетите;
- ✚ Спазването на етикет и терминологичен език при поднасянето на информацията съобразно аудиторията.

При изготвянето на отчети се препоръчва съблюдаването на утвърдени добри практики, онагледени на фиг. 3.8.

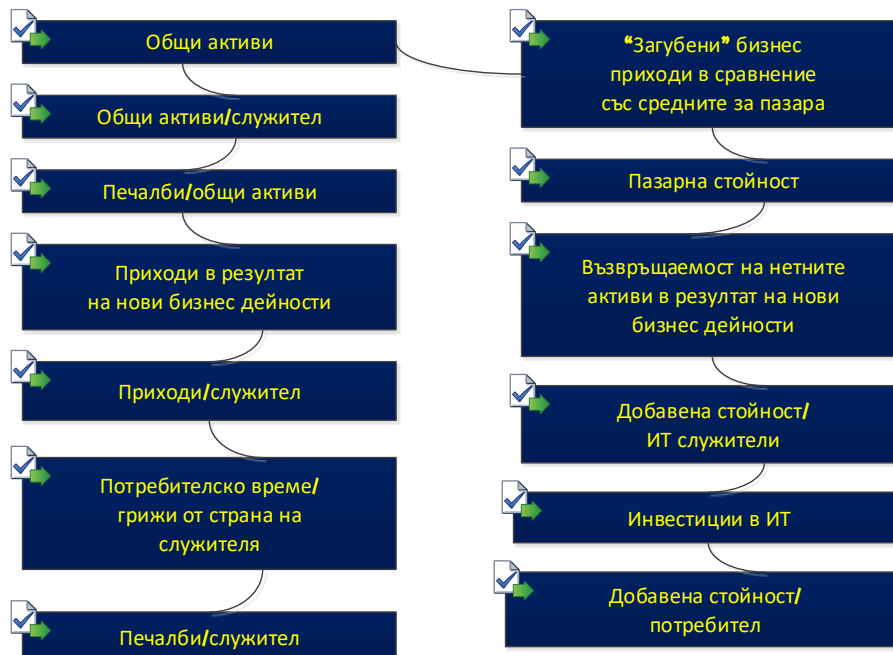


Фиг. 3.8. Добри практики, подпомагащи изготвянето на маркетингови отчети

Процесите на проучване, систематизиране, селектиране, обработка, анализ и оценка на данни при изготвянето на справки и отчети засягат няколко групи от фактори, указани на фигура от 3.9 до фиг. 3.14 [25-27].



Фиг. 3.9. Група фактори „А“ – „Потребители и персонал“



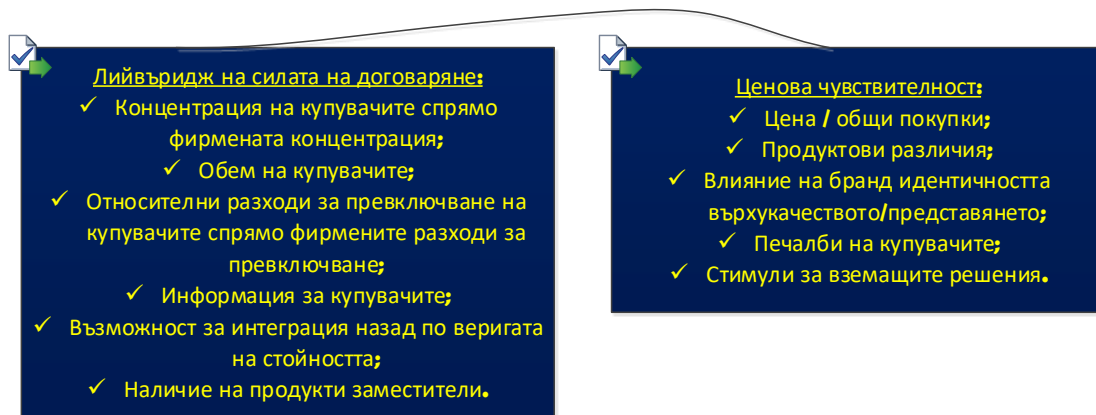
Фиг. 3.10. Група фактори „Б“ – „Финанси“



Фиг. 3.11. Група фактори „В“ – „Сила на доставчиците“



Фиг. 3.12. Група фактори „Г“ – „Интензивност на конкуренция“



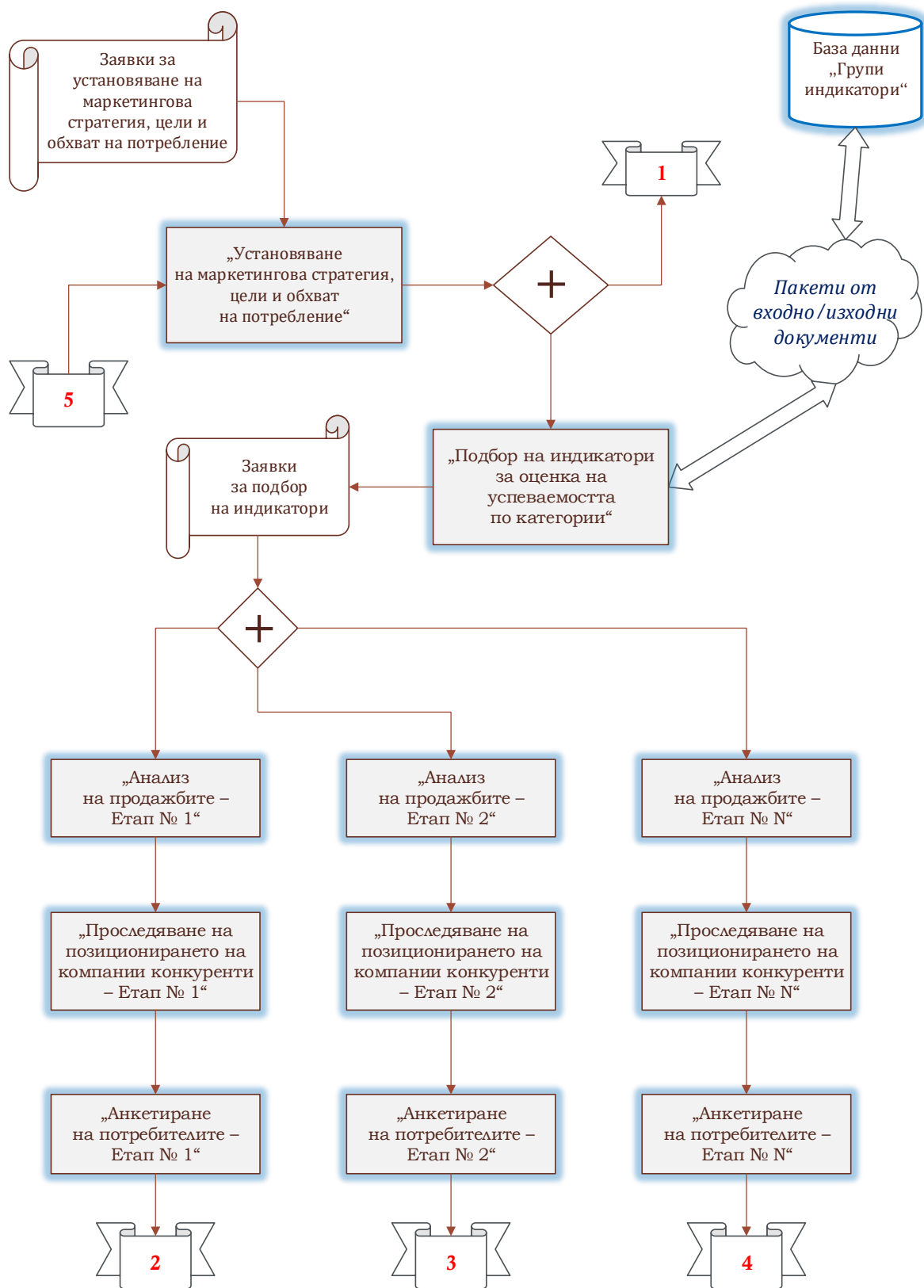
Фиг. 3.13. Група фактори „Д“ – „Сила на купувачите“



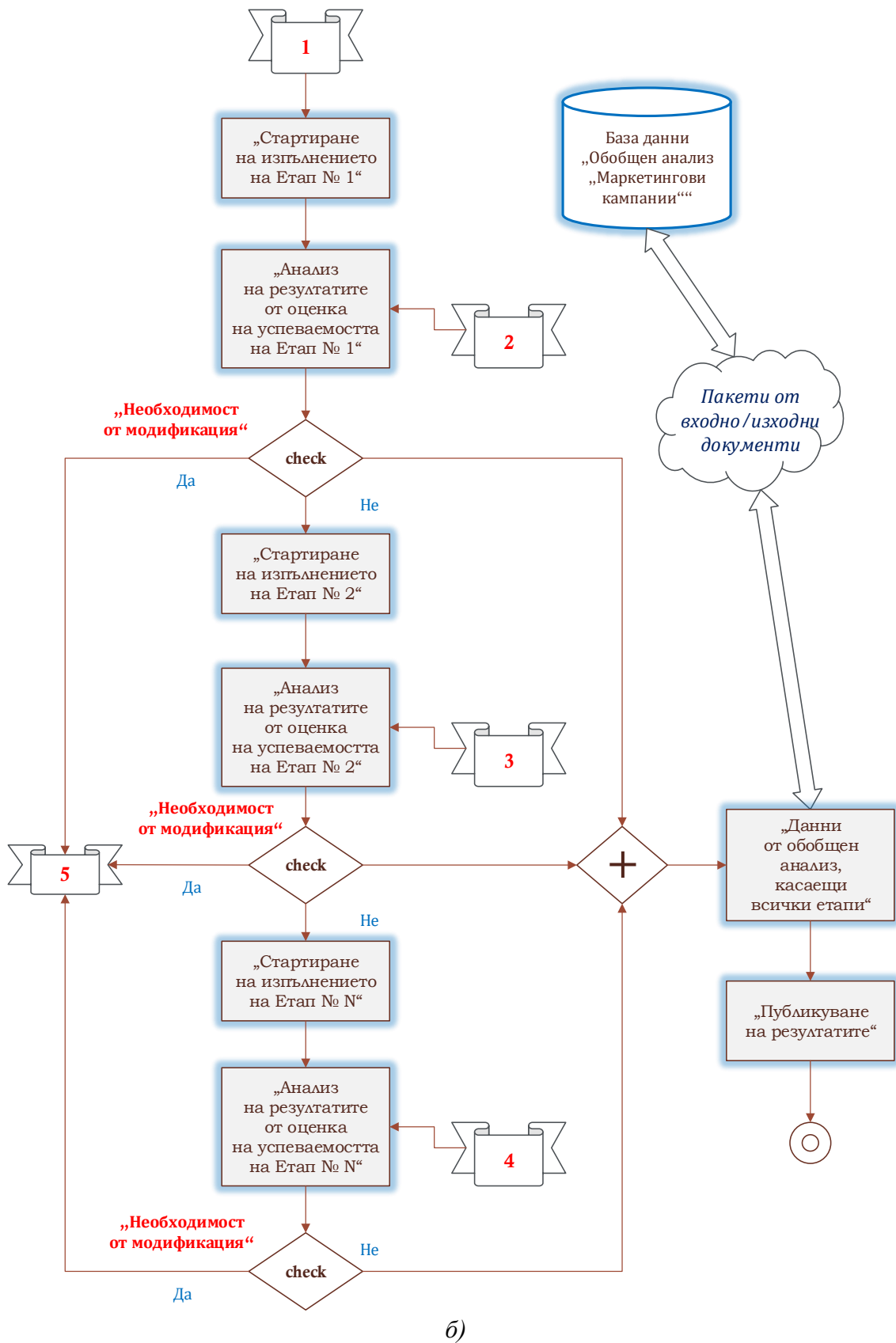
Фиг. 3.14. Група фактори „Е“ – „Заместители“

3.3. Задачи за изпълнение

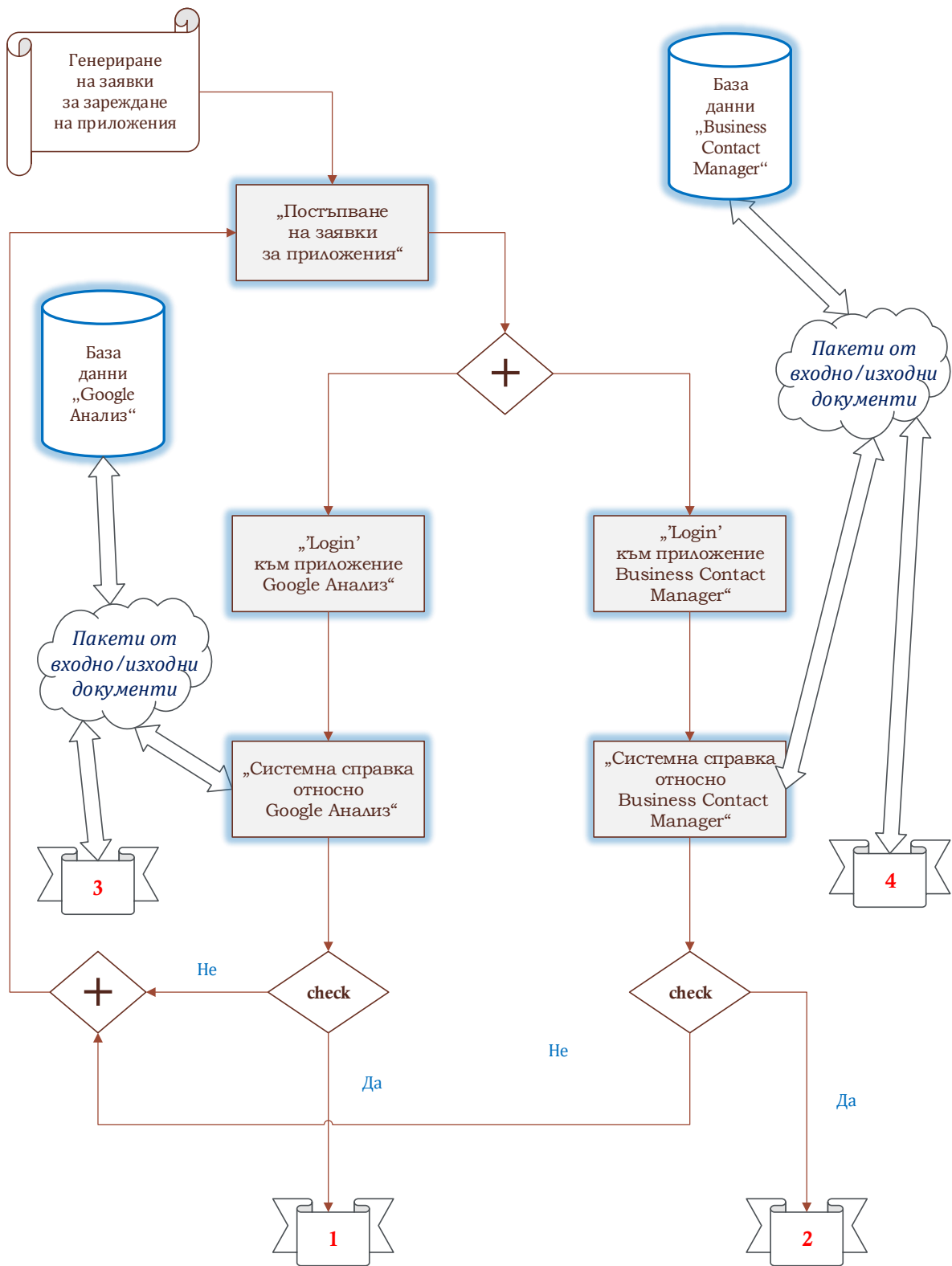
- ❖ Да се проектира модел спрямо концептуалната процедурна информационна диаграма на фиг. 3.15 по отношение на бизнес процеса „Организиране и отчитане на маркетингови кампании по изделия, региони и клиенти“ в графичната среда [Bizagi Modeler](#);
- ❖ Да се моделира бизнес функцията „Отворен интерфейс за импорт на данни и връзка с други системи“ според зададената функционална диаграма на фиг. 3.16 чрез платформата [Bizagi Modeler](#);
- ❖ Да се създаде модел, използвайки наличните инструменти на продукта [Bizagi Modeler](#), във връзка с бизнес функция „Справки и отчети, свързани с извършени бизнес активности“ съобразно функционалната диаграма на фиг. 3.17;
- ❖ Да се извърши моделиране на разглежданите бизнес процеси посредством програмен инструмент [PowerDesigner](#), анализ и оценяване на наличните ресурси с тези на средата [Bizagi Modeler](#) относно целите на изследването;
- ❖ Да се анализира същността и особеностите на заложените примерни концептуални диаграми за посочените бизнес функции и оптимизира тяхната процедурна структура във връзка с допълнителни персонални потребителски роли и бази данни за информационен обмен чрез входно-изходни документи.



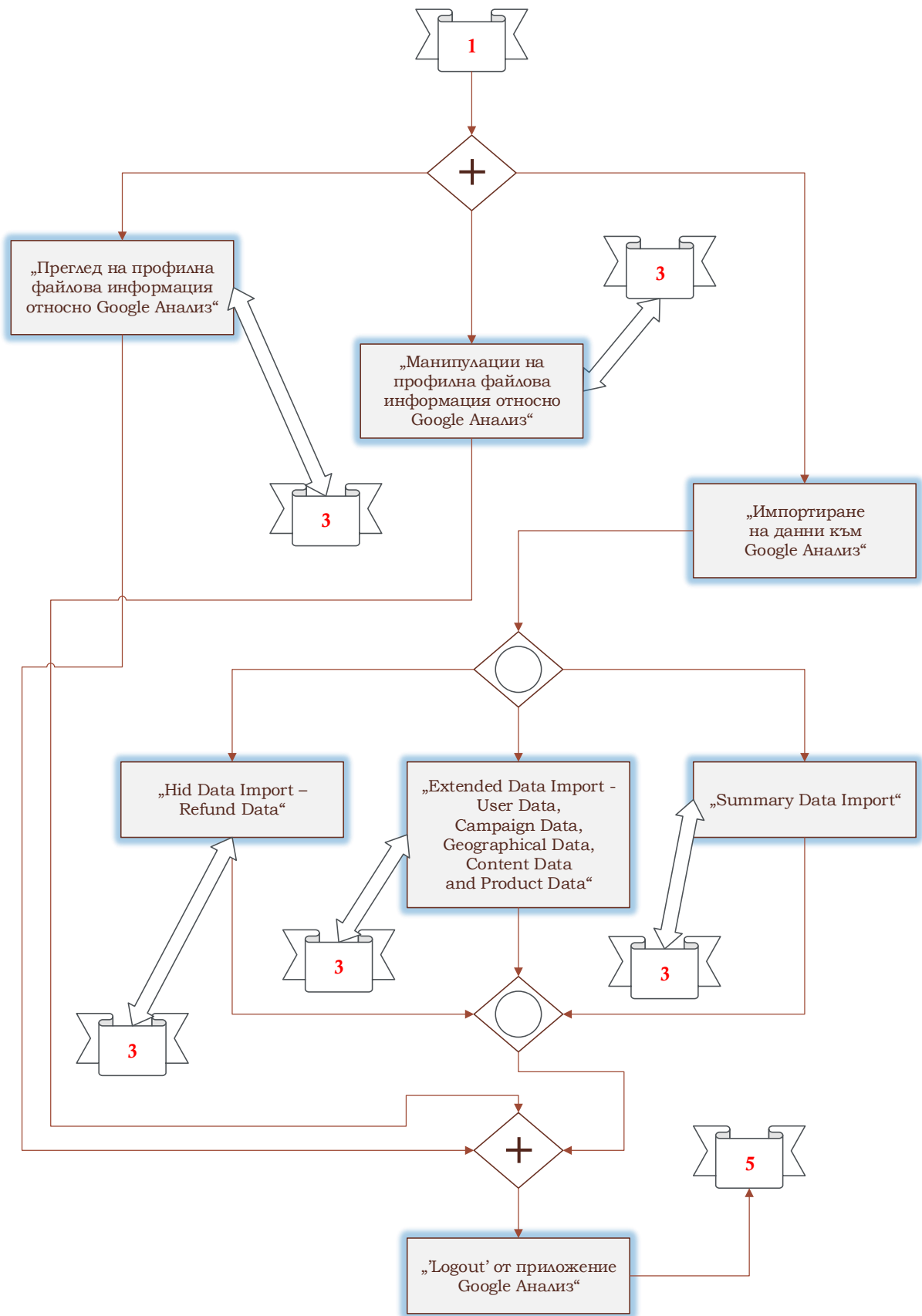
a)



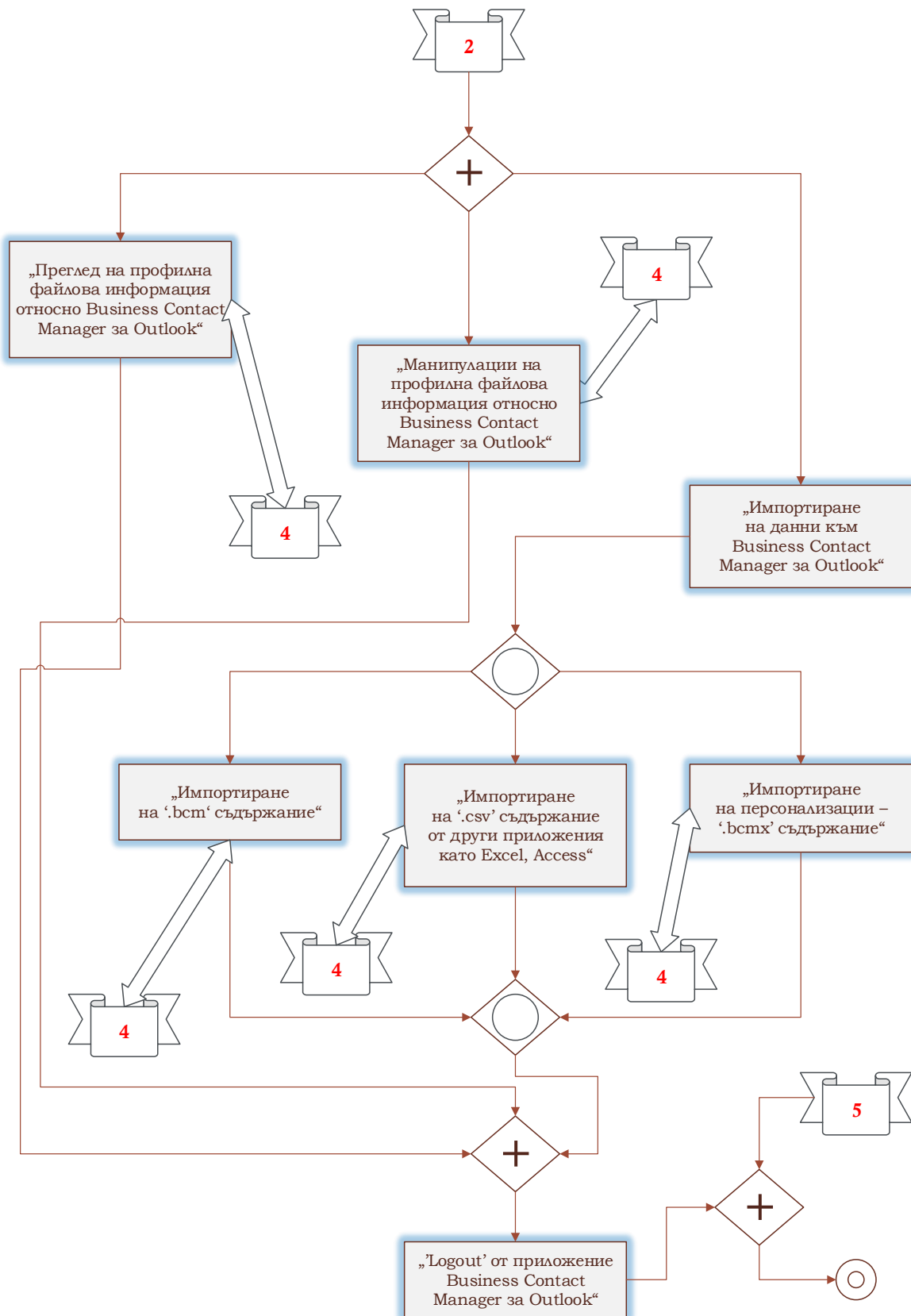
Фиг. 3.15. Адаптирана информационна диаграма на бизнес процеса „Организиране и отчитане на маркетингови кампании по изделия, региони и клиенти“ – модул №1 а) и модул №2 б)



a)

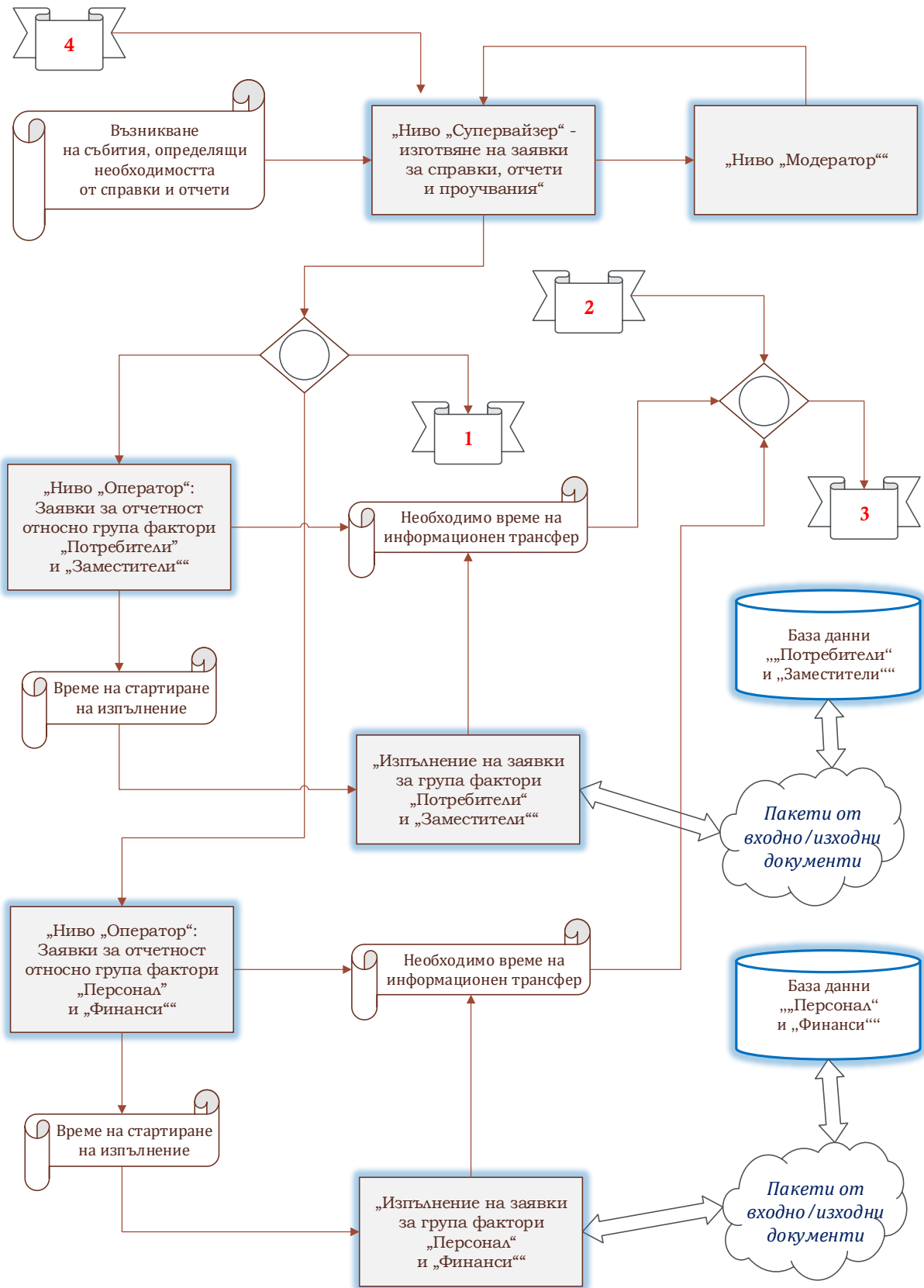


б)

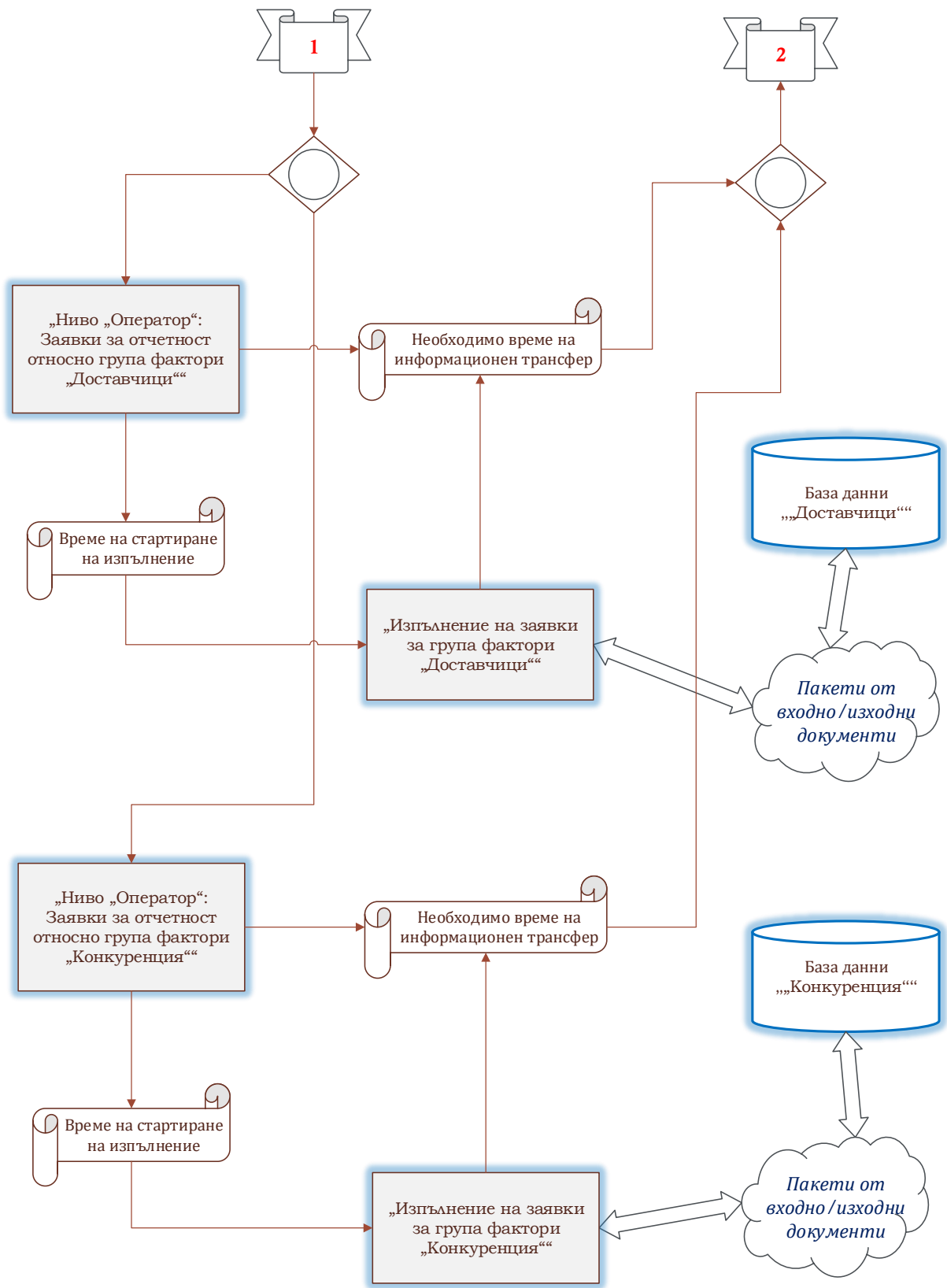


в)

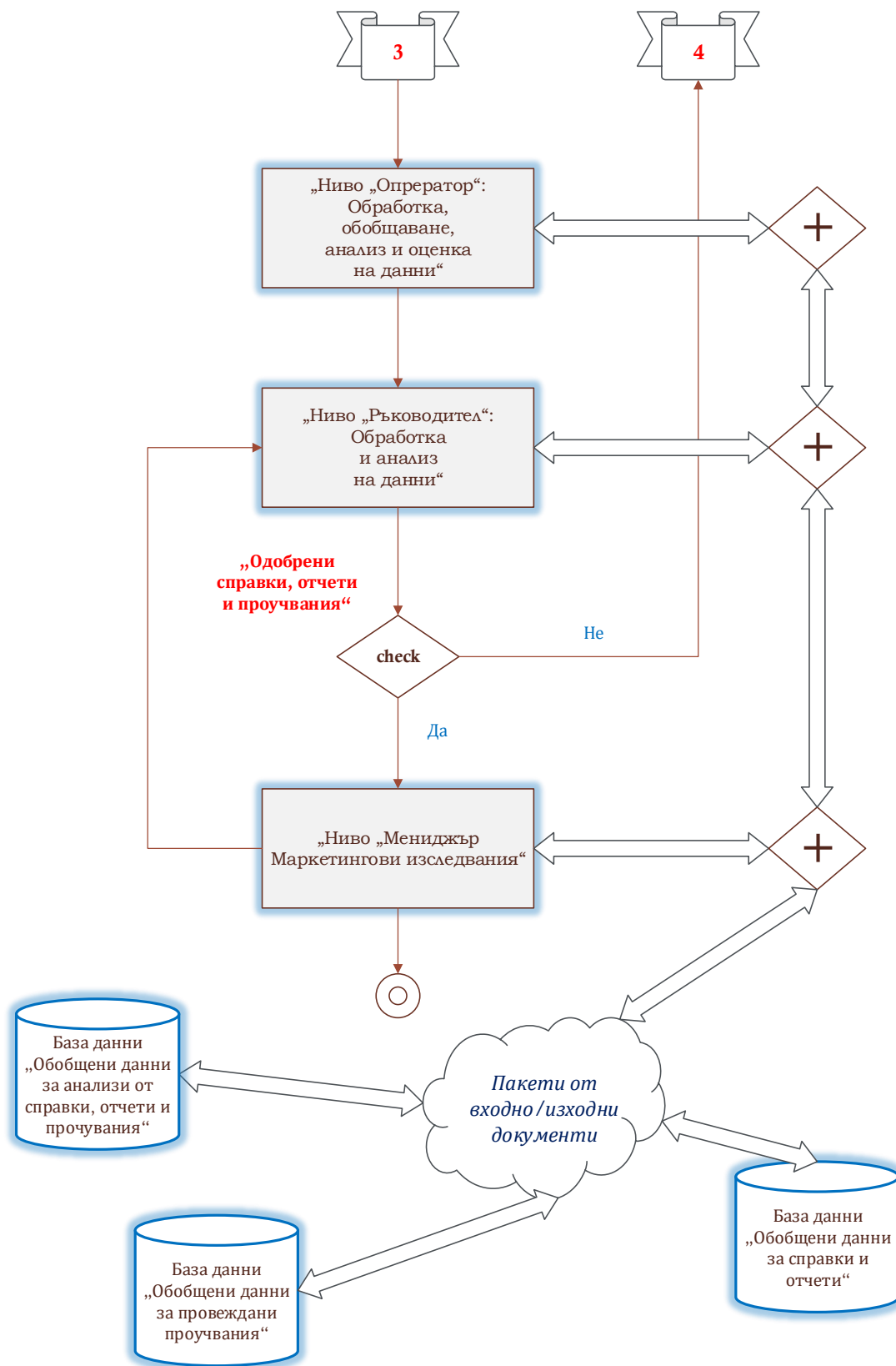
Фиг. 3.16. Адаптирана информационна диаграма на бизнес процеса „Отворен интерфейс за импорт на данни и връзка с други системи“ – модул №1 а), модул №2 б) и модул №3 в)



a)



б)



в)

Фиг. 3.17. Адаптирана информационна диаграма на бизнес процеса „Справки и отчети, свързани с извършени бизнес активности“ – модул №1 а), модул №2 б) и модул №3 в)

3.4. Контролни въпроси

1. Какви видове маркетингов контрол и съпътстващи основни дейности, които са обвързани с тях, познавате?
2. През какви етапи преминава стратегическото маркетинг планиране на основата на глобалната Интернет мрежа?
3. Какъв е типът на файловия формат, който се поддържа от Business Contact Manager за Outlook, при импортиране и манипулиране с категоризирани бизнес данни?
4. Какви типове разширени данни потребителите могат директно да интегрират при опериране с платформата Google Анализ?
5. Кои добри практики и подходи се прилагат при обезпечаване на дейностите по изготвяне на маркетингови отчети?
6. Какви групи фактори се обхващат при провеждане на проучване, систематизиране, селектиране, обработка, анализ и оценка на данни във връзка с изготвянето на справки и отчети с определени бизнес активности?

МОДУЛ II

КЛЪСТЕРЕН АНАЛИЗ, ПОДХОДИ ЗА РАЗПОЗНАВАНЕ И КЛАСИФИКАЦИЯ ВЪВ ВРЪЗКА С ДАННИ ПО ИКОНОМИЧЕСКИ РЕГИОНИ

Лабораторно упражнение №4

КЛЪСТЕРИЗАЦИЯ НА ИКОНОМИЧЕСКИ ДАННИ ПОСРЕДСТВОМ K-MEANS МЕТОД В СРЕДА МАТЛАВ

4.1. Цел на упражнението

Да се запознаят студентите с техники за формиране на информационни клъстери при обработка и изследване на взаимовръзките при различни категории от данни на основата на алгоритъм k-means.

4.2. Икономически обект на изследване

Изследванията в това и следващите лабораторни упражнения от модул II главно са насочени към възможността да бъде анализирана принадлежността на група показатели относно икономически приходи, разходи и процедури по превишен дефицит спрямо страната, за която са регистрирани, според доклад на Европейската комисия от пролетта на 2020 г. чрез синтезирани класификационни модели. Изследването се базира за отчетени годишни показатели за периода от 1995 г. до 2021 г за 10 страни – Belgium, Germany, Ireland, Greece, Spain, France, Italy, Latvia, Luxembourg и Matla.

При създаването и оценката на модели за целева идентификация с помощта на:

- ✦ Изкуствен интелект;
- ✦ Методи и алгоритми за машинно обучение,
- ✦

в хода на изследванията е предвидено да бъдат използвани различни конфигурации от набори от следните финансови показатели:

- ✦ Taxes on production and imports;
- ✦ Current taxes on income and wealth;
- ✦ Other current revenue, including sales;
- ✦ Total current revenue;
- ✦ Collective consumption;
- ✦ Intermediate consumption;
- ✦ Capital transfers received.

4.3. Теоретична постановка

Целта на клъстерния анализ се състои в обособяване на информационни групи, наречени клъстери, по отношение на n броя тестови наблюдения, формирани от стойности m признака, по намерени техни общи характеристики. В конкретния случай се прилага клъстерен анализ по метода k-means, където предварително се залагат

количеството кълстери. Методът може да бъде приложен за различни метрики между обектите (наблюденията) от еднородните информационни множества, представени в таблица 4.1 [28-32].

Таблица 4.1. Типове метрики при реализация на *k-means* кълстерен анализ

Тип	Описание
squclidean	Squared Euclidean distance (default). Each centroid is the mean of the points in that cluster.
cityblock	Sum of absolute difference. Each centroid is the component-wise median of the points in that cluster.
cosine	One minus the cosine of the included angle between points (treated as vectors). Each centroid is the mean of the points in that cluster, after normalizing those points to unit Euclidean length.
correlation	One minus the sample correlation between points (treated as sequences of values). Each centroid is the component-wise mean of the points in that cluster, after centering and normalizing those points to zero mean and unit standard deviation.
hamming	This metric is only suitable for binary data. It is the proportion of bits that differ. Each centroid is the component-wise median of points in that cluster.

В процеса на изследването с помощта на кълстерения анализ на данни е използван входящ набор от 108 еталона (наблюдения), включващи стойности на икономически индикатори:

- ✚ Taxes on production and imports;
- ✚ Current taxes on income and wealth,

за шест икономически региони Belgium, Germany, Ireland, Greece, Spain и France.

При изпълнение на анализа следва да бъдат определени:

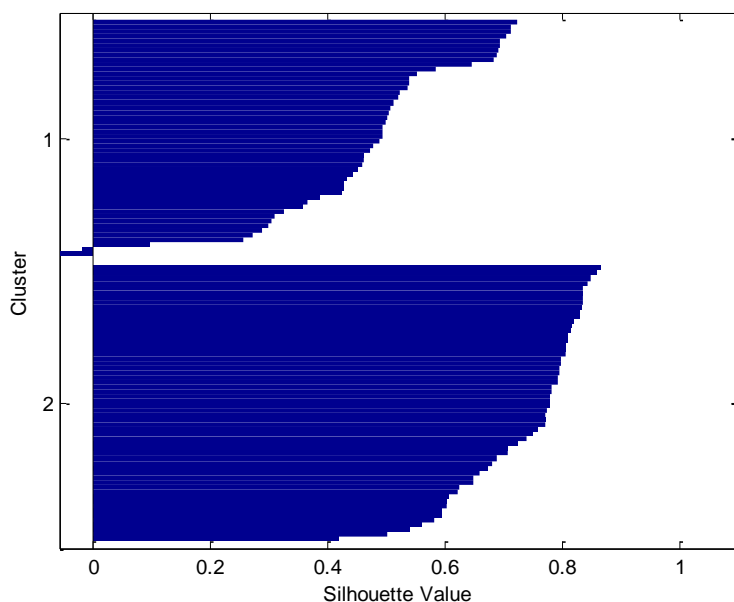
- ✚ оптималеният брой кълстери относно обработваните данни;
- ✚ най-подходящият тип метрично разстояние между центроидите и наблюденията в областите на кълстерите,

които се обособяват в околността им. Във връзка с това следва да бъде извършена оценка на профила на кълстерите като за целта се използва критерия *Silhouette Index (SI)*, чиято най-добра теоретична стойност е „1“. Тук се прави калкулация на средната стойност на коефициентите SI, получени по отношение на всички наблюдения от състава на зададените кълстерни зони.

