

Софийски университет „Св. Климент Охридски“  
Стопански факултет

## **АВТОРЕФЕРАТ**

На дисертационен труд на тема:

**АВТОМАТИЗИРАНИ ПОДХОДИ ЗА УПРАВЛЕНИЕ НА ОПЕРАЦИОННИЯ РИСК**

За придобиване на научна степен  
„доктор на науките“ в направление  
3.8 Икономика

Дисертант: **доц. д-р Антон Антонов Герунов**

София  
2020 г.

Дисертационният труд на тема „Автоматизирани подходи за управление на операционния риск“ се състои от 364 страници, от които 332 страници основен текст, библиография и три приложения. Основният текст е разделен в уводна част, пет глави и заключение. Изследването е илюстрирано с помощта на 141 графики и 79 таблици с основни резултати. Използваната литература е цитирана в библиографската справка, която включва 410 източника на български и чуждестранни автори.

По темата на дисертацията са направени десет публикации под научна рецензия, като осем от тях са в реферирани научни списания в България и чужбина. Сред публикациите са монографичен учебник, който се използва в рамките на преподаването на дисциплината Управление на риска в Стопанския факултет на СУ „Св. Климент Охридски“, три студии, както и пет статии, една от които е в индексирано в Scopus списание. От публикациите пет са на английски език, а останалите – на български език.

Дисертационният труд е преминал предварително обсъждане в рамките на катедра „Икономика и управление по отрасли“ на дата 26.05.2020 г. (Протокол № 120/26.05.2020), спрямо решение на Факултет съвет на Стопанския факултет на СУ „Св. Климент Охридски“ (Протокол № 6/07.05.2020 г., т. 3).

## Съдържание

Увод.....	3
<b>I. Обща характеристика на дисертационния труд.....</b>	<b>3</b>
1) Актуалност и значимост на изследването.....	4
2) Обект и предмет на изследването.....	4
3) Цел и изследователски задачи.....	5
4) Научноизследователска теза .....	6
5) Обхват на изследването.....	6
6) Методология на изследването.....	7
7) Информационно обезпечаване на изследването.....	8
8) Ползност и оригиналност на изследването .....	9
9) Ограничения на изследването .....	10
10) Насоки за бъдеща работа .....	11
<b>II. Структура на дисертационния труд .....</b>	<b>12</b>
<b>III. Основни резултати на дисертационния труд .....</b>	<b>17</b>
Глава 1: Стопанските рискове и подходи за управлението им.....	17
Глава 2: Нови подходи за управление на операционните рискове.....	22
Глава 3: Автоматизирано прилагане на класификационни алгоритми за управление на операционния риск.....	30
Глава 4: Автоматизирано прилагане на регресионни алгоритми за управление на операционния риск.....	41
Глава 5: Автоматизирана система за управление на операционния риск.....	55
<b>IV. Справка за научните и научно-приложните приноси .....</b>	<b>63</b>
<b>V. Публикации по темата на дисертационния труд .....</b>	<b>64</b>
<b>VI. Заключение.....</b>	<b>65</b>

## **Увод**

Съвременните икономики са икономики на риска. В тях огромната част от стопанските дейности се извършват в условия на риск, несигурност и неопределеност (Bullen et al., 2006). Това се дължи отчасти на информационните и властови асиметрии, които се наблюдават и отчасти на фундаменталната несигурност за бъдещето. Затова и управлението на риска се налага като основна задача в съвременните икономически изследвания. Традиционно, значителна част от разработките се фокусират върху финансовите рискове, но в последните години се наблюдава засилен интерес и към други типове риск, като все повече общата рискова експозиция на съвременните организации се разглежда в нейната пълнота. Сред основните групи рискове с потенциално голям ефект върху производителността са тези, които възникват в рамките на обичайната оперативна дейност. В този смисъл тяхното рационално управление с цел извличане на максимална оперативна ефективност е важна задача в рамките на общото изследователско направление.

Паралелно със задълбочаване на изследванията в областта на управлението на риска и внедряването на основни резултати в практиката, в съвременните организации тече процес на дигитална (цифрова) трансформация, която представлява значителната промяна на организационните процеси, структури и системи поради разширеното използване на цифрови технологии и платформи (Chaffey, 2015). Тенденцията за увеличаване на обхвата на автоматизацията на бизнес процеси следва да обхване и ключови дейности като управлението на операционните рискове. Настоящата интердисциплинарна дисертация има за цел да обоснове възможността за автоматизация на този процес, като предложи общ алгоритъм за управление на операционните рискове, набор от методи за тяхната количествена оценка, както и технологична архитектура, която успешно да ги интегрира в рамките на специализирана информационна система.

### **I. Обща характеристика на дисертационния труд**

Управлението на операционните рискове се налага като водеща задача в сферата на икономиката и бизнеса. В резултат от цифровата трансформация на съвременните организации този процес също се превръща в основен кандидат за цифровизация. За да може това да протече по успешен начин следва да се създаде общ алгоритъм за автоматизирано управление на операционните рискове, да се подберат подходящи методи за количествена оценка на риска и да се предложи подходяща архитектура на информационна система, която да интегрира всичко това. Основната задача на настоящата дисертация е да изгради тази система от различни артефакти и резултати, които едновременно да подпомогнат изследванията в областта, но да могат да бъдат полезни и в научно-приложен план. За постигане на това е осъществено интердисциплинарно изследване в изпълнение на поставените научноизследователски задачи, което има за цел да разшири познанието за подходите и методите за управление на операционния риск в дигитална среда.

## **1) Актуалност и значимост на изследването**

Растящата сложност на съвременната икономика води и до рязко нарастване на нивото на рискова експозиция на съвременните организации и оттам – на нуждата от ефективно и ефикасно управление на тези рискове (Chernobai, 2018). Увеличеният брой стопански трансакции и агенти на глобално интегрираните пазари увеличават възможностите за възникване на риск, а цифровата трансформация обуславя прехода от аналогови към дигитални рискови събития. В същото време теорията и практиката за взимане на решения в условия на несигурност и управлението на този процес стъпват до голяма степен върху идеята за аналогови дейности, извършвани от експерти в областта. Тази разлика между обективната реалност на съвременната икономическа среда и приложимата теория за нея обуславя нуждата от разширяване на изследванията в областта.

Допълнително отчитаме ръста на операционните рискове и ясно изразената необходимост за подобряване на тяхното управление, като някои автори дори говорят за „експлозия“ от операционни рискове (Power, 2005). Отчитайки, че този тип рискове произтичат от четири основни групи фактори – хора, процеси, системи и външни събития (Chernobai et al., 2012; Leone et al., 2018), то растящото значение на тези фактори обуславят и нарастващото значение на тяхното управление и рисковете, произтичащи от това. Намалявайки отрицателните последици от хората, процесите, информационните системи и външната среда и увеличавайки ползите от тях съдържа в себе си значителен потенциал за отключване на нова бизнес стойност и подобряване на конкурентното позициониране на организациите в съвременна среда.

На трето място, технологичните развития правят възможно използването на големи масиви от данни за подобряване на широк набор от бизнес процеси, включително тези по управление на операционните рискове (Davenport et al., 2012; Byrne & Corrado, 2017). Тази технологична промяна предполага и използването на нови статистически и иконометрични методи за решаване на важни за организациите икономически задачи. Недостатъчните изследвания на приложението на методи от сферата на машинното самообучение и тяхната успешна интеграция в цялостния управленски процес в сферата на риска също е важно основание за разширяване на изследванията в тази област.