Процедурата следва да бъде приложена при последователно увеличаване на броя на кълстерите като най-високата средна стойност на индекса определя оптималното количество кълстери. Действията могат да бъдат приложени спрямо различни метрики, за да се установи и най-подходящият тип разстояние.

4.4. Задачи за изпълнение

- ❖ Да се приложат процедури по k-means клъстерен анализ на данни като се използва базисен тип метрика “sqeuclidean” при формат за последователно обособяване на две, три, четири, пет и шест клъстерни зони. По отношение на всеки разглеждан случай да бъдат генерирани „профилна диаграма“ и “средна стойност на критерия SI” по начина, който е показан на фигура от 4.1 до фиг. 4.5;
- ❖ Да се анализират представените резултати като бъде установено наличието или отсъствието на закономерности, както и определи потенциалният най-добър брой клъстери, към които могат да бъдат отнесени опитните данни, на база на получените индексни стойности;



а)

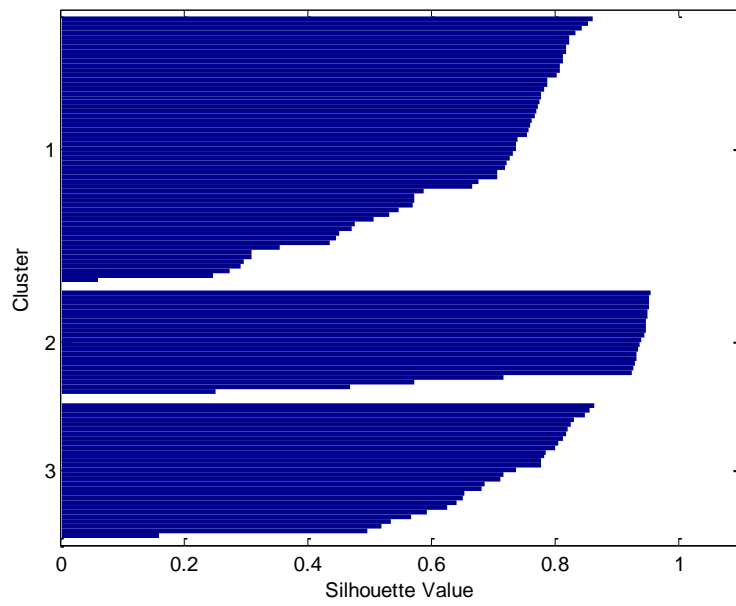
```
[silh2,h] = silhouette(economicdataset,idx2,'sqeuclidean');
cluster2 = mean(silh2)

cluster2 =

    0.6119
```

б)

Фиг. 4.1. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в два клъстера при метрично разстояние “sqeuclidean”



a)

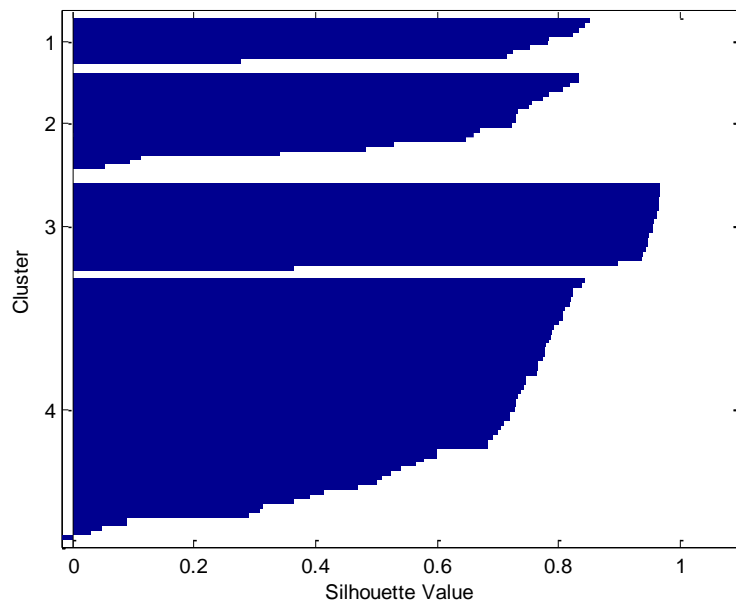
```
[silh3,h] = silhouette(economicdataset,idx3,'sqeuclidean');
cluster3 = mean(silh3)

cluster3 =

    0.7053
```

б)

Фиг. 4.2. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в три клъстера при метрично разстояние “sqeuclidean”



a)

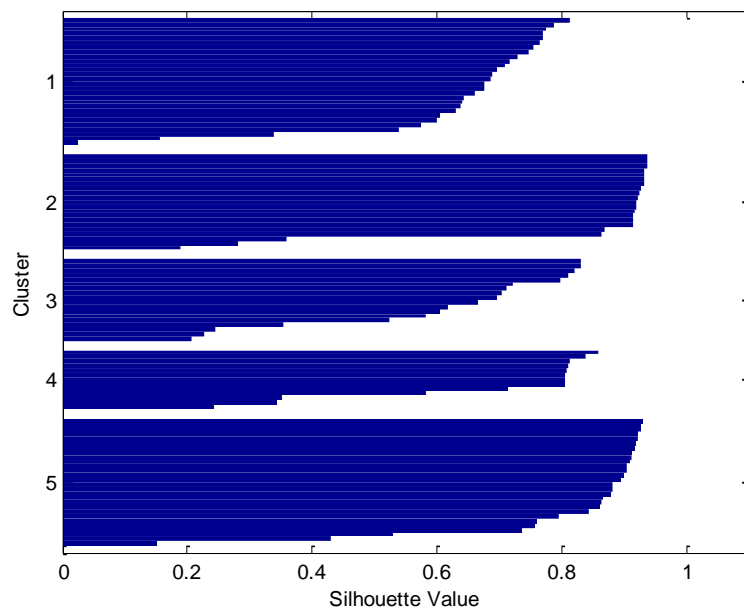
```
[silh4,h] = silhouette(economicdataset,idx4,'sqeuclidean');
cluster4 = mean(silh4)

cluster4 =

    0.6777
```

б)

Фиг. 4.3. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в четири клъстера при метрично разстояние “sqeuclidean”



а)

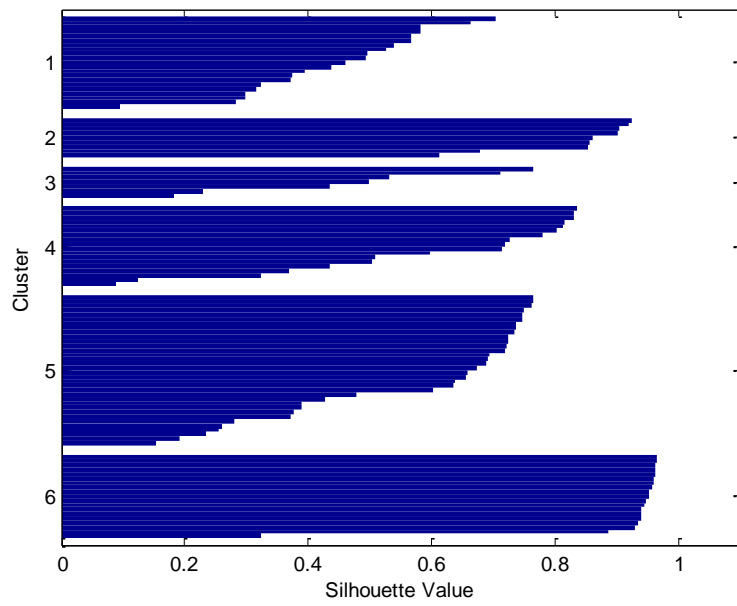
```
[silh5,h] = silhouette(economicdataset,idx5,'sqeuclidean');
cluster5 = mean(silh5)

cluster5 =

    0.7212
```

б)

Фиг. 4.4. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в пет клъстера при метрично разстояние “sqeuclidean”



a)

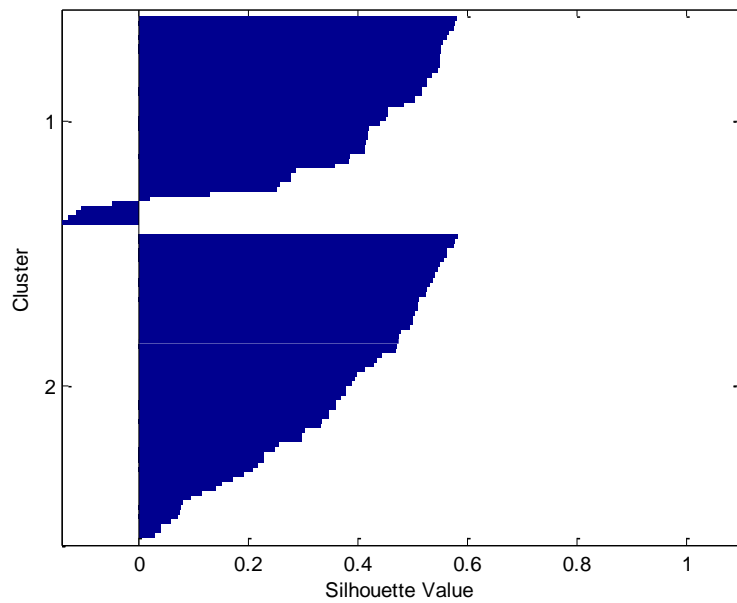
```
[silh6,h] = silhouette(economicdataset,idx6,'sqeuclidean');
cluster6 = mean(silh6)

cluster6 =

    0.6275
```

б)

Фиг. 4.5. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в шест клъстера при метрично разстояние “sqeuclidean”



a)

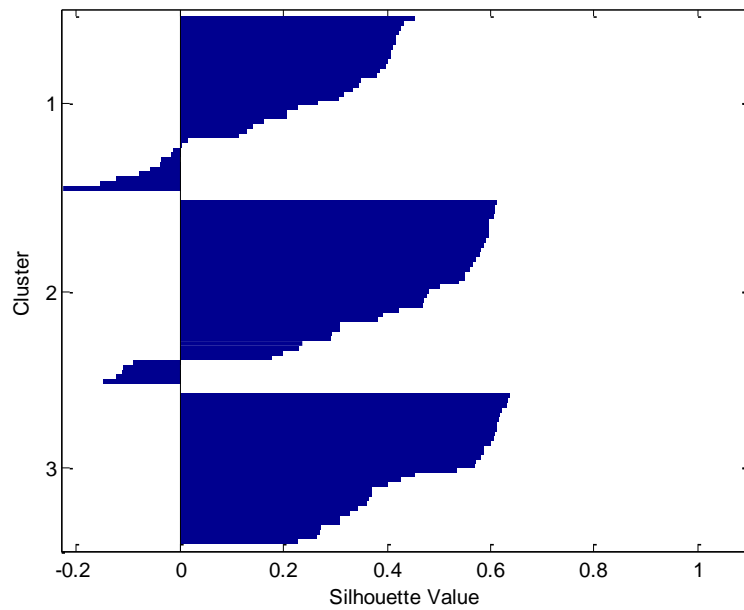
```
[silh2,h] = silhouette(economicdataset,idx2,'cityblock');
cluster2 = mean(silh2)

cluster2 =

    0.3650
```

б)

Фиг. 4.6. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в два клъстера при метрично разстояние “cityblock”



а)

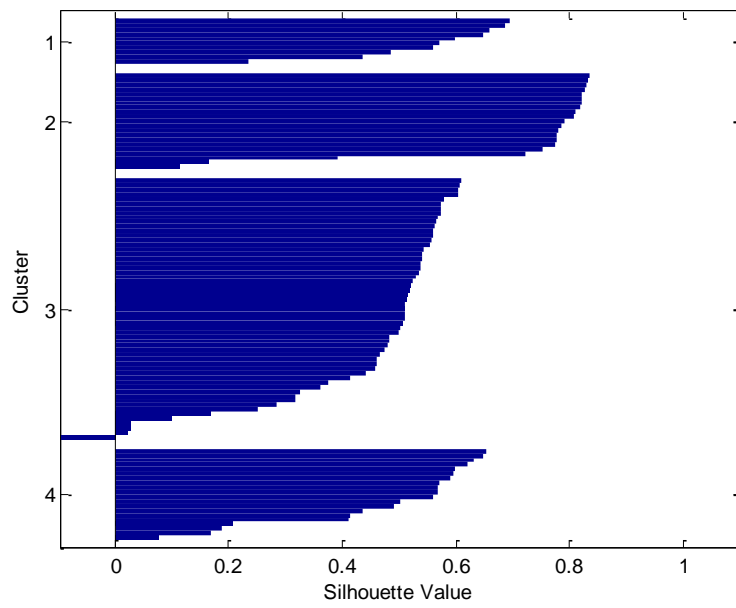
```
[silh3,h] = silhouette(economicdataset,idx3,'cityblock');
cluster3 = mean(silh3)

cluster3 =

    0.3573
```

б)

Фиг. 4.7. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в три клъстера при метрично разстояние “cityblock”



a)

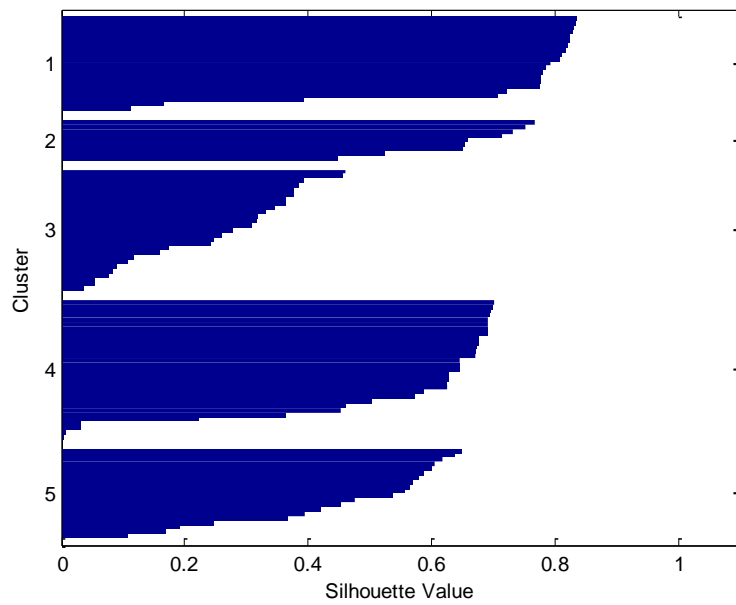
```
[silh4,h] = silhouette(economicdataset,idx4,'cityblock');
cluster2 = mean(silh4)

cluster2 =

    0.5161
```

б)

Фиг. 4.8. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в четири клъстера при метрично разстояние “cityblock”



a)

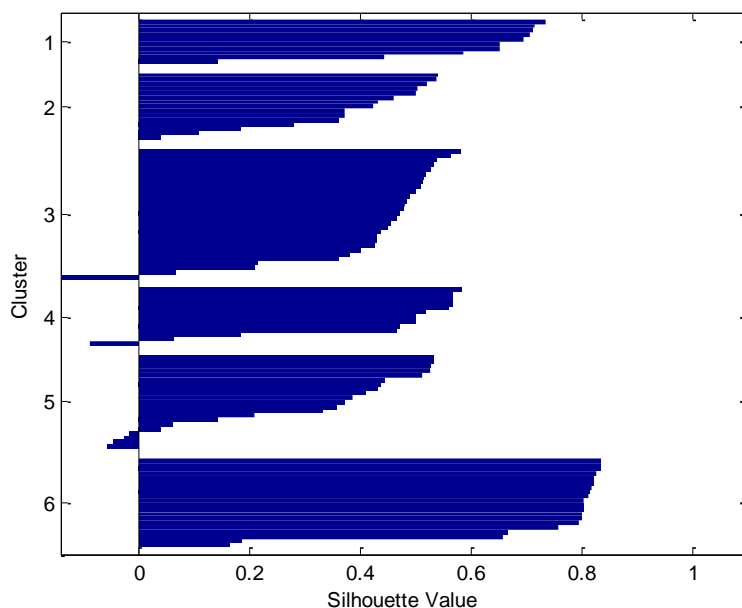
```
[silh5,h] = silhouette(economicdataset,idx5,'cityblock');
cluster5 = mean(silh5)

cluster5 =

    0.4960
```

б)

Фиг. 4.9. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в пет клъстера при метрично разстояние “cityblock”



а)

```
[silh6,h] = silhouette(economicdataset,idx6,'cityblock');
cluster2 = mean(silh6)

cluster2 =

    0.4646
```

б)

Фиг. 4.10. Профилна диаграма а) и SI средна стойност б) с разделяне на данните в шест клъстера при метрично разстояние “cityblock”

- ❖ Да се проведат аналогични на предходните дейности с приложение на тип метрика “cityblock” като бъдат оценени изходните „профилни диаграми“ и “SI средни стойности”, които следва да бъдат получени от фиг. 4.6 до фиг. 4.10;
- ❖ Да се определи и оцени характерът на изменение на “критерия SI” при нарастване на клъстерните зони и най-подходящото потенциално количество клъстерни зони, в които могат да се разпределят опитните данни съобразно метрика “cityblock”;

- ❖ Да се проведе сравнителен анализ във връзка с представените резултати и констатира **най-подходящият тип k-means метрика** и **оптималеният брой клъстери** на разделяне на опитните данни;
- ❖ Да се изведат координатите, определящи местоположението на центроидите спрямо входните променливи, както и изчислените разстояния от еталоните до тях в околността по отношение на всеки клъстер спрямо тестовите метрики “sqeuclidean” и “cityblock” (на фиг. 4.11 са дадени примерни параметрите за използвано разстояние “sqeuclidean”);

	1	2
1	15.8357	11.2500
2	13.2667	15.7476
3	11.9611	9.3722
4	8.6923	10.0154
5	10.9286	12.3000

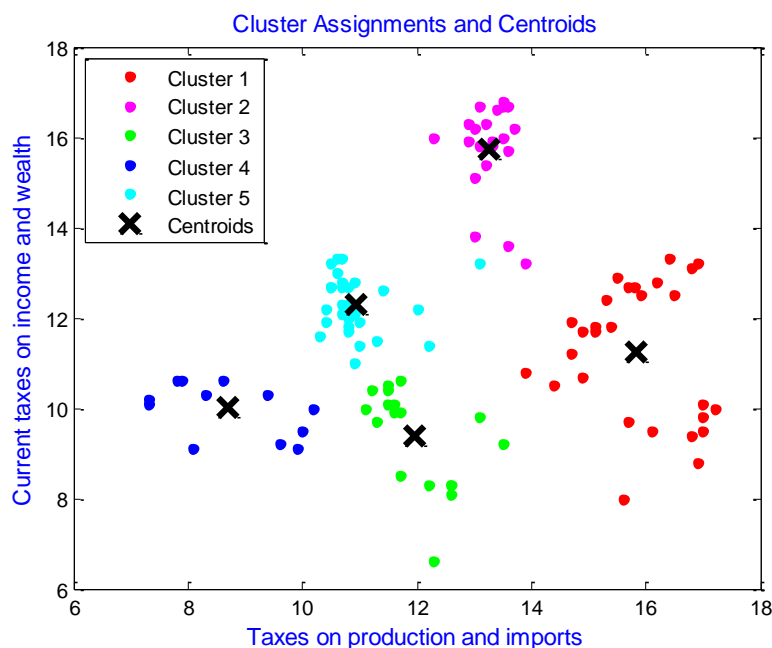
a)

	1	2
1	84.1343	
2	23.7790	
3	28.6589	
4	16.7662	
5	20.8171	

б)

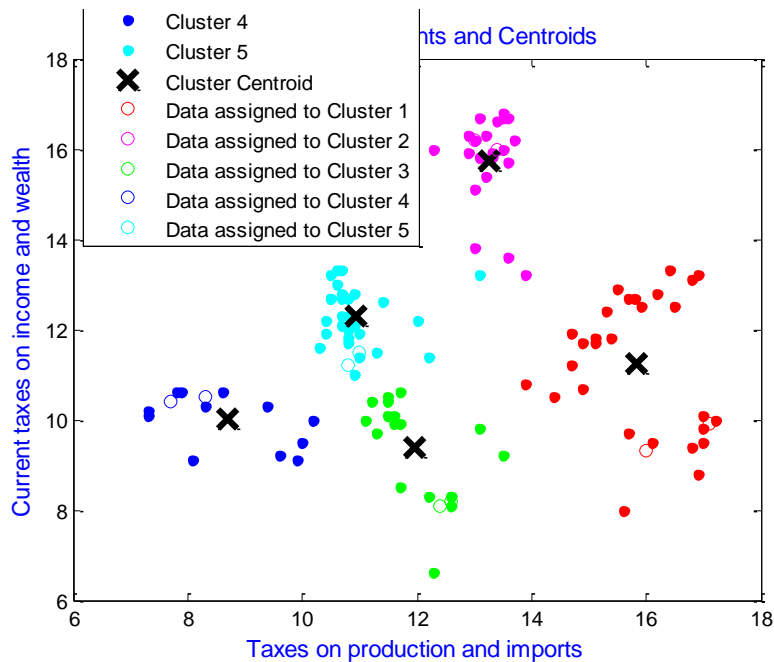
Фиг. 4.11. Местоположение на центроидите по отношение на всички информативни признаци а) и калкулирано разстояние от всяко наблюдение до центроидите в околността на съответните дефинирани клъстери б) за метрика „sqeuclidean“

- ❖ Да се генерират разпределения на наблюденията в околността на назначените клъстерни зони с обозначаване на центроидите в двумерното пространство, формирано от информативните признаци за разглежданите метрики “seuclidean” и “cityblock” (фиг. 4.12 дава представа за информационното разпределение за приложено разстояние “sqeuclidean”);



Фиг. 4.12. Разпределение на данните с обозначаване на центроидите в клъстерните зони – визуализация по отношение на информативните признаци за метрика „sqeuclidean“

- ❖ Да се извърши верификация на проведените дейности чрез назначаване на нови опитни еталони към обособените клъстерни региони в двумерното пространство, означени с „o“ относно заложените разстояния “sqeuclidean” и “cityblock” (на фиг. 4.13 е онагледен процеса на верификация при базисна метрика “sqeuclidean”);



Фиг. 4.13. Назначаване на нови данни в клъстерните зони – визуализация по отношение на информативните признаци за метрика „sqeuclidean“

- ❖ Да се направи повторение на изложената последователност от k-means клъстерен анализ при други типове метрики на основата на оценка на критерия SI.

4.5. Контролни въпроси

1. Кое налага необходимостта от клъстерен анализ на данни и относно целите на какви приложения се използва?
2. Какви типове метрики се залагат при изпълнение на k-means алгоритъм за клъстеризация?
3. По какъв начин се извършва изборът на оптимален брой клъстерни зони, в които опитните данни могат да бъдат групирани?
4. Какви инструменти за прилагане на способите на клъстерния анализ на данни познавате?

РАЗПОЗНАВАНЕ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ ПОСРЕДСТВОМ ИЗКУСТВЕН ИНТЕЛЕКТ В СРЕДА MATLAB

5.1. Цел на упражнението

Да се запознаят студентите с изкуствените невронни мрежи и алгоритмите за тяхното обучение като аналитично средство за решаване на задачи за разпознаване и класификация при опериране с масиви от икономически данни.

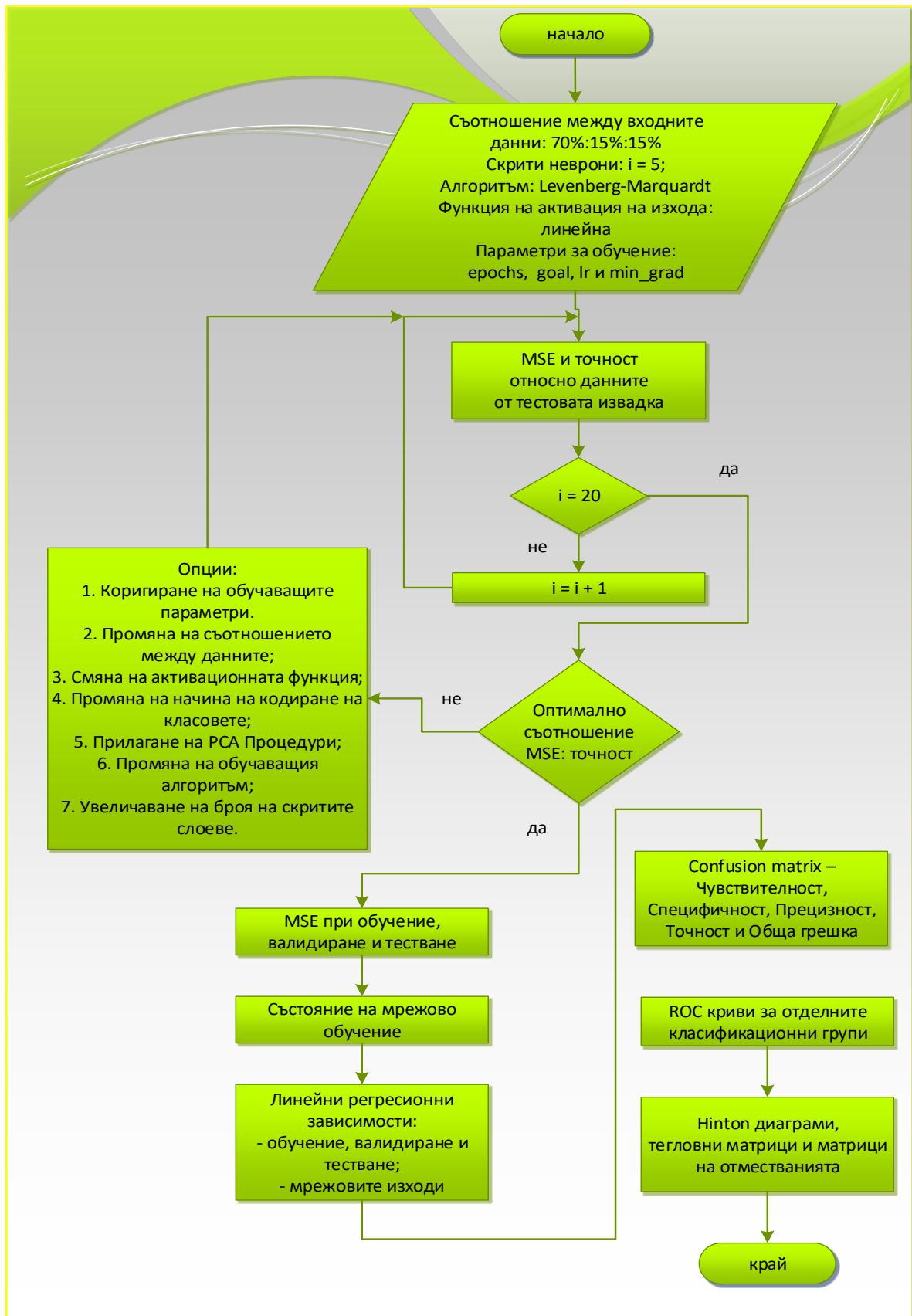
5.2. Теоретична постановка

Анализ, оценка и синтез на модели на идентификация на основата на изкуствен интелект

Предвижда се дефиниране на трислойни архитектури на изкуствени невронни мрежи с пряко разпространение на сигналите и обратно разпространение на грешката. Целевите информационни групи се кодират класово в отделни изходни неврони чрез дискретни кодови комбинации. Например при наличие на два класа „1 0“ за клас №1 и „0 1“ за клас №2. Това са теоретични стойности, които невронната мрежа трябва да върне като резултат в идеалния случай. Алгоритъмът на невронна приложимост за селекция на идентификационни модели е представен на фиг. 5.1. Първоначално се прави начална инициализация, включваща:

- ✚ нормализация на входните данни в границите от -1 до +1;
- ✚ дефиниране на съотношение между входните данни относно процесите на обучение, валидиране и тестване 70%:15%:15%;
- ✚ въвеждане на минимално количество неврони в скрития слой на мрежата като за конкретния случай може да бъде 5 неврона;
- ✚ задаване на обучаващ алгоритъм. Избира се алгоритъмът на Levenberg-Marquardt като най-бърз алгоритъм за обучение;
- ✚ избор на линеен тип на активационна функция в скрития слой на невронната мрежа;
- ✚ задаване на начални обучаващи параметри. Това са брой обучаващи цикли, целева грешка, скорост на обучение и минимален градиент (направление на минимизация на грешката).

Изчисляват се критериите средноквадратична грешка и точност на класификация, след което се преминава към увеличаване на скритите невронни единици и повторно определяне на критериите. Действията се повтарят циклично до момента достигане на 20 междинни неврона. Извършва се проверка за наличие на оптимално съотношение между ниска MSE и висока точност. Ако не бъде установено или параметричните техни стойности са незадоволителни, могат да бъдат направени следните промени до регистриране на краен модел с удовлетворяващи параметри, съответно [33-37]:



Фиг. 5.1. Алгоритъм на базата на ANN

- ✚ експериментална корекция на обучаващите параметри;
- ✚ използване на други процентни съотношения между входните данните при обучаващи, валидиращи и тестови процедури – 60%:20%:20% или 50%:25%:25%;
- ✚ замяна на изходната функция на активация с тангенс-сигмоидална, логаритмична-сигмоидална, softmax и др.;
- ✚ изменение на принципа на класово кодиране – запазване на дефинирането на изходните групи с отделни изходни неврони и замяна на дискретното с кодиране със стойности -1 и +1 (при три класа: „+1 -1 -1“ за първи, „-1 +1 -1“ при втори и „-1 -1 +1“ при трети клас) или използване на един неврон в изходния слой и константно кодиране на различни нива (при три класа: „-0.6“ за клас №1, „0“ за клас №2 и „+0.6“ за клас №3);
- ✚ приложение на принципен компонентен анализ за редукция на размерността на входните вектори;
- ✚ използване на друг алгоритъм за обучение, измежду BFGS Quasi-Newton, Resilient Backpropagation, Scaled Conjugate Gradient, Conjugate Gradient with Powell/Beale Restarts, Fletcher-Powell Conjugate Gradient, Polak-Ribiere Conjugate Gradient, One Step Second и Variable Learning Rate Backpropagation;
- ✚ прилагане на многослойни невронни архитектури.

Относно селектирания най-добър модел се прилагат действия по анализ на качеството, свързвани с:

- ✚ изменението на средноквадратичната грешка относно проведените процеси при обучение, валидиране и тестване (мрежовата производителност) – проверка за индикация за некоректност;
- ✚ състоянието на мрежово обучение – онагледяване на вариациите на градиента, стъпката на негово намаляване и извършените валидиращи проверки в итерационните рамки на обучаващия процес;
- ✚ характера на линейните регресионни зависимости по отношение на мрежово обучение, валидация, тестване и изходите на невронната мрежа, както и нивата на постигнатите корелационни коефициенти;
- ✚ оценка на класификационните параметри Чувствителност, Специфичност, Прецизност, Точност и Обща грешка въз основа на коректно и некоректно класифицираните тестови наблюдения/еталони по изходни групи;
- ✚ ROC кривите на база на изходната класификационна матрица a , изразяваща съотношението между Чувствителност и Специфичност;
- ✚ графична интерпретация на матриците с тегловни коефициенти и отмествания за междинния и изходния слой на невронния модел посредством диаграми на Hinton, както и тяхното числово извеждане.

Постановка на задачата за разпознаване и класификация

Изложеното изследване в следващите секции се състои в обучение, верификация и тестови процеси по отношение на селекция на изкуствени невронни мрежи с обратно разпространение:

- ✚ Group №1: Belgium и Germany;
- ✚ Group №2: Greece и Malta;
- ✚ Group №3: Italy и Luxembourg;
- ✚ Group №4: Spain и France;
- ✚ Group №5: Ireland и Latvia.

Предвидени са процедури по анализ на производителността на трислойни изкуствени невронни мрежи при различни функции на активация в изходните слоеве и алгоритми на обучение - линейна и тангенс-сигмоидална по алгоритъма на Levenberg-Marquardt (LM), softmax при Scaled Conjugate Gradient (SCG) обучаващ алгоритъм. Анализът се базира на оценка на критерийте:

- ✚ “Accuracy” и “Mean Squared Error (MSE)” при LM;
- ✚ “Accuracy” и “Cross-Entropy (CE)” за SCG,

като се следи тяхното изменение при фиксирана вариация на скритите неврони от 5 до 15 при зададена тангенс-сигмоидална активационна функция.

5.3. Задачи за изпълнение

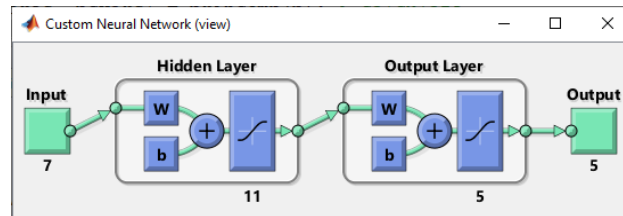
5.3.1. Изследване и синтез на изкуствени невронни мрежи с LM обучение

- ❖ Да се проведат процедури по обучение на трислойни невронни архитектури при 7 информативни признака (*точка 4.2*) на основата на LM обучаващ алгоритъм при фиксирана тангенс-сигмоидална функция на активация в изходния слой като се експериментира с количеството скрити неврони в интервала от 5 до 15 изчислителни единици. Резултатите, свързани с изследваните показатели за качество, да бъдат обобщени съгласно таблица 5.1;

Таблица 5.1. Резултати при синтез на изкуствени невронни мрежи с изходна тангенс-сигмоидална активация при обучение с LM алгоритъм

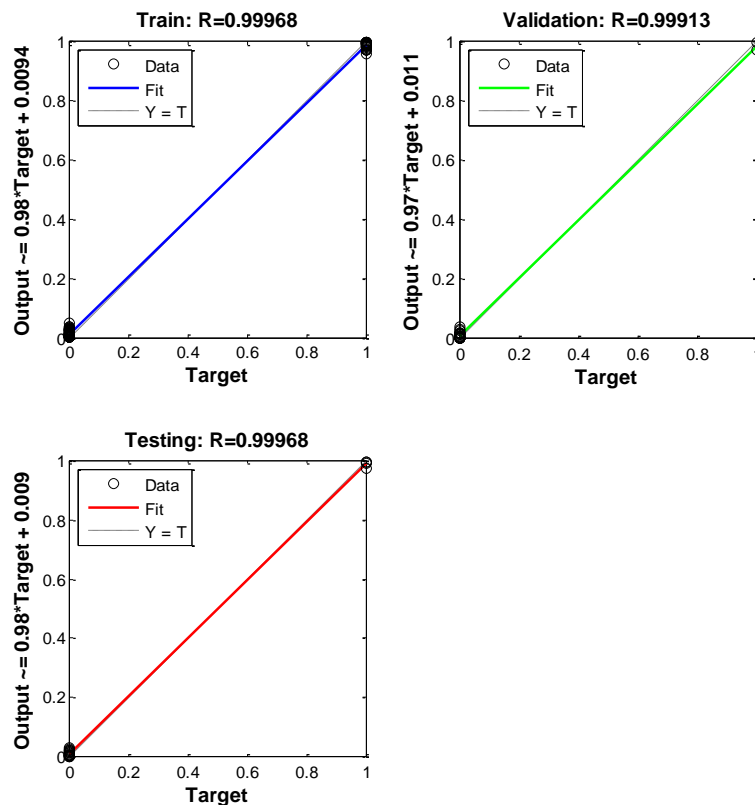
№	Скрити неврони	Точност, %	MSE
1.			
2.			
3.			
4.			
5.			
6.			
7.	11	100.00	0.0053
8.			
9.			
10.			
11.			

- ❖ Да се селектира и представи видът на намерена невронна архитектура според препоръчително изискване за оптималност, касаещо съотношението „максимална точност“ спрямо „минимална средноквадратична грешка“. Тук за конкретния случай е показана невронна мрежа с постигнати най-добри показатели за качество при задаване на 11 скрити неврона (фиг. 5.2);



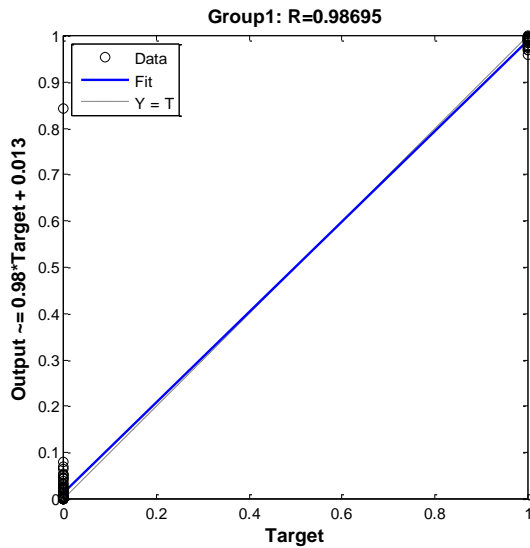
Фиг. 5.2. Синтезирана невронна мрежа с тангенс-сигмоидална изходна активация при LM обучение за разпознаване на географски регион на принадлежност на икономически показатели

- ❖ Да се построят линейни регресионни зависимости в хода на обучаващите, валидиращите и тестовите процеси, както е показано на фиг. 5.3. Да се направи оценка на постигнатите нива на коефициентите на корелация R и анализира характерът на получените емпирични линии на регресия;

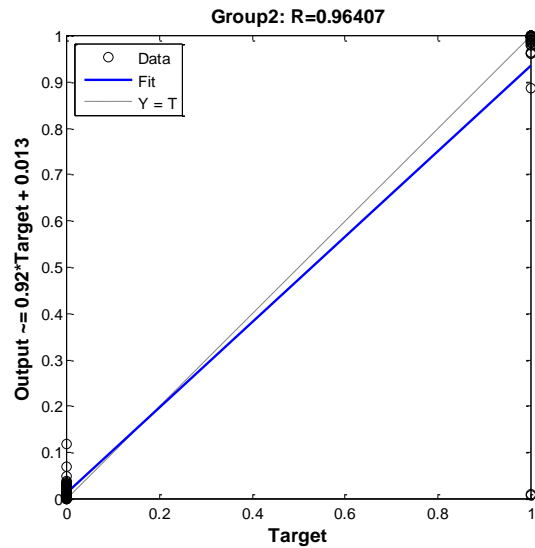


Фиг. 5.3. Регресионни зависимости от обучение, валидация и тестване на синтезираната архитектура с тангенс-сигмоидална изходна активация при LM обучение за разпознаване на географски регион на принадлежност на икономически показатели

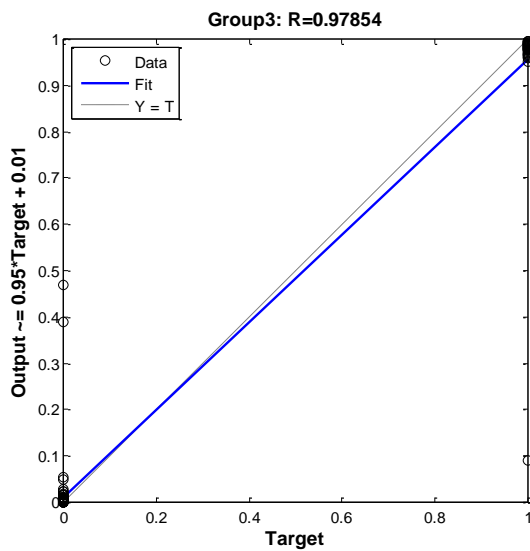
- ❖ По подобен начин да се генерират линейни регресионни зависимости за изходите на разглеждания невронен модел, представени на фиг. 5.4. Да се оцени степента на корелация и сходство между калкулираните с приложение на синтезираната архитектура и целеви резултати за всяка отделна класификационна група;



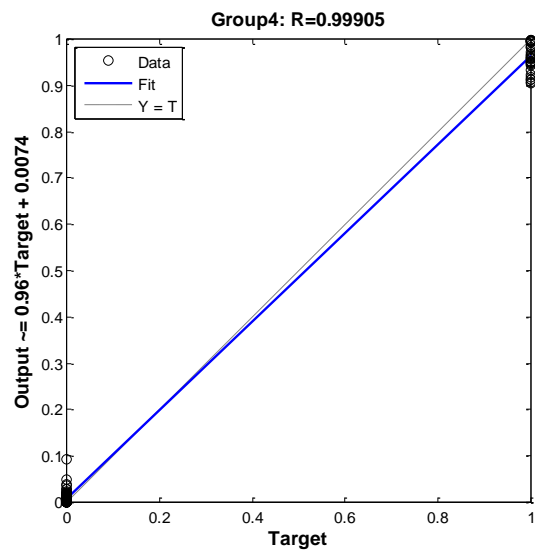
a)



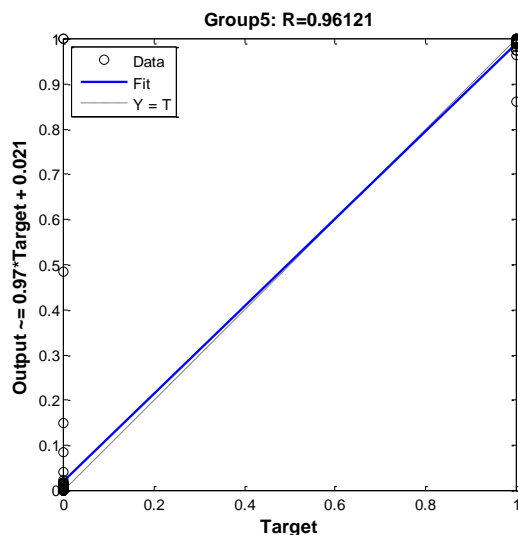
б)



в)



г)



д)

Фиг. 5.4. Регресионни зависимости относно мрежовите изходи за синтезираната архитектура с тангенс-сигмодална изходна активация при LM обучение за разпознаване на географски регион на принадлежност на икономически показатели – а) Group №1, б) Group №2, в) Group №3, г) Group №4 и д) Group №5

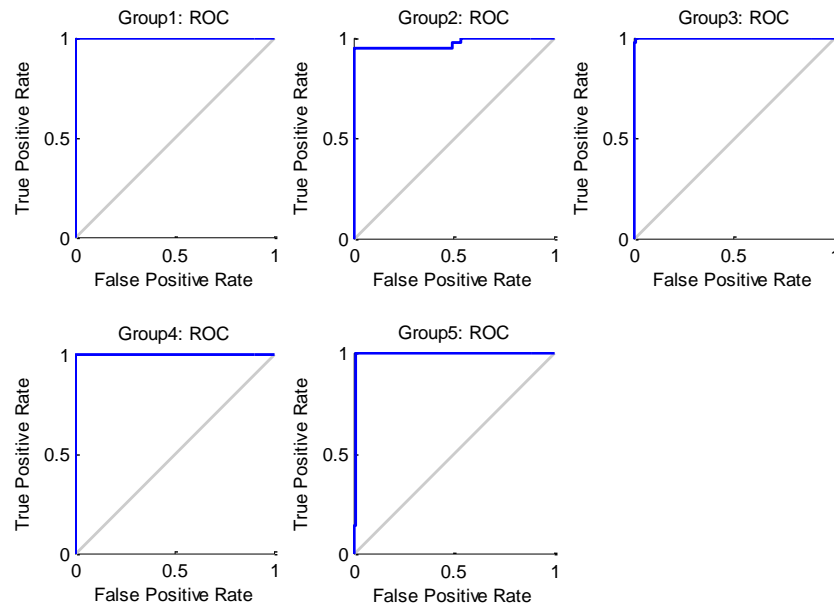
- ❖ Да се генерира матрица на коректните и некоректните класификации (фиг. 5.5) и определи набор от специфицирани класификационни параметри “Sensitivity”, “Specificity” и “Precision” за всеки отделен клас;

Confusion Matrix

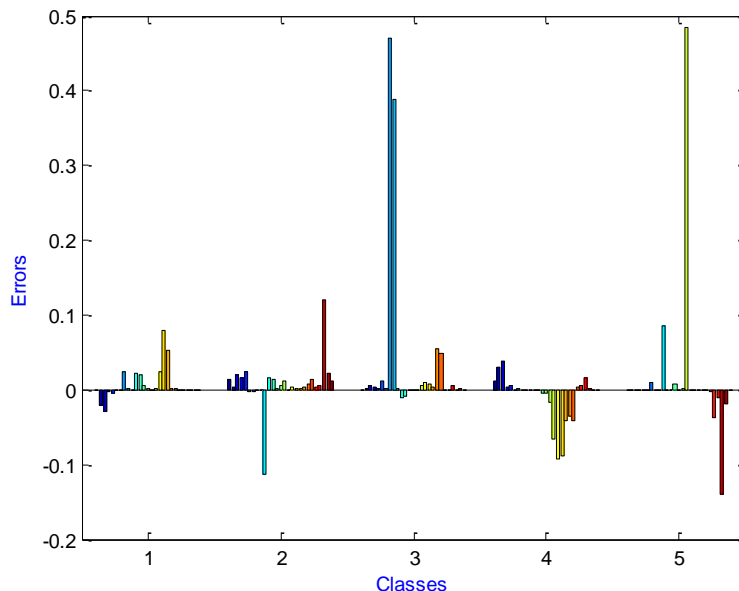
	1	2	3	4	5	
1	5 18.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	5 18.5%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	2 7.4%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	9 33.3%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	6 22.2%	100% 0.0%
	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	1	2	3	4	5	
	Target Class					

Фиг. 5.5. Класификационна матрица за синтезираната архитектура с tansig изходна активация при LM обучение за разпознаване на географски регион на принадлежност

- ❖ Да се изведат Receiver Operating Characteristics на фиг. 5.6, изразяващи зависимостта между „Sensitivity“ и „1-Specificity“, анализира и оцени качеството на класификация за отделните изходни групи за синтезирания невронен модел;



Фиг. 5.6. ROC криви относно мрежовите изходи за синтезираната архитектура с тангенс-сигмоидална изходна активация при LM обучение за разпознаване на географски регион на принадлежност на икономически показатели – а) Group №1, б) Group №2, в) Group №3, г) Group №4 и д) Group №5



Фиг. 5.7. Диаграма на грешките относно мрежовите изходи за синтезираната архитектура с тангенс-сигмоидална изходна активация при LM обучение за разпознаване на географски регион на принадлежност на икономически показатели

- ❖ Да се генерира диаграма на мрежовите грешки и установят вариационните диапазони по класификационни групи и общия за всички групи интервал на грешка, аналогично на представената на фиг. 5.7;
- ❖ Да се изведат тегловните матрици и матриците на отместванията за скрития и изходния слой на селектираната архитектурата на разпознаване и класификация;
- ❖ Да се направят изследвания при промяна на типа на изходната невронна активация като бъдат приложени „линейна“ и „логаритмична-сигмоидална“ функции като се направи подбор на модели с най-висока степен на адекватност.

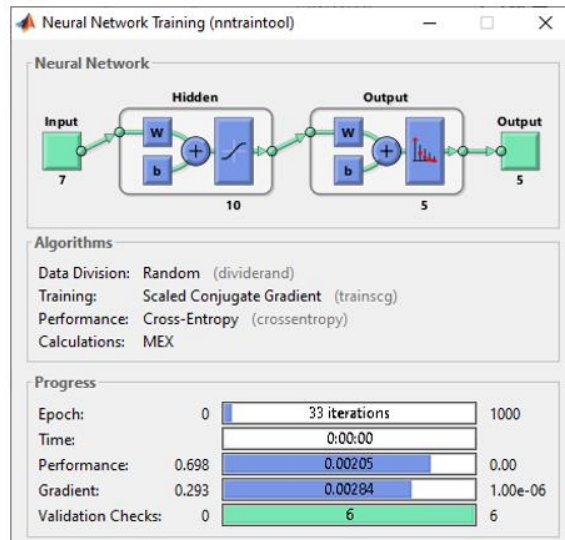
5.3.2. Изследване и синтез на модели на базата на изкуствен интелект при SCG алгоритъм на обучение

- ❖ Да се създадат трислойни архитектури на изкуствени невронни мрежи с изходна “softmax” функция на активация с изпълнение на SCG обучаващи процедури и обобщят резултати относно вариациите на „класификационната точност“ и индикатора “Cross-Entropy” при увеличаване на скритите неврони от 5 до 15 според вида на таблица 5.2;

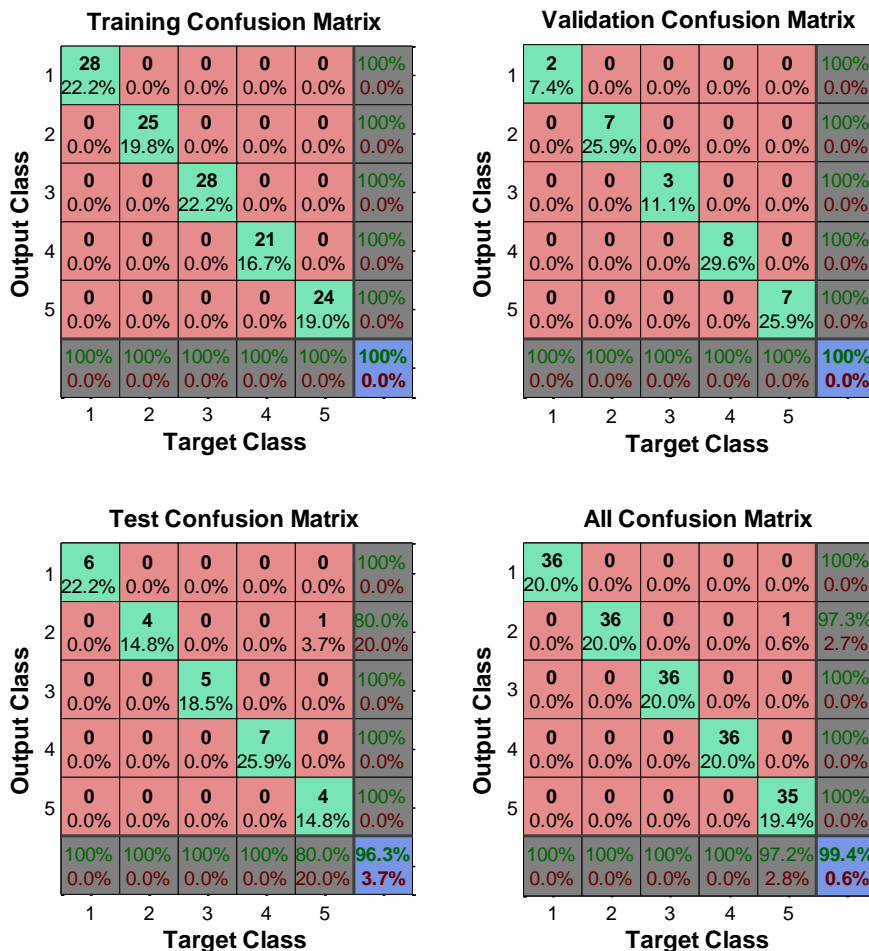
Таблица 5.2. Резултати при синтез на изкуствени невронни мрежи с изходна softmax активация при обучение с SCG алгоритъм

№	Скрити неврони	Точност, %	CE Testing
1.			
2.			
3.			
4.			
5.			
6.	10	99.4	6.54704e-0
7.			
8.			
9.			
10.			
11.			

- ❖ Да се представи ходът от обучение на синтезирания най-добър невронен модел при наличие на 10 структурни междинни неврона с приложение на Scaled Conjugate Gradient алгоритъм, показан на фиг. 5.8;
- ❖ Да се изследва тенденцията в изменението на показателя за качество “Cross-Entropy” при обучение, валидация и тестване за избрания невронен модел с на-добри показания за целите на определяне на райони на принадлежност на икономически показатели. Да се оцени мрежовата производителност;

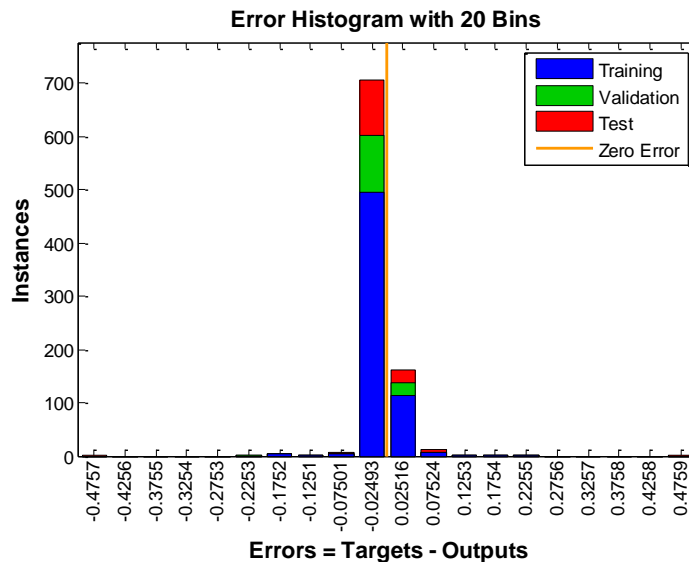


Фиг. 5.8. Ход от обучение на селектираната невронна мрежа при softmax изходна активация и SCG обучение за разпознаване на географски регион на принадлежност на икономически показатели



Фиг. 5.9. Класификационни матрици за селектираната архитектура с при softmax изходна активация и SCG обучение за разпознаване на географски регион на принадлежност на икономически показатели

- ❖ Да се визуализира набор от класификационни матрици от обучение, валидиране, тестване и обща за указаните процеси матрица на коректните и некоректните класификации, посочена на фиг. 5.9. Да се анализира разпределението на еталоните с правилно и неправилно дефинирана принадлежност по класове;



Фиг. 5.10. Хистограма на грешката за селектираната невронна мрежа при softmax изходна активация и SCG обучение за разпознаване на географски регион на принадлежност на икономически показатели

- ❖ Да се онагледи хистограма на грешките като на показаната на фиг. 5.10 и оценят нивата, в които попадат количествените опитни данни от процесите на обучение, валидиране и тестване;
- ❖ Да се изведат матриците на теглата и отместванията за междинния и изходния мрежови слой и техни графични интерпретации посредством диаграми на Hinton.

5.4. Контролни въпроси

1. Какви типове функции на активация могат да бъдат използвани в изходните слоеве на изкуствените невронни мрежи?
2. Коя е най-съществената особеност на алгоритъма за невронно обучение Levenberg-Marquardt?
3. По какъв начин се интерпретират изходните резултати при фиксиране на "softmax" тип на активация в изходите на изкуствените невронни мрежи?
4. Кой графичен признак определя качествата на синтезираните модели при анализа на ROC диаграмите?
5. Какви други видове невронни мрежи в предметна област „разпознаване и класификация“, освен разглежданите с обратно разпространение на грешката, се прилагат в практиката?

ИДЕНТИФИКАЦИЯ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ ЧРЕЗ ПРИЛОЖЕНИЕ НА АДАПТИВНИ НЕВРОННО-РАЗМИТИ ИНТЕРФЕЙСНИ СИСТЕМИ В СРЕДА MATLAB

6.1. Цел на упражнението

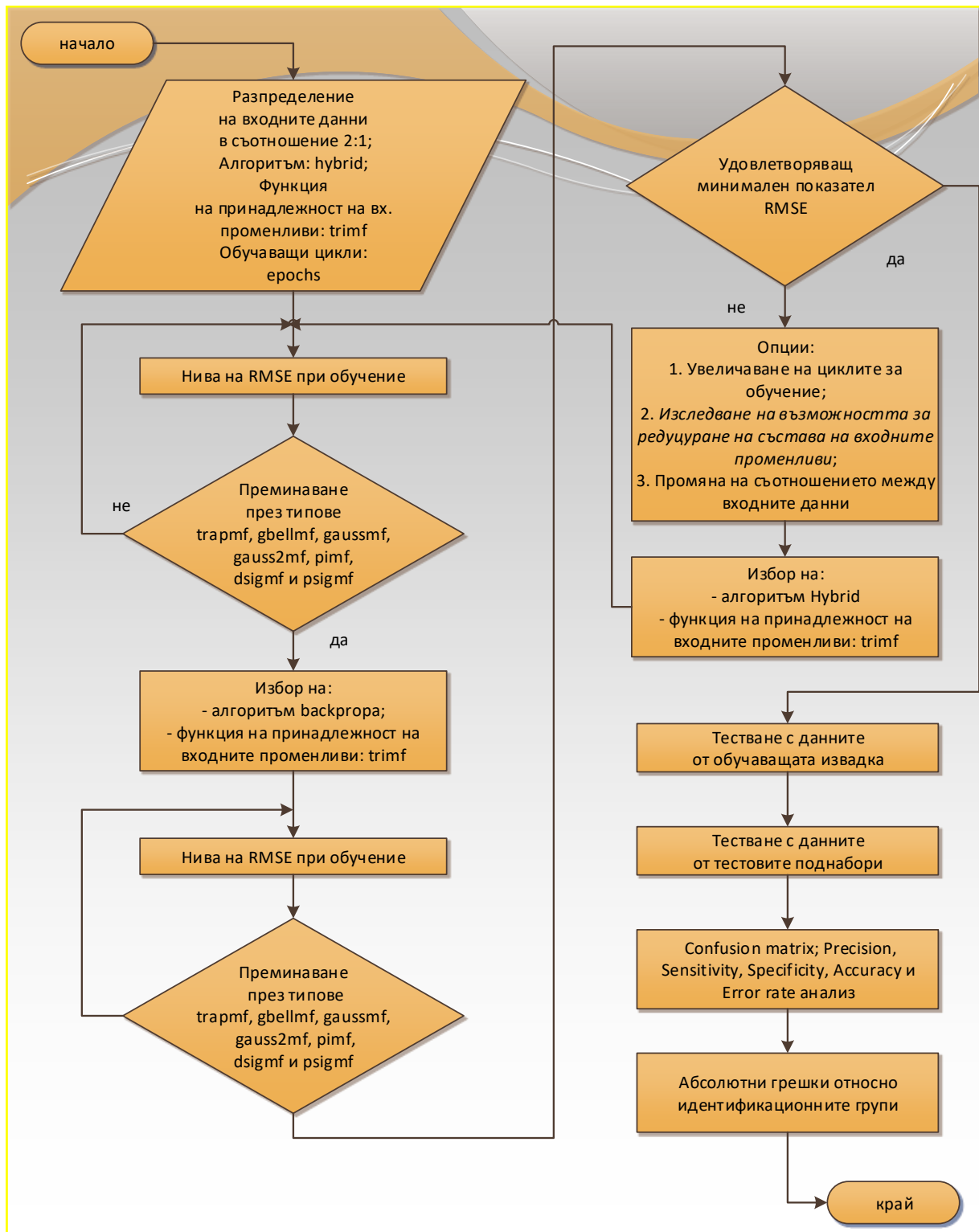
Да се запознаят студентите с процесите на обучение и селекция на адаптивни невронно-размити интерфейсни системи като инструмент при решаването на задачи за идентификация в средата MATLAB във връзка с анализа на масиви от икономически показатели по обособени географски региони.

6.2. Теоретична постановка

Алгоритъмът за изследване и синтезиране на невронно-размити модели за разпознаване и класификация на информационни групи е даден на фиг. 6.1.

Алгоритмичното описание е, както следва:

- ✚ извършва се разпределяне на входния набор от данни в два набора, съответно за обучение и тестови, в съотношение две към едно. Задават се хибриден обучаващ алгоритъм, реализиран посредством този на обратно разпространение на грешката и метода на най-малките квадрати, и функция на принадлежност на входните променливи от триъгълен тип. Изходните групи на принадлежност се дефинират чрез целочислен константен тип, съответно стойност „1“ за клас №1, „2“ при втора група и т.н.;
- ✚ изследва се показателят RMSE (средно квадратично отклонение) относно обучаващите данни. Процесът се извършва последователно за всички указани входни функции на принадлежност, съответно от трапецовиден, камбановиден, Гаусов, комбиниран Гаусов, р-образен, типове разлика и произведение между две сигмоидални функции;
- ✚ преминава се към избор на втори алгоритъм – алгоритъм с обратно разпространение на грешката. Калкулират се средно квадратичните отклонения за всички приложени функционални типове;
- ✚ оценяват се нивата на показателя по двата алгоритъма. В случай, че са констатирани нежелано високи нива на грешките от обучение се:
 - увеличава обучаващата продължителност и следва експериментално изследване на възможността за редуциране на входните променливи при последователно изключване и факторни комбинации с цел повишаване на невронно-размитата производителност;
 - извършва ново информационно наборно разделение на входните данни, след което се преминава към повторение на горепосочените процедури;
 - ако бъде установен модел с удовлетворяваща функционалност на следващ етап двукратно се оценява качеството му на класификация чрез еталони от обучаващата и тестовите извадки;



Фиг. 6.1. Алгоритъм на базата на ANFIS

- ✚ оценката се потвърждава чрез матрици на коректните и некоректните класификации и пресметнатите резултатни абсолютни грешки, даващи насоки за действителните отклонения на изхода на невронно-размития модел спрямо заложените теоретични целочислени нива [38-42].

6.3. Формиране на задачата за идентификация

Залага се следната задача за идентификация на географски региони на постигнати икономически показатели, както следва:

- ✚ Клас №1: Belgium;
- ✚ Клас №2: Germany;
- ✚ Клас №3: Greece;
- ✚ Клас №4: Malta,

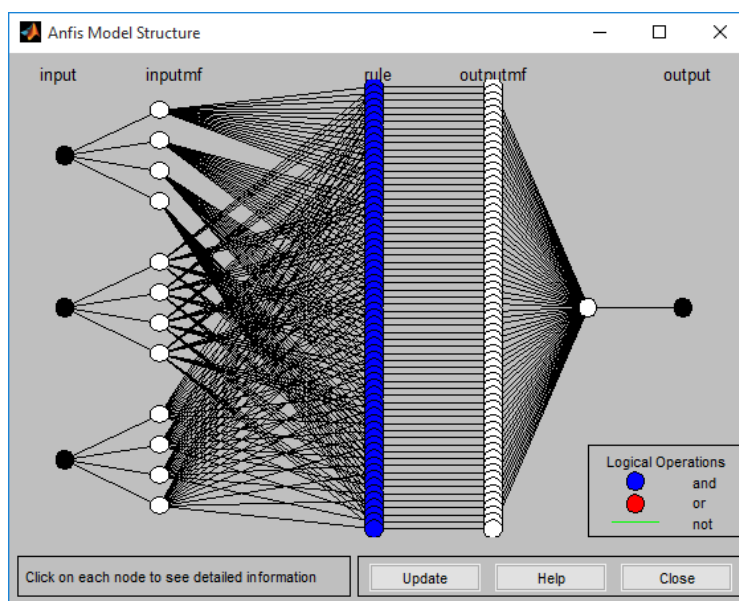
чрез използване на адаптивна невронно-размита интерфейсна система, теоретично задавани като „константен“ целочислен тип относно целевите групи от икономически показатели за идентификация („1“ за клас №1, „2“ за клас №2 и т.н.).

Въз основа на предварителна оценка на значимостта на информативност на използвания набор от 7 информативни признака (*точка 4.2*), за конкретния случай на изследване бяха избрани в качеството на входни променливи:

- ✚ Taxes on production and imports;
- ✚ Current taxes on income and wealth;
- ✚ Intermediate consumption.

6.4. Задачи за изпълнение

- ❖ Да се обособят отделни категории данни за обучаващи и валидиращи процедури по отношение на поставената задача за идентификация;



Фиг. 6.2. Невронно-размита архитектура за идентификация на постигнати икономически показатели по региони

- ❖ Да се импортират опитни набори за обучение и изследване на невронно-размита архитектура при подаване на три входни променливи, представена на фиг. 6.2 в работното поле на продукта MATLAB;

- ❖ Да се приложат обучаващи процеси за случаите на ограничение от 200 и 1200 итерации по хибриден алгоритъм “hybrid” и алгоритъм на обратно разпространение на грешката “backpropa” при задаване на различни типове функции на принадлежност на входните променливи. Да се обобщят данни във връзка с критерия RMSE - таблица 6.1 и таблица 6.2;

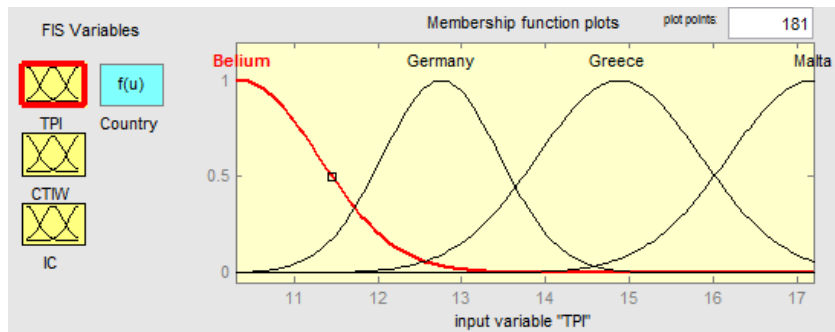
Таблица 6.1. Резултати при изследване на ANFIS архитектура при хибриден алгоритъм в рамките на 200 цикъла на обучение

RMSE индикатор	
Функция на принадлежност на входните променливи	Алгоритъм “hybrid”
trimf	
trapmf	
gbellmf	
gaussmf	0.0039411
gauss2mf	
pimf	
dsigmf	
psigmf	

Таблица 6.2. Резултати при изследване на ANFIS архитектура при алгоритъм на обратно разпространение на грешката при ограничение до 1200 обучаващи итерации

RMSE индикатор	
Функция на принадлежност на входните променливи	Алгоритъм “hybrid”
trimf	0.16283
trapmf	
gbellmf	
gaussmf	
gauss2mf	
pimf	
dsigmf	
psigmf	

- ❖ Да се анализират постигнатите средно квадратични отклонения за използваните подходи на обучение като се избере подходящ обучаващ алгоритъм и функционален тип на принадлежност относно входните информативни признаци, за които е налице най-ниска регистрирана стойност. За илюстриране на функционалността на изследването е установена примерна минимална грешка от обучение 0.0039411 при **Гаусов тип** функция на входните променливи за модел, обучен при **комбиниран подход от най-малките квадрати и обратно разпространение на грешката**;
- ❖ Да се представи и анализира видът на функциите на принадлежност на входните променливи, както и допълнителни специфични данни, за синтезираната примерна невронно-размита архитектура – фиг. 6.3 до фиг. 6.5 (зависимостта на „степената на принадлежност“, изменяща се в границите от 0 до 1, към „реални стойности на входните променливи“);



a)

Current Variable		Current Membership Function (click on MF to select)	
Name	TPI	Name	Belium
Type	input	Type	gaussmf
Range	[10.3 17.2]	Params	[0.9196 10.37]
Display Range	[10.3 17.2]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

б)

Current Variable		Current Membership Function (click on MF to select)	
Name	TPI	Name	Germany
Type	input	Type	gaussmf
Range	[10.3 17.2]	Params	[0.7011 12.75]
Display Range	[10.3 17.2]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

в)

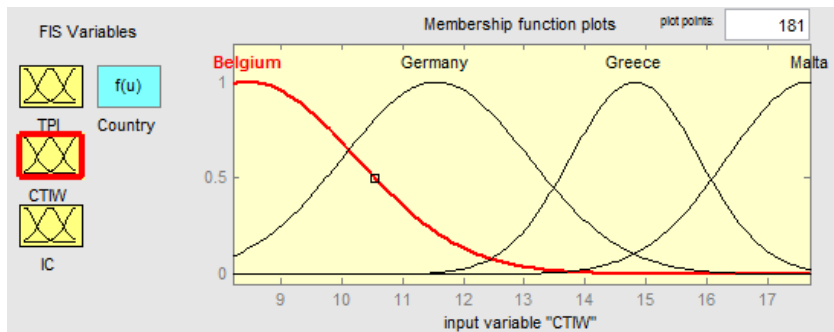
Current Variable		Current Membership Function (click on MF to select)	
Name	TPI	Name	Greece
Type	input	Type	gaussmf
Range	[10.3 17.2]	Params	[0.9745 14.88]
Display Range	[10.3 17.2]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

г)

Current Variable		Current Membership Function (click on MF to select)	
Name	TPI	Name	Malta
Type	input	Type	gaussmf
Range	[10.3 17.2]	Params	[1.001 17.19]
Display Range	[10.3 17.2]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

д)

Фиг. 6.3. Общ вид а) и параметрична информация относно функции на принадлежност б) “Belgium”, в) “Germany”, г) “Greece” и д) “Malta” на входна променлива “Taxes on production and imports (TPI)”



a)

Current Variable		Current Membership Function (click on MF to select)	
Name	CTIW	Name	Belgium
Type	input	Type	gaussmf
Range	[8.2 17.7]	Params	[1.777 8.451]
Display Range	[8.2 17.7]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

б)

Current Variable		Current Membership Function (click on MF to select)	
Name	CTIW	Name	Germany
Type	input	Type	gaussmf
Range	[8.2 17.7]	Params	[1.536 11.54]
Display Range	[8.2 17.7]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

в)

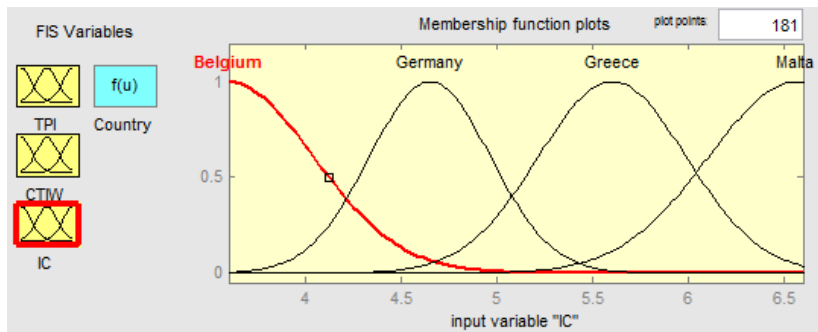
Current Variable		Current Membership Function (click on MF to select)	
Name	CTIW	Name	Greece
Type	input	Type	gaussmf
Range	[8.2 17.7]	Params	[1.058 14.82]
Display Range	[8.2 17.7]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

г)

Current Variable		Current Membership Function (click on MF to select)	
Name	CTIW	Name	Malta
Type	input	Type	gaussmf
Range	[8.2 17.7]	Params	[1.364 17.7]
Display Range	[8.2 17.7]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

д)

Фиг. 6.4. Общ вид а) и параметрична информация относно функции на принадлежност б) “Belgium”, в) “Germany”, г) “Greece” и д) “Malta” на входна променлива “Current taxes on income and wealth (CTIW)”



a)

Current Variable		Current Membership Function (click on MF to select)	
Name	IC	Name	Belgium
Type	input	Type	gaussmf
Range	[3.6 6.6]	Params	[0.4539 3.59]
Display Range	[3.6 6.6]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

б)

Current Variable		Current Membership Function (click on MF to select)	
Name	IC	Name	Germany
Type	input	Type	gaussmf
Range	[3.6 6.6]	Params	[0.317 4.652]
Display Range	[3.6 6.6]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

в)

Current Variable		Current Membership Function (click on MF to select)	
Name	IC	Name	Greece
Type	input	Type	gaussmf
Range	[3.6 6.6]	Params	[0.3877 5.603]
Display Range	[3.6 6.6]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

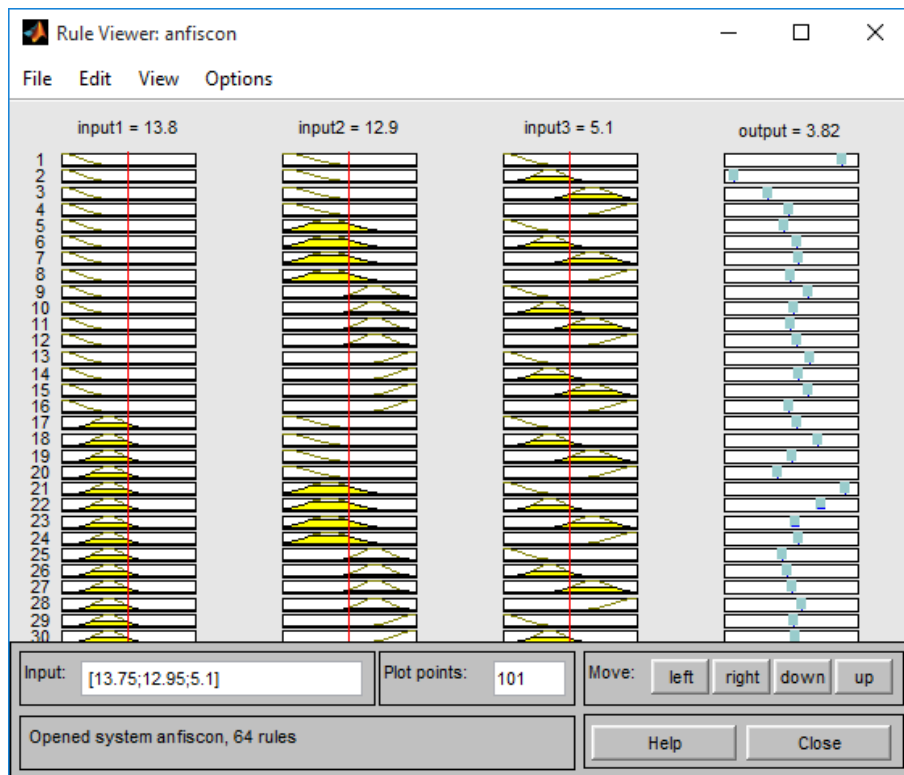
г)

Current Variable		Current Membership Function (click on MF to select)	
Name	IC	Name	Malta
Type	input	Type	gaussmf
Range	[3.6 6.6]	Params	[0.4636 6.568]
Display Range	[3.6 6.6]	<input type="button" value="Help"/> <input type="button" value="Close"/>	

д)

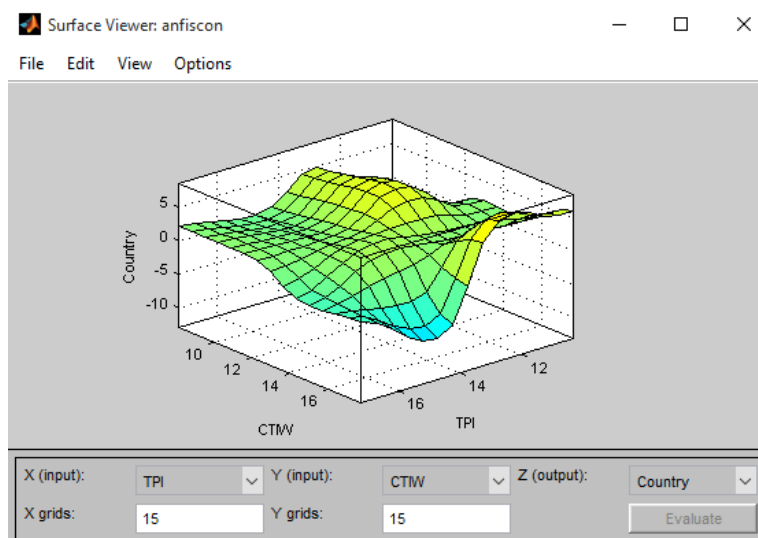
Фиг. 6.5. Общ вид а) и параметрична информация относно функции на принадлежност б) “Belgium”, в) “Germany”, г) “Greece” и д) “Malta” на входна променлива “Intermediate consumption (IC)”

- ❖ Да се изведе и обясни видът на графичната интерпретация на размитите (лингвистични) правила като трети слой на тестовата невронно-размита архитектура, показана на фиг. 6.6;

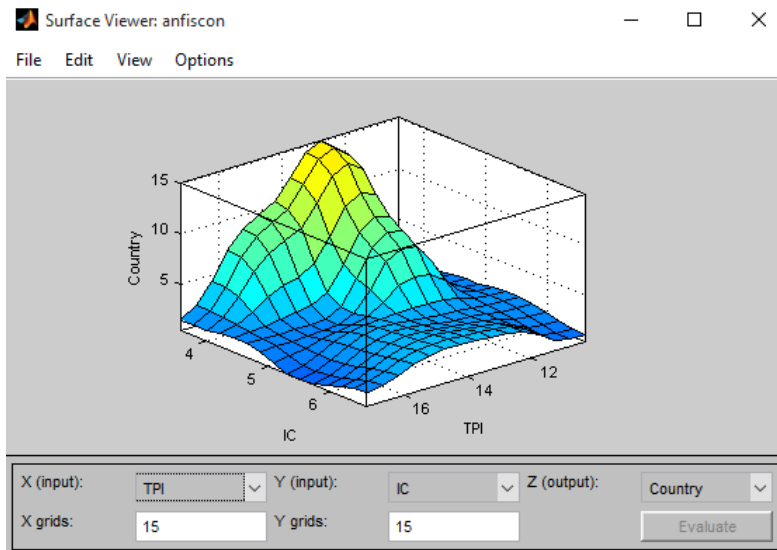


Фиг. 6.6. Лингвистични правила при невронно-размития класификатор за идентификация на постигнати икономически показатели

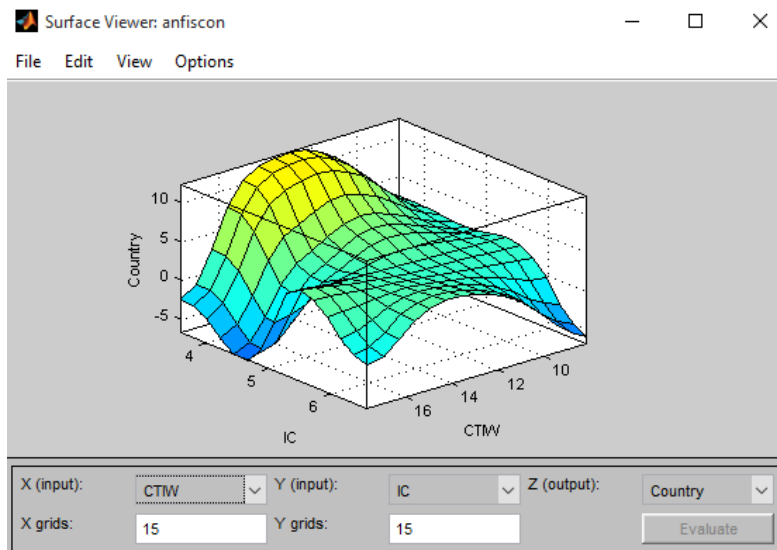
- ❖ Да се построят и анализират „повърхнини на решения (Decision Surfaces)“ при залагане на различни комбинации от два информативни признака, дадени на фиг. 6.7. Да се направи оценка относно всяка приложена комбинация от променливи;



a)



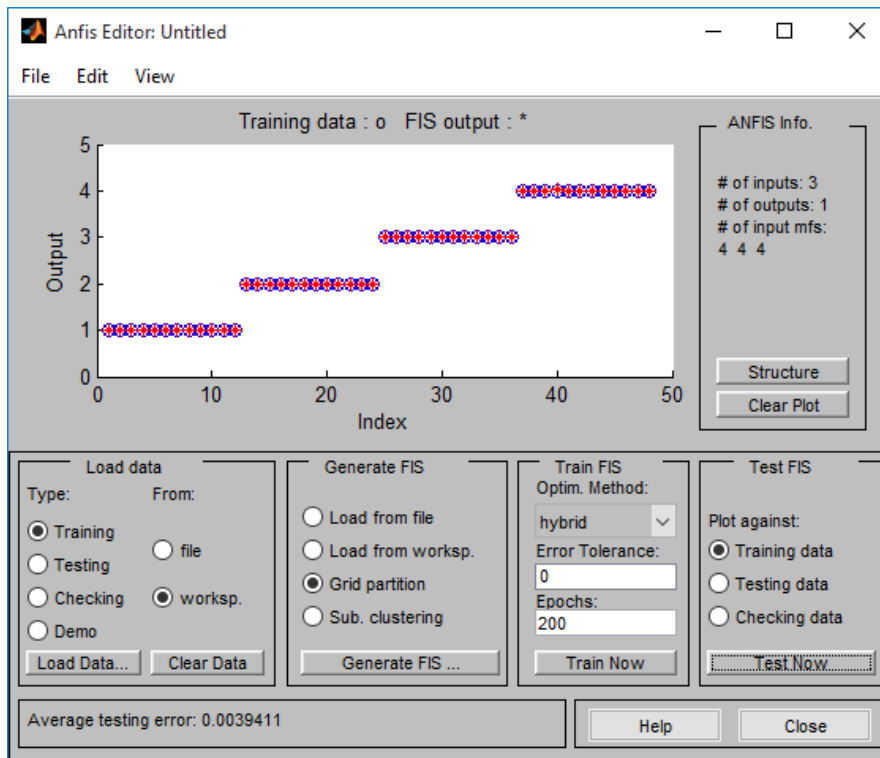
б)



в)

Фиг. 6.7. Повърхнини на решения при “TPI и CTIW”, б) “TPI и IC” и в) “CTIW и IC”

- ❖ Да се проведат тестови процедури с данните от състава на обучаващата извадка и оцени успеваемостта на отнасяне на опитни еталони към дефинираните изходни групи, представени на фиг. 6.8. Да се анализира нивото на грешката;
- ❖ Да се извърши импортиране на данни от формирани тестови поднабори (данни, които не са използвани за обучение на ANFIS) за всеки отделен клас и тестови процес с тяхно приложение (фиг. 6.9);



Фиг. 6.8. Тестване на невронно-размития класификатор за идентификация на постигнати икономически показатели с данните от обучаващата извадка

Name	Value	Min	Max
anfiscon	1x1 struct		
anfristrain	48x4 double		17.7000
ressampleclass3	3.0870	3.0870	3.0870
sampleclass3	[12.2000 13.4000 6.10...	6.1000	13.4000

a)

```
>> sampleclass3 = [12.2 13.4 6.1];
>> ressampleclass3 = evalfis(sampleclass3,anfiscon)

ressampleclass3 =

    3.0870
```

б)

Фиг. 6.9. Тестване на невронно-размития класификатор за идентификация на постигнати икономически показатели с произволен информационен еталон от тестовия набор за клас №3 – а) изглед на входните и изходните променливи и б) резултат, отнасящ се за произволен еталон

- ❖ Да се направи оценка на цялостната функционалност по отношение на синтезирания невронно-размит класификатор за идентификация на икономически данни.

6.5. Контролни въпроси

1. Обяснете накратко принципа на функциониране на адаптивните невронно-размити интерфейсни системи (ANFIS).

2. С какви приложения се свързват ANFIS в сферите на бизнеса на индустрията?
3. Кои алгоритми се използват при обучение и синтез на невронно-размити архитектури?
4. Каква е същността на параметъра „степен на принадлежност“?
5. По какъв начин могат да бъдат дефинирани функциите на принадлежност на изходните променливи при задачи за класификация и апроксимация на функции?