## **2) Обект и предмет на изследването**

Обект на изследването са операционните рискове, пред които е изправена съвременната организация. Минасян (2012) ги дефинира като рисковете от загуба, произтичаща от неадекватни или недобре функциониращи вътрешни процеси, хора и системи, или от външни събития, като в тази категория включва и правен риск. В този смисъл обектът на изследването е група от взаимосвързани рискове, като всеки от тях представлява отделна икономическа задача, предполагаща индивидуален подход.

Предмет на изследването е процесът по управление на операционните рискове, като от особен интерес е възможността за алгоритмизиране на този процес и превръщането му в набор от стандартизирани стъпки (дейности), които да подлежат на автоматизация в рамките на специализирана информационна система.

### **3) Цел и изследователски задачи**

Целта на настоящата работа е да се изгради напълно автоматизиран процес за управление на операционните рискове, който да може да се възползва от широк набор от количествени методи за оценка и да бъде подкрепен от съответна специализирана информационна система. Автоматизираното управление на рисковете има потенциал не просто да подобри организационната ефективност чрез подобряване качеството на взетите решения и скоростта на тяхното изпълнение, но и да намали разходи като облекчи част от работата на отговорниците по риска и експертите, които ги подкрепят.

За изпълнение на така поставената цел извеждаме следните изследователски задачи:

1. Преглед и критичен анализ на подходите и методите за управление на риска с фокус върху управлението на операционните рискове;
2. Извеждане на основните икономически, социални и технологични тенденции, които следва да бъдат взети под внимание при автоматизацията на управленския процес;
3. Операционализиране на дефиницията на понятието за операционен риск по начин, който може да се приложи в рамките на информационна система;
4. Извеждане на общ алгоритъм за управление на операционния риск, който подлежи на автоматизация;
5. Анализ на методите за оценка на операционния риск при задачи от класификационен тип и извеждане на най-оптимални от гледна точка на прогностичната точност алгоритми, които могат да обслужат нуждите на алгоритъма;
6. Анализ на методите за оценка на операционния риск при задачи от регресионен тип и извеждане на най-оптимални от гледна точка на прогностичната точност алгоритми, които могат да обслужат нуждите на алгоритъма;
7. Изграждане на референтна архитектура на управленска информационна система, която да позволи прилагането на предложения алгоритъм и методи по автоматизиран начин, както и да обезпечи безпроблемно включване в бизнес процесите и общата организационна архитектура;
8. Анализ на социалните и етични аспекти от внедряването на информационна система за управление на операционните рискове с възможност за автономно взимане на решения.

Изпълнението на поставените изследователски задачи води и до изпълнение на общата цел на дисертацията, като обуславя натрупването на достатъчно резултати, които да позволят изграждането и внедряването на автоматизиран процес за управление на операционен риск.

#### **4) Научноизследователска теза**

Основната тези на представеното изследване, е че процесът по управление на операционните рискове подлежи на пълна автоматизация чрез прилагане на алгоритмизиран управленски подход, набор от алгоритми, подходящи за анализ на големи масиви от данни и изграждане на подходяща информационна система, надграждаща известни в литературата референтни архитектури (пр. Klein et al., 2016).

Работните хипотези на изследването са:

1. Всички дейности, извършвани от експерти в процеса по управление на операционните рискове могат да бъдат автоматизирани.
2. Приложението на класификационни алгоритми от сферата на машинното самообучение за количествена оценка на рисковете може да подобри прогностичната точност спрямо традиционно използваните иконометрични методи и оттам да доведе до по-високо качество на резултатите и икономическа стойност.
3. Приложението на регресионни алгоритми от сферата на машинното самообучение за количествена оценка на рисковете може да подобри прогностичната точност спрямо традиционно използваните иконометрични методи и оттам да доведе до по-високо качество на резултатите и икономическа стойност.
4. Използването на няколко критерия за оценка на нивото на риск в рамките на автоматизиран алгоритъм ще доведе до групиране (клъстериране), а не до разнопосочност на резултатите. По този начин се гарантира надеждността на предложената система.

Тестването на работните хипотези се извършва спрямо тяхната специфика. Хипотеза 1 се тества качествено, като към нея се прилага подходът на проектиращата наука (Hevner et al., 2004; Arnott & Pervan, 2016)<sup>1</sup>. Създаването на артефакт, който да отговаря на определени изисквания отхвърля нулевата хипотезата в този случай. Хипотези 2 и 3 се тестват количествено, като се сравняват прогностичните точности на предложените тук методи с прогностичната точност на наивната прогноза и на тази на класическите регресионни иконометрични методи. Хипотеза 4 се тества с комбинация между количествени и качествени методи, като се отчита доколко има съответствие между различните критерии за нива на риск и се отчете статистическото определение за аномална (екстремна стойност).

#### **5) Обхват на изследването**

Изследването обхваща разглеждане на дванадесет основни ситуации на операционен риск, към които се прилагат предложеният общ алгоритъм, както и набор от специализирани методи от областта на машинното самообучение. Тези дванадесет ситуации по същество представляват дванадесет различни задачи, като те са описани и в съответния набор от данни. Те се отнасят до управление на операционния риск при следните ситуации:

---

<sup>1</sup> От англ. – design science.

1. Провеждане на кампания по директен маркетинг
2. Провеждане на операции с кредитни карти
3. Отпускане на кредити
4. Управление на отношенията с външни партньори
5. Дейности по електронна търговия
6. Прекомерни отсъствия от работното място
7. Онлайн комуникация
8. Оценка на цените на активи (недвижими имоти)
9. Резки промени на пазарното търсене
10. Обработка на сигнали от клиенти
11. Маркетингова комуникация чрез социалните мрежи
12. Управление на търсенето в електронен магазин

От времева гледна точка, данните са преобладаващо от последните 10 години, като това е вярно за 10 от разгледаните 12 масива. Приложението на предложените подходи и алгоритми към сравнително голям набор от оперативни ситуации както в аналоговия, така и в дигиталния свят, обезпечава тяхната гъвкавост и възможността за трансфер към широк спектър от други задачи в сферата на операционния риск.

## **6) Методология на изследването**

Настоящата дисертация използва както общонаучни, така и специализирани методи за изпълнение на поставените изследователски задачи. Общата методология се базира на системния подход и следва общо-научни принципи като обективност, преход от конкретно към абстрактно, конкретизация и единство между теория и практика. В рамките на литературния обзор и критичната оценка на съществуващите типологии на риска, както и методите и подходите за неговото управление са използваните методи на анализ и синтез, както и индуктивни и дедуктивни методи за достигане до основните изводи. Този методологически инструментариум е подходящ за такъв тип задачи, тъй като позволява обзор и обобщение на широк набор от литературни източници, а анализа и синтеза позволяват извеждане на нови научни резултати от анализ на вече съществуващи изследвания. Системният и интердисциплинарен подход, от друга страна, спомага за обединението и интегрираното разбиране на резултати, идващи от различни дисциплини и приложни полета – икономика, управление на риска, иконометрични и статистически методи и управленски информационни системи.

При изграждането на общия алгоритъм за управление на операционния риск, както и при извеждането на референтна архитектура на подходяща информационна система е удачно да се използва не само анализ и синтези, но и умения за проектиране на нови артефакти. Поради тази причина тук са използвани подходите и методите на проектиращата наука (Hevner et al., 2004; Arnott & Pervan, 2016), като особено внимание е обърнато на методите за прототипизиране. Този подход позволява да се изградят нови полезни артефакти с висока научно-приложна стойност на базата на набор от определени изисквания. Тези изисквания



могат да са както директно следващи от функцията на артефакта (пр. изисквания управленската информационна система да изпълнява всяка дейност от алгоритъм за управление), така и нефункционални изисквания, следващи от други източници (пр. технологични ограничения или етични съображения). В този смисъл предложените алгоритъм и референтна архитектура на система са резултат от комбинацията между общонаучната и специализираната методология на проектиращата наука.