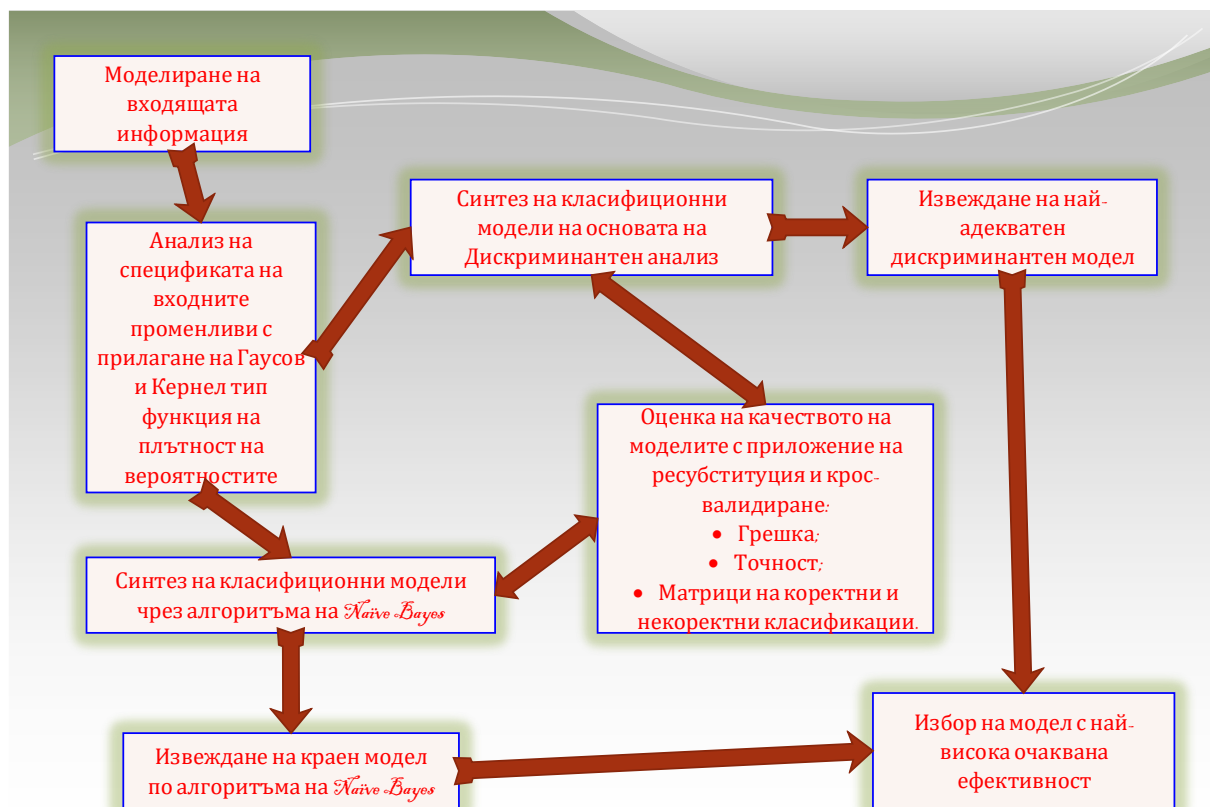
ИДЕНТИФИКАЦИЯ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ С ПОМОЩТА НА ДИСКРИМИНАНТЕН АНАЛИЗ И БЕЙС КЛАСИФИКАЦИЯ В СРЕДА MATLAB

7.1. Цел на упражнението

Да се запознаят студентите с процедурите при изследване и обработка на опитни данни относно икономически показатели при различни типове разпределения, прилагане на инструментите на дискриминантния анализ и алгоритъма на Naïve Bayes за разпознаване и класификация на обособени географски региони в среда MATLAB.

7.2. Теоретична постановка

Фигура 7.1 представя подход, комбиниращ апаратите на Дискриминантния анализ и алгоритъма на Naïve Bayes, за първоначален базов анализ на спецификата на входните променливи – например това могат да бъдат конкретни финансови индикатори, по отношение на степента на случаен характер на разпределение и определяне на функционален вероятностен тип, с който се дава по-ниска или по-висока степен на описание на моделирана информация.



Фиг. 7.1. Подход за анализ на принадлежността на икономически данни на основата на Дискриминантен анализ и алгоритъм на Naïve Bayes

След което се преминава към постояване на две категории класификатори, както следва:

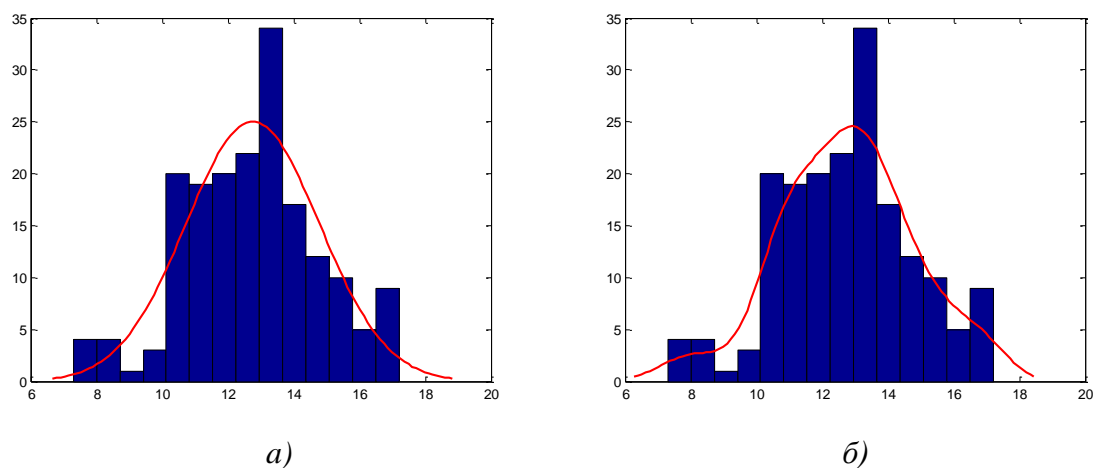
- ✚ Линейни и квадратични дискриминантни класификатори;
- ✚ Класификатори с помощта на алгоритъма на Бейс при Гаусово и Кернел разпределение на входящите променливи.

Също така се извършва и оценка на прогнозните резултати с помощта на подходите ресубституция и крос-валидиране, следвайки подобна аналогия на изследване. Въз основа на анализ на резултатите се извършва селекция на модели по двата математически апарата и избор на най-подходящ инструментариум относно конкретиката на решавания вид задача [43-49].

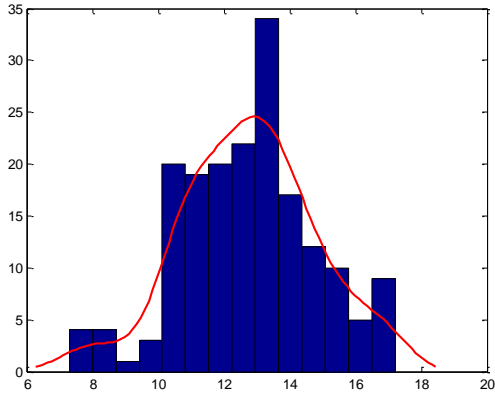
7.3. Задачи за изпълнение

7.3.1. Анализ на характера на икономически данни чрез функция на плътност на вероятността от Нормален (Гаусов) и Кернел тип

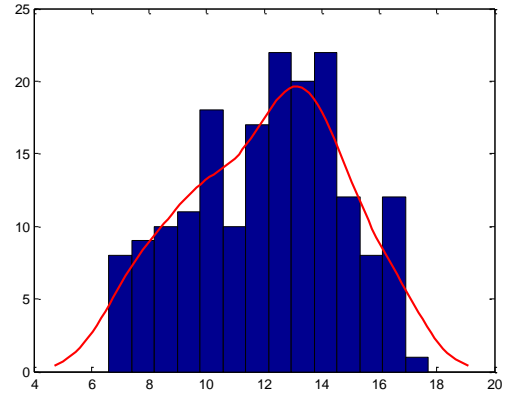
- ❖ Да се приложат процедури по Нормално (Гаусово) и Кернел разпределение спрямо входни икономически променливи (точка 4.2) като се построят и представят съответни типове графични хистограми, чиито вид е изложен от фиг. 7.2 до фиг. 7.7;



Фиг. 7.2. Нормално (Гаусово) а) и Кернел б) разпределение на входна променлива “Taxes on production and imports”

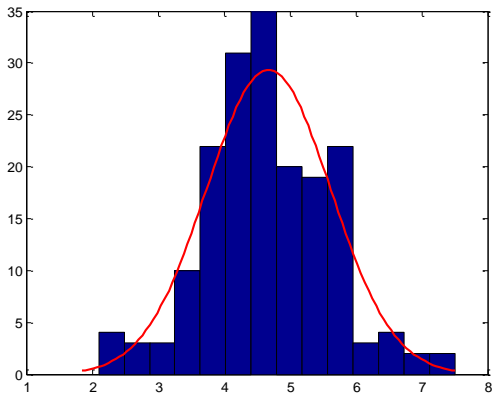


a)

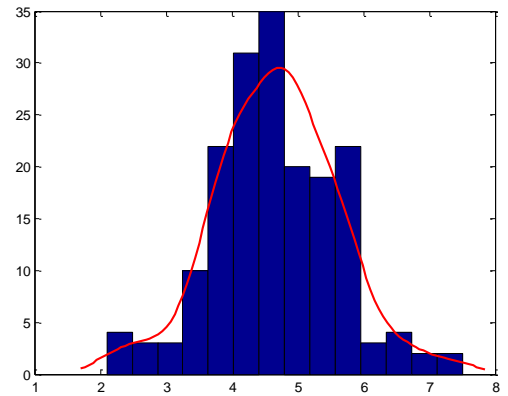


б)

Фиг. 7.3. *Нормално (Гаусово) а) и Кернел б) разпределение на входна променлива “Current taxes on income and wealth”*

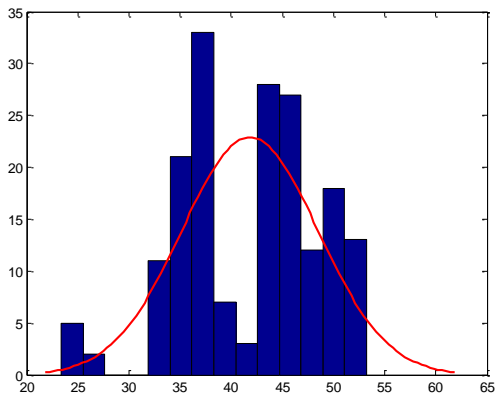


a)

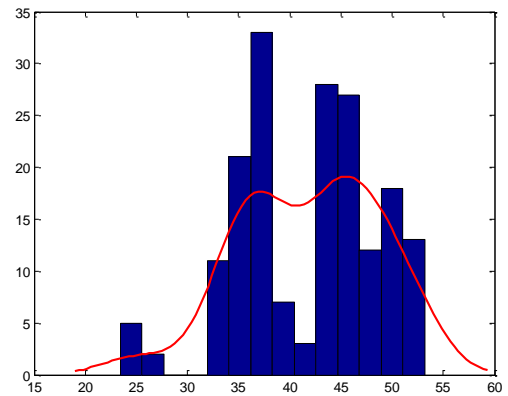


б)

Фиг. 7.4. *Нормално (Гаусово) а) и Кернел б) разпределение на входна променлива “Other current revenue, including sales”*

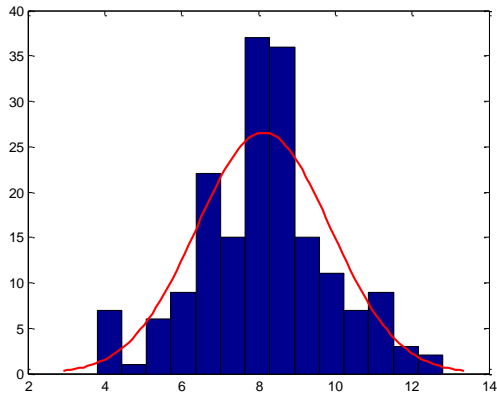


a)

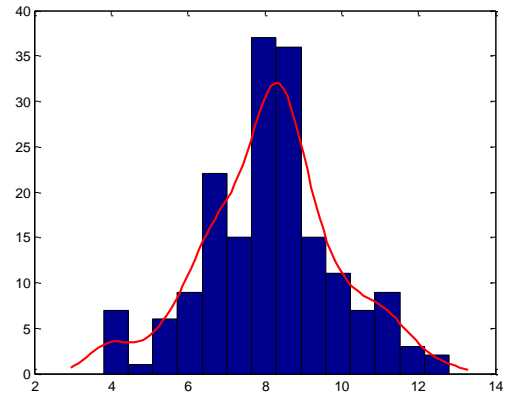


б)

Фиг. 7.5. *Нормално (Гаусово) а) и Кернел б) разпределение - “Total current revenue”*

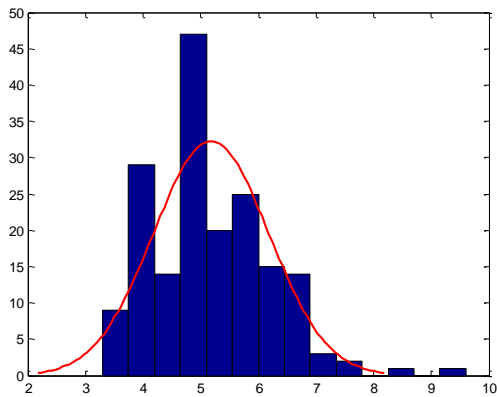


а)

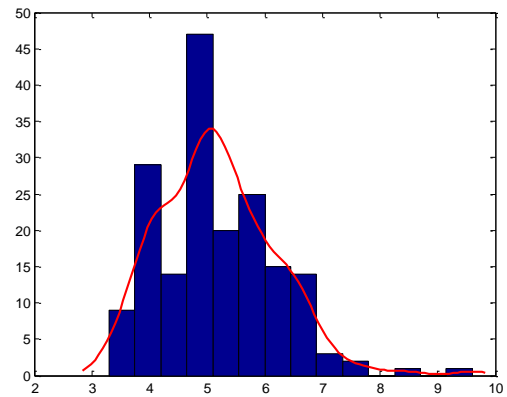


б)

Фиг. 7.6. *Нормално (Гаусово) а) и Кернел б) разпределение на входна променлива “Collective consumption”*

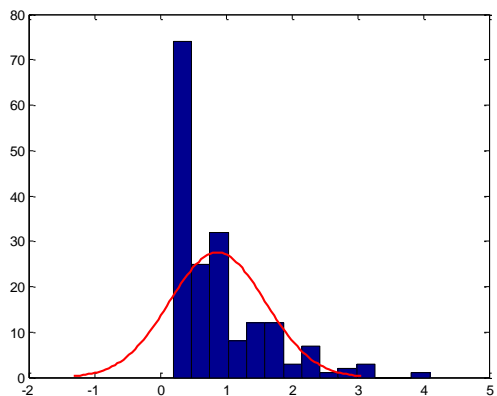


а)

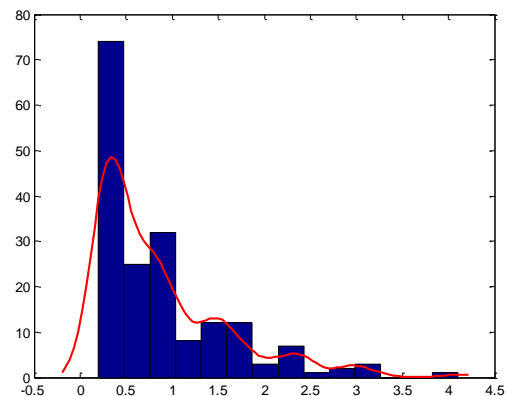


б)

Фиг. 7.7. *Нормално (Гаусово) а) и Кернел б) разпределение на входна променлива “Intermediate consumption”*



а)



б)

Фиг. 7.8. *Нормално (Гаусово) а) и Кернел б) разпределение - “Capital transfers received”*

- ❖ Да се анализира характерът на изменение на информационните еталони, отнасящи се за всеки изследван икономически показател, спрямо приложените разпределения, т.е. да се оцени степента на приближение на данните до теоретичните функции на плътност на вероятностите;
- ❖ Да се определи и обоснове изборът на тип разпределение, описващо в по-пълна степен анализиранияте опитни данни;
- ❖ Да се формират входни информационни набори, които да бъдат използвани в хода на обучение и синтез на класификатори на основата на Дискриминантен анализ и алгоритъма на Бейс във връзка със следните групи региони:
 - ✚ Group №1: Belgium и Germany;
 - ✚ Group №2: Greece и Malta;
 - ✚ Group №3: Italy и Luxenbourg;
 - ✚ Group №4: Spain и France;
 - ✚ Group №5: Ireland и Latvia.

7.3.2. Синтез на модели за определяне на географски райони на принадлежност при анализ на икономически показатели с приложение на Дискриминантен анализ

- ❖ Да се създадат линейни и квадратични типове дискриминантни класификационни модели и изследва качеството на класификация съобразно критерии за „грешка“, „точност“ и „некоректни класификации“ на базата на подходите „ресубституция и крос-валидиране“. Да се приложи процентно разделяне на данните 75% към 25 % във връзка с използването на крос-валидиращите модели. Резултатите да бъдат представени съгласно таблица 7.1 и таблица 7.2;

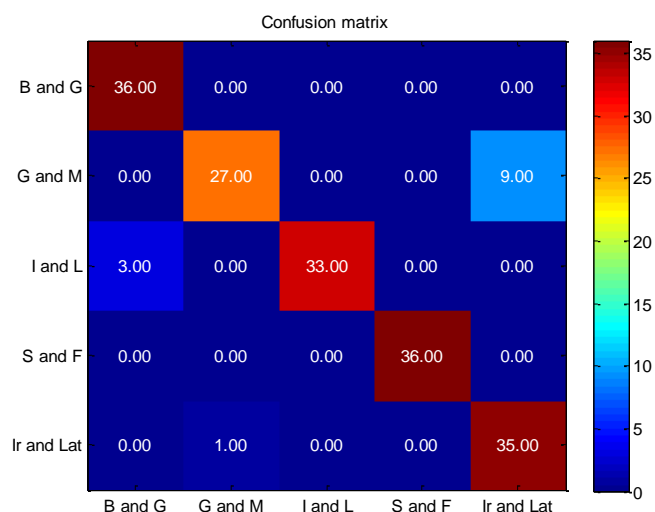
Таблица 7.1. Резултати от приложение на ресубституция при оценка на дискриминантни модели

Тип класификатор	Грешка	Точност, %	Некоректни класификации
linear	0.0722	92.78	13
diaglinear			
pseudolinear			
quadratic	0.0056	99.44	1
diagquadratic			
pseudoquadratic	0.0056	99.44	1

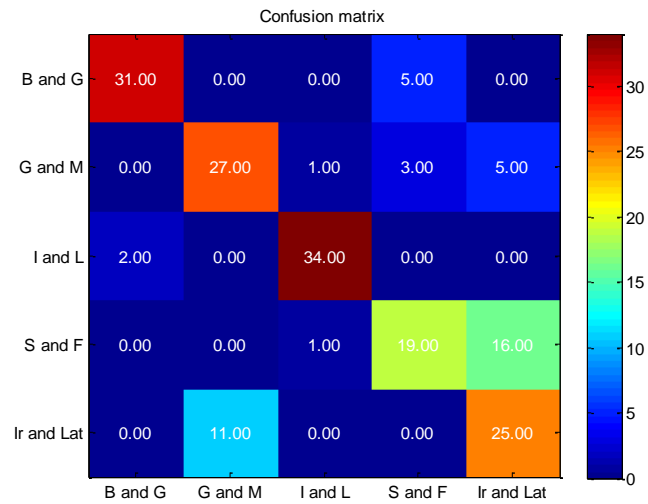
Таблица 7.2. Резултати от приложение на крос-валидиране при оценка на дискриминантни модели

Тип класификатор	Грешка	Точност, %	Некоректни класификации
linear	0.0889	91.11	16
diaglinear			
pseudolinear			
quadratic	0.0389	96.11	7
diagquadratic			
pseudoquadratic	0.0389	96.11	7

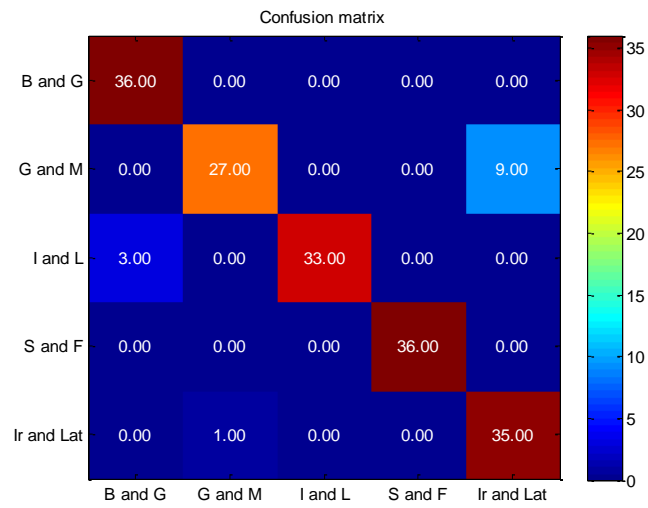
- ❖ Да се установят основни тенденции и селектират дискриминантни модели с най-добри индикатори за качество относно икономическата задачата на изследване;
- ❖ Да се генерират матрици на коректни и некоректни класификации съобразно подходите за оценка на качество на тестовите дискриминантни класификатори, показани на фиг. 7.9 и фиг. 7.10;
- ❖ Да се оцени и обобщи разпределението на еталоните с правилно и неправилно определена принадлежност по изходни групи по отношение на „ресубституция и „крос-валидиране“;



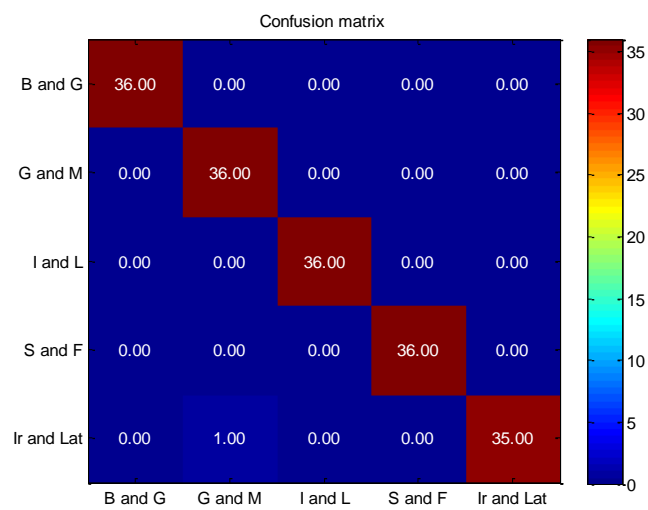
a)



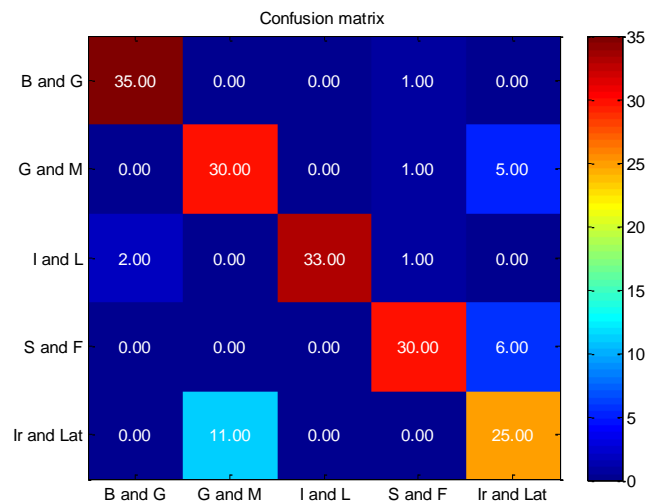
a)



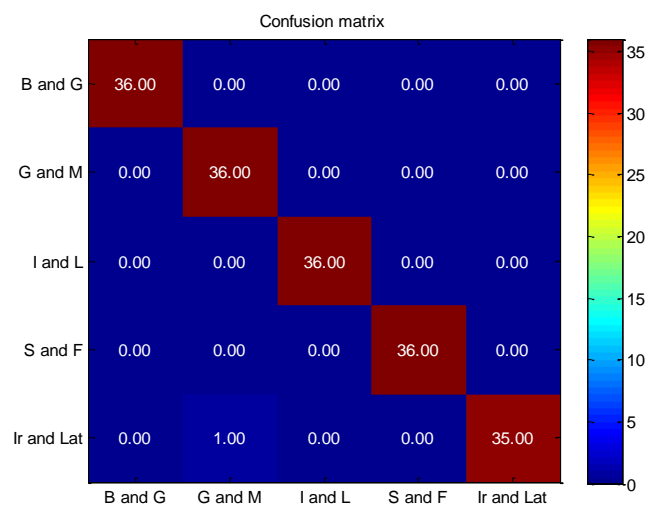
b)



c)

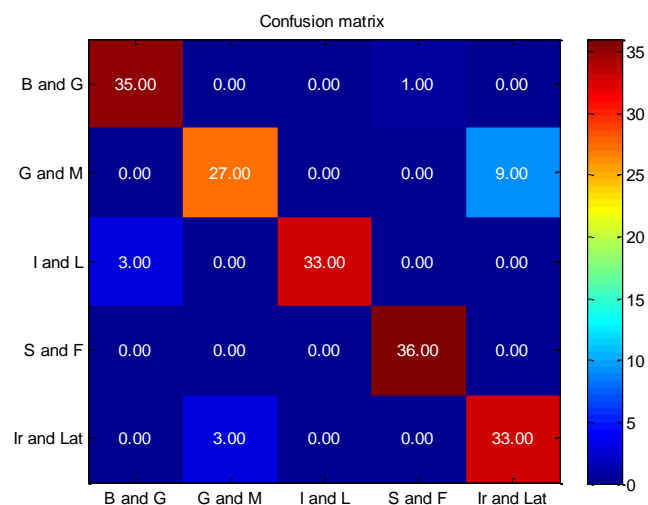


д)

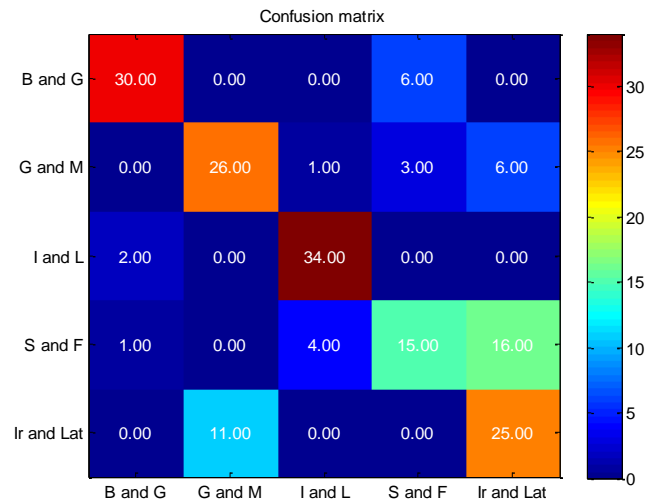


е)

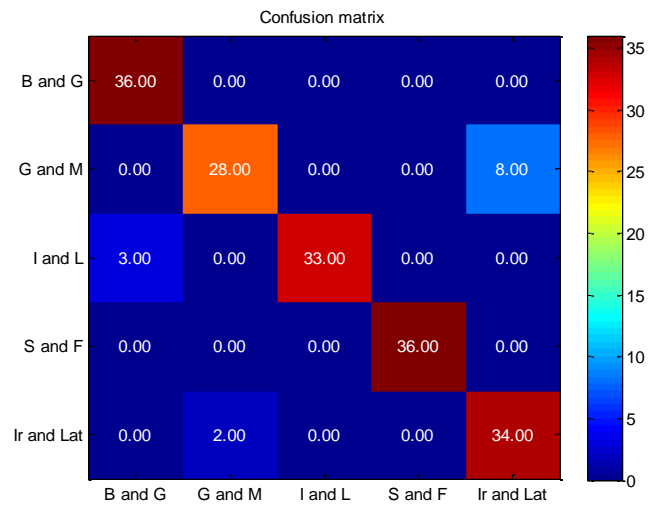
Фиг. 7.9. Класификационни матрици от ресубституция при а) linear, б) diaglinear, в) pseudolinear, г) quadratic, д) diagquadratic и е) pseudoquadratic



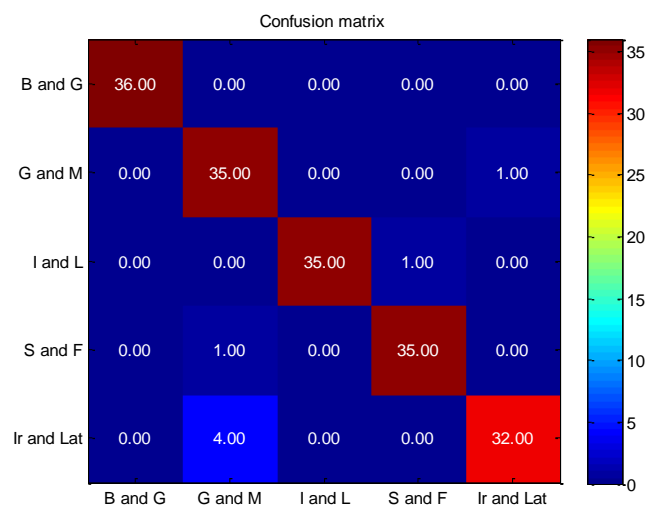
а)



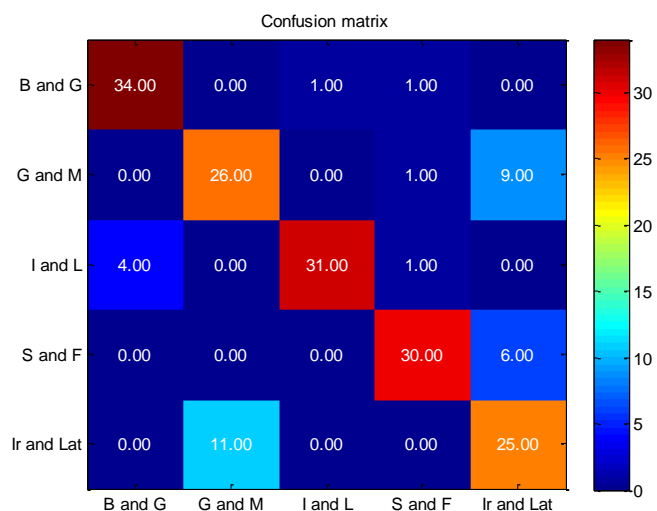
a)



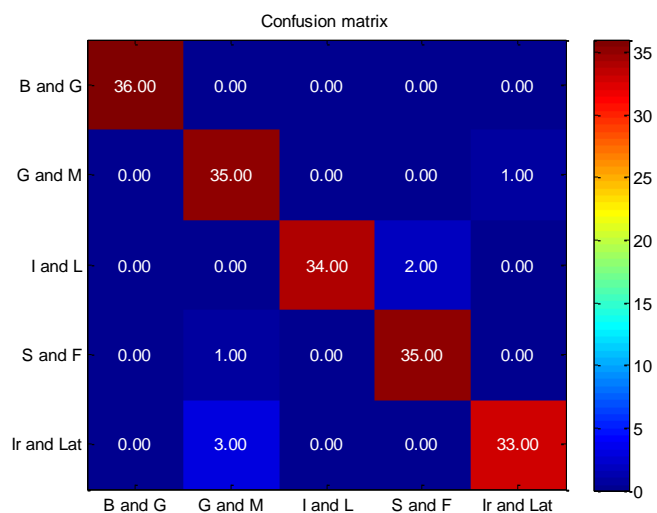
b)



c)



д)



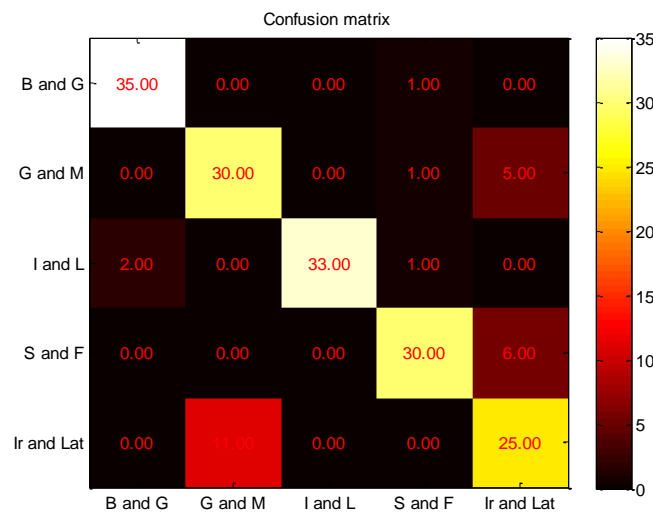
е)

Фиг. 7.10. Класификационни матрици от крос-валидиране при а) linear, б) diaglinear, в) pseudolinear, г) quadratic, д) diagquadratic и е) pseudoquadratic

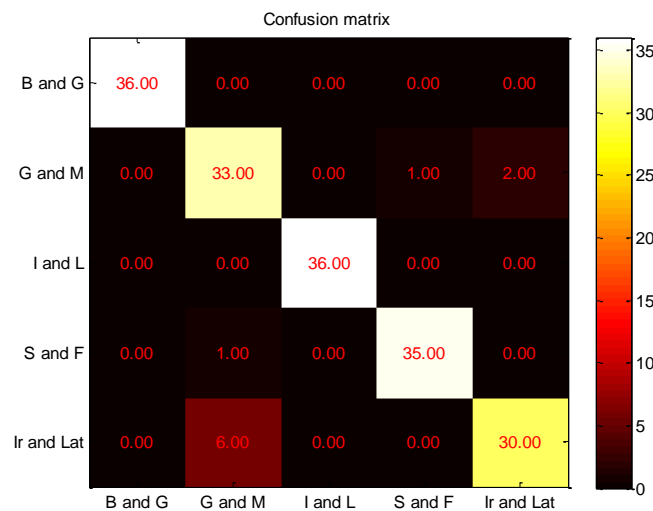
Таблица 7.3. Очаквани точности при обработка на нови данни с помощта на дискриминантни модели

Тип класификатор	Точност, %
linear	91.945
diaglinear	
pseudolinear	
quadratic	97.775
diagquadratic	
pseudoquadratic	97.775

- ❖ Да се проведат последващи аналогични на предходните процедури по оценка на създадените Бейсови класификатори на базата на подходи „ресубституция“ и „крос-валидиране“ чрез прилагане на разработени програмни скриптове в командния прозорец на продукта MATLAB, представени на фиг. 7.11 и фиг. 7.12;
- ❖ Да се построят матрици на коректните и некоректни класификации, отразяващи разпределянето на тестовите информационни еталони, по отношение на Бейсови класификатори при обучение с опитни данни при тяхното Гаусово и Кернел разпределение, дадени на фиг. 7.13;



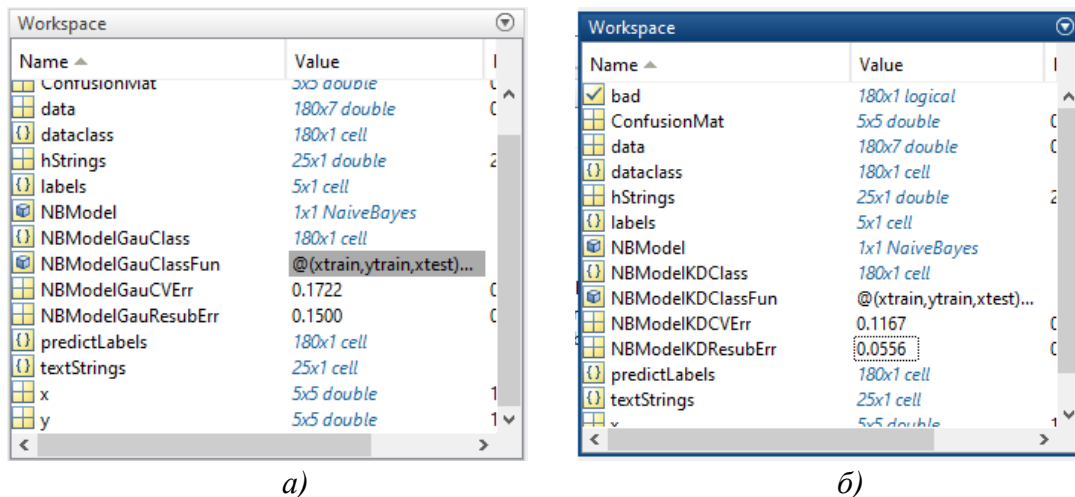
a)



б)

Фиг. 7.13. Класификационни матрици на модели по алгоритъма на Naïve Bayes за случая при а) Гаусово и б) Кернел разпределение на входните данни

- ❖ Да се направи сравнителен анализ на получените резултати и избере тип класификационен модел за целите на идентификация с приложение на алгоритъма на Naïve Bayes;



Фиг. 7.14. Променливи при изследване на модели по алгоритъма на Naive Bayes за случая при а) Гаусово и б) Кернел разпределение на входните данни

- ❖ Да се представят наборите от входни и изходни параметри при изследване и селекция на класификационни модели на основата на процедури по предварителна обработка на входните данни, свързана с описание на входните данни чрез Гаусова и Кернел функции на плътност на вероятността, показани на фиг. 7.14;
- ❖ Да се определи адекватността на използвания математически инструмент спрямо този на Дискриминантния анализ.

7.4. Контролни въпроси

1. Потвърждава ли се експериментално дефиницията за по-добрия описателен характер на входните данни при приложение на Кернел спрямо Гаусови функции?
2. Какви типове линейни и квадратични модели на базата на Дискриминантен анализ познавате?
3. Какво е характерно за прилаганите подходи за оценка на качеството на модели за разпознаване и класификация „ресубституция“ и „крос-валидиране“?
4. По какъв начин се интерпретира разпределянето на еталоните с правилна и некоректна групова принадлежност при анализ на класификационните матрици?

**РАЗПОЗНАВАНЕ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ
НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ НА БАЗАТА НА СТАТИСТИЧЕСКИ
МЕТОДИ ДЪРВО НА РЕШЕНИЯТА И К – НАЙ-БЛИЗКИ СЪСЕДИ
В СРЕДА НА ПРОДУКТИТЕ STATISTICA И MATLAB**

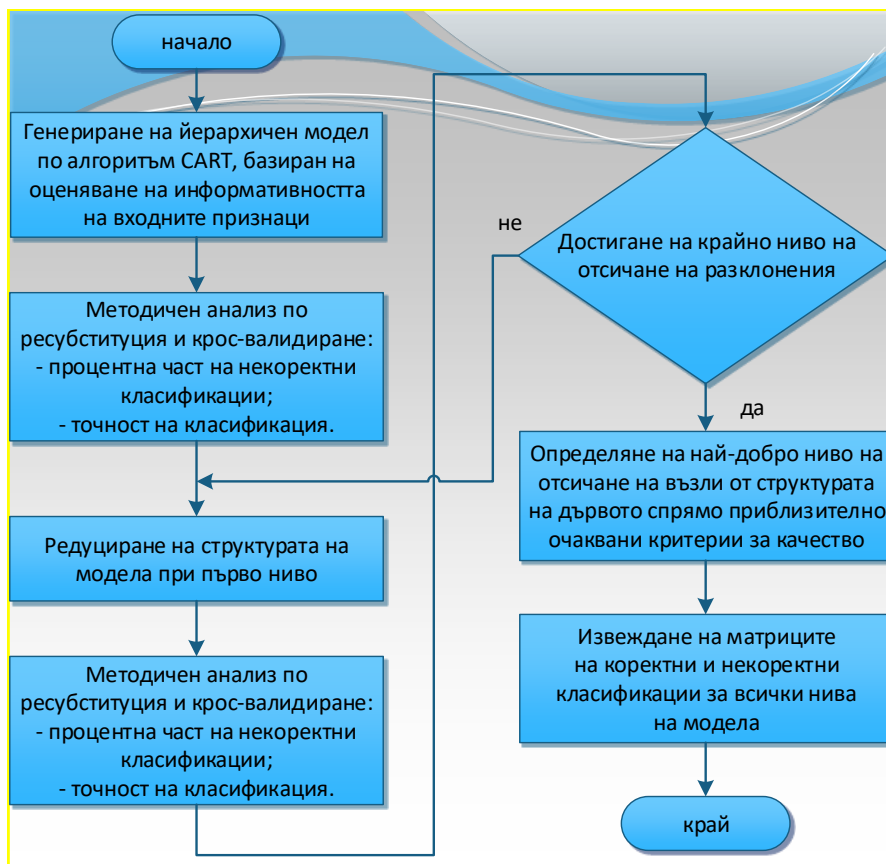
8.1. Цел на упражнението

Да се запознаят студентите с процедури при подбора и изследването на модели на основата на методите за машинно обучение Дърво на решенията (Decision Tree - DT) и k – най-близки съседни (k – Nearest Neighbors - k-NN) във връзка с разпознаване на постигнати икономически показатели по обособени региони чрез продуктите STATISTICA и MATLAB.

8.2. Теоретична постановка

Подход за синтез на модели с използване на метод Дърво на решенията

На фиг. 8.1 е представен блоков алгоритъм за идентификация на информационни групи по метода Дърво на решенията. Същността на подхода се свежда до построяване на йерархичен модел за многовариационен избор на решение, прилагайки CART алгоритъм.



Фиг. 8.1. Алгоритъм на базата на метода Дърво на решенията

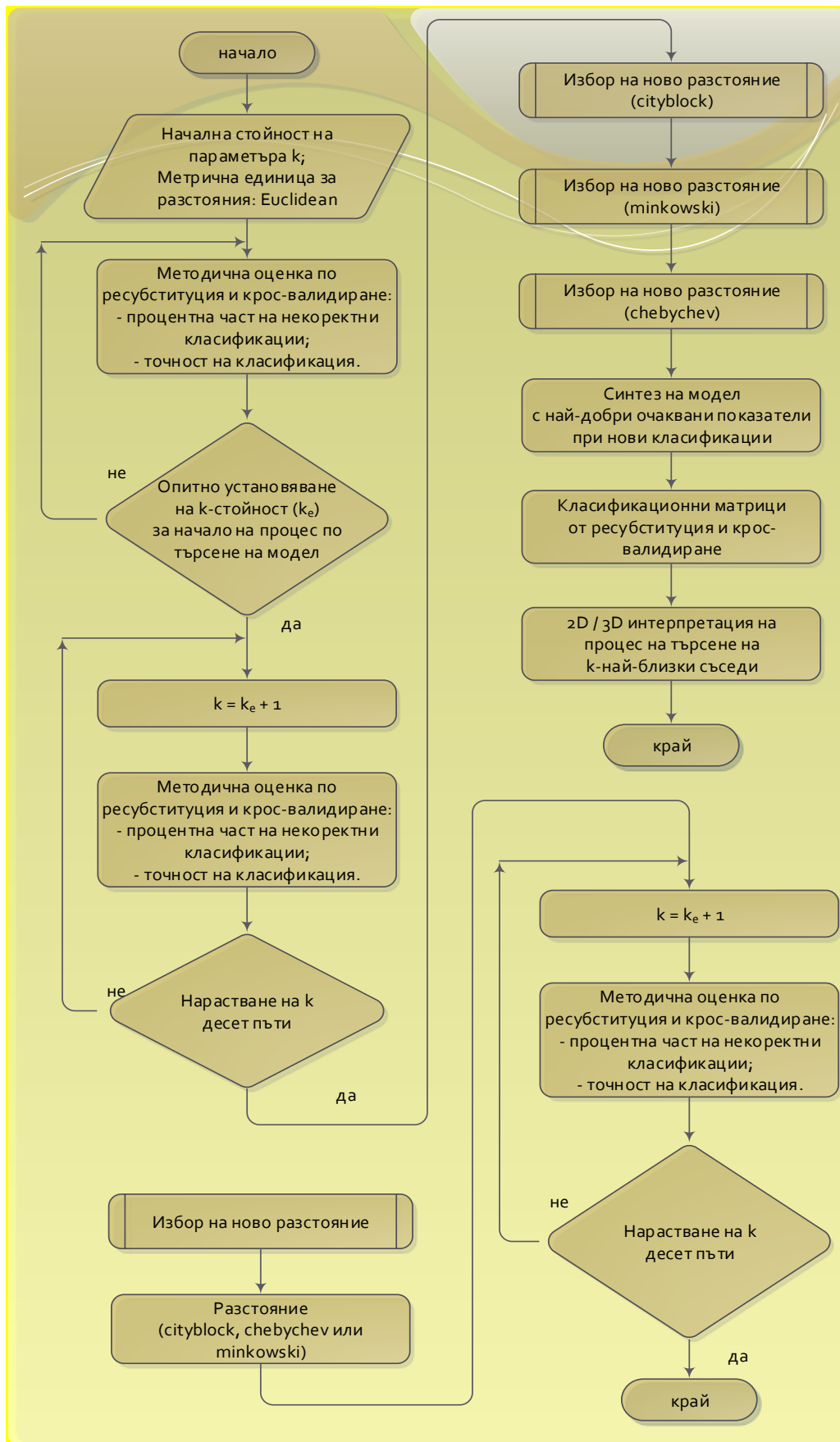
Анализира се качеството на генерирания модел, използвайки същите подходи по аналогия с приложението на метод k – най-близки съседи. По отношение на крос-валидация се използва разделяне на входния набор от данни на четири поднабора, съответстващо на 75% за обучение и 25% данни за тестване на крос-валидиращите модели. Извършва се премахване на възлови разклонения от класификационното дърво, след което отново оценка на качеството на класификация. Процедурата се повтаря до преминаване през всички нива, докато остане само наличен един възел. Броят на нивата се определя индивидуално за всеки конкретен модел. Определят се приблизително очаквани точности при обработка и анализ на новопостъпващи данни и синтезира най-адекватен модел за идентификация. Представят се матрици на коректни и некоректни класификации, илюстриращи разположението на участващите еталони по изходни групи при всички нива [50-54].

Подход за синтез на модели за разпознаване на основата на метода k – най-близки съседи

Синтезирането на класификационни модели по метода k – най-близки съседи се основава на прилагането на набор от метрични единици за разстояния като Евклидово, Cityblock (Manhattan), разстояния на Минковски и Чебишев. Стойността на параметъра k се определя опитно в процеса на изследване като се взема базово Евклидово разстояние (фиг. 8.2). Целта е да бъде намерен потенциален брой k -съседни точки, при които се наблюдават високи стойности на показателите процентна част на некоректни класификации и класификационна точност, резултат от оценка на качеството от техническите подходи ресубституция и крос-валидация. При намерена удовлетворяваща k -стойност се извършва пресмятане на указаните показатели за всяко опитно разстояние.

Селектиращият процес на краен идентификационен модел се базира на изчислените приблизително очаквани точности при идентификация на данни, които не са били използвани при синтез. Относно модела с най-добри качества се извеждат класификационни матрици по отношение на подходите за оценка. Също така в двумерно или тримерно дефинирано класификационно пространство се изследва синтезиращият модел при зададено разстояние и k -съседни точки при търсене на нови заявени еталони в околността на k [55-59].

Показан е и алгоритъм на подпроцес при селекция на метрично разстояние и анализ на поведението на целеви модели в установения диапазон от k – съседи (фиг. 8.2). Алгоритъмът е идентичен за всички изследвани разстояния. Особеност при дефиниране на зависимите предсказващи променливи (идентификационните групи) е, че тук не се използва числово кодиране, а категориално задаване на целевите изходни променливи. Други потенциални разстояния, които биха могли да бъдат използвани са Cosine, Mahalanobis, Correlation, Hamming, Jaccard, Seclidean, и Spearman.



Фиг. 8.2. Алгоритъм на базата на k -NN

8.3. Задачи за изпълнение

8.3.1. Определяне на принадлежността на група икономически показатели чрез статистически метод „Дърво на решенията“ в програмна среда STATISTICA

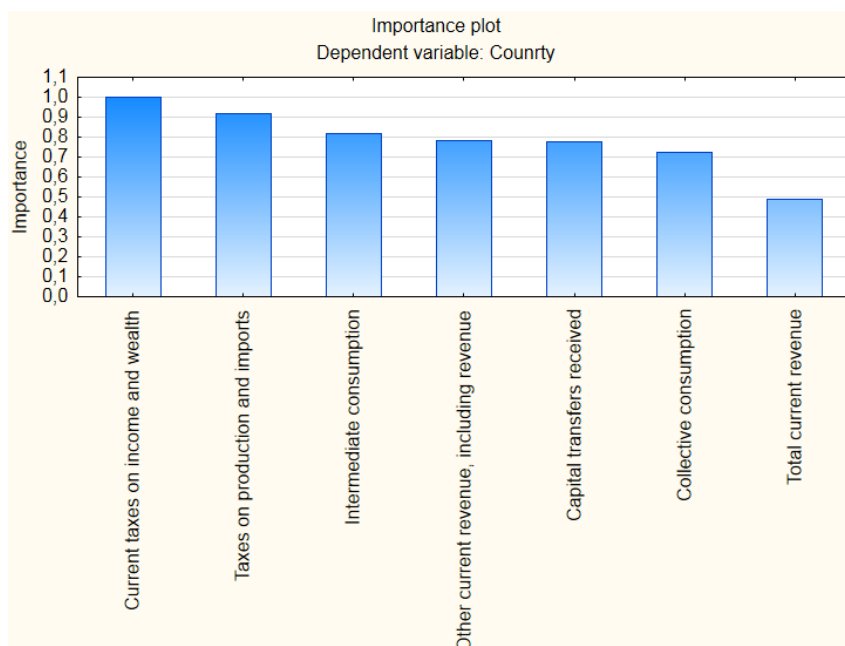
- ❖ Да се разгледа задачата за определяне на принадлежността на анализирани икономически показатели спрямо 10 заложенни страни, съответно Belgium, Germany, Ireland, Greece, Spain, France, Italy, Latvia, Luxembourg и Matla;
- ❖ Да се реализира импортиране на опитните данни в програмна среда STATISTICA, като посоченото на фиг. 8.3;

	1	2	3	4	5	6	7	8
	Taxes on	Current i	Other cu	Total cur	Collectiv	Intermed	Capital tra	Country
1	12,3	16	3,4	47,9	8,6	3,6	0,3	Belgium
2	13,1	16,7	3,5	48,9	8,1	3,7	0,4	Belgium
3	13	16,2	3,8	48,4	8,4	4	0,6	Belgium
4	12,9	15,9	3,7	48,1	8,2	3,7	0,6	Belgium
5	12,9	16,3	4	49,2	8,5	3,9	0,5	Belgium
6	13	15,1	4,2	48,8	9	4,3	0,4	Belgium
7	13,2	15,4	4,4	49,2	8,8	4,2	0,6	Belgium
8	13,3	15,9	4,7	50,4	8,8	4,2	0,6	Belgium
9	13,7	16,2	4,8	51,4	8,9	4,3	0,8	Belgium
10	13,6	16,7	4,8	52	8,8	4,2	1	Belgium
11	13,5	16,7	4,9	51,6	8,8	4,3	0,9	Belgium
12	13,2	16,3	4,6	50,5	8,4	4,1	0,8	Belgium
13	13,5	16	4,8	49,9	8,1	4,1	0,8	Belgium
14	13,4	16,6	4,8	50,4	8	4	0,8	Belgium
15	13,5	16,8	4,9	50,7	8	4,1	0,7	Belgium
16	13,6	15,7	4,8	49,7	8	4,1	0,6	Belgium
17	13,1	15,8	4,9	49,9	9,2	4,7	0,8	Belgium
18	13,3	15,8	4,8	49,1	8,5	4,3	0,8	Belgium
19	10,4	11,9	4,7	45,2	7,9	3,9	0,5	Germany
20	10,8	12,7	4,1	45,8	7,7	3,8	0,4	Germany

Фиг. 8.3. Входни икономически показатели и целеви класови – държави (региони) в продукта STATISTICA

Predictor importance 3 (datacountry)	
Dependent variable: Country	
Options: Categorical response, Tree number 3	
Variable rank	Importance
Current taxes on income and wealth	100 1,000000
Taxes on production and imports	92 0,919279
Intermediate consumption	82 0,818663
Other current revenue, including revenue	78 0,781952
Capital transfers received	78 0,779374
Collective consumption	72 0,722279
Total current revenue	49 0,487168

a)



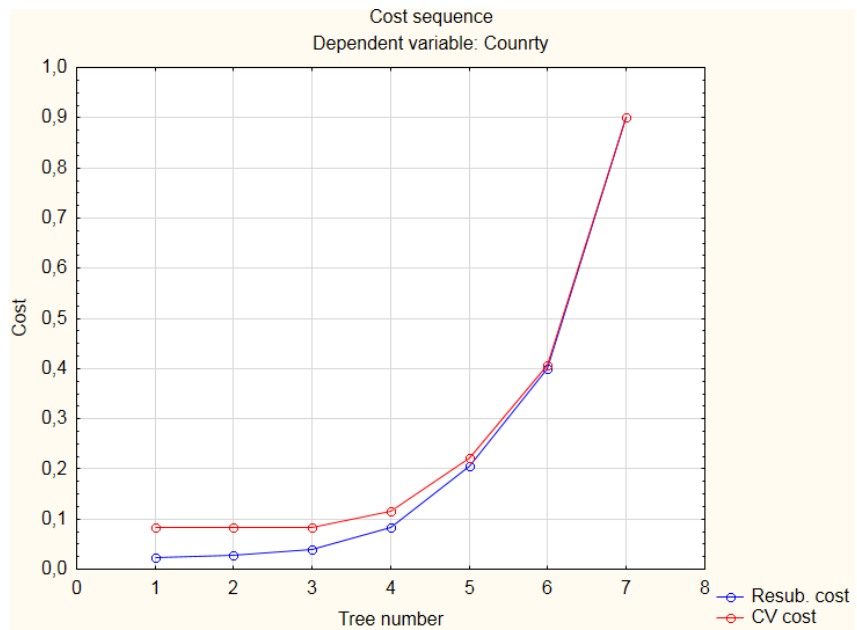
б)

Фиг. 8.4. Числова а) и графична б) интерпретация на резултатите при определяне на значимостта на входните променливи при изследването на модел по метода „Дърво на решенията“

- ❖ Преди построяване на първоначален модел по метода да се направи оценка на значимостта на входните информативни признаци като се представят графична числова интерпретация на процеса (фиг. 8.4). Да се извърши оценка на наблюдаваните нива на значимост;
- ❖ Да се създадат и анализира качеството на модели по метода Дърво на решенията чрез приложение на „ресубституция“ и „крос-валидиране“ спрямо експериментални нива на премахване на възлови разклонения от структурата на генерирания базов модел „Tree 1“. Да се дадат числови и графични резултати от указаните действия (фиг. 8.5). Грешките от подходите за оценка на качеството варират в интервала от 0.0000 или еквивалент на точност 100.0 % до единица, съответстваща на точност 0.0 %;

Tree sequence (datacountry)					
Dependent variable: Country					
Optimal tree denoted by *					
	Terminal nodes	CV cost	CV std. error	Resubstitution cost	Node complexity
Tree 1	14	0,083333	0,020601	0,022222	0,000000
Tree 2	13	0,083333	0,020601	0,027778	0,005556
*Tree 3	12	0,083333	0,020601	0,038889	0,011111
Tree 4	10	0,116667	0,023928	0,083333	0,022222
Tree 5	8	0,222222	0,030987	0,205556	0,061111
Tree 6	6	0,405556	0,036597	0,400000	0,092222
Tree 7	1	0,900000	0,022361	0,900000	0,100000

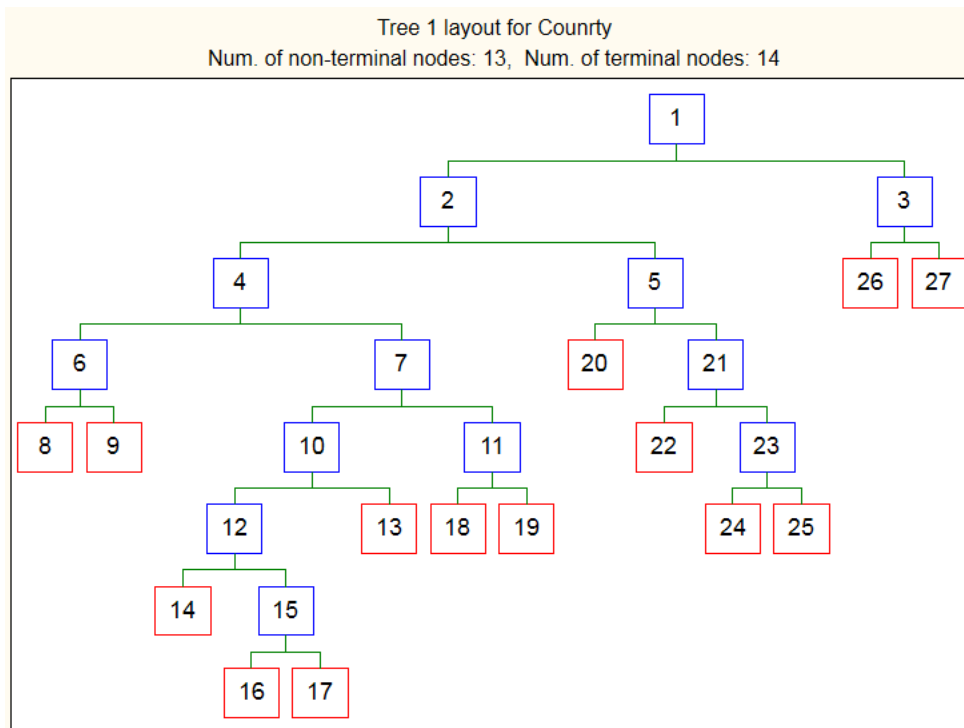
а)



б)

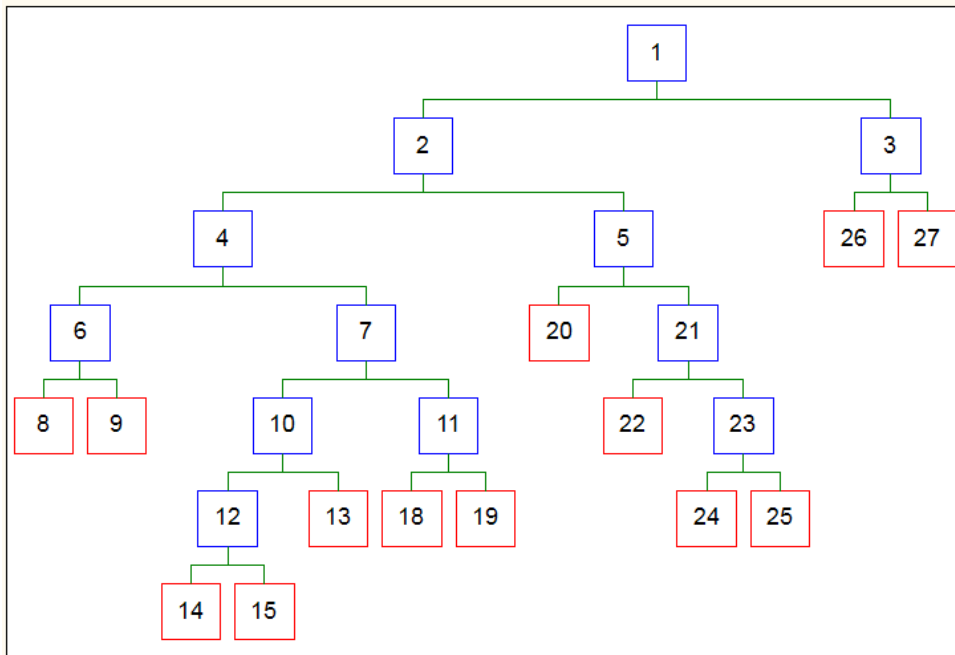
Фиг. 8.5. Числова а) и графична б) интерпретация на резултатите
Оценка на качеството на структурите по метода „Дърво на решенията“

- ❖ Да се представят и анализира видът на архитектурите за многовариантен избор на решение от “Tree 1” и “Tree 7”, показани на фиг. 8.6;



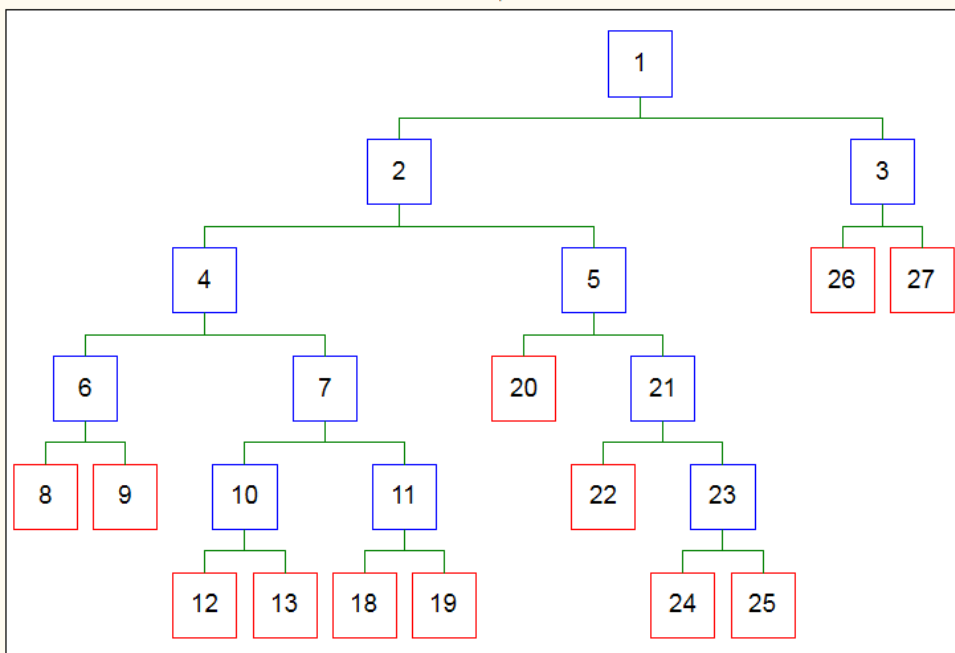
а)

Tree 2 layout for County
 Num. of non-terminal nodes: 12, Num. of terminal nodes: 13

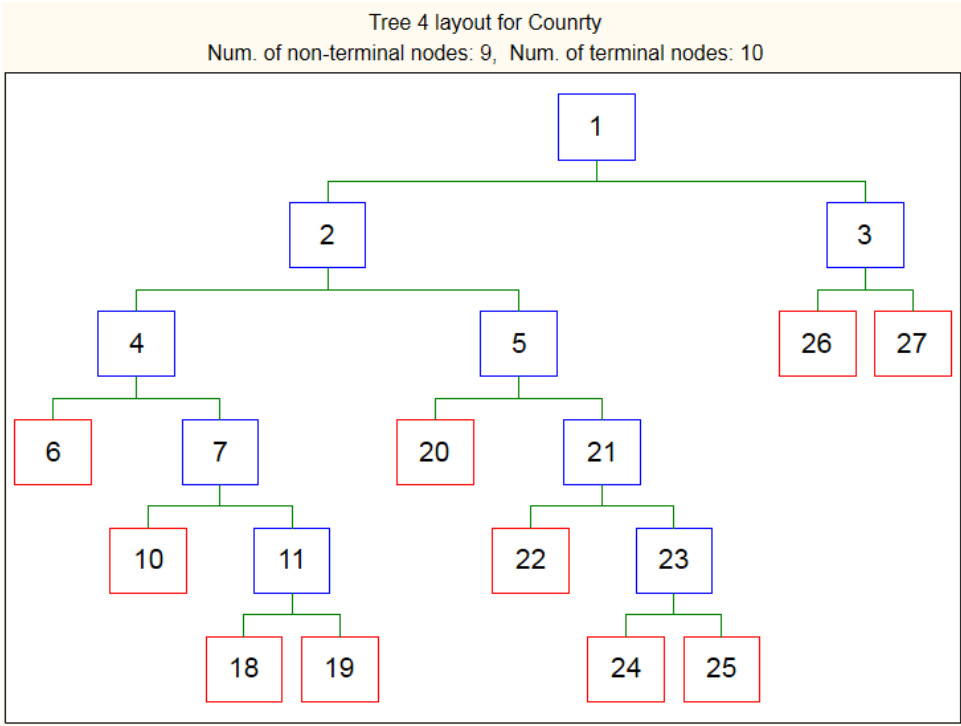


b)

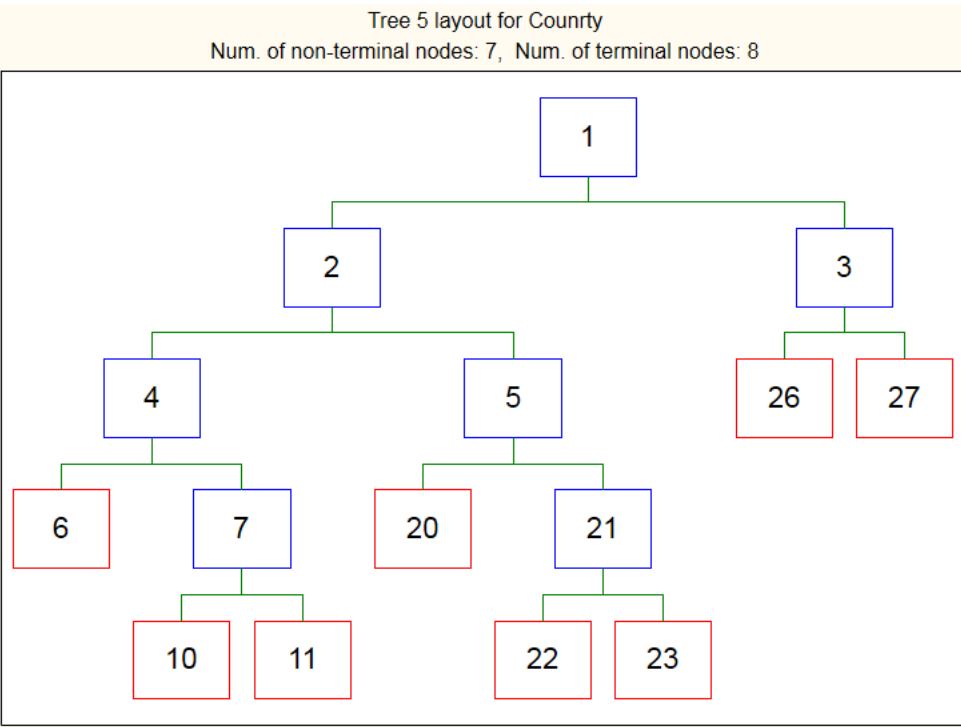
Tree 3 layout for County
 Num. of non-terminal nodes: 11, Num. of terminal nodes: 12



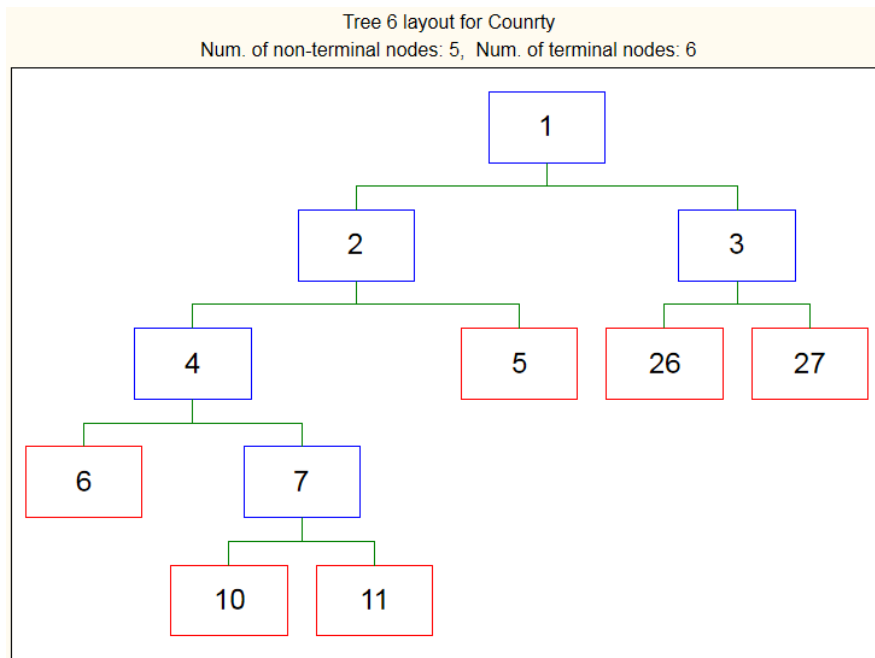
b)



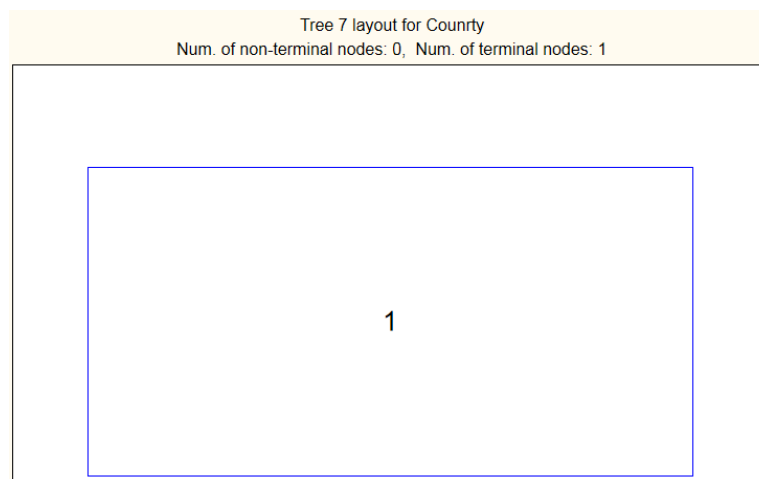
e)



d)



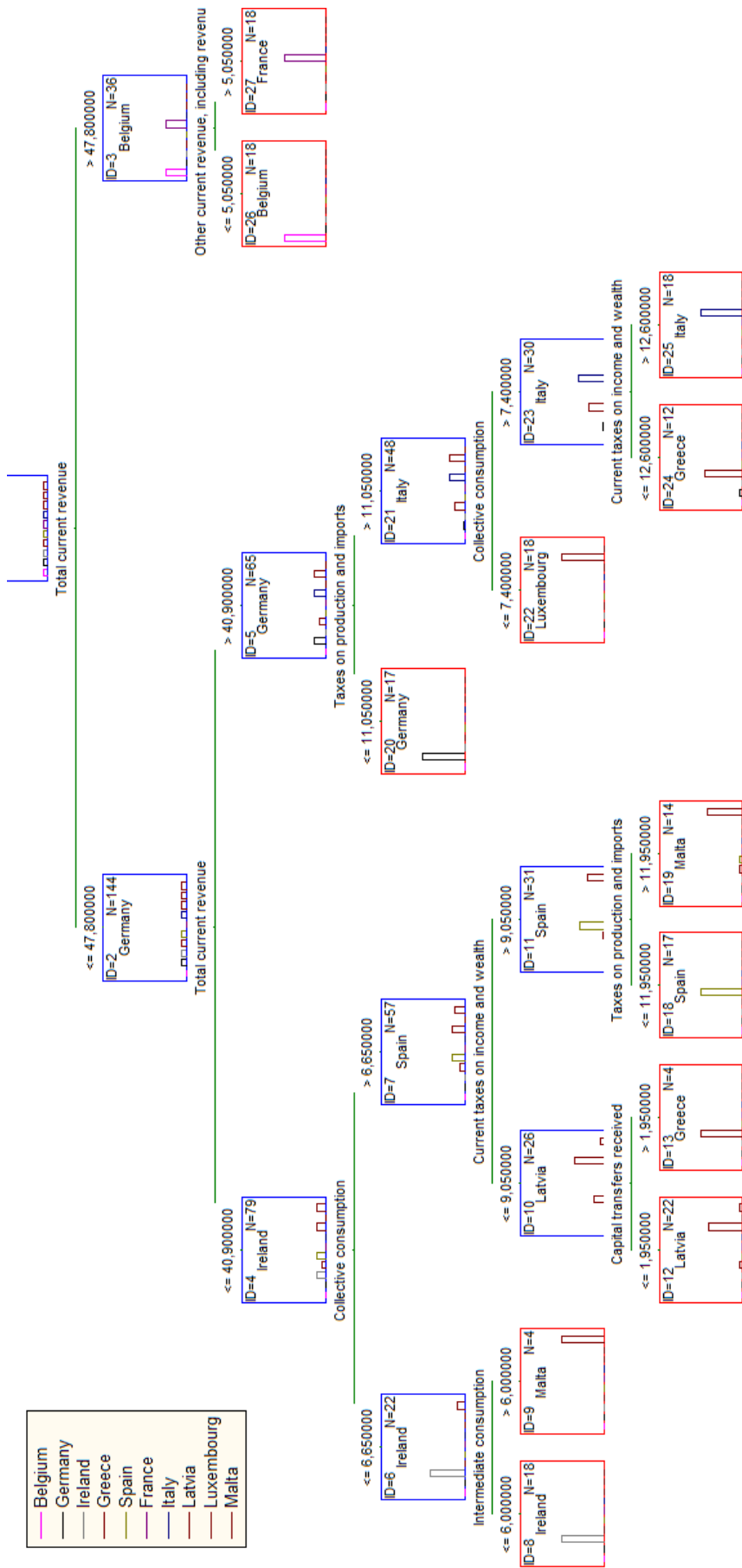
е)



ж)

Фиг. 8.6. Модели а) Tree 1, б) Tree 2, в) Tree 3, г) Tree 4, д) Tree 5, е) Tree 6 и ж) Tree 7 по метода „Дърво на решенията“ – блокова интерпретация

- ❖ Да се направи оценка на нивата на грешката, респективно вариациите на точността, от използваните подходи за оценка на качеството и избере оптимален модел относно целите на задачата за разпознаване и класификация. За конкретния случай с оглед на изискването за оптималност, състоящо се в поддържане на съотношение „гарантирана по-голяма точност“ спрямо модел с „възможно по-редуциран брой структурни възли“, може да бъде селектиран “Tree 2”. Да се генерира графична интерпретация на класификатора и да се онагледят разпределенията на еталони от съответни класове за всеки отделен възел от архитектурата на модела, показани на фиг. 8.7;



a)

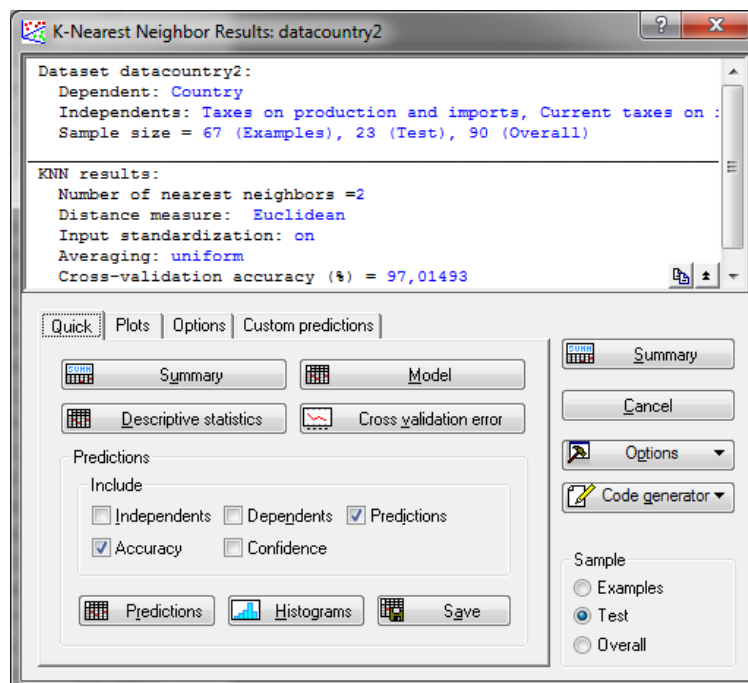
Result of terminal nodes 3 (datacountry)										
Dependent variable: Country										
Options: Categorical response, Tree number 3										
Node #	Class Belgium	Class Germany	Class Ireland	Class Greece	Class Spain	Class France	Class Italy	Class Latvia	Class Luxembourg	Class Malta
8	0	0	18	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	4
12	0	0	0	2	0	0	0	18	0	2
13	0	0	0	4	0	0	0	0	0	0
18	0	0	0	0	17	0	0	0	0	0
19	0	0	0	1	1	0	0	0	0	12
20	0	17	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	18	0
24	0	1	0	11	0	0	0	0	0	0
25	0	0	0	0	0	0	18	0	0	0
26	18	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	18	0	0	0	0

б)

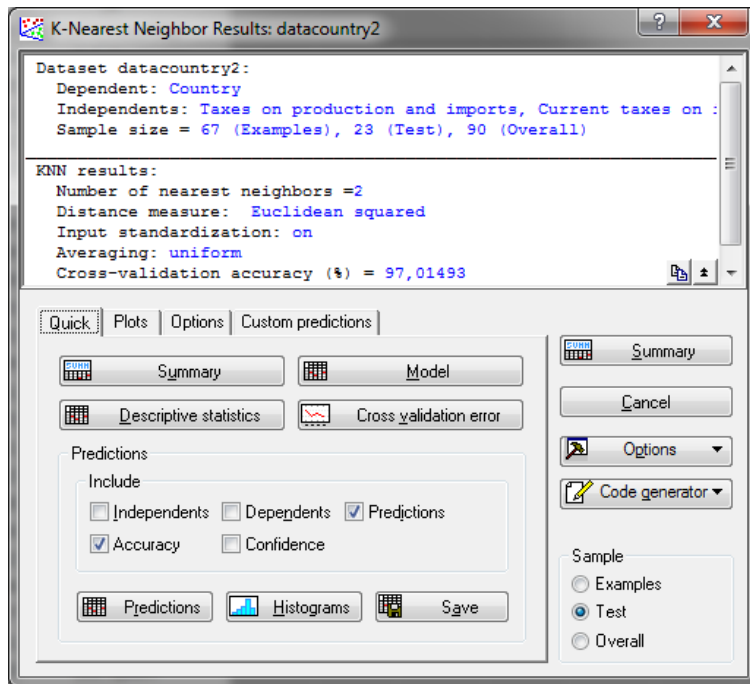
Фиг. 8.7. Архитектура на оптимален модел по метода „Дърво на решенията“ Tree 2 а) и разпределение на опитните еталони към всеки клас по структурни възлови разклонения

8.3.2. Определяне на принадлежността на група икономически показатели с помощта на статистически метод „Дърво на решенията“ чрез софтуера STATISTICA

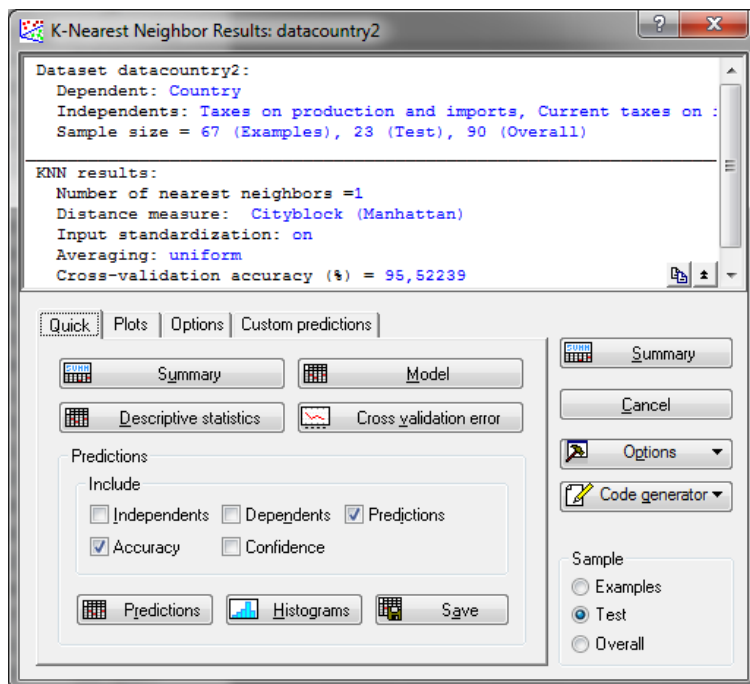
- ❖ Да се разгледа възможността за идентификация на принадлежността на обработвани показатели във връзка със следните региони Germany, Spain, France, Italy и Luxenbourg;



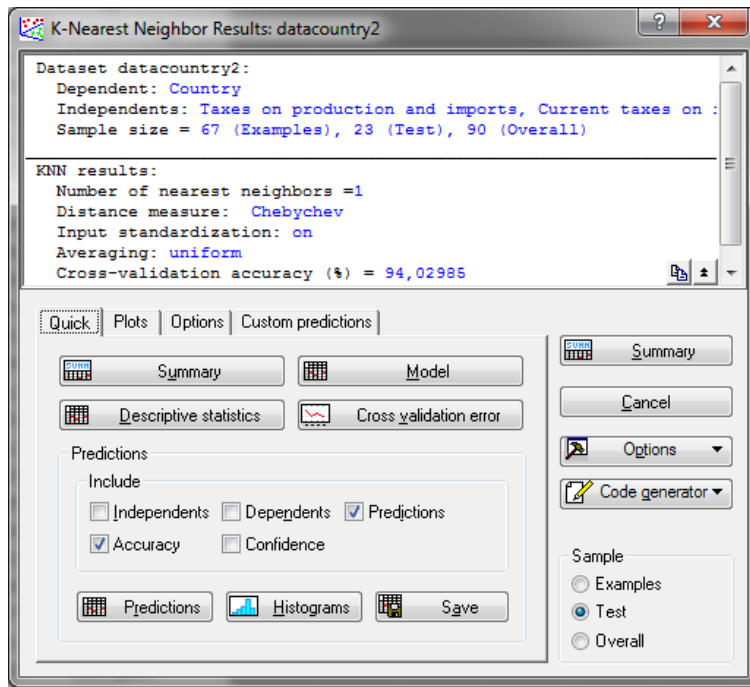
а)



b)

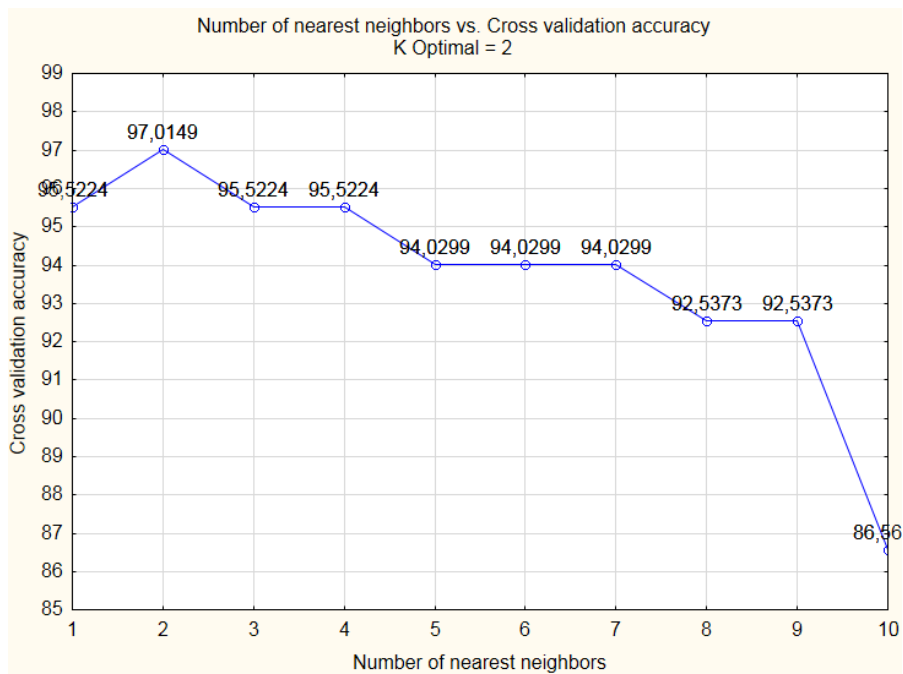


b)

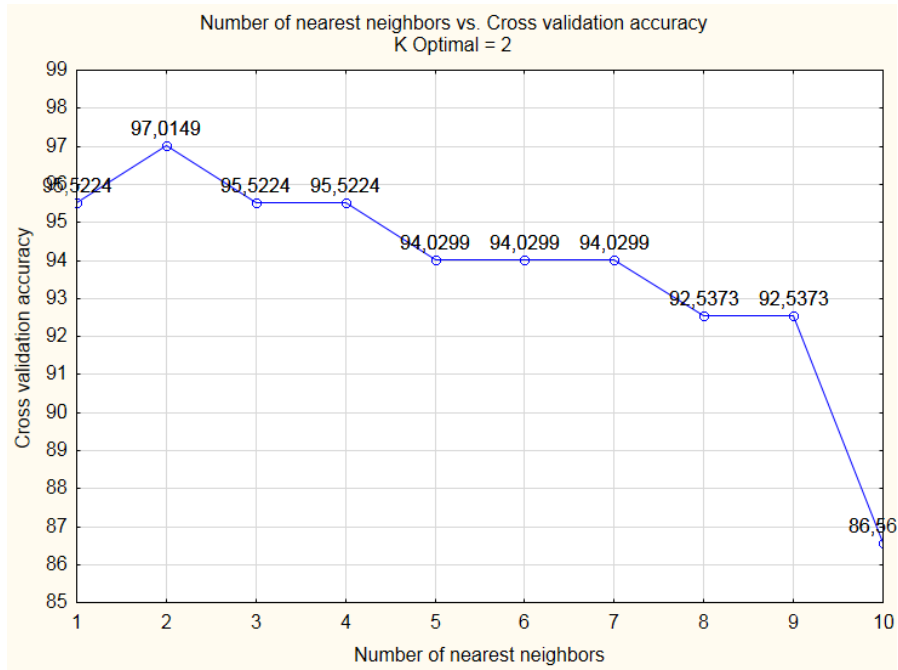


2)

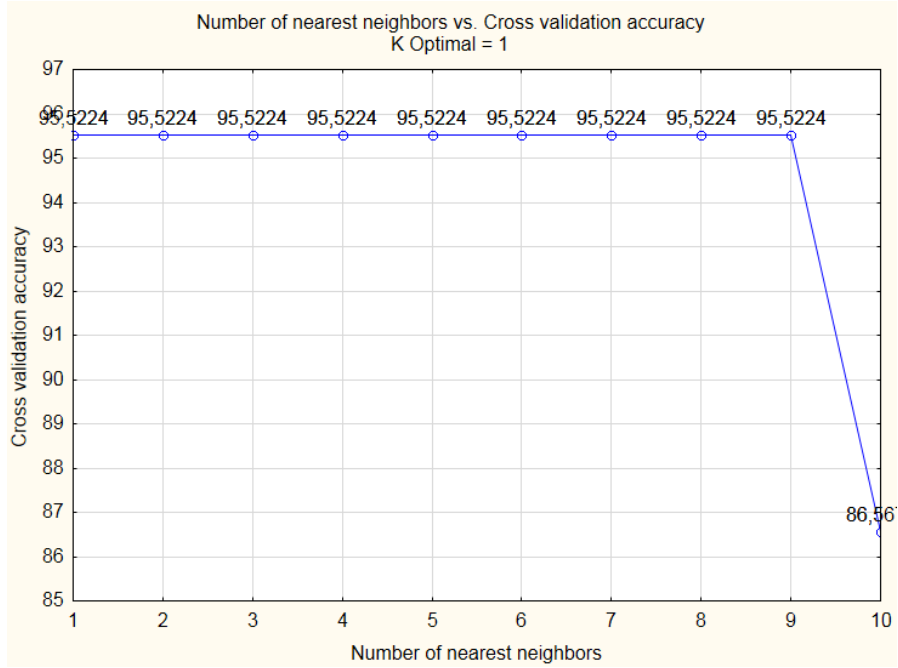
Фиг. 8.8. Оценка на качеството при модели по метода „ k – най-близки съседни“ при а) Eucliden, б) Euclidean squared, в) Cityblock и г) Chebychev



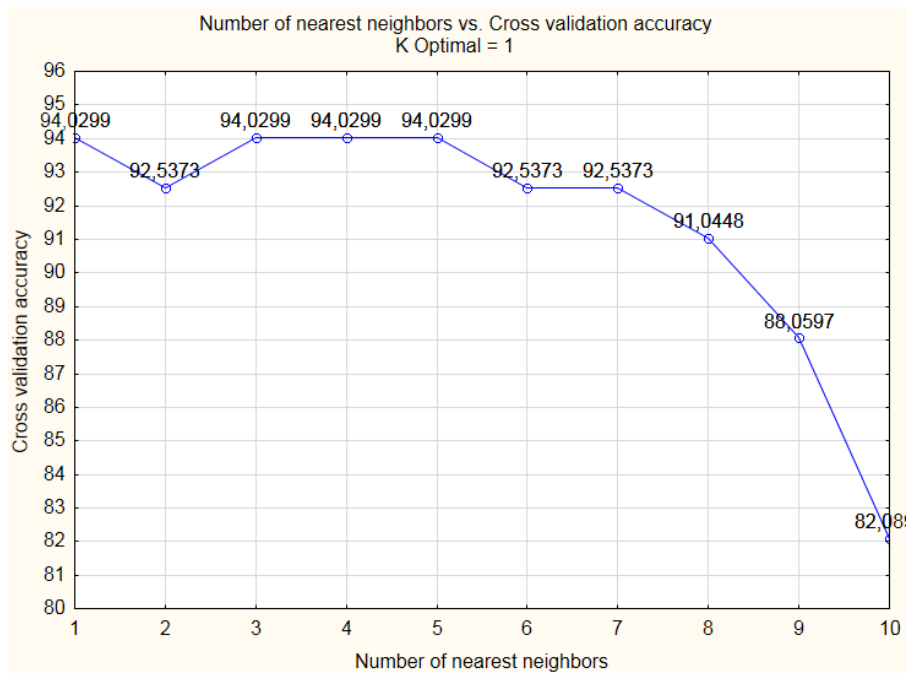
а)



b)



b)



з)

Фиг. 8.9. Изменение на точността от крос-валидация при изследването на модели с метрично разстояние а) Eucliden, б) Euclidean squared, в) Cityblock и г) Chebychev

- ❖ Да се проведат последователни изследвания, свързани с определянето на оптимални k-NN модели при фиксиране на различни метрични единици за разстояние, респективно “Euclidean”, “Euclidean squared”, “Cityblock” и “Chebychev”, т.е. намиране на оптимална стойност на параметъра k, за която е налице минимална грешка / максимална класификационна точност. Изследванията да се извършат съобразно максимално количествено ограничение за $k = 10$ с приложение на „крос-валидация“ – фиг. 8.8;
- ❖ Да се онагледи изменението на точността от „крос-валидиране“ в рамките на нарастване на параметъра k от 1 до 10, дадено на фиг. 8.9, и анализира поведението на всеки модел при съответно метрично разстояние.