Анализът на различните типове статистически алгоритми за машинно самообучение и подбора на най-точните сред тях за нуждите на автоматизираното управление предполага специализирана методология. За целта са използвани стандартни методи и подходи от сферата на статистиката и иконометрията. При моделирането на ситуациите на операционен риск са използвани както познати и сравнително стандартни иконометрични модели (корелации, линейна и логистични регресии и др.), така и набор от авангардни методи от сферата на машинното самообучение (невронни мрежи, случайни гори, бейсови методи, машини с подкрепящи вектори и др.). За оценка на важността и приноса на определени променливи към моделите са изведени статистически хипотези и те са проверени формално чрез изчисляване на тестови статистики и точни нива на значимост. За оценката на алгоритмите са дефинирани и приложени статистически критерии за прогностична точност и те са сравнени количествено. Количествените статистически методи позволяват прецизно тестване на поставените хипотези (основно Хипотези 2 и 3, а отчасти – Хипотеза 4), като в същото време разширяват методологическото познание за оптималните методи за управление на операционния риск. Всички изчисления са направени с помощта на езика за статистическо програмиране R и пакетите към него (Kuhn, 2008).

## **7) Информационно обезпечаване на изследването**

Информационното обезпечаване на настоящето изследване е ключово за успешното му провеждане. Използваните количествени данни в дисертацията са три основни типа. Първите са статистически данни, идващи или от първичен източник или цитирани от други автори. И в двата случая произходът им е ясно маркиран. Вторите са данните, използвани при оценка на ситуациите на операционен риск, което в общия случай представлява анализ на вторични данни. Третите са симулирани данни, като за целта се използва вградения генератор на случайни числа в програмния език R.

Процесът по управление на рисковете има редица характеристики, които носят предизвикателства при моделирането му. Първо, той е силно информационно-интензивен и изисква определен обхват и ниво на качество на данните. Нещо повече, успешното прилагане на някои от потенциалните методи за количествена оценка предполага наличието на сравнително големи извадки, които да позволят коректното и неизместено изчисление на техните параметри. На трето място, операционният риск има редица различни проявления в различните ситуации, които могат да се опишат от различни по формат и обхват данни, като техният формат, структура, тип и вид не могат да се стандартизират – т.е. няма една задача за

оценка и управление, а поредица от различни задачи. Всичко това предполага и по-прецизен подбор на използваните данни.

Мнозинството от извършените анализи използват данни за различни онлайн и офлайн ситуации на операционен риск, предоставени от изследователи в тази и сходни области. Пълният набор от използвани данни е представен в *Таблица 1*. Извън обезпечаване на качеството и необходимия размер на извадките, тези данни са тествани най-често при анализ и разработка на количествени методи. Това позволява резултатите от дисертацията директно да надградят утвърдени разработки, но и да бъдат съпоставени с резултатите, получени от съответните автори, като се прецени дали предложените тук алгоритми представляват подобрене спрямо оригинално използваните такива. Това е така при всеки един от случаите.

*Таблица 1: Източници на данни за представените анализи*

№	Тип ситуация на операционен риск	Източник на данни
1	Провеждане на кампания по директен маркетинг	Moro et al., 2014
2	Провеждане на операции с кредитни карти	Yeh & Lien, 2009
3	Отпускане на кредити	Eggermont et al., 2004; Hofmann, 1994
4	Управление на отношенията с външни партньори	Zieba et al., 2016
5	Дейности по електронна търговия	Sakar et al., 2018
6	Прекомерни отсъствия от работното място	Martinian et al., 2012
7	Онлайн комуникация	Fernandes et al., 2015
8	Оценка на цените на активи (недвижими имоти)	Yeh & Hsu, 2018
9	Резки промени на пазарното търсене	Ferreira et al., 2016
10	Обработка на сигнали от клиенти	Amaral et al., 2018
11	Маркетингова комуникация чрез социалните мрежи	Dehouche & Wongkitrungrueng, 2018
12	Управление на търсенето в електронен магазин	Chen et al., 2012

## **8) Ползност и оригиналност на изследването**

Ползността на изследването произтича от постигнати резултати в научен и научно-приложен план, като следва да се отбележи, че те могат да използват в повече от едно поле на научен интерес. Приложимостта им е в полетата на икономическата наука и науката за управление на риска, бизнес управлението и внедряването на управленски информационни системи. Първо, разработеният общ алгоритъм за автоматизирано управление позволява да се приложи единен управленски подход към всички ситуации на операционен риск, с което да стандартизира и оптимизира този бизнес процес в съвременната организация. Нещо повече, алгоритъмът е изведен с изричната идея да може да бъде приложен като част от усилията по внедряване на инструменти на дигиталната трансформация и поради това предвижда възможности за частична или пълна автоматизация както на основни дейности, както и на процеса в неговата цялост. Поради това този подход може да бъде използван както за рамка или основа на бъдещи изследвания в областта на управление на рисковете и цифровата трансформация, така и в научно-приложен контекст при внедряване на такива решения в структурите на частния и публичния сектор.

Второ, анализът на алгоритмите за количествена оценка на рисковете показва кои методи е удачно да се прилагат в решаването на задачи от сферата на управлението на риска. Подобряването на прогностичната точност на различни аналитични дейности и системи има потенциала да подобри взимането на организационни решения и оттам да генерира подобрения на производителността, спад на разходите и потенциални нови конкурентни предимства. Допълнително, резултатите относно разгледаните алгоритми могат сравнително лесно да бъдат генерализирани и трансферирани към други полета, разчитащи на анализа на големи масиви от данни, което да е от полза за съответните научни изследвания и приложни разработки в тези полета. От методологична гледна точка сравнението на различни алгоритми и методи за решаване на сходни задачи насочва изследователите към оптималните такива, което позволява както тяхното по-широко използване, така и потенциално развитие.

Трето, изведената референтна архитектура на управленска информационна система за управление на операционните рискове в научен план може да бъде използвана като отправна точка за бъдещи изследвания, а в приложен – като артефакт, който да улесни процеса по разработване и внедряване на подобно решение. Четвърто, в рамките на дисертацията се извеждат и операционализират редица дефиниции, параметри и критерии, чието поясняване има подчертана научно-приложна полза. Очертават се подходящи критерии за автоматичен избор на оптимален алгоритъм между различни алтернативи. Показва се и приложението на алгоритми за ненадзиравано самообучение за решаване на задачи от сферата на риска, като на базата на консенсусен критерий се демонстрира тяхното съвместяване. В литературата такива подходи са ограничени предимно до рискове от финансови измами и рискове в областта на информационната сигурност (кибер-рискове), но дисертацията подчертава как те могат да бъдат приложени и към операционни рискове, произтичащи от потребителското поведение.

## **9) Ограничения на изследването**

Представеното изследване има и някои ограничения, които следва да бъдат взети под внимание, така че получените резултати да бъдат тълкувани коректно в тези рамки. На първо място следва да се има предвид, че се изследва голям, но все пак ограничен набор от ситуации, включващи операционен риск. Дисертацията не покрива изчерпателно всички възможни операционни рискове, нито всички възможни съответстващи набори от данни, които ги описват. Поради това е възможно при прилагането на алгоритъма, набора от методи за количествена оценка и при референтната архитектура на системите към нова и неизследвана ситуация да се наложи тяхната адаптация или разширение.