8.3.3. Оценяване и селектиране на модели за идентификация на географски региони на принадлежност във връзка с постигнати икономически показатели на основата на статистически метод k-NN в MATLAB

- ❖ Да се разгледа задачата за обучение и оценка на икономически показатели по обособени групи региони:
 - ✚ Група №1: Belgium и Germany;
 - ✚ Група №2: Greece и Malta;
 - ✚ Група №3: Italy и Luxenbourg;
 - ✚ Група №4: Spain и France;
 - ✚ Група №5: Ireland и Latvia.

- ❖ Да се изследват и анализират количествените изменения на точността на създадени k -NN модели за k от 2 до 10 при залагане на метрични разстояния „Euclidean”, „Cityblock”, „Minkowski” и „Chebychev”, като резултатите бъдат обобщени съгласно изложените таблици от 9.1 до 9.4;

Таблица 9.1. Резултати при изследване на k -NN модели с разстояние „Euclidean“

k параметър	Точност при ресубституция, %	Точност при крос- валидиране, %	Точност при анализ на нови данни, %
2	99.44	97.78	98.610
3			
4			
5			
6			
7			
8			
9	97.78		94.44
10			

Таблица 9.2. Резултати при изследване на k -NN модели с разстояние „Cityblock“

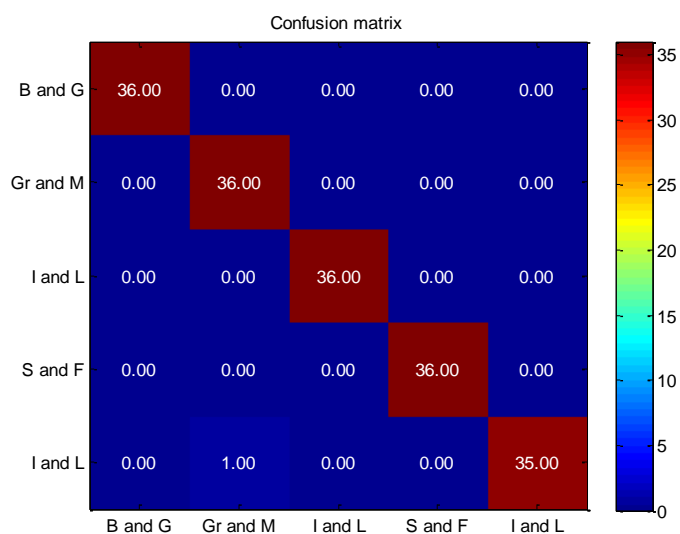
k параметър	Точност при ресубституция, %	Точност при крос- валидиране, %	Точност при анализ на нови данни, %
2	99.44	98.33	98.885
3			
4			
5			
6			
7			
8			
9			
10	97.78	95.00	96.390

Таблица 9.3. Резултати при изследване на k -NN модели с разстояние „Minkowski“

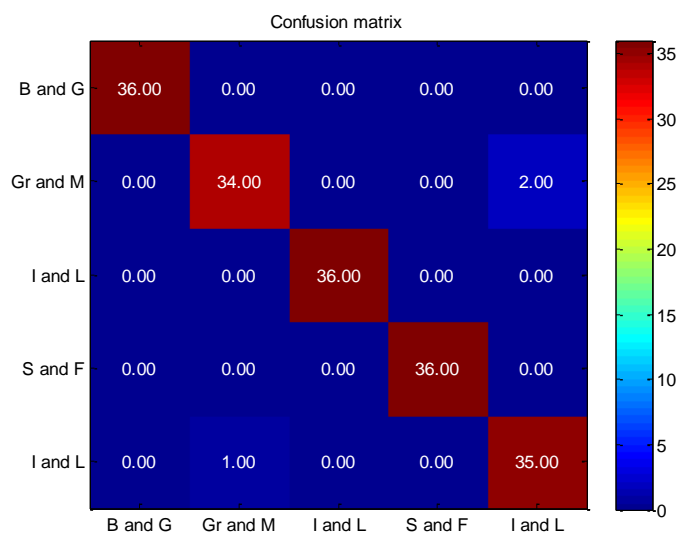
k параметър	Точност при ресубституция, %	Точност при крос- валидиране, %	Точност при анализ на нови данни, %	
2	99.44	97.78	98.610	
3				
4				
5				
6				
7				
8	97.22		95.56	96.390
9				
10				

Таблица 9.4. Резултати при изследване на k -NN модели с разстояние „Chebyshev“

к параметър	Точност при ресубституция, %	Точност при крос-валидиране, %	Точност при анализ на нови данни, %
2	98.89	96.11	97.945
3			
4			
5			
6		93.33	95.000
7			
8			
9			
10			



а)



б)

Фиг. 8.10. Матрици на коректни и некоректни класификации относно идентификация на икономически географски региони при а) ресубституция и б) крос-валидиране

- ❖ Да се селектира оптимална стойност на параметъра k и респективно тип метрична единица за разстояние с най-добри индикации за качество при анализ на нови данни. В случая може да бъде избран класификационен модел при разстояние “Cityblock” за $k = 2$;
- ❖ Да се построят и анализира видът на класификационните матрици от „ресубституция“ и „крос-валидация“ за синтезирания модел, показани на фиг. 8.10;

Name	Value	Min	Max
classes	5x1 cell		
confmat_cross	5x5 double	0	36
confmat_res	5x5 double	0	36
cvmdl	1x1 classreg.learning...		
data	180x7 double	0.2000	53.2000
dataclass	180x1 cell		
hStrings	25x1 double	203.00...	227.00...
kloss	0.0167	0.0167	0.0167
KNNModel	1x1 ClassificationKNN		
rloss	0.0056	0.0056	0.0056
textStrings	25x1 cell		
x	5x5 double	1	5
y	5x5 double	1	5

Фиг. 8.11. Променливи във връзка със селектирания k -NN класификатор за определяне на принадлежността на икономически показатели към установени географски региони

- ❖ Да се представят променливите в хода на изследванията в MATLAB по статистически метод k – най-близки съседи, дадени на фиг. 8.11.

8.4. Контролни въпроси

1. На каква процедура се основава изборът на оптимален модел по метода Дърво на решенията при манипулации с данни във връзка със задачи за класификация и регресионен анализ?
2. Какви типове метрични единици за разстояние могат да бъдат задавани в процесите на обучение и селекция на класификационни модели по метода k -NN?
3. Кои подходи се използват за оценка на качествените параметри и ефективността на класификационните модели на основата на двата статистически метода?
4. Какви са приликите и различията при дейности по метода k – най-близки съседи от гледна точка на използваните програмни продукти относно тяхната приложимост?
5. С какви подходи за предварителна обработка на данни по отношение на извличане на характеристики от информационни масиви могат да бъдат съчетани способите на методите за машинно обучение DT и k -NN?

МОДУЛ III

ПРОГНОЗЕН АНАЛИЗ НА ФИНАНСОВИ ДАННИ ЧРЕЗ ИЗКУСТВЕН ИНТЕЛЕКТ, ПРИЛОЖНА СТАТИСТИКА И МАШИННО ОБУЧЕНИЕ

Лабораторно упражнение №9

ИЗКУСТВЕН ИНТЕЛЕКТ КАТО ИНСТРУМЕНТ ЗА ПРОГНОЗЕН АНАЛИЗ ВЪВ ВРЪЗКА С ИНВЕСТИЦИОННИ ПОКАЗАТЕЛИ В СРЕДА MATLAB

9.1. Цел на упражнението

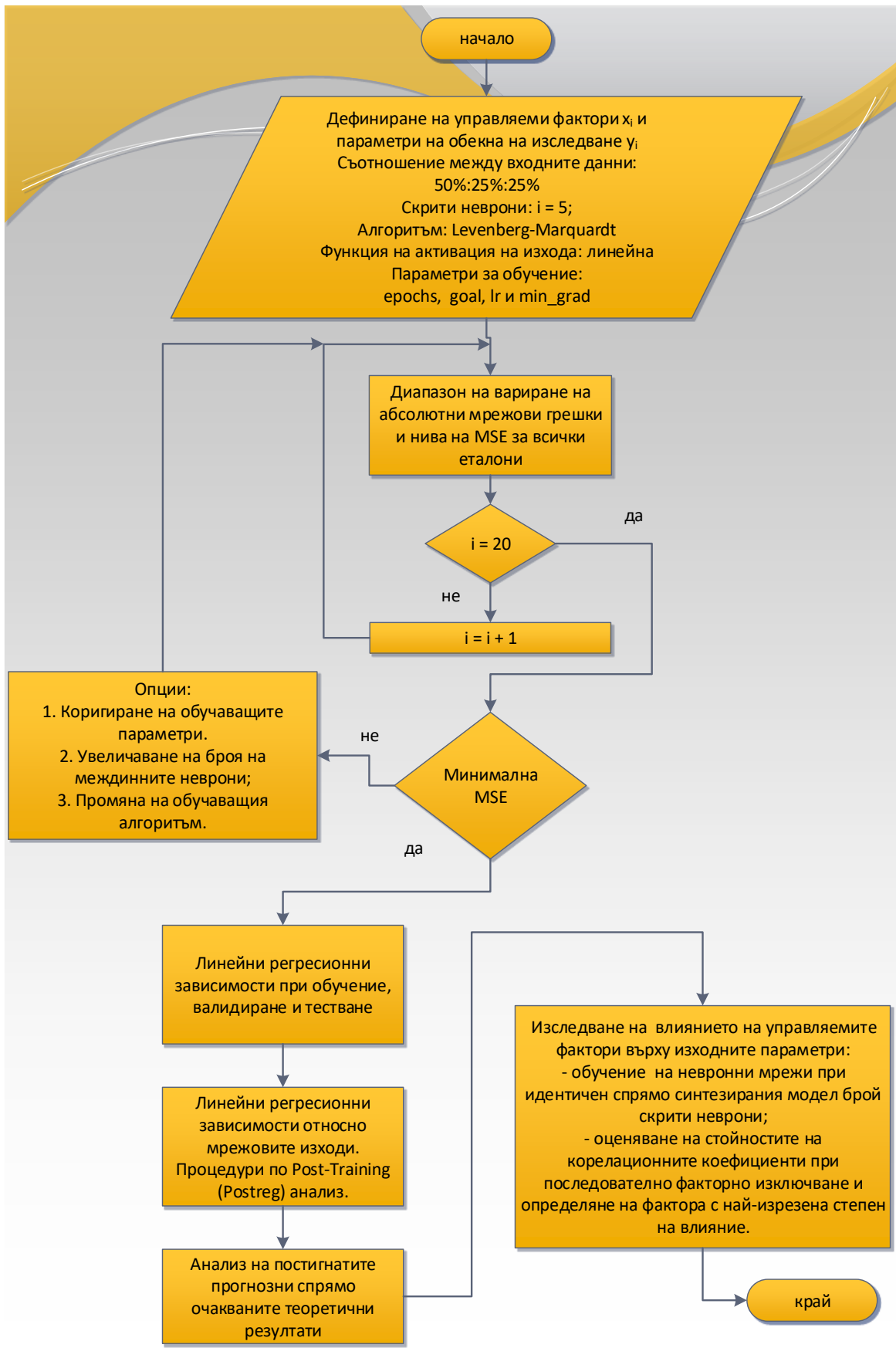
Да се запознаят студентите с процедурите при изследването и подбора на изкуствени невронни мрежи с обратно разпространение на грешката в качеството им на аналитично средство за апроксимация на икономически функции на преобразуване в програмен продукт MATLAB.

9.2. Теоретична постановка

Подход за синтез на невронни модели за прогнозен анализ на икономически данни

Апроксимацията на функции при анализа на бизнес процеси се решава въз основа на представения блоков алгоритъм на фиг. 9.1. Тук дефинираните управляеми фактори на обекта се подават като входни параметри към изследваните невронни архитектури, докато отклиците на обекта (фактори на бизнес обекти и бизнес събития, подлежащи на прогнозиране) се въвеждат като целеви числови стойности с използването на един неврон в изходния слой при твърдо зададена линейна активационна функция. Използваното съотношение между входните данни при инициализация е съответно 50% за обучение, 25% за валидиране и 25% за тестване на мрежата. Синтезирането на прогнозни ANNs (Artificial Neural Networks), базирано на изследване на измененията на MSE (Mean Squared Error) и вариационните диапазони на абсолютните мрежови грешки с промяна на невроните от 5 до 20 в скрития мрежови слой. В случай, че не бъде постигнат минимален MSE критерий се извършва промяна на стойностите на параметрите при невронно обучение, увеличаване на количеството на скритите невронни единици или смяна на обучаващия алгоритъм. Относно селектирания модел се изследват линейни регресионни зависимости при основни мрежови процеси, както и прилагат процедури по Post-Training (Postreg) анализ за мрежовия изход. Анализира се коректността на получените резултати. Изследва се факторното влияние върху изходните целеви прогнозни параметри при последователно изключване на входните мрежови независими променливи, базирано на:

- ✚ обучение на невронни архитектури с количество скрити неврони, равняващи се на невронните единици в междинните слоеве на синтезирани ANNs;
- ✚ калкулация и сравнение на корелационните коефициенти и абсолютните мрежови грешки относно мрежовите изходи [60-63].



Фиг. 9.1. Подход за синтез и селекция на невронни модели за прогнозен анализ

Задача за прогнозен анализ във връзка с инвестиционни показатели относно американски капиталови дружества

Обект на изследване е възможността за получаване на модели за прогнозиране на инвестиционни индикатори на икономическа ефективност във връзка с американски капиталови дружества, както следва:

- ✚ Y_1 : Annual Return;
- ✚ Y_2 : Excess Return;
- ✚ Y_3 : Systematic Risk,
- ✚ Y_4 : Total Risk (процентно изменение);
- ✚ Y_5 : Rel. Win Rate (изменение на нормализирани нива),

на основата на регресионни дървета на решенията. В задачата се включва още изследване на характера на изменение на указаните индикатори спрямо влиянието на входните независими предсказващи информационни компоненти (weights of the stock-picking concepts), които са изброени по-долу:

- ✚ X_1 : The weight of the Large B/P concept;
- ✚ X_2 : The weight of the Large ROE concept;
- ✚ X_3 : The weight of the Large S/P concept;
- ✚ X_4 : The weight of the Large Return Rate in the last quarter concept;
- ✚ X_5 : The weight of the Large Market Value concept;
- ✚ X_6 : The weight of the Small systematic Risk concept.

9.3. Задачи за изпълнение

- ❖ Да се изследват трислойни архитектури на изкуствени невронни мрежи с обучение по алгоритъма на Levenberg-Marquardt при линейна изходна активационна функция за прогнозен анализ на параметри на обекта “ Y_1 ”, “ Y_2 ” и “ Y_3 ”;
- ❖ Да се систематизират изходни данни в таблици от 9.1 до 9.3 от процедури по синтез на невронни модели по отношение на критериите „средноквадратична грешка (Mean Squared Error - MSE)“ и „вариационен интервал на мрежовите грешки“ относно данните от тестовия поднабор. Критериите да бъдат регистрирани при плавно нарастване на невроните в междинните слоеве на мрежите от 10 до 20 изчислителни единици;
- ❖ Да се анализират обобщените таблични резултати и селектират най-подходящи невронни модели във връзка с прогнозен анализ на заложените тестови инвестиционни показатели;

Таблица 9.1. Резултати от процедури по селекция на невронна мрежа за прогнозиране на процентното изменение на индикатор „Y₁: Annual Return“

№	Неврони в междинния слой	Диапазон на абсолютните мрежови грешки	Средноквадратична грешка
1.	10	-1.3055 до 1.0431	0.5550
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.	20	-1.7700 до 2.7058	2.2554

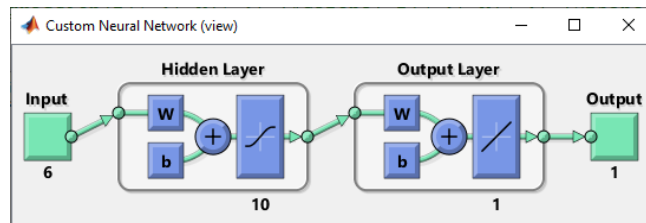
Таблица 9.2. Резултати от процедури по селекция на невронна мрежа за прогнозиране на процентното изменение на индикатор „Y₂: Excess Return“

№	Неврони в междинния слой	Диапазон на абсолютните мрежови грешки	Средноквадратична грешка
1.	10	-0.4526 до 0.3515	0.0728
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.	20	-0.5421 до 0.5793	0.1135

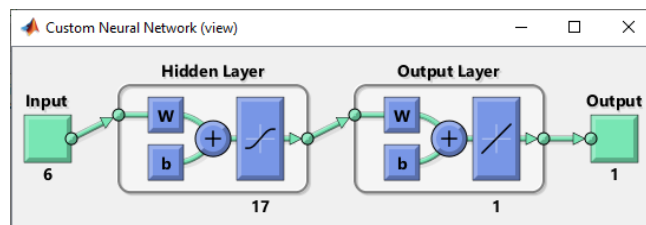
- ❖ Да се изведат синтезираните модели, представени на фиг. 9.2 и фиг. 9.3, за функционална апроксимация и приложат дейности по допълнителни анализ и оценка на тяхната ефективност и функционалност;
- ❖ Да се направи анализ на производителността на селектираните изкуствени невронни мрежи на базата на изведени зависимости на „средноквадратичната грешка“ за процесите на обучение, валидиране и тестване, дадени на фиг. 9.4 и фиг. 9.5. Да се направи визуална проверка за наличие на индикации за преобучение на изследваните невронни структури;

Таблица 9.3. Резултати от процедура по селекция на невронна мрежа за прогнозиране на изменението на индикатор „Y₃: Systematic Risk“

№	Неврони в междинния слой	Диапазон на абсолютните мрежови грешки	Средноквадратична грешка
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.	17	-0.0289 до 0.0377	5.0069e-04
9.			
10.			
11.	20	-0.1194 до 0.2984	0.0147



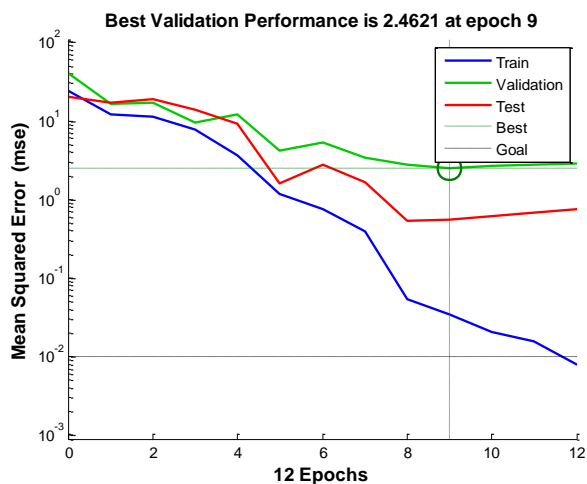
Фиг. 9.2. Селектирана невронна мрежа за прогнозиране на процентното изменение на индикатори „Y₁: Annual Return“ и „Y₂: Excess Return“



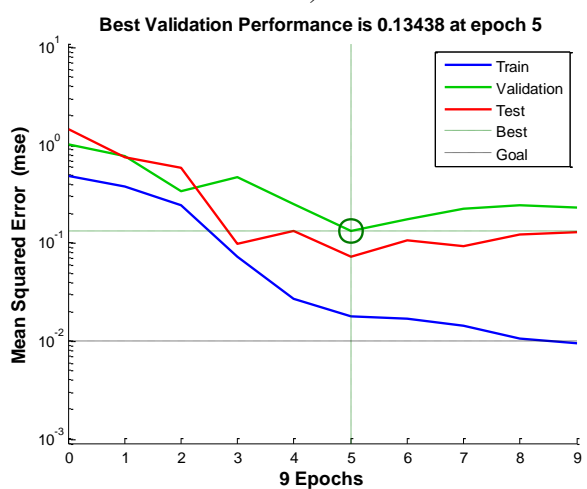
Фиг. 9.3. Селектирана невронна мрежа за прогнозиране на изменението на индикатор „Y₃: Systematic Risk“

❖ Да се анализира състоянието на мрежово обучение, онагледено на фиг. 9.6 и фиг. 9.7, по отношение на:

- направлението на минимизация на грешката след изпълнение на всеки обучаващ цикъл;
- параметъра, участващ в стъпката на редукция на грешката;
- броя на извършените валидиращи проверки по отношение на последователното изпълнение от обучаващи итерации.

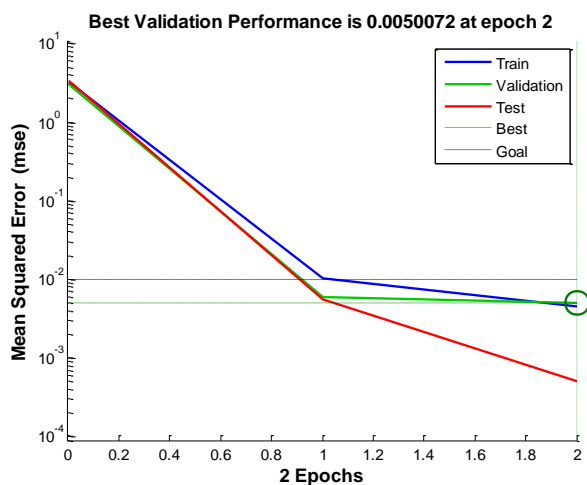


a)

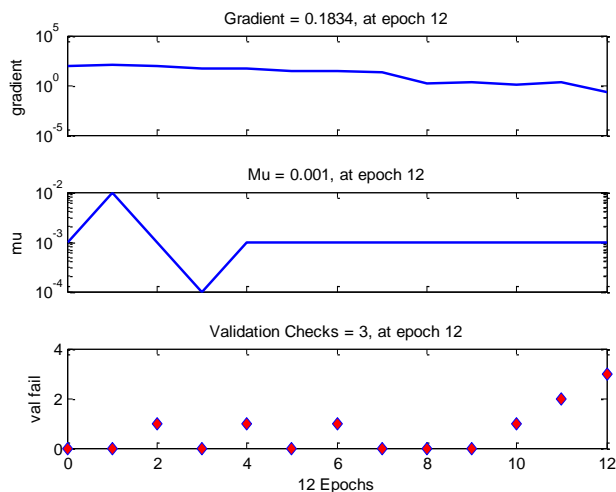


б)

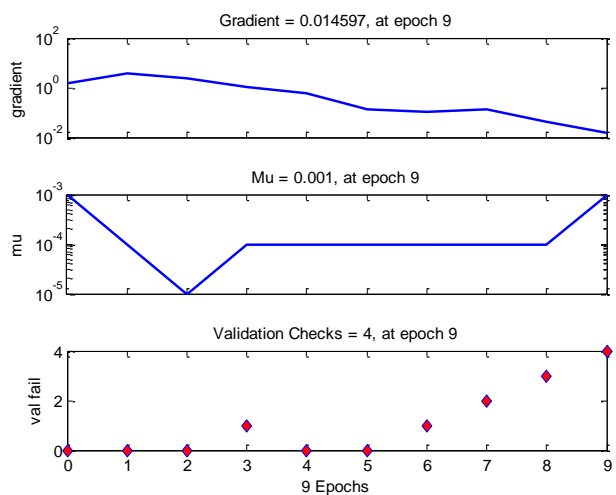
Фиг. 9.4. MSE за процесите на обучение, валидиране и тестване по отношение на селектираните невронни мрежи за прогнозиране на процентното изменение на индикатори за а) „ Y_1 : Annual Return“ и б) „ Y_2 : Excess Return“



Фиг. 9.5. Изменение на средноквадратичната грешка за процесите на обучение, валидиране и тестване по отношение на селектираната невронна мрежа за прогнозиране на изменението на индикатора „ Y_3 : Systematic Risk“

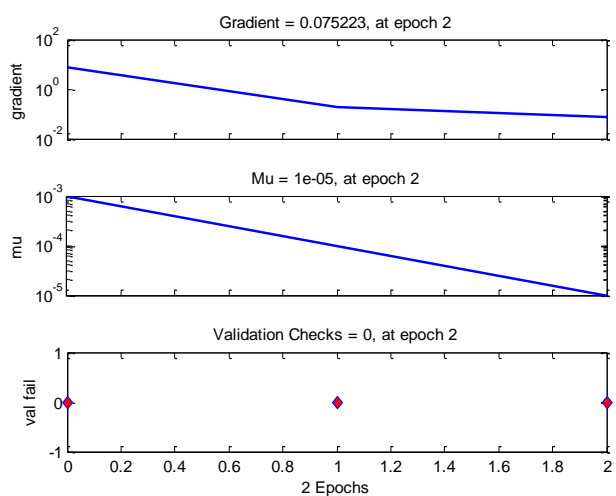


a)

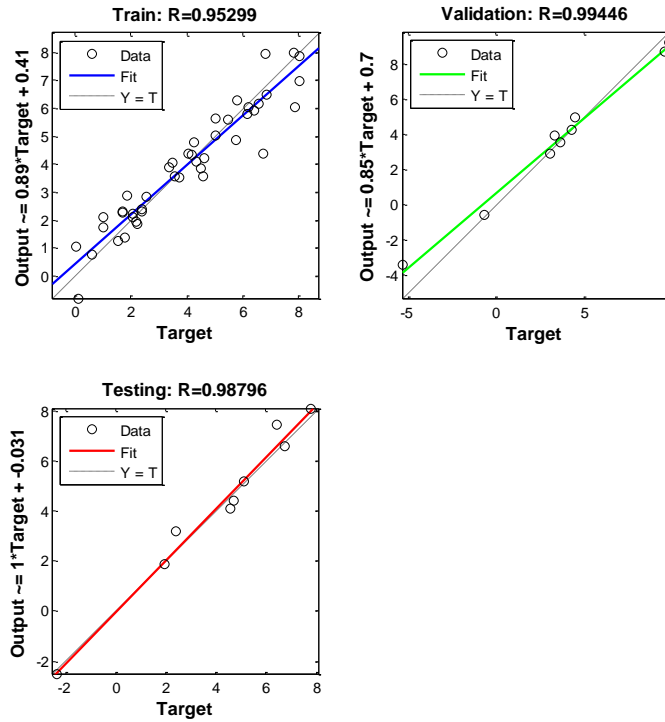


б)

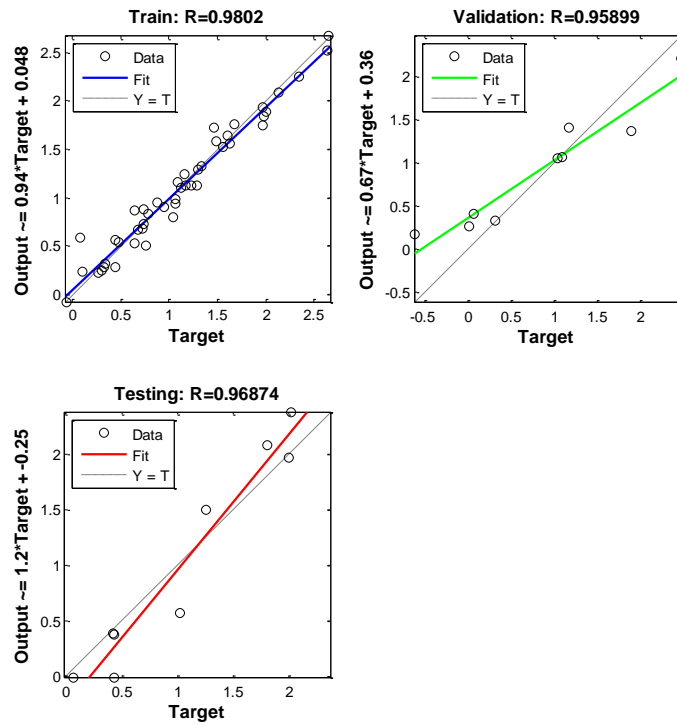
Фиг. 9.6. Състояние на мрежово обучение по отношение на селектираните невронни мрежи за прогнозиране на а) „ Y_1 : Annual Return“ и б) „ Y_2 : Excess Return“



Фиг. 9.7. Състояние на мрежово обучение по отношение на селектираната невронна мрежа за прогнозиране на изменението на индикатора „ Y_3 : Systematic Risk“



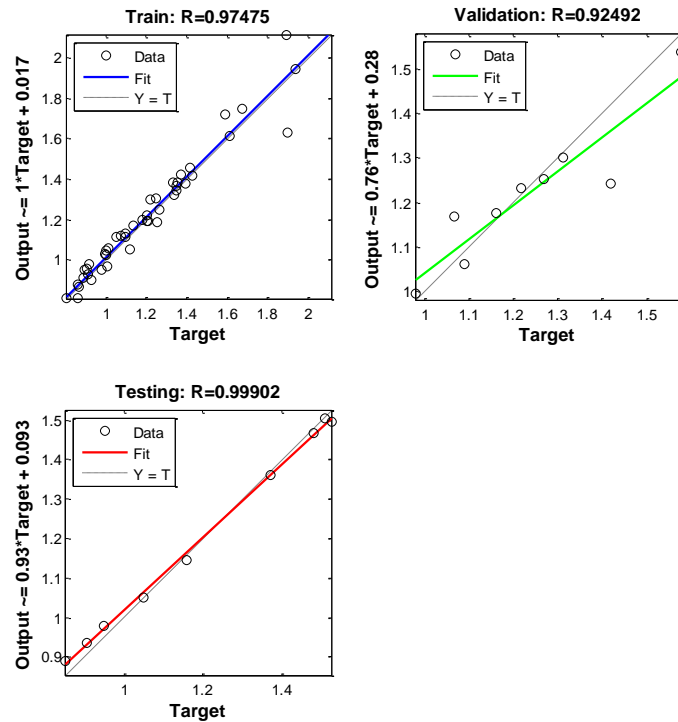
a)



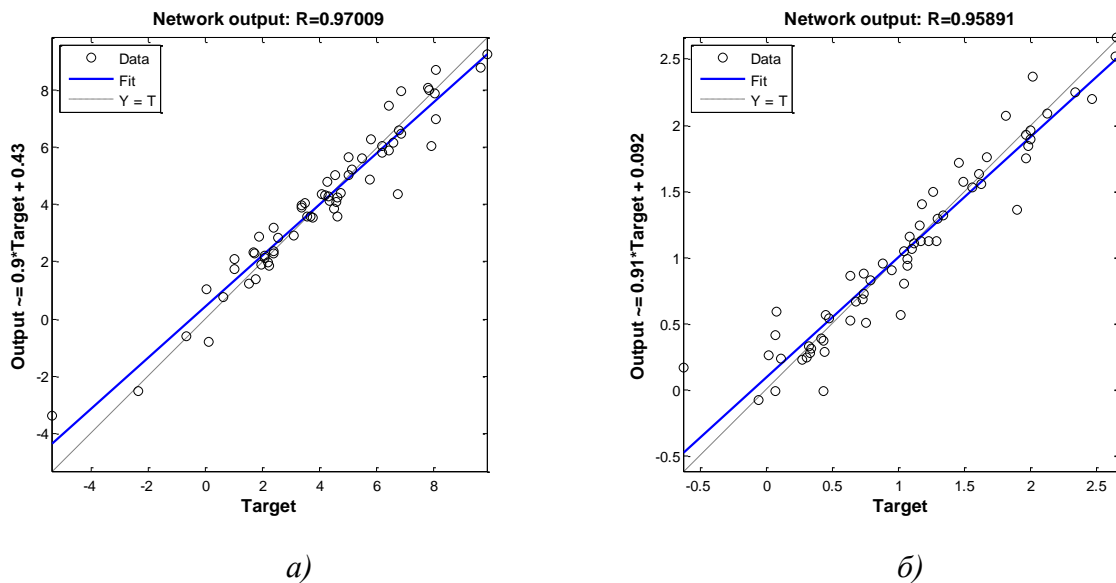
б)

Фиг. 9.8. Линейни регресионни зависимости за процесите на обучение, валидиране и тестване по отношение на селектираните невронни мрежи за прогнозиране на процентното изменение на индикатори за а) „ Y_1 : Annual Return“ и б) „ Y_2 : Excess Return“

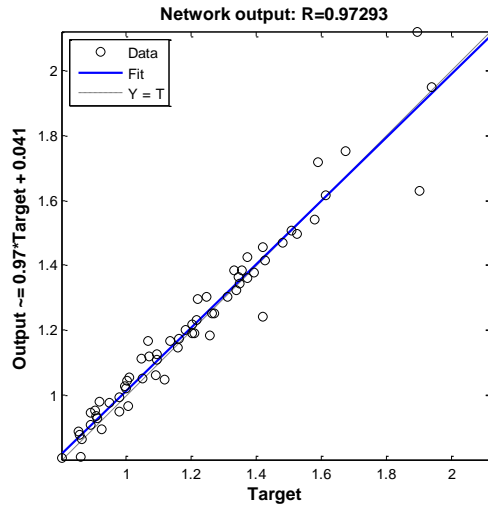
- ❖ Да се построят групи от линейни регресионни зависимости във връзка обучаващи, валидиращи и тестови процедури (фиг. 9.8 и фиг. 9.9), както и линии на регресия за изходите на невронните модели от фиг. 9.10 и фиг. 9.11. Да се оценят коефициентите на корелация R, степента на приближение между теоретичните и прогнозните резултати и групирането на опитните данни;



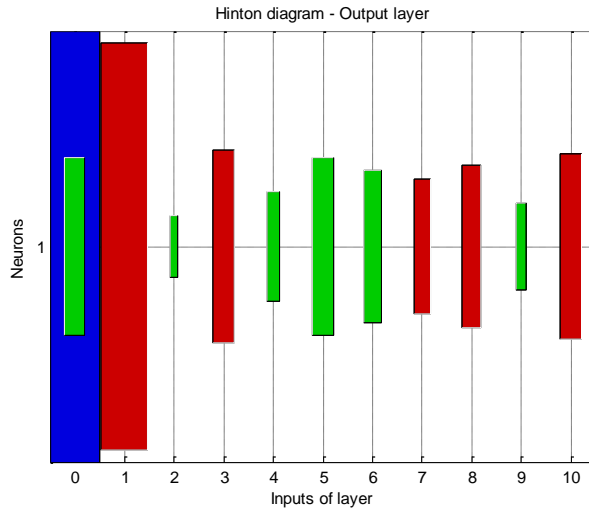
Фиг. 9.9. Линейни регресионни зависимости за обучение, валидиране и тестване на селектираната невронна мрежа за прогнозиране на изменението на индикатора „ Y_3 “



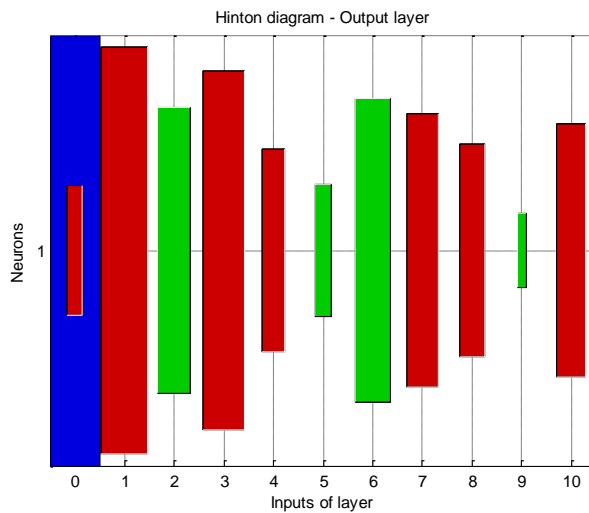
Фиг. 9.10. Линейни регресионни зависимости за изходите на селектираните невронни мрежи за прогнозиране на процентното изменение на индикатори за а) „ Y_1 : Annual Return“ и б) „ Y_2 : Excess Return“



Фиг. 9.11. Линејна регресионна зависимост за изхода на селектираната невронна мрежа за прогнозиране на изменението на индикатора „ Y_3 : Systematic Risk“

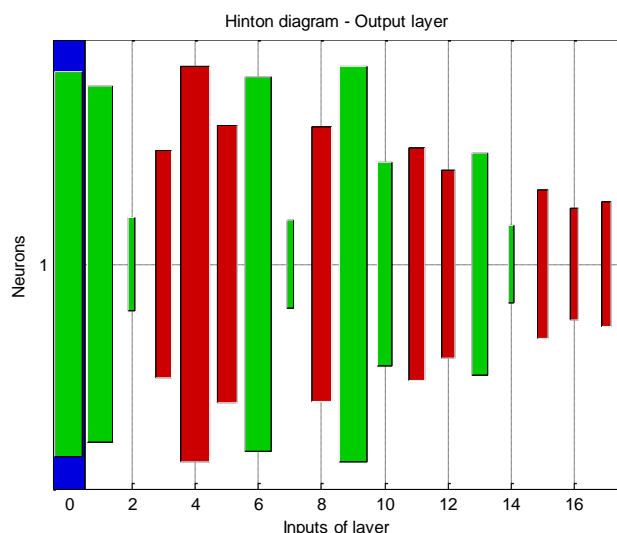


a)



б)

Фиг. 9.12. Диаграми на Hinton за изходите на селектираните невронни мрежи за прогнозиране на а) „ Y_1 : Annual Return“ и б) „ Y_2 : Excess Return“



Фиг. 9.13. Диаграма на Hinton за изхода на селектираната невронна мрежа за прогнозиране на изменението на индикатора „ Y_3 : Systematic Risk“

- ❖ Да се генерират и обясни вида на Hinton диаграми като графична интерпретация на изобразяване на тегловните матрици и матриците на отместванията за изходите на синтезираните невронни мрежи като аналитични средства прогнозиране на изменението на инвестиционни индикатори “ Y_1 ”, “ Y_2 ” и “ Y_3 ”, показани на фиг. 9.12 и фиг. 9.13.