На второ място, с цел обезпечаване на сравнимост между различните алгоритми, при тяхното обучение са използвани едни и същи стойности на сходни параметри. Това важи с особена сила при оценката на алгоритмите за ненадзиравано самообучение DBSCAN и LOF. Макар това да обезпечи възможност за съпоставка на дадените алгоритми, то е удачно да се осъществи и формален процес на калибрация, който итеративно да търси оптималните стойности на тези

параметри при зададен определен критерий (пр. процент потенциално рискови наблюдения). Възможно е при промяна на параметрите на някои модели да настъпят и малки промени на получените резултати, като това е удачно да се изследва в повече дълбочина.

На трето място следва да се отчете сравнително ограниченият обхват на модула за управление на бизнес правила. Той се ангажира с автоматизацията на действията, които се извършват за управление на операционните рискове и в настоящата си форма предполага наличие на конкретни правила (бизнес логика) за извършване на определени дейности при идентифициране на потенциален риск (пр. при откриване на рискова транзакция в реално време да се изпрати заявка до обработващата система за нейното прекратяване). Това предполага индивидуално дефиниране на правилата в рамките на всяка организация и за всеки отделен тип операционен риск. Макар това да е стандартния подход на настоящето ниво на технологично развитие и организационна зрялост, то с развитието на технологиите може да се очаква този модул да се замени с по-пълноценна версия на ограничен изкуствен интелект, така че поне част от правилата да се генерират автоматично.

Този въпрос е от особена важност от гледна точка и на това как системата би реагирала при наличие на сериозни външни шокове. Тези шокове условно могат да бъдат разделени на две групи. При първата наблюдаваме случаен шок или извънредно отклонение, като този тип шокове могат идеално да бъдат идентифицирани от предложеното решение. Вторият тип шокове са по своята същност първите наблюдения след превключване на моделираната система към нов режим и предполагат известно време за адаптация и обучение на моделите спрямо това ново базисно ниво. Наличието на задълбочени бизнес правила или приложението на специализиран изкуствен интелект може да съкрати този срок на адаптация.

На четвърто място като потенциално ограничение може да се отбележи сравнително ограничената оценка на етичните аспекти при въвеждане на автоматизираната управленска информационна система. Оценката на етични аспекти в случая на автономни дигитални агенти е още в зародиша си и макар тя да е в съзвучие със стратегическите насоки за развитие на изкуствения интелект в рамките на Европейския съюз (European Commission, 2019), методологиите и подходите за целта са сравнително ограничени и подлежат на допълнителна разработка и разширение.

## **10) Насоки за бъдеща работа**

Насоките за бъдеща работа в областта на изследваната проблематика в голяма степен следва от идентифицираните ограничения. На първо място е удачно да се разшири обхвата на разгледаните ситуации на операционен риск, като се запази баланса между ситуации, типични за аналоговия свят (физически или присъствени бизнес процеси) и такива, които са типични за дигиталния свят (цифрови бизнес процеси). Разглеждането на повече ситуации и приложението на изведения алгоритъм и количествени методи за оценка към него има

потенциала да доведе до допълнителна валидация на представените тук резултати или да ги надгради и разшири, така че те да са от още по-голяма полза за научните среди и бизнеса.

На второ място има значителен потенциал да се работи в посока разширяване на модула за управление на бизнес правила. Това е именно дейността, която подлежи най-трудно на автоматизация поради спецификата както на всяка индивидуална ситуация на операционен риск, така и на начина, по който различните организации адресират тези ситуации. Нещо повече, не винаги пълният набор от бизнес правила е ясно изведен, като някои от тях могат да са в имплицитна форма. Поради това е реалистично като първа стъпка от работата по разширяване на този модул да се анализират данни за предприети действия спрямо даден идентифициран операционен риск и на тази база да започне автоматично извеждане на правила (Raad et al., 2017; Mutlu et al., 2018). Този подход има потенциала да автоматизира генерирането на правила, а на по-късен етап може да се интегрира и по-пълноценен агент с ограничен изкуствен интелект.

Като трета насока за бъдеща работа се откроява разработването на методология за икономическа и етична оценка на управленски информационни системи с възможности за взимане на автономни решения. Предвид нарастващото значение на информационните активи за конкурентното позициониране на съвременните организации и постоянно увеличаващия се обхват на действие на информационните системи е удачно да се изведе формална методология за оценка на техните икономически, социални, технологични и етични аспекти.

## **II. Структура на дисертационния труд**

Дисертацията е структурирана в пет глави, увод, заключение, библиография и три приложения. Всяка от главите се фокусира върху постигане на една или повече от поставените цели, като представя обобщения и нови резултати в отговор на потребността от по-добро управление на операционния риск.

### **Съдържание**

<b>Увод</b> .....	<b>5</b>
1) Актуалност и значимост на изследването .....	6
2) Обект и предмет на изследването .....	6
3) Цел и изследователски задачи.....	7
4) Научноизследователска теза .....	8
5) Обхват на изследването.....	8
6) Методология на изследването.....	9
7) Информационно обезпечаване на изследването.....	10
8) Структура на изследването.....	11

<b>Глава 1: Стопанските рискове и подходи за управлението им .....</b>	<b>13</b>
1.1. Въведение .....	13
1.2. Типологии на рисковете .....	17
1.3. Стандартни подходи за управление на риска .....	35
1.4. Разширен процес за управление на риска.....	42
1.5. Практики при управление на операционния риск .....	46
1.6. Качествени подходи за оценка на риска .....	49
1.7. Количествени подходи за управление на риска.....	57
1.8. Общи стратегии за управление на риска .....	69
1.9. Основни изводи .....	71
<b>Глава 2: Нови подходи за управление на операционните рискове .....</b>	<b>73</b>
2.1. Въведение .....	73
2.2. Структурни промени на икономическата среда.....	73
2.3. Технологични промени на икономическата среда .....	79
2.4. Алгоритми за оценка на операционните рискове .....	85
2.5. Нови тенденции при управлението на операционните рискове.....	119
2.6. Общ подход за управление на операционните рискове.....	124
2.7. Основни изводи .....	136
<b>Глава 3: Автоматизирано прилагане на класификационни алгоритми за управление на операционния риск.....</b>	<b>139</b>
3.1. Въведение .....	139
3.2. Критерии за избор на оптимален алгоритъм .....	140
3.3. Избор на оптимален алгоритъм за управление на риска при директен маркетинг.....	145
3.4. Избор на оптимален алгоритъм за управление на риска при операции с кредитни карти ..	156
3.5. Избор на оптимален алгоритъм за управление на риска при отпускане на кредити.....	167
3.6. Избор на оптимален алгоритъм за управление на риска при управление на отношенията с външни партньори.....	178
3.7. Избор на оптимален алгоритъм за управление на риска при електронна търговия .....	189
3.8. Основни изводи .....	200
<b>Глава 4: Автоматизирано прилагане на регресионни алгоритми за управление на операционния риск.....</b>	<b>203</b>
4.1. Въведение .....	203
4.2. Критерии за избор на оптимален алгоритъм .....	204
4.3. Критерии за определяне на дадено наблюдение за рисково .....	207
4.4. Избор на оптимален алгоритъм за управление на риска от прекомерно отсъствие от работното място .....	209
4.5. Избор на оптимален алгоритъм за управление на риска при онлайн комуникация .....	220
4.6. Избор на оптимален алгоритъм за управление на риска при оценка на недвижими имоти	233

4.7.	Избор на оптимален алгоритъм за управление на риска при промени на пазарното търсене	243
4.8.	Избор на оптимален алгоритъм за управление на риска при обработка на сигнали от клиенти .....	252
4.9.	Основни изводи .....	264
<b>Глава 5: Автоматизирана система за управление на операционния риск .....</b>		<b>268</b>
5.1.	Въведение .....	268
5.2.	Автоматизация на управленския процес чрез използване на информационни системи .....	269
5.3.	Информационно обезпечаване на системата .....	279
5.4.	Обща архитектура на автоматизирана информационна система за управление на операционния риск .....	290
5.5.	Модул за автоматично търсене на рискови наблюдения на ниво индивидуални транзакции	302
5.6.	Модул за автоматично търсене на рискови наблюдения на ниво индивидуални агенти .....	310
5.7.	Етични аспекти при въвеждането в продуктивен режим на информационната система .....	320
5.8.	Основни изводи .....	327
<b>Заключение .....</b>		<b>329</b>
1)	Обобщение на основните резултати .....	329
2)	Научни и научно-приложни приноси .....	331
3)	Полезност и оригиналност на изследването .....	332
4)	Ограничения на изследването .....	334
5)	Насоки за бъдеща работа .....	335
<b>Използвана литература .....</b>		<b>337</b>
<b>Приложение 1: Апробирани класификационни алгоритми .....</b>		<b>357</b>
<b>Приложение 2: Апробирани регресионни алгоритми .....</b>		<b>360</b>
<b>Приложение 3: Разпределение на извадките спрямо изчислените модели .....</b>		<b>363</b>

**Първа глава** представлява преглед на литературата в областта на управлението на риска, като впоследствие акцентът се поставя върху операционните рискове. Разгледани са основните типологии на рисковете като цяло и на операционните рискове в частност, като е направена критична оценка на всяка от тях. Прегледани са количествените и качествените методи за управление на операционния риск, както и общите стратегии и добри практики за този процес. В тази глава е застъпена тезата, че макар количествените методи да предлагат най-висока прецизност и ефективност на процеса, то в настоящия си вариант те са прекомерно зависими от наличието на сравнително голям брой квалифицирани експерти, за да бъдат успешно приложени. Това и подчертава нуждата от пълна или частична автоматизация на този управленски процес с помощта на специализирана информационна система.

**Втора глава** се фокусира върху преглед на новите развития в областта на управлението на операционните рискове и на база на проведения преглед на литературата тя очертава четири основни тенденции в актуалните изследвания. На първо място, се наблюдава засилен фокус върху конкретни индустрии и специфични ситуации на управление на операционния риск, което отчасти отразява глобалната мега-тенденция за персонализиране на стоки, услуги и процеси. На второ място, в литературата се забелязва и задълбочаване на изследванията във връзка с риска и навлизането на информационните и комуникационни технологии, като броя на изследванията в областта на за рисковете в областта на информационната сигурност бележи ускорен растеж. В актуалните изследвания е видимо и навлизане на нови и авангардни методи за оценка на операционния риск и взимане на адекватни управленски решения в условия на несигурност. На последно място се отчита и тенденция към постигане на интегрирана представа за общата рискова експозиция на организацията и намален фокус върху синтетичното разделяне на определените рискови групи.

В контекста на новите развития в областта на глобалната среда, технологиите и изследванията в областта на операционните рискове и стъпвайки върху утвърдени методологии за анализ, в тази глава се извежда и общият алгоритъм за автоматизирано управление на операционните рискове. Той се състои от осем стъпки, са отделени в пет основни управленски етапа - дефиниция на проблема, информационно обезпечаване, обучение на модел за управление на операционния риск, приложение на модела и действия за управление. Така представените етапи отчитат основните фази в управленския процес, но и позволяват алгоритмизиране на действията, така че те да могат да бъдат приложени в рамките на специализирана информационна система.

**Трета глава** си поставя за цел да идентифицира подходящи алгоритми за количествена оценка на операционните рискове в случаите на дискретен избор. За тази цел се тестват 136 различни алгоритъма от сферата на статистиката и машинното самообучение, като те се прилагат за решаване на пет класификационни задачи от сферата на операционния риск. Тези ситуации и съответстващите им бази данни отчитат следните случаи, предполагащи управление на риска: провеждане на кампания по директен маркетинг, провеждане на операции с кредитни карти, отпускане на кредити, управление на отношенията с външни партньори, дейности по електронна търговия. За всеки от тези масиви данни са изчислени тези 136 алгоритъма и на база на прогностичната им точност, измерена чрез площта под кривата на работната характеристики, са подбрани десетте метода с най-висока точност.

Сред различните задачи се наблюдава преобладаващо добро представяне на методите от семейството на случайната гора, а от класическите иконометрични методи висока точност се постига от линейния дискриминантен анализ. Допълнително тук се докладват и мярка за комплексност, отчитаща изчислителните ресурси, необходими за оценката на всеки от алгоритмите. Резултатите показват, че алгоритмите с най-висока прогностична точност не се характеризират задължително с най-висока ресурсоемкост. Това дава основание да се изкаже тезата, че изборът на оптимален алгоритъм следва да бъде извършена чрез постигане на



баланс между разходите за него, измерени чрез необходимите изчислителни ресурси и ползите от него, измерени чрез неговата прогностична точност.

**Четвърта глава** разглежда и идентифицира подходящи алгоритми за количествена оценка на риска в случаите, когато целевата променлива на задачата е продължителна. Това предполага тестването на алгоритми с регресионен характер, като тук отново са анализирани както традиционни иконометрични методи, така и такива от сферата на машинното самообучение. Общо се тестват 109 различни алгоритъма, като те се прилагат за решаване на пет ситуации от сферата на операционния риск с регресионен характер. Тези ситуации включват управлението на прекомерни отсъствия от работното място, онлайн комуникацията, оценка на цените на активи (недвижими имоти), резки промени на пазарното търсене и обработка на сигнали от клиенти.

Отново всичките алгоритми са оценени за всеки един от масивите данни и се докладва мярка за комплексност, отчитаща ресурсната интензивност на всеки от разгледаните методи. Тук също се наблюдава много добро представяне на методите от семейството на случайната гора, но така също и на невронните мрежи. За разлика от класификационните задачи, в този случай класическите иконометрични методи на линейна регресия имат чувствително по-слабо представяне спрямо методи от областта на машинното самообучение. При задачите с регресионен характер отново наблюдаваме сравнително слаба връзка между прогностична точност и ресурсна интензивност, което позволява да се направи оптимизационен избор между ползи от дадения алгоритъм и разходите за неговата оценка.

**Пета глава** разглежда възможностите за интегриране на предложения общ алгоритъм за управление на операционните рискове и съответните методи за количествена оценка на риска в рамките на специализирана информационна система. Като начало се разглеждат типовете информационни системи и различни референтни архитектури на групата на управленските информационни системи. Стъпвайки върху утвърдени архитектури и добри практики, с помощта на методи от проектиращата наука се извежда референтна архитектура на информационна система за автоматизирано управление на операционния риск. Системата съдържа четири основни подсистеми, които целят да автоматизират предложените в алгоритъма дейности – това са подсистемата за съхранение и обработка, подсистемата за моделиране, подсистемата за анализи и подсистемата за управление. Допълнително три хоризонтални модула (сигурност и контрол на достъпа, управление на системата и интеграция) допълват системата и обезпечават изпълнението на нефункционалните изисквания към нея.