9.4. Контролни въпроси

1. Кои други видове невронни мрежи се използват като инструменти за прогнозиране на параметрични изменения?
2. Кое определя настъпването на процеса „преобучение“ в хода на итерациите на изменение на MSE грешка от обучение, валидиране и тестване?
3. Какви установени нива на коефициента на корелация при опитните линии на регресия се приемат за удовлетворяващи от гледна точка на задачите за прогнозен анализ?
4. По какъв начин са кодирани положителните и отрицателните стойности на матричните тегла и отмествания при диаграмите на Hinton?
5. Какви други критерии освен „средноквадратична грешка“ могат да послужат за теоретична база при избора на модели за прогнозен анализ?

КОРЕЛАЦИОНЕН И РЕГРЕСИОНЕН АНАЛИЗ, РЕГРЕСИОННИ ДЪРВЕТА НА РЕШЕНИЯ ЗА ИЗСЛЕДВАНЕ НА ВЗАИМОВРЪЗКИТЕ И ПРОГНОЗИРАНЕ НА ИНВЕСТИЦИОННИ ПОКАЗАТЕЛИ В STATISTICA И MATLAB

10.1. Цел на упражнението

Да се запознаят студентите с процедурите при изследване на взаимовръзките, проверка на адекватността на аналитични модели и синтез на регресионни дървета на решения за прогнозен анализ на инвестиционни показатели във връзка с капиталови дружества чрез продуктите STATISTICA и MATLAB.

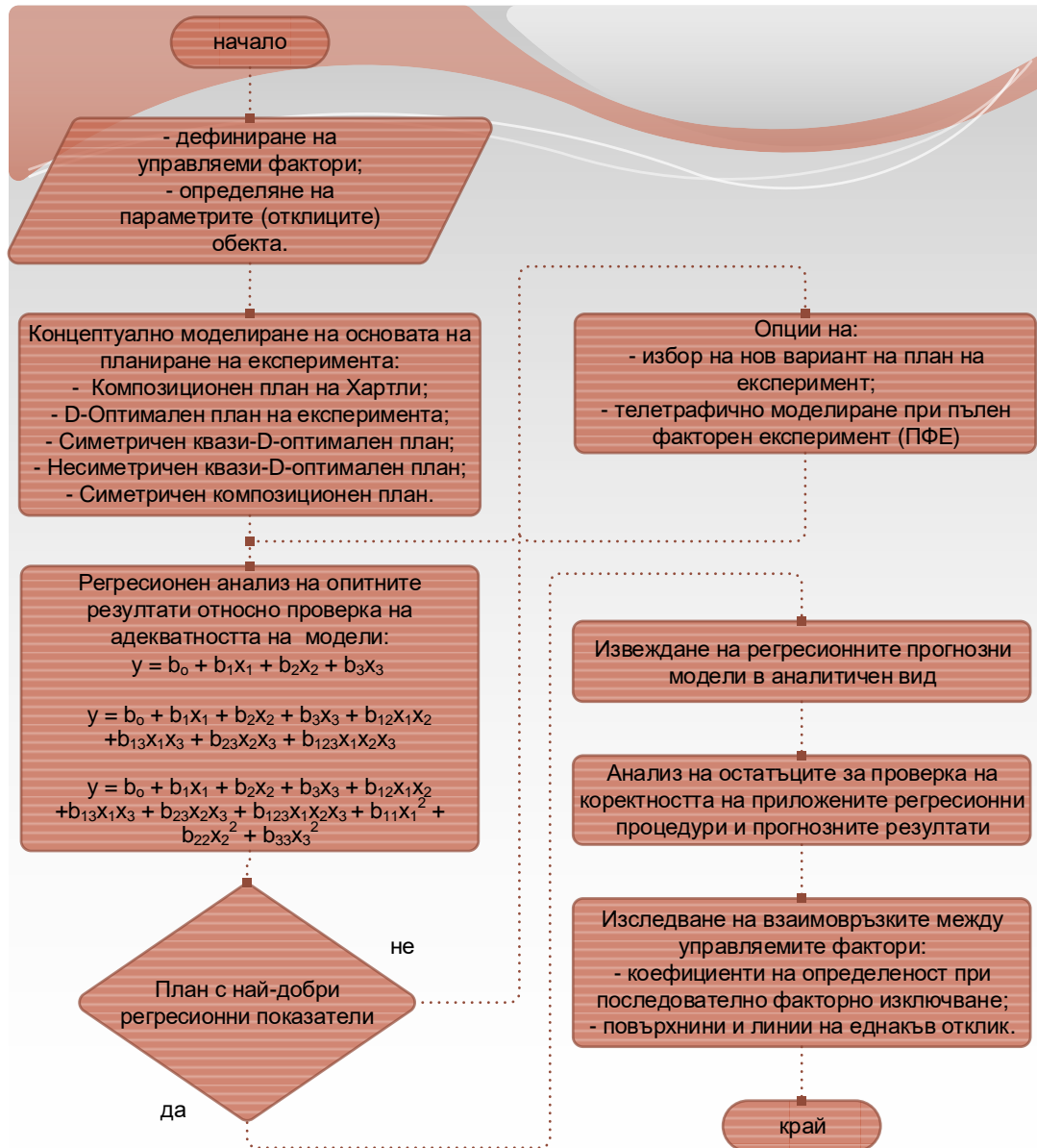
10.2. Теоретична постановка

Подход за извеждане на математически регресионни модели за прогнозен анализ

Изследва се възможността за извеждане на математически **регресионни модели** за прогнозиране на вариациите на селективни фактори относно маркетинг бизнес обекти и бизнес събития във връзка с конкретни бизнес процеси. На фиг. 10.1 е показан алгоритъм на дейностите по синтез на аналитични модели за прогнозен анализ, включващ:

- ✚ дефиниране на управляеми фактори и параметри на обекта на изследване на база на планиране на експеримента;
- ✚ формиране и представяне на разширени матрици на експеримента за регресионни модели от най-висока степен;
- ✚ прилагане на апарата на регресионния анализ върху опитните данни за всеки план на експеримента относно проверка на адекватността на модели от нулева, първа, втора или по-висока степени;
- ✚ анализ на резултатите и избор на план и регресионни модели от съответни степени с най-добри показатели по отношение на коефициентите на определеност. При наличие на ниски нива на регресионните показатели се предвижда прилагане на нов вариант на план на експеримент или търсене на модели на базата на пълен факторен експеримент при моделирането на разглежданите бизнес процеси;
- ✚ оценяване на значимостта на коефициентите на регресия при прието равнище на значимост и адекватността на моделите от съответните степени съгласно критериите на Фишър и съответните им вероятности за сектирания план на експеримента;
- ✚ избор на регресионен модел от съответна степен с най-добри качествени показатели. Извеждане на прогнозните параметрични модели в аналитичен вид;
- ✚ проверка относно коректността на приложение на регресионния анализ посредством нормални вероятностни диаграми на остатъците, както и по отношение на постигнатите прогнозни спрямо теоретично заложените резултати съобразно минимална, максимална стойности на остатъците, стандартно отклонение и други;

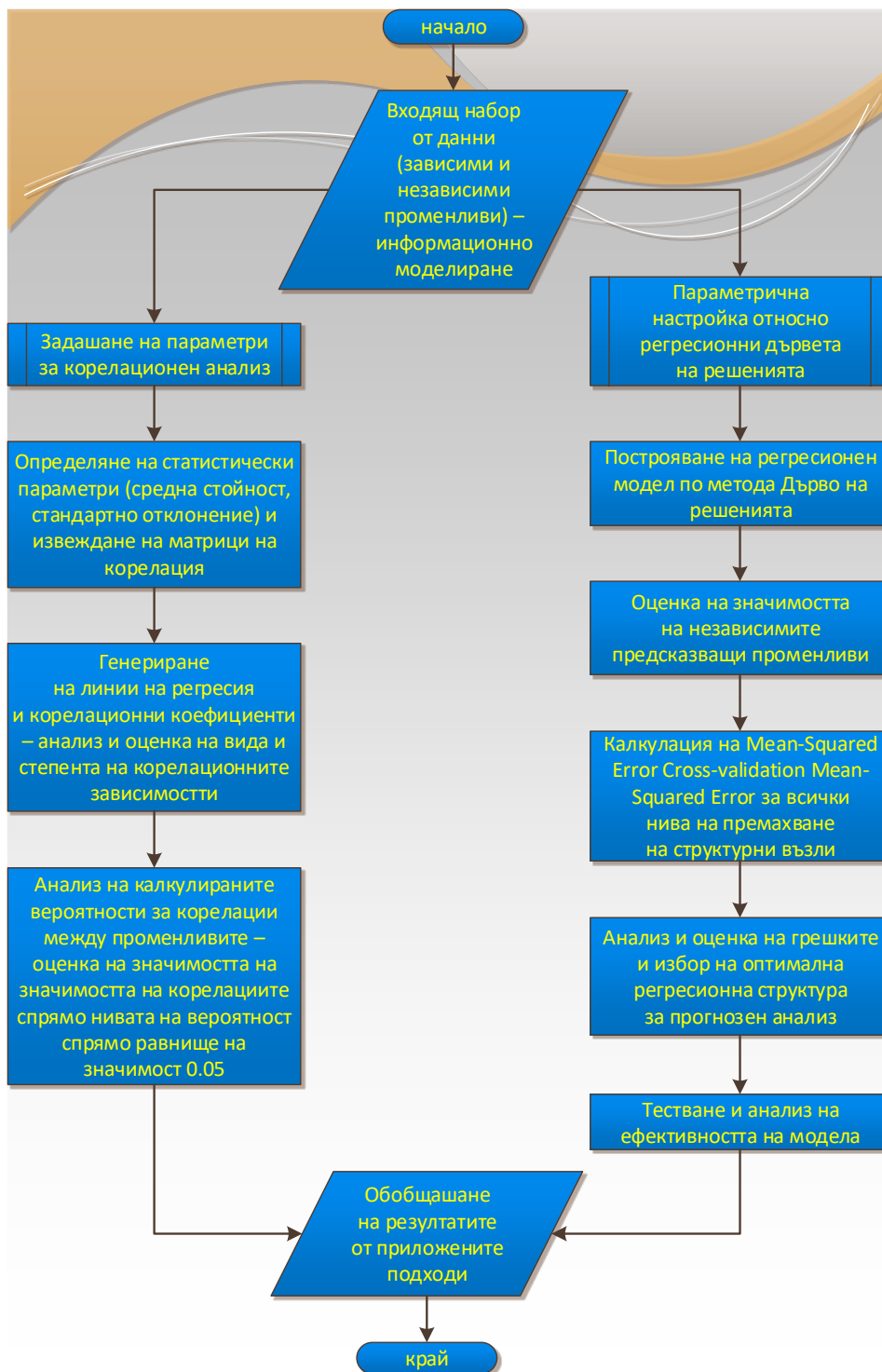
- изучаване на влиянието на управляемите фактори върху параметрите на обекта, свързано с: коефициентите на определеност при последователно факторно изключване и повърхнините и линиите на еднакъв отклик при последователно факторно изключване [64-69].



Фиг. 10.1. Подход за получаване на аналитични прогнозни модели на базата на регресионен анализ

Подход за изследване на взаимовръзките и прогнозиране на изменението на целеви индикатори посредством корелационен анализ и регресионни дървета на решения

Фигура 10.2 онагледява процедурите в основата на подход, съчетаващ статистическите апарати „Корелационен анализ“ и „Дърво на решенията на базата на алгоритъм CART“.



Фиг. 10.2. Алгоритъм за корелационен анализ и синтез на регресионни структури по метода Дърво на решенията за прогнозен анализ на икономически данни

Приложимостта на корелационната диагностика се състои в:

- ✚ определяне на базови статистически показатели (средна стойност и стандартно отклонение);
- ✚ установяване на възходящи или низходящи тенденции спрямо зависимите и независимите променливи, както и наличието и силата на причинно-следствените връзки между тях;

- ✚ анализ на значимостта на определените корелации според намерените вероятности, съпоставени с ниво на значимост 0.05.

Процедурите по синтез на регресионна архитектура на базата на метода Дърво на решенията се свеждат до:

- ✚ калкулация и последователен анализ на средноквадратичните грешки при подходите за оценка на ресубституция и крос-валидиране в рамките на всички възможни нива на премахване на възлови разклонения;
- ✚ дефиниране на очаквани нива на MSE при прилагане на нови входни данни и селекция на модел с най-добра ефективност.

Друг вариант на подхода се свързва с построяване, изследване, анализ на базов и търсене на оптимален модел, построен с помощта на оценените информативни признаци с най-висока степен на значимост [70-77].

10.3. Задачи за изпълнение

10.3.1. Извеждане на прогнозни модели на основата на регресионен анализ

- ❖ Да се разгледа задачата за получаване на регресионни модели за прогнозен анализ на нормализирани нива на инвестиционен показател “y: Systematic Risk” при залагане на следните управляеми фактори на обекта, означени с:
 - x_1 – Large B/P;
 - x_2 – Large S/P;
 - x_3 – Large Market Value;
 - x_4 – Small Systematic Risk.
- ❖ Да се реализира проверка на адекватността на базови аналитични модели от нулева, първа и втора степен, дадени по-долу:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 \quad (10.1)$$

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{134}x_1x_3x_4 + b_{234}x_2x_3x_4 + b_{1234}x_1x_2x_3x_4 \quad (10.2)$$

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{134}x_1x_3x_4 + b_{234}x_2x_3x_4 + b_{1234}x_1x_2x_3x_4 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{44}x_4^2 \quad (10.3)$$

- ❖ Да се импортират опитните данни в графичната среда STATISTICA и построи „разширена матрица на експеримента“, показана на фиг. 10.3, чрез въвеждане на допълнителни променливи, които отразяват взаимното влияние между комбинациите от управляеми фактори;

STATISTICA - [Data: stockdatasetnew3 (20v by 63c)]

File Edit View Insert Format Statistics Data Mining Graphs Tools Data Enterprise Window Help

Add to Workbook Add to Report Add to MS Word

Calibri 11 B I U

C:\Users\George\Desktop\UNSS_proekt\Otcheti\GG_otcheti\December2020\stockdatasetnew3.xlsx: Sheet1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
	x1	x2	x3	x4	x12	x13	x14	x23	x24	x34	x123	x124	x134	x234	x1234	x11	x22	x33	x44	y	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1,902608
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,263287
3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1,894339
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,425454
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1,094579
6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,859995
7	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	0	0	1,392192
8	0,5	0,5	0	0	0,25	0	0	0	0	0	0	0	0	0	0	0	0,25	0,25	0	0	1,939118
9	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	0	1,354462
10	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	0	0	1,524957
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,247376
12	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	0	1,591409
13	0,5	0	0,5	0	0	0,25	0	0	0	0	0	0	0	0	0	0	0,25	0	0,25	0	1,420759
14	0	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	1,091537
15	0	0,5	0,5	0	0	0	0	0,25	0	0	0	0	0	0	0	0	0	0,25	0,25	0	1,419546
16	0	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0	1,203244
17	0,5	0	0	0,5	0	0	0,25	0	0	0	0	0	0	0	0	0	0,25	0	0	0,25	0,977325
18	0	0	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,25	0,864122
19	0	0,5	0	0,5	0	0	0	0	0,25	0	0	0	0	0	0	0	0	0,25	0,25	0,25	1,066938

Фиг. 10.3. Разширена матрица на експеримента при приложение на регресионен анализ за получаване на модел за прогнозен анализ относно индикатор “Systematic Risk”

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,89015945 R²= ,79238385 Adjusted R²= ,77806550						
F(4,58)=55,340 p<0,0000 Std.Error of estimate: ,12806						
N=63	b*	Std.Err. of b*	b	Std.Err. of b	t(58)	p-value
Intercept			1,191831	0,045925	25,95173	0,000000
x1	0,375056	0,066892	0,511560	0,091237	5,60692	0,000001
x2	0,417814	0,066892	0,569881	0,091237	6,24614	0,000000
x3	-0,162613	0,066892	-0,221798	0,091237	-2,43100	0,018169
x4	-0,565109	0,066892	-0,770785	0,091237	-8,44814	0,000000

a)

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,94616461 R²= ,89522747 Adjusted R²= ,86178943						
F(15,47)=26,773 p<,00000 Std.Error of estimate: ,10106						
N=63	b*	Std.Err. of b*	b	Std.Err. of b	t(47)	p-value
Intercept			1,13417	0,03846	29,48691	0,000000
x1	0,497826	0,072531	0,67901	0,09893	6,86360	0,000000
x2	0,508411	0,072531	0,69345	0,09893	7,00955	0,000000
x3	-0,037492	0,072531	-0,05114	0,09893	-0,51691	0,607642
x4	-0,282492	0,072531	-0,38531	0,09893	-3,89477	0,000310
x12	0,047988	0,068108	0,29999	0,42576	0,70459	0,484542
x13	-0,048397	0,068108	-0,30255	0,42576	-0,71060	0,480846
x14	-0,253095	0,068108	-1,58217	0,42576	-3,71607	0,000537
x23	-0,046925	0,068108	-0,29334	0,42576	-0,68898	0,494223
x24	-0,172322	0,068108	-1,07724	0,42576	-2,53013	0,014812
x34	-0,123721	0,068108	-0,77341	0,42576	-1,81653	0,075669
x123	0,031382	0,065506	1,46274	3,05330	0,47907	0,634112
x124	-0,003295	0,065506	-0,15359	3,05330	-0,05030	0,960095
x134	-0,055025	0,065506	-2,56476	3,05330	-0,83999	0,405164
x234	-0,074541	0,065506	-3,47440	3,05330	-1,13792	0,260923
x1234	-0,011358	0,064128	-5,43688	30,69584	-0,17712	0,860174

б)

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= .97042971 R²= .94173382 Adjusted R²= .91598830						
F(19,43)=36,579 p<0,0000 Std.Error of estimate: .07879						
N=63	b*	Std. Err. of b*	b	Std. Err. of b	t(43)	p-value
Intercept			1,30611	0,04283	30,49516	0,000000
x1	-0,065577	0,147233	-0,08944	0,20082	-0,44539	0,658270
x2	0,034149	0,147233	0,04658	0,20082	0,23194	0,817689
x3	-0,436322	0,147233	-0,59512	0,20082	-2,96347	0,004942
x4	-0,932756	0,147233	-1,27224	0,20082	-6,33522	0,000000
x12	0,229000	0,067834	1,43155	0,42405	3,37586	0,001570
x13	0,114849	0,067834	0,71795	0,42405	1,69307	0,097676
x14	-0,030634	0,067834	-0,19150	0,42405	-0,45159	0,653832
x23	0,095328	0,067834	0,59592	0,42405	1,40530	0,167117
x24	0,029146	0,067834	0,18220	0,42405	0,42966	0,669590
x34	0,059982	0,067834	0,37497	0,42405	0,88424	0,381483
x123	0,028632	0,051078	1,33456	2,38078	0,56055	0,578010
x124	-0,007284	0,051078	-0,33952	2,38078	-0,14261	0,887264
x134	-0,058642	0,051078	-2,73336	2,38078	-1,14810	0,257276
x234	-0,077719	0,051078	-3,62254	2,38078	-1,52158	0,135436
x1234	-0,011919	0,049998	-5,70511	23,93204	-0,23839	0,812713
x11	0,357826	0,106727	0,69869	0,20839	3,35272	0,001677
x22	0,276814	0,106727	0,54050	0,20839	2,59366	0,012930
x33	0,208258	0,106727	0,40664	0,20839	1,95131	0,057554
x44	0,436768	0,106727	0,85283	0,20839	4,09238	0,000184

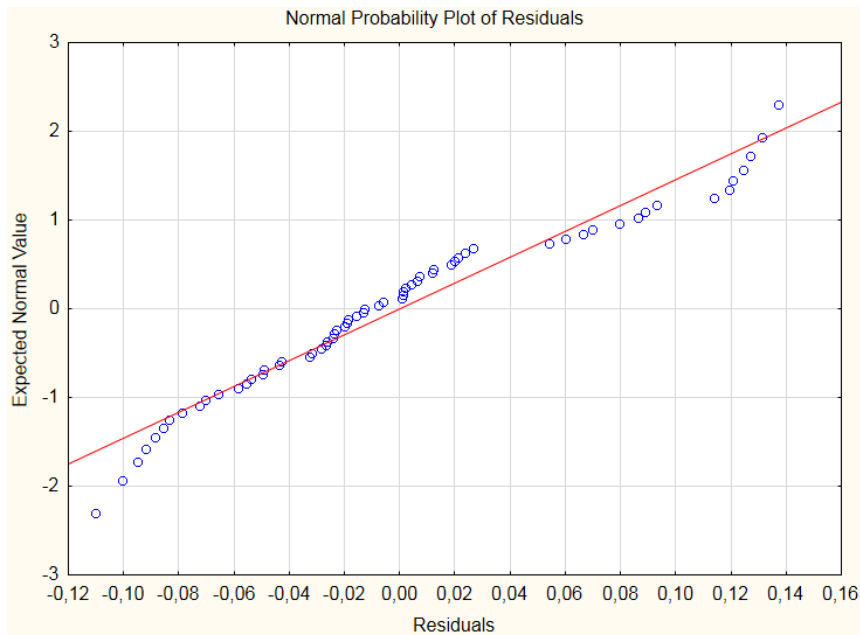
в)

Фиг. 10.4. Регресионни резултати във връзка с проверката на адекватността на модели от а) линеен тип, б) първа и в) втора степен

- ❖ Да се приложат процедури по регресионна диагностика във връзка с базовите аналитични резултати и представят обобщени регресионни резултати, съответстващи на показаните на фиг. 10.4;
- ❖ Да се оценят получените:
 - „коэффициенти за определеност R“ и свързаното с критерия процентно описание на опитните данни, чието изменение е резултат от влиянието на управляемите и това, дължащо се на случайни фактори;
 - критериите на Фишър и съответните им вероятности;
 - значимите и незначимите „опитни коэффициенти на регресия b_i “ спрямо прието равнище на значимост $\alpha = 0.05$.
- ❖ Да се определи моделът с най-подходяща експериментално констатирана степен на адекватност и изведе в краен аналитичен вид на база на значимите коэффициенти b_i – модел (10.4):

$$y = 1.306111 - 0.59512x_3 - 1.27224x_4 + 1.43155x_1x_2 + 0.69869x_1^2 + 0.54050x_2^2 + 0.85283x_4^2 \quad (10.4)$$

- ❖ Да се генерира „Нормална вероятността графика“ относно получен модел за прогнозиране на изменението на целевия инвестиционен показател и анализира разположението на „остатъците“ по направление на линията от 45°, посочена на фиг. 10.5. Процедурата е необходима като обосновка на коректността на приложение на апарата на регресионния анализ за конкретния случай;



Фиг. 10.5. Нормална вероятностна графика на остатъците за изведения модел за прогнозен анализ относно индикатор “Systematic risk”

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,88789289 R ² = ,78835379 Adjusted R ² = ,74765259						
F(10,52)=19,369 p<.00000 Std.Error of estimate: ,13656						
N=63	b*	Std.Err. of b*	b	Std.Err. of b	t(52)	p-value
Intercept			1,42683	0,048746	29,27052	0,000000
x2	0,18008	0,193407	0,24562	0,263799	0,93108	0,356117
x3	-0,48971	0,193407	-0,66795	0,263799	-2,53203	0,014400
x4	-1,22577	0,193407	-1,67190	0,263799	-6,33778	0,000000
x23	0,04846	0,101906	0,30293	0,637047	0,47552	0,636405
x24	0,02082	0,101906	0,13018	0,637047	0,20435	0,838881
x34	0,07549	0,101906	0,47190	0,637047	0,74075	0,462174
x234	-0,08132	0,082422	-3,79018	3,841744	-0,98658	0,328418
x22	0,11737	0,155927	0,22917	0,304460	0,75271	0,455017
x33	0,19193	0,155927	0,37476	0,304460	1,23089	0,223904
x44	0,59054	0,155927	1,15307	0,304460	3,78727	0,000397

a)

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,87858805 R ² = ,77191696 Adjusted R ² = ,72805484						
F(10,52)=17,599 p<.00000 Std.Error of estimate: ,14176						
N=63	b*	Std.Err. of b*	b	Std.Err. of b	t(52)	p-value
Intercept			1,43925	0,050604	28,44162	0,000000
x1	0,09894	0,200777	0,13494	0,273851	0,49276	0,624255
x3	-0,51024	0,200777	-0,69594	0,273851	-2,54130	0,014066
x4	-1,12523	0,200777	-1,53476	0,273851	-5,60435	0,000001
x13	0,06461	0,105790	0,40389	0,661322	0,61072	0,544043
x14	-0,06876	0,105790	-0,42984	0,661322	-0,64997	0,518571
x34	0,04691	0,105790	0,29322	0,661322	0,44339	0,659322
x134	-0,06172	0,085563	-2,87693	3,988133	-0,72137	0,473912
x11	0,17763	0,161868	0,34684	0,316062	1,09737	0,277534
x33	0,19994	0,161868	0,39040	0,316062	1,23519	0,222312
x44	0,51398	0,161868	1,00359	0,316062	3,17529	0,002516

б)

Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,95185611 R ² = ,90603005 Adjusted R ² = ,88795890						
F(10, 52)=50,137 p<0,0000 Std.Error of estimate: ,09099						
N=63	b*	Std. Err. of b*	b	Std. Err. of b	t(52)	p-value
Intercept			1,19893	0,032481	36,91166	0,000000
x1	0,099155	0,128873	0,13524	0,175777	0,76940	0,445138
x2	0,159775	0,128873	0,21793	0,175777	1,23978	0,220622
x4	-0,903649	0,128873	-1,23254	0,175777	-7,01193	0,000000
x12	0,212899	0,067903	1,33090	0,424484	3,13533	0,002822
x14	-0,059123	0,067903	-0,36959	0,424484	-0,87069	0,387925
x24	0,001880	0,067903	0,01175	0,424484	0,02769	0,978014
x124	-0,009163	0,054920	-0,42711	2,559869	-0,16685	0,868136
x11	0,303050	0,103899	0,59173	0,202871	2,91678	0,005210
x22	0,250798	0,103899	0,48970	0,202871	2,41387	0,019341
x44	0,484782	0,103899	0,94658	0,202871	4,66591	0,000022

в)

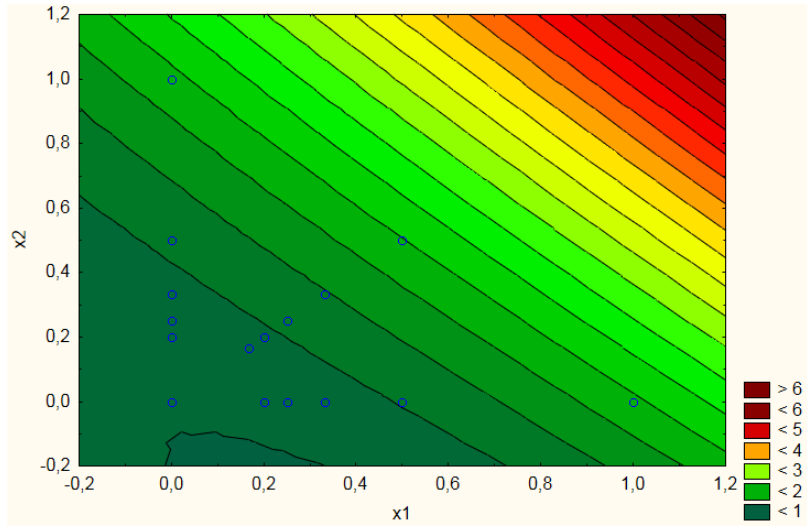
Regression Summary for Dependent Variable: y (stockdatasetnew3)						
R= ,79448739 R ² = ,63121021 Adjusted R ² = ,56028910						
F(10, 52)=8,9002 p<,00000 Std.Error of estimate: ,18026						
N=63	b*	Std. Err. of b*	b	Std. Err. of b	t(52)	p-value
Intercept			1,069607	0,064347	16,62258	0,000000
x1	-0,124195	0,255304	-0,169397	0,348223	-0,48646	0,628684
x2	0,057494	0,255304	0,078420	0,348223	0,22520	0,822706
x3	-0,390940	0,255304	-0,533225	0,348223	-1,53127	0,131764
x12	0,312705	0,134520	1,954811	0,840922	2,32460	0,024033
x13	0,174051	0,134520	1,088047	0,840922	1,29387	0,201427
x23	0,129321	0,134520	0,808423	0,840922	0,96135	0,340824
x123	0,027326	0,108800	1,273688	5,071222	0,25116	0,802680
x11	0,531449	0,205828	1,037698	0,401897	2,58200	0,012679
x22	0,394629	0,205828	0,770546	0,401897	1,91727	0,060705
x33	0,314571	0,205828	0,614227	0,401897	1,52832	0,132494

г)

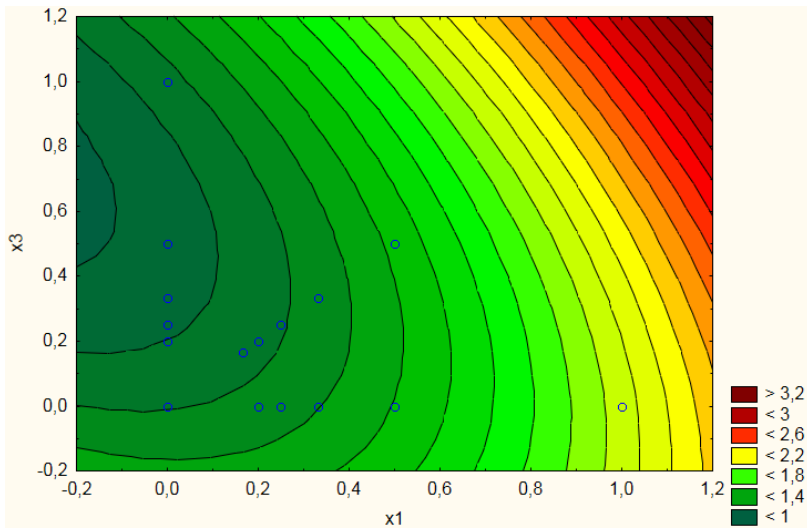
Фиг. 10.6. Изключване на фактор а) x₁, б) x₂, в) x₃ и г) x₄ при изследване на влиянието върху индикатор “Systematic risk”

- ❖ Да се изследва влиянието на управляемите фактори върху прогнозния икономически показател чрез стъпки на последователно тяхно индивидуално изключване и изключване на онези комбинации от фактори, в които те имат участие. Да се приложат текущи протоколни резултати от проведената регресионна диагностика, представени на фиг. 10.6;
- ❖ Да се оценят и систематизират получените нива на „коэффициентите на определеност R“ по степен на влияние върху изходния параметър (факторът с най-ниска стойност на R притежава най-голямо влияние върху параметъра “y”);
- ❖ Да се построят „линии на еднакъв отклик“ относно изведения математически регресионен модел за прогнозен анализ спрямо потенциални комбинации от управляеми фактори, както следва от фиг. 10.7:

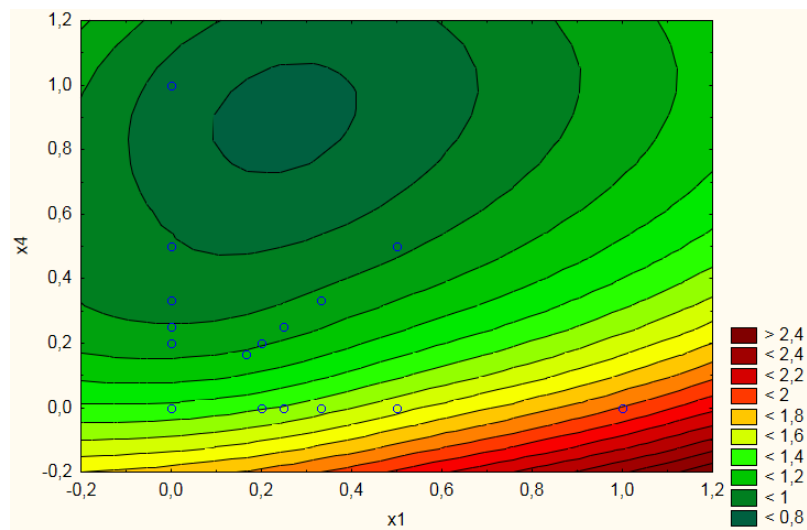
- $y = f(x_1, x_2)$;
- $y = f(x_1, x_3)$;
- $y = f(x_1, x_4)$;
- $y = f(x_2, x_3)$;
- $y = f(x_2, x_4)$;
- $y = f(x_3, x_4)$.



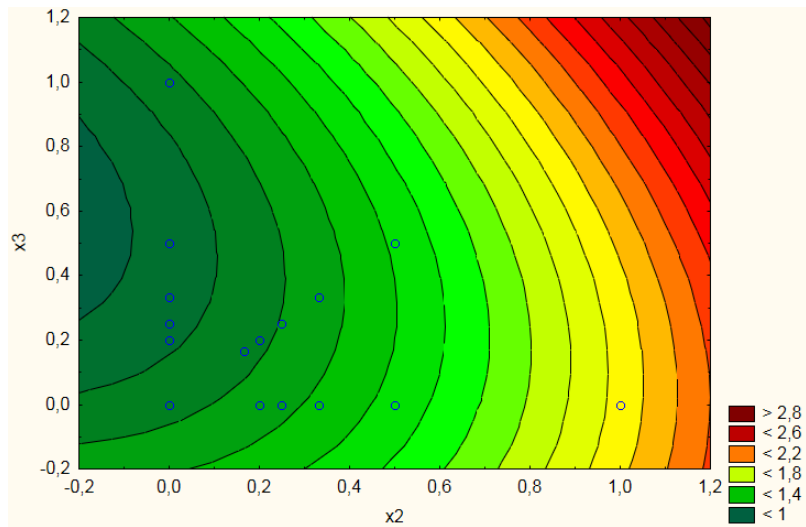
a)



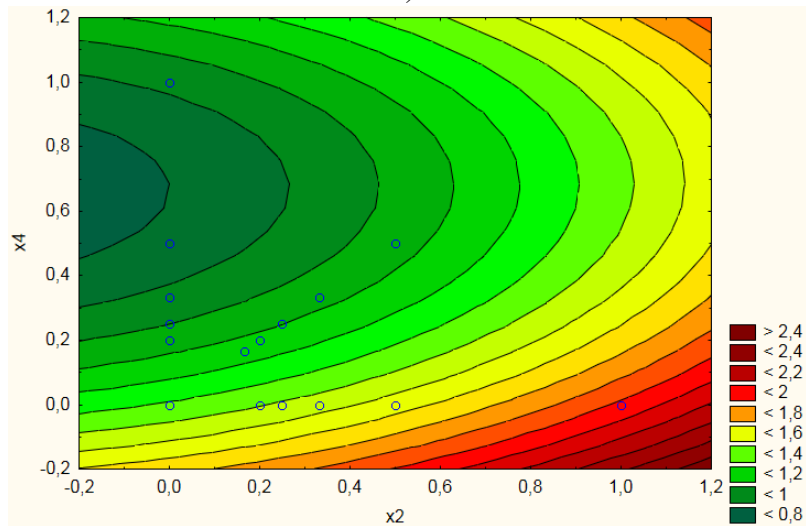
b)



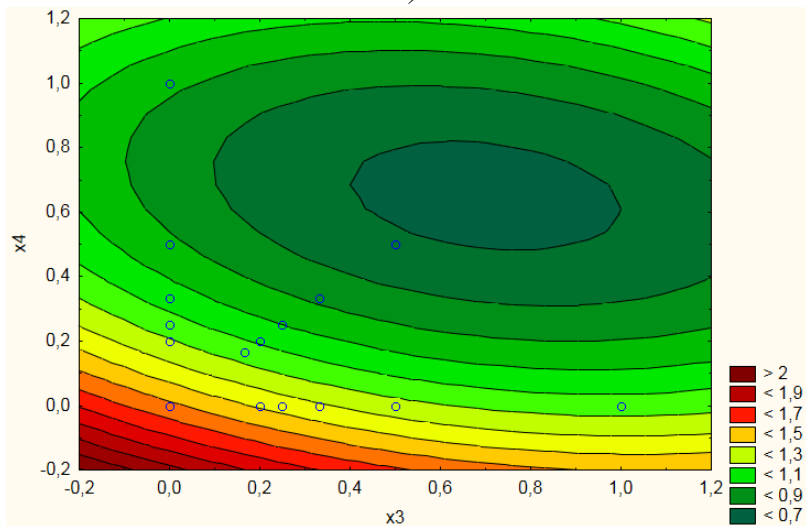
c)



з)



д)



е)

Фиг. 10.7. Линии на еднакъв отклик а) $y = f(x_1, x_2)$, б) $y = f(x_1, x_3)$,
в) $y = f(x_1, x_4)$, з) $y = f(x_2, x_3)$, д) $y = f(x_2, x_4)$ и е) $y = f(x_3, x_4)$

- ❖ Да се определят областите за всяка двойка от управляеми фактори – ниски, средни или високи нива, предизвикващи най-големи изменения на анализирувания прогнозен инвестиционен показател;
- ❖ Да се построят аналогични триизмерни „повърхнини на отклик“ за по-добро онагледяване на влиянието на управляемите фактори върху изходния параметър;
- ❖ Да се извърши повторение на изложените процедури за извеждане на прогнозни модели при залагане на нови специфицирани инвестиционни показатели.

10.3.2. Изследване на взаимовръзките между инвестиционни показатели с приложение на Корелационен анализ

- ❖ Да се изследва характерът и степента на връзка по отношение на набор от фактори от X_1 до X_6 (точка 9.2) спрямо параметъра “Y: Total Risk”;
- ❖ Да се направи базов анализ във връзка с придобиване на статистически „средни стойности“, „стандартни отклонения“ и „матрици на корелация“ за всяка изследвана целева параметрична двойка (фиг. 10.8), съответно:

- „ X_1 и Y“;
- „ X_2 и Y“;
- „ X_3 и Y“;
- „ X_4 и Y“;
- „ X_5 и Y“;
- „ X_6 и Y“.

Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Large B/P	Total Risk
Large B/P	0,16662	0,199304	1,000000	0,410119
Total Risk	12,51651	3,168730	0,410119	1,000000

a)

Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Large ROE	Total Risk
Large ROE	0,16662	0,199304	1,000000	-0,128951
Total Risk	12,51651	3,168730	-0,128951	1,000000

б)

Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Large SP	Total Risk
Large SP	0,16662	0,199304	1,000000	0,334141
Total Risk	12,51651	3,168730	0,334141	1,000000

в)

Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Large Return Rate	Total Risk
Large Return Rate	0,16662	0,199304	1,000000	0,160323
Total Risk	12,51651	3,168730	0,160323	1,000000

г)

Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Large Market Value	Total Risk
Large Market Value	0,16662	0,199304	1,000000	-0,176429
Total Risk	12,51651	3,168730	-0,176429	1,000000

д)

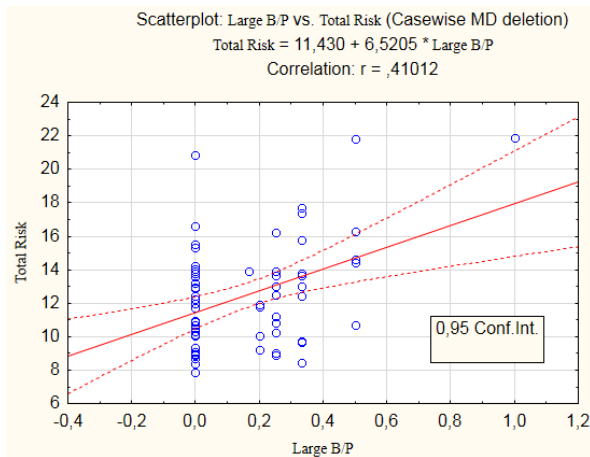
Correlations (daracorr)				
Marked correlations are significant at $p < .05000$				
N=63 (Casewise deletion of missing data)				
Variable	Means	Std.Dev.	Small Systematic Risk	Total Risk
Small Systematic Risk	0,16662	0,199304	1,000000	-0,599094
Total Risk	12,51651	3,168730	-0,599094	1,000000

е)

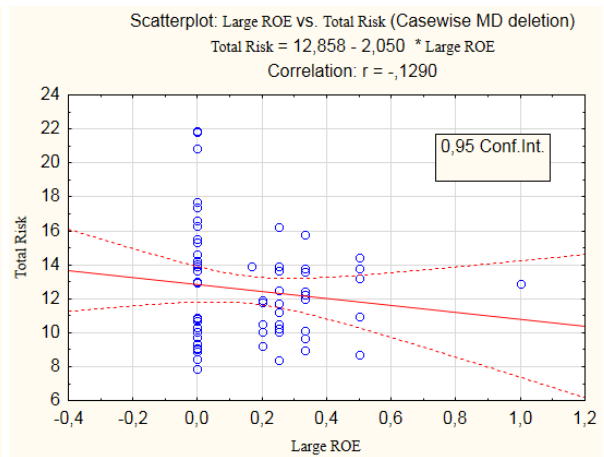
Фиг. 10.8. Статистически параметри и корелационни матрици относно индикаторите а) „Large BP u Total Risk“, б) „Large ROE u Total Risk“, в) „Large SP u Total Risk“, г) „Large Return Rate u Total Risk“, д) „Large Market Value u Total Risk“ и е) „Small Systematic Risk u Total Risk“

❖ Да се построят линии на регресия относно дефиниран доверителен интервал 0.95 (фиг. 10.9), даващи информация за:

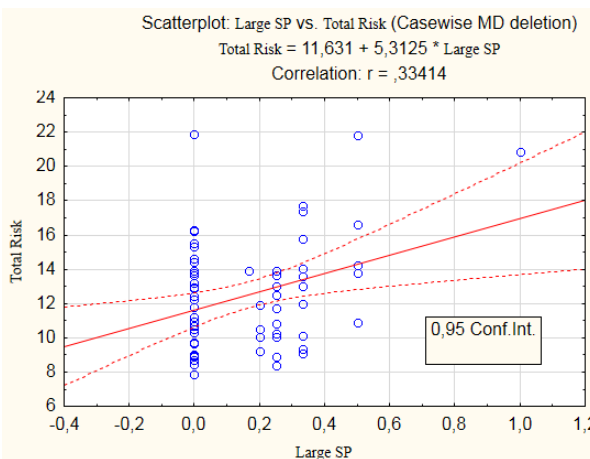
- коефициентите на корелация и посоката на зависимостта между изследваните двойки индикатори;
- процедурно получените аналитични линейни модели за калкулация на зависимата спрямо независимата променлива.