Предлага се и нарочен модул за автоматизация на процеса по определяне на целевата променлива като за целта се използват четири алгоритъма за ненадзиравано самообучение. Извежда се консенсусен критерий за аномалност на базата на тези четири алгоритъма, като наблюденията, класифицирани като аномални от три от тях се определят за рискови, а тези класифицирани от четири – за високо-рискови. Този подход е приложен и тестван в две допълнителни нови ситуации на операционен риск – при маркетингова комуникация чрез

социалните мрежи и при управление на търсенето в електронен магазин, като неговата полезност е показана. Разширяването на информационната система с възможности за автономно взимане на решения я доближава до някои характеристики на ограничения изкуствен интелект. Поради това е направена и оценка на съответствието с етични изисквания за автономни агенти, като основният извод е, че така предложената система има много добро ниво на съответствие с тези изисквания.

### **III. Основни резултати на дисертационния труд**

#### **Глава 1: Стопанските рискове и подходи за управлението им**

Последните два века стопанска история се характеризират с експоненциален растеж на световната икономика – както в абсолютно изражение, така и на глава от устойчиво нарастващото население. Макар често тези фактори да се възприемат като еднозначно положителни, нарастващата взаимосвързаност и зависимост на различните сектори и ускореното темпо на внедряване на иновации в икономическите дейности обуславят еволюцията на съвременната икономика като комплексна и трудно предвидима система.

Макар стопанският риск винаги да е бил неразделна част от икономическите дейности, неговите ефекти са много по-силно изразени в условия на комплексни системи с ускорена обработка на информацията каквито са много сектори от съвременната икономика. Затова не е учудващо, че формалното изучаване на риска придобива все по-голямо значение през последния век и към настоящия момент управлението на рисковете е не просто предмет на академични изследвания, но и важна част от съвременните управленски практики в частния и публичния сектор. Научното разбиране за риск е силно обвързано с областта на изследване и приложение, но при всички случаи е в значителна степен обусловено от моделирането и управлението на случайни събития чрез инструментариума на теорията на вероятностите.

Докато дефиницията за риска е относително устойчива и широко възприета, то определенията за различните типове риск варират в значителна степен. Фокусирайки се върху операционния<sup>2</sup> риск, следва да отчетем, че той се дефинира по различен начин от различните автори. Някои от алтернативните дефиниции са представени в Таблица 2: Определения за операционен риск.

*Таблица 2: Определения за операционен риск*

<b>Година</b>	<b>Автор</b>	<b>Дефиниция на операционен риск</b>
1993	Group of Thirty	Несигурност, свързана със загубите, произтичащи от неподходящи системи или контроли, човешки грешки или управленски проблеми.

<sup>2</sup> В българската литература за управление на риска се налагат паралелно преводите „операционен риск“ и „оперативен риск“. В настоящата работа се придържаме към първия, тъй като той е залегнал не само в научните изследвания, но и в стопанската практика, както и в нормативни документи.

<b>1998</b>	Crouhy et al.	Рискът, че външни събития или недостатъци във вътрешните контроли или информационни системи ще доведат до загуба – независимо дали загубата е очаквана до известна степен или напълно неочаквана.
<b>1999</b>	The Commonwealth Bank of Australia	Всички рискове с изключение на кредитния и пазарния риск, които биха довели до колебания в приходите, разходите и в стойността на банковите дейности.
<b>2000</b>	Jorion	Рискът, произтичащ от човешки и технологични грешки или инциденти.
<b>2000</b>	King	Мярка за връзката между бизнес дейностите (процесите) на организацията и нейните резултати.
<b>2001</b>	Crouhy et al.	Рисковете, свързани с дейностите на даден бизнес.
<b>2001</b>	Basel Committee	Рисковете, произтичащи от неподходящи или неуспешни вътрешни процеси, служители или системи, както и от външни събития.
<b>2002</b>	Lopez	Всеки неколкоквентим риск, с който се сблъсква банката.
<b>2003</b>	Securities and Exchange Commission	Рисковете от загуба, произтичаща от неуспешни контроли в рамките на организацията, неизчерпателно включващи неидентифицирано надвишаване на лимити, неоторизирани търговски операции, измами при търговските операции или при поддържащи функции, неопитен персонал и нестабилни или лесни за достъп компютърни системи.
<b>2005</b>	Deutsche Bank	Потенциала за реализиране на загуби, произтичащите от действията на служителите, договорните отношения и документацията, наличните технологични решения, нефункциониране на инфраструктурата, инциденти, външни влияния и отношения с клиентите.
<b>2005</b>	Vinella & Jin	Рискът, че дадена дейност няма да постигне една или повече от оперативните си цели, като дейността може да бъде свързана със хора, технология, процеси, информация или инфраструктурата, която подпомага бизнес процесите.
<b>2006</b>	Basel Committee on Banking Supervision	Рисковете, произтичащи от неподходящи или неуспешни вътрешни процеси, служители или системи, както и от външни събития.
<b>2009</b>	Solvency II (Directive 2009/138/EC)	Рискът от промяна в стойността, предизвикан от факта, че реализираните загуби от неуспешни вътрешни процеси, хора и системи, или от външни събития (включително юридически рискове) се различават от очакваните загуби.
<b>2012</b>	Минасян	Рискът от загуба, произтичаща от неадекватни или недобре функциониращи вътрешни процеси, хора и системи, или от външни събития, като включва в себе си правен риск.
<b>2012</b>	Chernobai et al.	Несистематичен риск, свързан с четири основни групи фактори: (1) Хора; (2) Процеси; (3) Системи; (4) Външни събития.
<b>2018</b>	Leone et al.	Рискът от загуба, произтичащ от неадекватни вътрешни процеси, човешки грешки или неуспешно функциониране на определени системи.

Chernobai et al. (2012) подчертават, че операционния риск е несистематичен и строго специфичен за дадената организация. В този смисъл неговата диверсификация е силно

затруднена, ако не и невъзможна. Leone et al. (2018) правят подробен преглед на алтернативните дефиниции на понятието операционен риск, като стъпват върху практиката в банковата сфера, където операционен риск се дефинира като рискът от загуба, произтичащ от неадекватни вътрешни процеси, човешки грешки или неуспешно функциониране на определени системи. Подобна е и дефиницията на Jorion (2000). Към тази дефиниция Банката за международни разплащания добавя и риска, произтичащ от външни събития (вж. BIS, 2001). В практиката се налага много по-широко разбиране за операционен риск, като Deutsche Bank (2005) го дефинира като потенциала за реализиране на загуби, произтичащите от действията на служителите, договорните отношения и документацията, наличните технологични решения, нефункциониране на инфраструктурата, инциденти, външни влияния и отношения с клиентите.

Vose, 2008 представя детайлна методология за количествено управление на рисковете, като си струва да отбележим, че при нея са изрично очертани двете ключови роли на участващите в процеса:

- **Анализатор** – експерт, който моделира и оценява рисковете, като изгражда формален модел, произвеждащ резултати в услуга на формалната управлние;
- **Управляващ процеса** – най-често мениджър или група мениджъри, отговорни за управлението на рисковата експозиция.

Хронологично процесът на управление започва като управляващия процеса формулира организационните проблеми и контекст и извежда списък от потенциалните рискове и възможностите за тяхното управление. Това може да бъде направен както от определен индивид, така и от групи ръководители и експерти в зависимост от нуждите и практиките в съответната организация. Като важен момент тук отчитаме необходимостта от избор на количествени критерии за избор на подходящи модели, подходи и стратегии за управлението на риска. Всичко дотук може да се разглежда като входни данни за последващия аналитичен процес. Аналитичният процес се състои от следните основни стъпки (Vose, 2008, стр. 5), които обичайно се извършват от експерт или екип от експерти по управлението на рисковете (вж. *ibid.*, стр. 23-26, както и Kroenke et al., 2012; Larose et al., 2014; Lu, 2018):

- **Преглед и анализ на наличните данни** – първата стъпка е преглед и анализа на наличните данни. Това най-често включва описателен (дескриптивен) анализ, като и извеждане на връзките между различните променливи, които имат отношение към изграждане на общия модел на риск. Връзките могат да бъдат проследени както визуално (пр. чрез топлинна карта) или аналитични (пр. чрез корелационна матрица).
- **Дизайн на модела** – като следваща стъпка експертите следва да изградят обща архитектура и структура на модела, който ще бъде използван за управление на риска. Тук е удачно да се избере както типът модел (ниво на формалност, използване на емпирични данни, симулация), така и конкретните алгоритми, които ще се използват (пр. корелационен или регресионен анализ, методи за машинно самообучение, алгоритми за оптимизация и други).

- **Създаване на модела** – тази стъпка включва изграждането на модела, като количествените модели често се изграждат и впоследствие изпълняват в конкретна софтуерна среда. На тази стъпка се осъществяват избраните в предишната такава подходи, като се прилагат различните подходящи алгоритми.
- **Извеждане на вероятностни разпределения на параметри** – след създаването на модела се оценяват параметрите включени в него. При използване на подхода за оценка на параметри на база на предишни данни, тези параметри се изчисляват и се определя тяхното ниво на несигурност (най-често под формата на доверителен интервал). Извън оценката на основни характеристики на променливите понякога се налага да се изведе и тяхното статистическо разпределение, обобщено във функция на разпределение или плътност на вероятностите.
- **Провеждане на симулации** – използвайки изчислените оценки и изведените вероятностни разпределения на ключовите променливи, анализаторът може да направи симулации (пр. по метода Монте Карло), с което да отчете очакванията разпределения на резултатите, както и да оцени необходимите проценти на очакваните реализации на риска за нуждите на управлението му. На този етап може да се направи и сценариен анализ, който да покаже чувствителността на резултатите към промяна на някои или всички включени в него параметри. Подобен подход позволява да се околичества моделната несигурност, която
- **Валидация на модела** – след изграждането на количествения модел, той следва да се валидира, като тази валидация може да бъде при сравнение с други модели (пр. на базата на информационен критерий, точност, обяснителна сила и др.) или на база на експертна оценка на получените резултати. Резултатите от валидирания модел могат да бъдат използвани на последващи стъпки от управленския процес.
- **Преглед на резултатите** – на тази стъпка се обсъждат получените от модела резултати и тяхната полза и приложение в рамките на общия процес. При този преглед участват както анализаторите, така и ръководителите на процеса и от организационна гледна точка това е момент не само на одобрение на получените резултати, но и на комуникация и изглаждане на потенциални различия.
- **Доклад на резултатите от модела** – след постигане на общо разбиране относно получените резултати от моделирането, те се оформят и се представят в подходящ формат в съответен доклад. Целта е визуализациите и аналитичното представяне да подпомогнат в максимална степен управленския процес.
- **Поддръжка на модела** – последната важна стъпка от аналитичния процес е поддръжката на модела, която включва както актуализация на данните и оценките на използваните параметри, така и техническата му поддръжка. Последното е особено в

сила когато се ползват значителни по размер модели, изискващи комплексна изчислителна инфраструктура.

Като алтернативен подход за управление на операционния риск се налага използването на количествени данни и тяхното моделиране с помощта на статистически и иконометрични методи. Редица автори (Vose, 2008) препоръчват този подход като по-надежден и формално проследим. Отбелязваме, че означението количествени подходи на практика съдържа в себе си широк набор от различни аналитични методи, използващи разнообразни източници на данни. Количественото управление на операционния риск е обусловено в значителна степен от изискванията за управлението му във финансовия сектор, но си струва да отбележим и алтернативни подходи в други сфери на икономиката рисковете (Galloppo & Regoga 2011; McNeil et al., 2005; Frachot et al., 2001). За целта първоначално отчитаме изискванията на Базелските споразумения и след това очертаваме принципите на използването на Монте Карло методи при оценката на операционния риск (Akkizidis & Kalyvas, 2018).

В рамките на изследванията, стандартния икономически подход е подчертаване на принципа на методологическия индивидуализъм и използване на основни резултати от теория на полезността за анализ на проблемите с взимане на рискови решения (вж. пр. Damodaran, 2007). Голяма част от литературата се доминира от изследване и анализ на околностими финансови рискове, като те често са поставени в рамките на регулативните изисквания към финансовите институции (пр. Eliot, 2012; Leone et al, 2018; Guegan & Hassani, 2018). Този фокус води до факта, че управлението на финансовите рискове е в значителна степен прецизирано, като фокусът е върху неговото приложение и еволюционно развитие, предимно задвижвано от регулативните изисквания. Прави впечатление, че и в нашата научна литература се отразява засиленото внимание върху управлението на финансовите рискове, като това води до редица научни и научно-приложни приноси на родни учени.

В научната литература се забелязва подчертана тенденция за синтетично разграничение между различните типове рискове и съответните конкретни стратегии за тяхното управление. Това дава началото на различни класификации на рисковите групи, някои от които са представени и в настоящата разработка (Frame, 2003; Crouhy et al., 2006; Dionne, 2013; Балабанов, 1996; Зафирова, 2016 и др.). Отчитаме, че подобен подход е полезен от аналитична гледна точка, но от гледна точка на изследователските и научно-приложните аспекти е често проблематичен. По-специално подчертаваме високата свързаност между различни типове рискове и значителното припокриване между някои от тях. Това предполага интегрирано разглеждане на рисковата експозиция на съвременната организация и единен подход за управлението ѝ. Като основен растящ риск отбелязваме операционния риск, който има значителен потенциал да доведе до критични загуби, като в същото време някои аспекти от управлението му подлежат на чувствително подобрене. В този смисъл управлението на операционният риск се налага като важна и актуална тема на научни изследвания.

Тук разгледахме стандартна дефиниция на операционния риск като произтичащ от хората, процесите, информационните системи и външните събития (Chernobai et al., 2012), като тази

дефиниция се налага като консенсусна и служи като основа на различните таксономии на операционните рискове (Embrechts et al., 2003; Jarrow, 2008).

В контекста на увеличената комплексност на стопанската среда се променят и типовете рискове, пред които са изправени организации и индивиди, както и интензитета им. Като пример, Chernobai et al. (2018) отбелязват, че с растящата сложност на финансовата среда се увеличават значително и операционните рискове пред банковите финансови институции. Тази тенденция допълнително засилва актуалността на изследванията в тази област.

В практиката се налага в значителна степен използването на качествени методи за управление на операционните рискове поради това, че те са лесни за приложение и подлежат на адаптация към широк спектър от различни ситуации (Pritchard, 2014). Тези преимущества се компенсират от тяхната значителна непрецизност. Най-забележително, поставените качествени етикети за размера и вероятността на риска не са семантично еквиваленти и в този смисъл са несравними (пр. два риска със „средна“ вероятност нямат една и съща вероятност за реализация). Те не позволяват да се вземе и формално оптимизационно решение съпоставящо ползите и разходите от управлението им, нито в общия случай да се изследва несигурността при тези оценки. Затова и редица изследователи (вж. пр. McNeil et al., 2005; Crouhy et al., 2006; Vose, 2008; Akkizidis & Kalyvas, 2018) предпочитат количественото управление на операционните рискове, което включва конкретни числени стойности на ключовите параметри и извеждане на формални индикатори за нивото на риск – пр. бета коефициенти, стандартни отклонения, коефициенти на Шарп, метрики за стойност под риск или очаквана крайно тежка загуба. Този подход се характеризира с това, че е особено интензивен от гледна точка на информационните си потребности – необходими са сравнително дълги времеви редове от данни или сравнително прецизни експертни оценки на основните параметри.