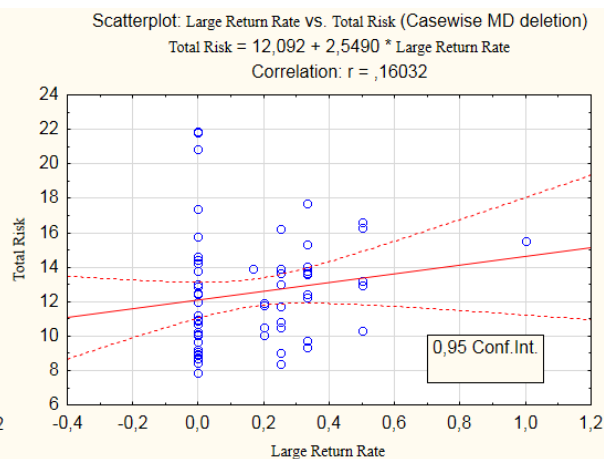
a)



б)

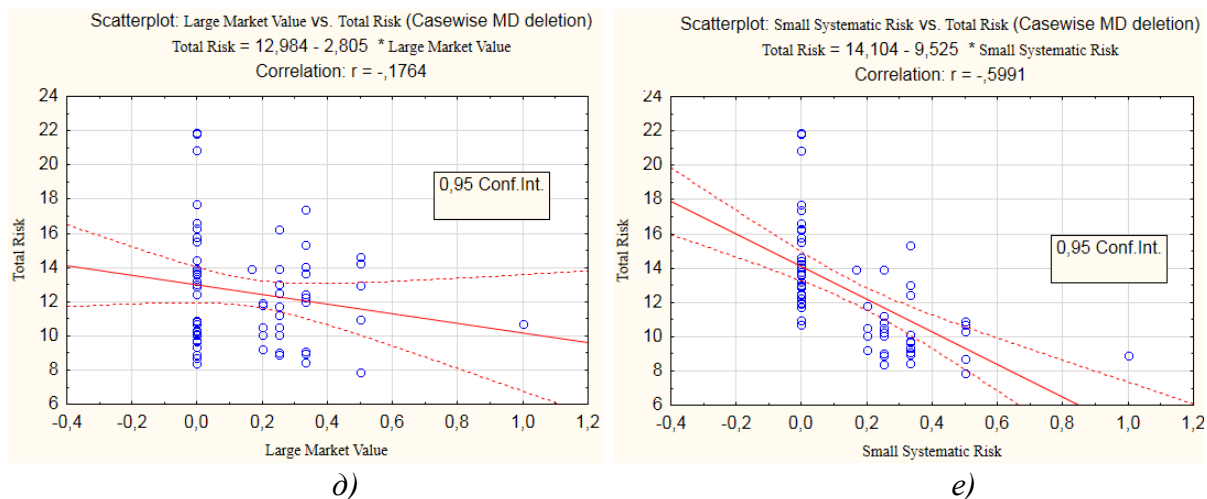


в)



г)

❖ Да се направи оценка на характера на зависимостта и степента на корелация по отношение на изследваните показатели като за целта се използва таблица 10.1;



Фиг. 10.9. Линии на регресия при доверителен интервал 0.95 във връзка с колелационен анализ между индикатори а) „Large BP u Total Risk“, б) „Large ROE u Total Risk“, в) „Large SP u Total Risk“, г) „Large Return Rate u Total Risk“, д) „Large Market Value u Total Risk“ и е) „Small Systematic Risk u Total Risk“

Таблица 10.1. Скали за оценяване на вида и степента на корелация

Скала №1 вариации на r	Скала №2 вариации на r	Интерпретация
възходяща (положителна) зависимост		
0 до 0.3	0 до 0.2	слаба корелация
0.3 до 0.5	0.2 до 0.4	умерена корелация
0.5 до 0.7	0.4 до 0.6	значителна корелация
0.7 до 0.9	0.6 до 0.8	силна корелация
по-големи от 0.9	по-големи от 0.8	много силна корелация
низходяща (отрицателна) зависимост		
0 до -0.3	0 до -0.2	слаба корелация
-0.3 до -0.5	-0.2 до -0.4	умерена корелация
-0.5 до -0.7	-0.4 до -0.6	значителна корелация
-0.7 до -0.9	-0.6 до -0.8	силна корелация
по-ниски от -0.9	по-ниски от -0.8	много силна корелация

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Large B/P Total Risk
Large B/P	0,001
Total Risk	0,001

а)

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Large ROE Total Risk
Large ROE	0,314
Total Risk	0,314

б)

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Large SP Total Risk
Large SP	0,007
Total Risk	0,007

в)

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Large Return Rate Total Risk
Large Return Rate	0,209
Total Risk	0,209

з)

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Large Market Value Total Risk
Large Market Value	0,167
Total Risk	0,167

д)

Color map of p-values for correlations (daracorr)	
N=63 (Casewise deletion of missing data)	
p<= 0,001 0,010 0,025 0,050 0,100 0,150 0,200 0,350 0,500 1	
Variable	Small Systematic Risk Total Risk
Small Systematic Risk	0,000
Total Risk	0,000

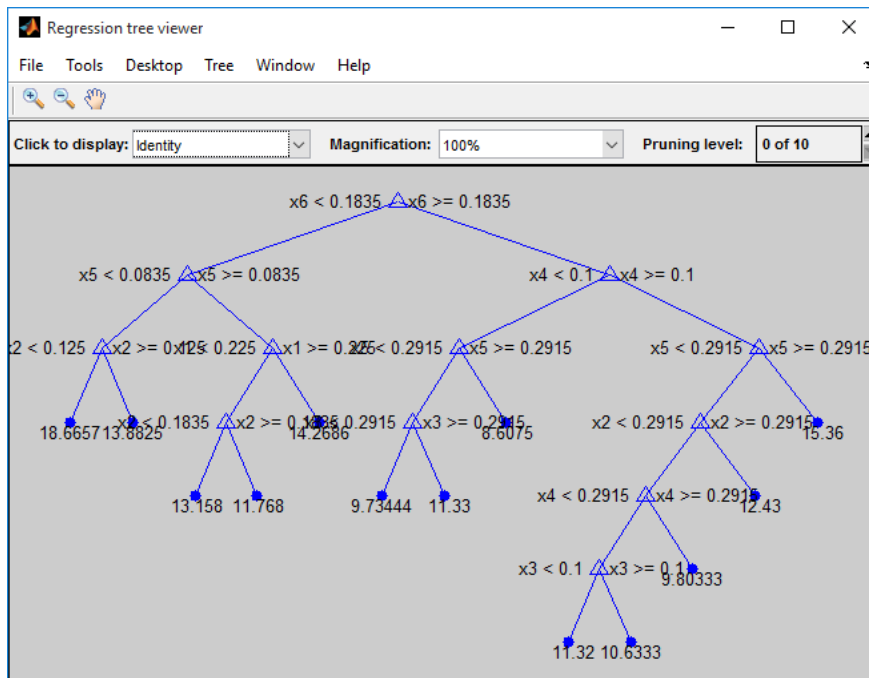
е)

Фиг. 10.10. Вероятности за корелации между индикатори а) „Large BP u Total Risk“, б) „Large ROE u Total Risk“, в) „Large SP u Total Risk“, з) „Large Return Rate u Total Risk“, д) „Large Market Value u Total Risk“ и е) „Small Systematic Risk u Total Risk“-

- ❖ Да се потвърди наличието на установените вид и степен на корелация между независимите и зависимата променлива на базата на анализ на вероятностите за разглежданите двойки параметри, както е показано на фиг. 10.10.

10.3.3. Изследване на регресионни структури по метода Дърво на решенията за прогнозен анализ на изменението на инвестиционни показатели за американски капиталови дружества с помощта на продукта MATLAB

- ❖ Да се създаде регресионна структура за многовариантен избор на решение на база на набор от фактори от X_1 до X_6 (точка 9.2) за прогнозен анализ във връзка с параметъра “Y: Total Risk” – фиг. 10.11;



a)

```

Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

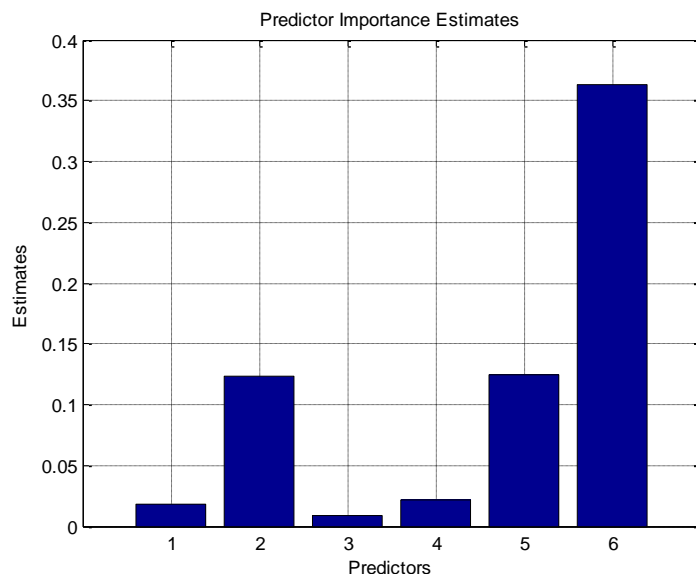
Decision tree for regression
1 if x6<0.1835 then node 2 elseif x6>=0.1835 then node 3 else 12.5165
2 if x5<0.0835 then node 4 elseif x5>=0.0835 then node 5 else 14.5697
3 if x4<0.1 then node 6 elseif x4>=0.1 then node 7 else 10.3971
4 if x2<0.125 then node 8 elseif x2>=0.125 then node 9 else 16.1147
5 if x1<0.225 then node 10 elseif x1>=0.225 then node 11 else 13.2065
6 if x5<0.2915 then node 12 elseif x5>=0.2915 then node 13 else 9.75188
7 if x5<0.2915 then node 14 elseif x5>=0.2915 then node 15 else 11.0853
8 fit = 18.6657
9 fit = 13.8825
10 if x2<0.1835 then node 16 elseif x2>=0.1835 then node 17 else 12.463
11 fit = 14.2686
12 if x3<0.2915 then node 18 elseif x3>=0.2915 then node 19 else 10.1333
13 fit = 8.6075
14 if x2<0.2915 then node 20 elseif x2>=0.2915 then node 21 else 10.78
15 fit = 15.36
16 fit = 13.158
17 fit = 11.768
18 fit = 9.73444
19 fit = 11.33
20 if x4<0.2915 then node 22 elseif x4>=0.2915 then node 23 else 10.6531
21 fit = 12.43
22 if x3<0.1 then node 24 elseif x3>=0.1 then node 25 else 10.908
23 fit = 9.80333
24 fit = 11.32
25 fit = 10.6333

```

b)

Фиг. 10.11. Базова регресионна архитектура по метода Дърво на решенията за прогнозиране на процентното изменение на индикатор „Y: Total Risk“ – графична a) и текстова интерпретация б)

- ❖ Да се проведе оценка и анализира значимостта на използваните при построяването на модела на предсказващи променливи от X_1 до X_6 (точка 9.2) - фиг. 10.12;



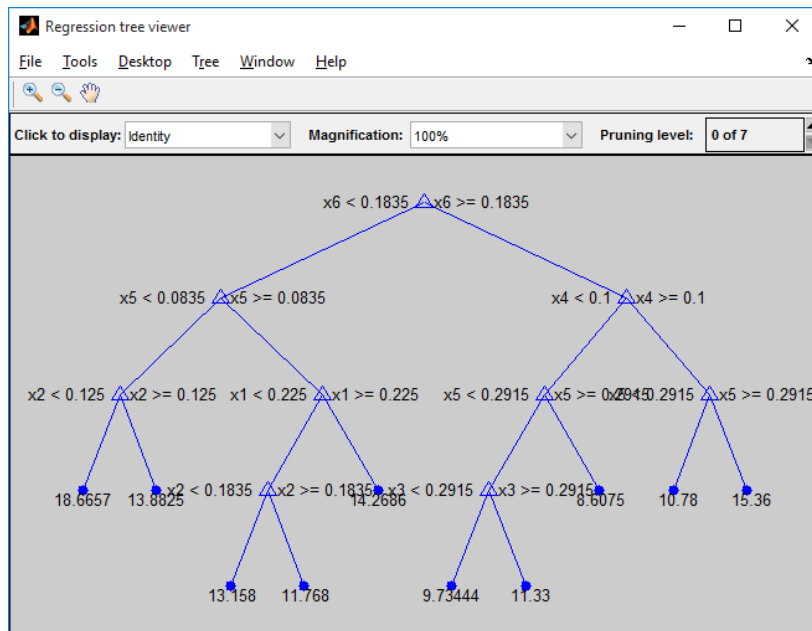
Фиг. 10.12. Оценка на информативните признаци относно регресионни модели по метода Дърво на решенията за прогнозиране на изменението на индикатор „Y: Total Risk“

- ❖ Да се генерират последователни регресионни дървета на решения при редуциране на количество структурни възли за съответни опитни нива и обобщят данните за случаите на вариации на „Mean-Squared Error“, „Cross-validation Mean-Squared Error“ и „приети нива на Mean-Squared Error при прогнозен анализ при опериране с нови данни“ – таблица 10.2;

Таблица 10.2. Изследване на MSE вариациите при редуциране на възлите на базовата на регресионна архитектура по метод Дърво на решенията за индикатор „Y: Total Risk“

Нива на премахване на разклонения	Mean-Squared Error	Cross-validation Mean-Squared Error	Приети нива на Mean-Squared Error при анализ на нови данни
0	1.9743	8.0370	5.00565
1			
2			
3	2.0835	6.4419	4.26270
4			
5			
6			

- ❖ Да се анализират представените резултати от оценка на качеството на експерименталните регресионни архитектури по метода Дърво на решенията. Да се определи и визуализира **оптимален модел за прогнозна диагностика** (фиг. 10.13);



a)

```

Command Window
New to MATLAB? Watch this Video, see Examples, or read Getting Started.

>> view(tree3,'Mode','graph');
>> view(tree3)

Decision tree for regression
 1 if x6<0.1835 then node 2 elseif x6>=0.1835 then node 3 else 12.5165
 2 if x5<0.0835 then node 4 elseif x5>=0.0835 then node 5 else 14.5697
 3 if x4<0.1 then node 6 elseif x4>=0.1 then node 7 else 10.3971
 4 if x2<0.125 then node 8 elseif x2>=0.125 then node 9 else 16.1147
 5 if x1<0.225 then node 10 elseif x1>=0.225 then node 11 else 13.2065
 6 if x5<0.2915 then node 12 elseif x5>=0.2915 then node 13 else 9.75188
 7 if x5<0.2915 then node 14 elseif x5>=0.2915 then node 15 else 11.0853
 8 fit = 18.6657
 9 fit = 13.8825
10 if x2<0.1835 then node 16 elseif x2>=0.1835 then node 17 else 12.463
11 fit = 14.2686
12 if x3<0.2915 then node 18 elseif x3>=0.2915 then node 19 else 10.1333
13 fit = 8.6075
14 fit = 10.78
15 fit = 15.36
16 fit = 13.158
17 fit = 11.768
18 fit = 9.73444
19 fit = 11.33

```

б)

Фиг. 10.13. Оптимална регресионна архитектура по метода Дърво на решенията за прогнозиране на процентното изменение на индикатора „Y1: Total Risk“ – графична а) и текстова интерпретация б)

- ❖ Да се представи структурна информация за количественото разпределение на опитните еталони, отнасящи се за крайните възли в структурата на базовия генериран и констатиран оптимален модел за конкретния случай на изследване, съгласно таблица 10.3;

Таблица 10.3. Разпределение на опитните наблюдения за базовия и оптималния регресионен модел по метода Дърво на решенията за индикатора „Y: Total Risk“

Базова архитектура		Оптимална архитектура	
Краен възел	Брой наблюдения	Краен възел	Брой наблюдения
1	63	1	63
2		2	
3		3	
4		4	
5		5	
6		6	
7		7	
8		8	
9		9	
10		10	
11		11	
12		12	
13		13	
14		14	
15		15	
16		16	
17		17	
18		18	
19		19	
20			
21			
22			
23			
24			
25			

```
>> tree3predictres = predict(tree3,[0 0.3333 0.3333 0 0.3333 0; 0.3333 0 0 0.3333 0.3333 0])
tree3predictres =
    11.7680
    14.2686
```

Фиг. 10.14. Прогнозиране на процентното изменение на индикатора „Y: Total Risk“ за произволни еталони с приложение на оптималната регресионна архитектура

- ❖ Да се извърши проверка на пригодността на синтезирания оптимален модел за регресионна диагностика по метода Дърво на решенията за произволни информационни еталони, както е дадено на фиг. 10.14;
- ❖ Да се изследва възможността за прилагане на указаните процедури по отношение на синтеза на регресионни дървета на решенията за целите на прогнозния анализ за други икономически параметри.

10.4. Контролни въпроси

1. Какви типове планове на експеримента познавате и по какво правило се определят възможните комбинации в техния състав въз основа на

количествената измеримост на управляемите фактори и нивата на тяхното вариране?

2. По какъв начин се определя адекватността и каква е връзката с описанието на опитните данни на приложените линейни или модели от по-висока степен в процеса на регресионен синтез?
3. Кои са приетите теоретични по дефиниция нива на критерия „равнище на значимост“, спрямо които се оценяват опитните коефициенти на регресия?
4. На базата на какви нива на коефициента на определеност при последователно изключване на управляеми фактори се оценява степента на тяхното влияние върху отклиците на обекта?
5. Как се тълкуват цветовите зони при построяване на „линиите на еднакъв отклик“ относно влиянието на управляемите фактори върху изходните прогнозирани параметри?
6. По какъв начин могат да бъдат тълкувани негативен и положителен корелационен коефициент във връзка с корелацията между две променливи?

ИЗПОЛЗВАНА ЛИТЕРАТУРА

- [1] Wisniewski P., Kluza K., and Ligeza A., “An Approach to Participatory Business Process Modelling: BPMN Model Generation Using Constant Programming and Graph Composition”. Applied Science, vol. 8, pp. 1-22, 2018.
- [2] „Инструменти за бизнес моделиране и характеристики на приложението му. Какво е бизнес модел? Какви бизнес модели съществуват?“, <https://gazkomplekt-saratov.ru/bg/instrumenty-biznes-modelirovaniya-i-osobennosti-ego.html>., 2018.
- [3] „Вътрешни правила за прилагане за технически и организационни мерки в допустимия вид на защита на личните данни в „Плади Компютърс“ ООД“, 1-10 стр.
- [4] Altuhhova O., Matulevičius R., and Ahmed N., “Towards Definition of Secure Business Processes”. Conference “Lecture Notes in Business Information Processing”, pp. 1-15, 2012.
- [5] “Автоматизиране на бизнес процесите във фирмата. Автоматизация на бизнес процесите в CRM. Сравнение на подходите.”, <https://oddagipermarket.ru/bg/biznes-idei/avtomatizaciya-biznes-processov-v-kompanii-avtomatizaciya-biznes.html>, 2019.
- [6] Babiano L., Mayol E., and Pastor J., “Role and Importance of Business Processes in the Implementation of CRM Systems”, pp. 1-8.
- [7] „Управление на бизнес процесите“, 1-8 стр., 2019.
- [8] Краева В., Емилова П., Маринова Н., Маринова К., „Управлението на бизнес процеси – важен инструмент за интеграция на информационната инфраструктура на организацията“. Списание „Диалог“, ИНИ, Извънреден тематичен брой I, 42-50 стр., 2013.
- [9] Деминг У., „Наръчник за прилагане на методологията за усъвършенстване на работните процеси за предоставяне на административни услуги“, 1-67 стр.
- [10] Димитров В., „BPMN за спецификация на бизнес процеси“. Съюз на учените в България, Годишник на секция „Информатика“, vol. 6, 168-174 стр., 2013.
- [11] Vaglini G., Cimino M., “An Interval-Valued Approach to Business Process Simulation Based on Genetic Algorithms and BPMN”. Open Access – Information, vol. 5, pp. 319-356, 2014.
- [12] „Управление и мониторинг на склада /Warehouse Management/ с BORA BS“, <https://www.bora-bg.com/bg/За-нас/Блог/Post/2133/Управление-и-мониторинг-на-склада-Warehouse-Management-с-BORA-BS>, 2020.
- [13] „Как да оптимизирате бизнес процесите чрез интелигентно управление на документи“, FTS sharing your digital future, <http://blog.fts-eu.com/2019/03/11/как-да-оптимизирате-бизнес-процесите/>, 2020.
- [14] Банка ДСК, „Годишно оповестяване на съгласно разпоредбите на регламент 575/2013 относно пруденциалните изисквания за кредитните институции и инвестиционните посредници“, 1-69 стр., 2017.
- [15] Райфайзен Асет Мениджмънт (България) ЕАД, “Правила за управление на риска ДФ „Райфайзен консервативен фонд България“, организиран и управляван от Райфайзен Асет Мениджмънт (България) ЕАД”, 1-35 стр.
- [16] „Управление на вземанията на клиенти“, <http://tmfc-bg.com/en/управление-на-вземанията-от-клиенти>, 2020.
- [17] Стоянов М., „Счетоводно-информационно осигуряване на вземанията от клиенти“, Годишен алманах „Научни изследвания на докторанти“, Брой IX, 340-357 стр., 2016.

- [18] Маринова Н., „Международен маркетинг“. Издателство „Нов български университет“, 1-378 стр., 2017.
- [19] IBS Center for Management Research, “Management Control Systems”, 2th Edition, pp. 1-528.
- [20] Янева Р., „Стратегическо маркетингово планиране в онлайн среда“. МАТТЕХ, Том №1, 376-382, 2016.
- [21] Google, „Всичко за импортирането на данни“, <https://support.google.com/analytics/answer/3191589?hl=bg>, 2020.
- [22] Ledford J., Teixeira J, and Tyler M., “Google Analytics – Third Edition”, Wiley Publishing, pp. 1-406.
- [23] Microsoft, “Импортиране на данни в Business Contact Manager”, <https://support.office.com/bg-bg/article/Импортиране-да-данни-в-business-contact-manager-aa93e491-60ca-441d-8adc-539f4f572a05>, 2013.
- [24] Microsoft, „Какво е Business Contact Manager за Outlook?“, <https://support.microsoft.com/bgbg/office/%d0%ba%d0%b0%d0%ba%d0%b2%d0%be%d0%b5-business-contact-manager-%d0%b7%d0%b0-outlook-47128874-e5d6-4383-9196-17d1d59e5cbd?ui=bg-bg&rs=bg-bg&ad=bg>, 2013.
- [25] “Without Excellent Reports, Your BPM Software is Uncooked”, <https://kissflow.com/bpm/without-excellent-reports-your-bpm-software-is-uncooked/>.
- [26] “How to Build, Run, & Analyze Marketing Reports”, <https://blog.hubspot.com/marketing/running-marketing-reports-ht>.
- [27] Василева Б., „Маркетинг. Практическо ръководство“. Издателство „eM Студио Адвъртайзинг енд Криейтив“ ООД, Варна, 1-137 стр., 2012.
- [28] Morissiette L., Chartier S., “The k-Means Clustering Technique: General Considerations and Implementation in Mathematica”. *Tutorials in Quantitative Methods for Psychology*, vol. 9, issue 1, pp. 15-24, 2013.
- [29] Ali H., Kadhum K., “K- Means Clustering Algorithm Applications in Data Mining and Pattern Recognition”. *International Journal of Science and Research*, vol. 2015, pp. 1577-1854, 2015.
- [30] Rezanikova H., “Cluster Analysis of Economic Data”. *STATISTIKA*, vol. 94, issue 1, pp. 73-86, 2014.
- [31] Boyko N., Komarnytska H., Kryvenchuk Y., Malynovskyy Y., “Clustering Algorithms for Economic and Psychological Analysis of Human Behavior”. *International Workshop on Conflict Management in Global Information Networks*, pp. 1-13, 2020.
- [32] Brauksa I., “Use of Cluster Analysis in Exploring Economic Indicator Differences among Regions: The Case of Latvia”. *Journal of Economics, Business and Management*, vol. 1, № 1, pp. 42-45, 2013.
- [33] Gallo C., “Encyclopedia of Information Science and Technology - Artificial Neural Networks: Tutorial”. *Information Science Reference*, pp. 179-189, 2015.
- [34] Falat L., Pancikova L., “Quantitative Modelling in Economics with Advanced Artificial Neural Networks”. *ELSEVIER, Business Economics and Management Conference*, vol. 34, № 2015, pp. 194-201, 2015.
- [35] Farhat D., “Artificial Neural Networks and Aggregate Consumption Patterns in New Zealand”. *University of Otago Economics Discussion Papers*, No. 1404, pp. 1-35, 2014.
- [36] Vesely A., “Economic Classification and Regression Problems and Neural Networks”. *Ministry of Education, Youth and Sports of the Czech Republic (Project No. MSMT 6046070904)*, pp. 150-157, 2011.

- [37] Safe M., Barouni M., Saif S., "Health Impact on Economy by Artificial Neural Network and Dynamic Ordinary Least Squares". *Journal of Health Managements & Informatics*, vol. 4, № 4, pp. 107-113, 2017.
- [38] Trillas F., "Fuzzy Logic and Modern Economics". *Studies in Fuzziness and Soft Computing*, vol. 325. pp. 55-64, 2015.
- [39] Barbara D, Morillas A., "Some Experiences Applying Fuzzy Logic to Economics". Diaz, Barbara & Morillas, Antonio. (2012). *Studies in Fuzziness and Soft Computing, Soft Computing in Humanities and Social Sciences*, pp. 347-379, 2012.
- [40] Salleh M., Talpur N., Hussian H., "Adaptive Neuro-Fuzzy Inference System: Overview, Strengths, Limitations, and Solutions". *International Conference on Data Mining and Big Data*, pp. 1-8, 2017.
- [41] Hussian H., Slusarczyk B., Kamarudin F., Thaker H., Szczepanska-Woszczyna K., "An Investigation of an Adaptive Neuro-Fuzzy Inference System to Predict the Relationship among Energy Intensity, Globalization, and Financial Development in Major ASEAN Economies". *MDPI, Energies*, vol. 13, issue 850, pp. 1-17, 2020.
- [42] Ambukege G., Justo G., Mushi J., "Neuro-Fuzzy Modeling for Prediction of Customer Price Index". *International Journal of Artificial Intelligence and Applications*, vol. 8, № 5, pp. 33-44, 2017.
- [43] Egbo M., Bartholomew D., "A Discriminant Function Analysis Approach to Country's Economy Status". *Journal of Advanced Statistics*, vol. 2, № 2, pp. 125-136, 2017.
- [44] Yakymova L., Kuz V., "The Use of Discriminant Analysis in the Assessment of Municipal Company's Financial Health". *Economics and Sociology*, vol. 12, № 2, pp. 64-78, 2019.
- [45] Kočišová K., Mišanková M., "Discriminant Analysis as a Tool for Forecasting Company's Financial Health". *ELSEVIER, Procedia - Social and Behavioral Sciences*, vol. 110, pp. 1148-1157, 2013.
- [46] Berrar R., "Bayes' Theorem and Naive Bayes Classifier". *ELSEVIER, Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, pp. 403-412, 2018.
- [47] Krichene A., "Using a Naive Bayesian Classifier Methodology for Loan Risk Assessment: Evidence from a Tunisian commercial bank". *Journal of Economics, Finance and Administrative Science*, vol. 22, № 42, pp. 3-24, 2017.
- [48] Redjeki S., Guntara M., Anggoro P., "Naive Bayes Classifier Algorithm Approach for Mapping Poor Families Potential". *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, № 12, pp. 29-33, 2015.
- [49] Lohar P., Chowdhury K., Afli H., Hasanuzzaman M., Way A., "ADAPT at IJCNLP-2017 Task 4: A Multinomial Naive Bayes Classification Approach for Customer Feedback Analysis Task". *Proceedings of the 8th International Joint Conference on Natural Language Processing, Shared Task*, pp. 161-169, 2017.
- [50] Zebda A., "Decision Trees And Quality Control Decisions". *Journal Of Business & Economics Research*, vol. 2, № 2, pp. 71-80.
- [51] Bucea-Matea-Tonis R., Bucea-Matea-Tonis T., "Decision Tree Method Applied in Economics and Statistics". *Buletinul AGIR*, № 1, pp. 144-148, 2014.
- [52] Tyrychtr J., "Analytical System with Decision Tree for Economic Benefit". *Agris on-line Papers in Economics and Informatics*, vol. 9, № 4, pp. 123-129, 2017.
- [53] Tofan C., "Decision Tree Method Applied in Cost-Based Decisions in an Enterprise". *ELSEVIER, Procedia Economics and Finance*, vol. 32, pp. 1088-1092, 2015.

- [54] Rotim S., Dopsa J., Krakar Z., "Using Decision Trees for Identification of Most Relevant Indicators for Effective ICT Utilization". *Cybernetics and Information Technologies*, vol. 13, № 1, pp. 83-94, 2013.
- [55] Zhang Zh., "Introduction to Machine Learning: K-Nearest Neighbors". *Annals of Translational Medicine*, vol. 4, № 11, pp. 1-7, 2016.
- [56] Hassanat A., Abbadi M., Altarawneh G., Alhasanat A., "Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach". *International Journal of Computer Science and Information Security*, vol. 12, № 8, 2014.
- [57] Imandoust S., Bolandraftar M., "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background". *Int. Journal of Engineering Research and Applications*, vol. 3, issue 5, pp. 605-610, 2013.
- [58] Vainionpää I., Davidsson S., "Stock Market Prediction using the K-Nearest Neighbours Algorithm and a Comparison with the Moving Average Formula", pp. 1-31, 2014.
- [59] Fan G., Guo Y., Zheng J., Hong W., "Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting". *MDPI, Energies*, vol. 12, pp. 1-19, 2019.
- [60] Issah M., Antwi S., "Role of macroeconomic variables on firms' performance: Evidence from the UK". *Cogent Economics & Finance*, pp. 1-18, 2017.
- [61] Kumar V., "Predictive Analytics: A Review of Trends and Techniques". *International Journal of Computer Applications*, vol. 182, № 1, pp. 31-37, 2018.
- [62] Bihi T., Young W., Weckman M., "Artificial Neural Networks for Business Analytics". *Book: Encyclopedia of Business Analytics and Optimization*, pp. 193-208, 2014.
- [63] Huang H., Chang A., Ho Ch., "Using Artificial Neural networks to Establish a Customer-cancellation Prediction Model". *Przegląd Elektrotechniczny*, 89(1), pp.178-180, 2013.
- [64] Rusov J., Misita M., Milanovic D., Milanovic D., "Applying Regression Models to Predict Business Results". *FME Transaction*, 45, pp. 198-202, 2017.
- [65] Eckerson W., "Predictive Analytics: Extending the Value of Your Data Warehousing Investment". *TDWI Best Practices Report*, pp. 1-36, 2007.
- [66] Chandrashekar G., Naidu N., Prakash G., "Predictive Analytics Tools and Techniques". *Global Journal of Finance and Management*, vol. 6, № 1, pp. 59-66, 2014.
- [67] Kalechofsky H., "A Simple Framework for Building Predictive Models". *M Squared Consulting*, pp. 1-19, 2016.
- [68] Bakiya J., "A Non-Technical Introduction to Regression". *Department of Economics Williams College Williamstown*, pp. 1-43, 2013.
- [69] Ekelof L., "A Multiple Linear Regression Model To Assess The Effects of Macroeconomic Factors On Small and Medium-Sized Enterprises. Royal Institute of Technology School of Engineering Sciences, EXAMENSARBETE INOM TEKNIK, GRUNDNIVÅ, 15 HP STOCKHOLM, SVERIGE, pp. 1-62, 2019.
- [70] Akoglu H., "User's Guide to Correlation Ccoefficients". *Turkish Journal of Emergency Medicine*, 18(3), pp. 91-93, 2018.
- [71] Constantin A., Madalina-Gabriela A., Capusneanu S., "Econometric Model Used for GDP Correlation Analysis and Economic Aggregates". *Economic Computation and Economic Cybernetics Studies and Research*, vol. 53, issue 1, pp. 183-197, 2019.
- [72] Prus B., Krol K., Chrobot K., "Analysis of the Correlation Between Socio-Economic Development and Land Prices - A Study of the Zagnańsk Municipality". *Acta Scientiarum Polonorum Formatio Circumiectus*, 2(2), pp. 87-94, 2018.

- [73] Simoin D., Stanciu M., Armaselu S., “Correlation Analysis between Structure Financial System and Economic Growth in Romania”. ELSEVIER, *Procedia Economics and Finance* 32, pp. 1332-1341, 2015.
- [74] Schiltz F., Masci Ch., Agasisti T., Horn D., “Using Regression Tree Ensembles to Model Interaction Effects: A Graphical Approach”. Routledge Taylor & Francis Group, *Applied Economics*, vol. 50, № 58, pp. 6341-6354, 2018.
- [75] Ozhegov E., Ozhegova A., “Regression Tree Model for Prediction of Demand with Heterogeneity and Censorship”. Wiley, *Journal of Forecasting*, vol. 39, issue 3, 2019.
- [76] Woloshco N., “Adaptive Trees: A New Approach to Economic Forecasting”. Organisation for Economic Co-operation and Development, pp. 1-43, 2020.
- [77] Dumitrescu E., Hue S., Hurlin Ch., Tokpavi S., “Machine Learning or Econometrics for Credit Scoring: Let’s Get the Best of Both Worlds”, *HAL Archives-Ouvertes Journal*, pp. 1-43, 2021.

СЪДЪРЖАНИЕ

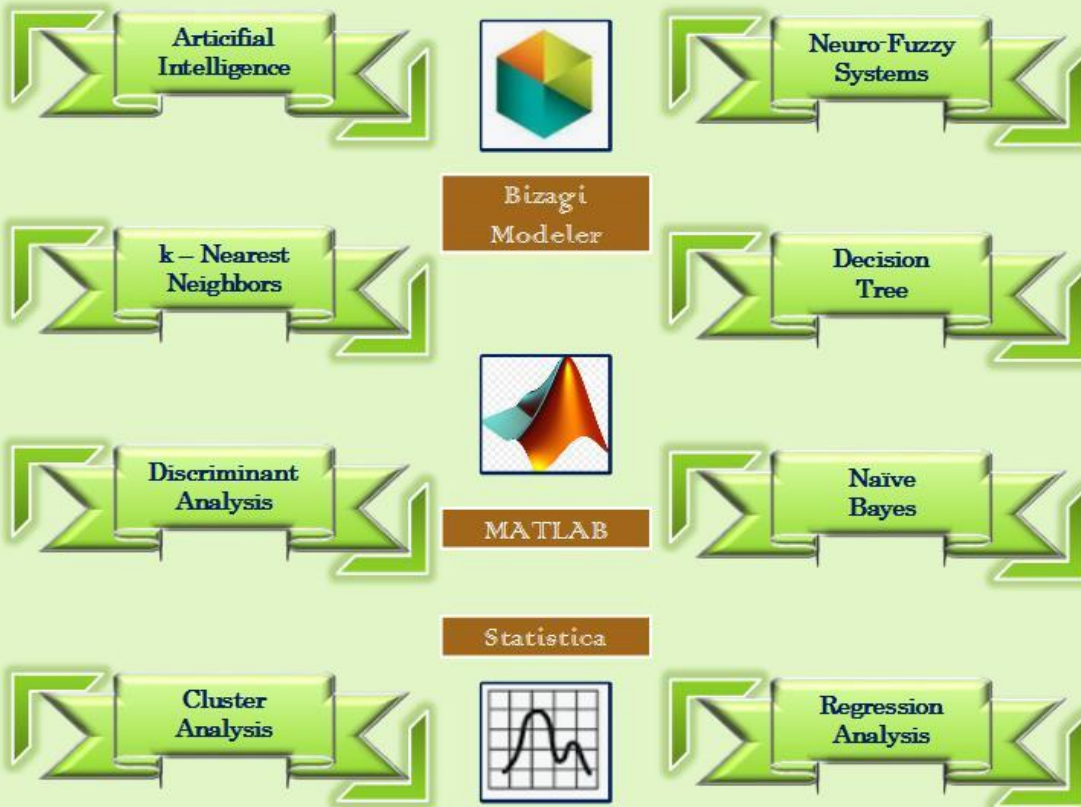
ПРЕДГОВОР	3
МОДУЛ I. КОНЦЕПТУАЛНО МОДЕЛИРАНЕ НА АДМИНИСТРАТИВНИ МАРКЕТИНГОВИ БИЗНЕС ПРОЦЕСИ	5
ЛАБОРАТОРНО УПРАЖНЕНИЕ №1 „БИЗНЕС ПРОЦЕСИ ЗА РЕГИСТРАЦИЯ НА КЛИЕНТИ, ОБРАБОТКА НА КЛИЕНТСКИ ПОРЪЧКИ И ЗАЯВЛЕНИЯ“	5
ЛАБОРАТОРНО УПРАЖНЕНИЕ №2 „БИЗНЕС ПРОЦЕСИ ЗА ЕКСПЕДИЦИЯ И ФАКТУРИРАНЕ, УПРАВЛЕНИЕ НА ДАНЪЦИТЕ И КРЕДИТЕН КОНТРОЛ, ПРОСЛЕДЯВАНЕ НА ВЗЕМАНИЯ И БЛОКИРАНЕ НА НЕКОРЕКТНИ ПЛАТЦИ“	18
ЛАБОРАТОРНО УПРАЖНЕНИЕ №3 „БИЗНЕС ПРОЦЕСИ ЗА МАРКЕТИНГОВИ КАМПАНИИ, ОТВОРЕН ИНТЕРФЕЙС ЗА ИМПОРТИРАНЕ НА ДАННИ И ИЗГОТВЯНЕ НА СПРАВКИ И ОТЧЕТИ“	33
МОДУЛ II. КЛЪСТЕРЕН АНАЛИЗ, ПОДХОДИ ЗА РАЗПОЗНАВАНЕ И КЛАСИФИКАЦИЯ ВЪВ ВРЪЗКА С ДАННИ ПО ИКОНОМИЧЕСКИ РЕГИОНИ	53
ЛАБОРАТОРНО УПРАЖНЕНИЕ №4 „КЛЪСТЕРИЗАЦИЯ НА ИКОНОМИЧЕСКИ ДАННИ ПОСРЕДСТВОМ K-MEANS МЕТОД В СРЕДА МАТЛАВ“	53
ЛАБОРАТОРНО УПРАЖНЕНИЕ №5 “РАЗПОЗНАВАНЕ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ ПОСРЕДСТВОМ ИЗКУСТВЕН ИНТЕЛЕКТ В СРЕДА МАТЛАВ”	64
ЛАБОРАТОРНО УПРАЖНЕНИЕ №6 „ИДЕНТИФИКАЦИЯ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ ЧРЕЗ ПРИЛОЖЕНИЕ НА АДАПТИВНИ НЕВРОННО-РАЗМИТИ ИНТЕРФЕЙСНИ СИСТЕМИ В СРЕДА МАТЛАВ”	75
ЛАБОРАТОРНО УПРАЖНЕНИЕ №7 “ИДЕНТИФИКАЦИЯ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ С ПОМОЩТА НА ДИСКРИМИНАНТЕН АНАЛИЗ И БЕЙС КЛАСИФИКАЦИЯ В СРЕДА МАТЛАВ”	86
ЛАБОРАТОРНО УПРАЖНЕНИЕ №8 „РАЗПОЗНАВАНЕ НА РЕГИОНИ НА ПРИНАДЛЕЖНОСТ НА ИКОНОМИЧЕСКИ ПОКАЗАТЕЛИ НА БАЗАТА НА СТАТИСТИЧЕСКИ МЕТОДИ ДЪРВО НА РЕШЕНИЯТА И K – НАЙ-БЛИЗКИ СЪСЕДИ В СРЕДА НА ПРОДУКТИТЕ STATISTICA И МАТЛАВ”	99
МОДУЛ III. ПРОГНОЗЕН АНАЛИЗ НА ФИНАНСОВИ ДАННИ ЧРЕЗ ИЗКУСТВЕН ИНТЕЛЕКТ, ПРИЛОЖНА СТАТИСТИКА И МАШИННО ОБУЧЕНИЕ	117
ЛАБОРАТОРНО УПРАЖНЕНИЕ №9 „ИЗКУСТВЕН ИНТЕЛЕКТ КАТО ИНСТРУМЕНТ ЗА ПРОГНОЗЕН АНАЛИЗ ВЪВ ВРЪЗКА С ИНВЕСТИЦИОННИ ПОКАЗАТЕЛИ В СРЕДА МАТЛАВ“	117

ЛАБОРАТОРНО УПРАЖНЕНИЕ №10 „КОРЕЛАЦИОНЕН И РЕГРЕСИОНЕН АНАЛИЗ, РЕГРЕСИОННИ ДЪРВЕТА НА РЕШЕНИЯ ЗА ИЗСЛЕДВАНЕ НА ВЗАИМОВРЪЗКИТЕ И ПРОГНОЗИРАНЕ НА ИНВЕСТИЦИОННИ ПОКАЗАТЕЛИ В STATISTICA И MATLAB”	128
ИЗПОЛЗВАНА ЛИТЕРАТУРА	147



BUSINESS PROCESS MODELING

BUSINESS PROCESS MODELING



DATA MINING TECHNIQUES

DATA MINING TECHNIQUES