Макар количествените подходи да се смятат за най-усъвършенстваните методи за управление на риска на този етап, те страдат от редица недостатъци. По-конкретно, те не винаги могат да очертаят конкретните източници на риск, липсва единна методология за приложението им и те не водят до конкретни препоръки за стратегия или дейности по управление на риска. Нещо повече, на този етап количествената оценка на риска е силно зависима от използването на високо-квалифицирани експерти, които във взаимодействие с управленския персонал оценяват експозицията и начертават стратегии за справяне с несигурността. Предвид нарастващото значение на информационните системи и растящата нужда от взимане на по-бързи (дори реалновремени) решения, този подход е незадоволителен. В този смисъл съществува ясно изразена нужда от подобряване и автоматизиране на подходите за управление на операционните рискове, така че те да отговорят на нуждите на съвременните организации.

## **Глава 2: Нови подходи за управление на операционните рискове**

Бурното развитие на технологиите през последните две десетилетия и съответстващите промени в социално-икономически контекст на съвременните организации значително

промениха възможностите за, но и необходимостта от, нови модели за управление на операционния риск. Направения дотук преглед на основните тенденции в областта ясно очертава посоката на развитие на новите подходи, като също така отбелязва липсата на пълноценен и универсално приет подход за дигитална (цифрова) трансформация на управлението на организационните рискове. От една страна наблюдаваме редица отключващи технологии, които следва да бъдат приложени към този процес. На първо място отбелязваме наличието на големи масиви от данни ("big data"), които се генерират постоянно и в реално време от редица оперативни и транзакционни системи. Развитието на технологичните подходи за тяхното набиране, съхраняване, обработка и анализ предоставят редица възможности на бизнеса за предлагане на нова потребителска стойност и дори за съвместно създаване на стойност, но и разкриват нови рискове и предизвикателства в безпрецедентен мащаб. Експоненциалният ръст на данните и ускореното внедряване на цифрови продукти, услуги и процеси допринася в значителен план не само за дигитална революция в съвременните организация, но и за разширяване на техните рискови експозиции.

Технологичните развития са съпроводени и от социално-икономически такива. Най-забележително отбелязваме растящото разделение на пазара на труда, при което се наблюдава недостиг на високо-квалифицирани кадри, подпомагащи процесът по прехвърляне на нискоквалифицирана работа към автоматизирани и роботизирани системи. От друга страна този процес поставя под съмнение работните места на нискоквалифицираната работна сила.



Фигура 1: Основни тенденции и изисквания към новите подходи за управление на операционните рискове

Допълнително, наблюдаваме увеличаване на сложността на бизнес процесите и усложняване на управлението на съвременните организации под влияние не само на ИТ пробивите, но и на глобализацията и използване на глобални вериги на доставка и създаване на стойност.



Тези основни двигатели променят чувствително контекста на управлението на операционните рискове, като едновременно дават възможността, но и поставят императива за автоматизацията на този процес. Тази тенденция става все по-видима и в новите изследователски направления и подходи в сферата на операционните рискове. На база на представения преглед отчитаме четири основни тенденции в изследванията на операционните рискове. На първо място се забелязва засилен фокус върху конкретни приложни полета и проблеми, с което процесът по управление на риска става персонализиран за дадения проблем по начин сходен с персонализацията на цифровите продукти и услуги към отделния потребител. На второ място отбелязваме, че рисковете в областта на информационната сигурност се превръщат в основен клас операционни рискове, което се дължи на ускореното навлизане на информационни системи в почти всички аспекти на съвременния бизнес.

На трето място отчитаме ускореното проучване, тестване и отчасти навлизане на авангардни алгоритми за анализ на данни и оценка на рисковете, като тези алгоритми са пригодени да функционират ефективно в рамките на съвременната ИТ архитектура, като позволяват бързодействие, мащабируемост и разпределено изчисление. На четвърто място се откроява тенденцията за интегриран поглед върху общата рискова експозиция на съвременната организация. За разлика от предишния фокус върху прецизното аналитично разделение между типовете рискови групи, съвременните подходи залагат върху общ холистичен преглед на риска, като отчитат високата корелация между различните рискови групи в реална среда и подчертават, че често разделението им е трудно. Така описаните тенденции задават нови възможности и изисквания към съвременните подходи за управление на операционните рискове и очертават контурите на нова парадигма в областта. На Фигура 1: Основни тенденции и изисквания към новите подходи за управление на операционните рискове е схематично представено как промяната в контекста, технологичната обусловеност и новите изследвания водят до нови изисквания към съвременния управленски процес.

Отбелязваме пет основни изисквания към новите подходи за управление на операционните рискове, както следва:

- **Базиран на големи масиви от данни** – експоненциалният ръст на наличната информация предполага, че организациите разполагат с по-големи масиви от данни, които съдържат в себе си познание за техните операции. Разширяването на информационното множество е възможно и желателно с цел по-прецизна дефиниция и оценка на операционните рискове. Допълнителната стойност от анализа на големи масиви от данни е вече установен факт и управлението на операционните рискове следва да се възползва в максимална степен от наличното организационно познание. Нещо повече, съществува технологичната възможност широкият спектър от нови данни в различни формати да бъде успешно интегриран в управленския процес, така че да генерира бизнес стойност (McAfee et al., 2012). Разходите за подобна интеграция също намаляват значително (Burne & Corrado, 2017), което я прави финансово изгодна за множество организации.

- **Използване на усъвършенствани статистически алгоритми** – отбелязахме наличието и навлизането на авангардни методи от областта на машинното обучение за нуждите на бизнес анализа, взимането на решения и управлението на риска. Редица изследвания (пр. Makridarkis et al., 2019) показват значителни различия между резултатите на различни статистически алгоритми, като обосновават, че най-оптималните сред тях са чувствително по-добри спрямо техните подгласници. Тъй като дори малко подобрене в прогностичната точност на алгоритмите за управление на операционните рискове може да доведе до осезателно увеличение на генерираната от тях стойност, то следва в пълна степен да се изследват оптималните алгоритми за различни типове под-задачи от управлението на операционните рискове и те да се приложат в практиката.
- **Интегрирани с информационни системи** – ключова промяна в операциите на съвременния бизнес се корени в нарастващото използване на автоматизирани информационни системи, както и необходимостта от тяхното интегриране с цел получаване на единна среда за управление на организацията. Именно алгоритмите за взимане на решения в контекста на риск и несигурност биха спечелили в най-голяма степен от интегриране с информационните масиви на дадената структура, откъдето алгоритмите за управление могат да бъдат директно захранвани с данни. Допълнителната интеграция с оперативни и транзакционни системи ще позволи автоматизирано предприемане на съответните действия в резултат от анализа на риска, с което може да се ускори този процес и оттам да се достави значителна стойност (вж. пр. Raghupathi & Raghupathi, 2014).
- **Автоматични действия с минимална човешка намеса** – цифровата трансформация предполага преминаването на все по-голям брой процеси и решения извън човешки контрол и тяхната автоматизация с помощта на подходящи системи. В действителност автоматизираното взимане на решения се налага като водеща практика, която спестява време, генерира последователни решения и спомага за преодоляването на когнитивните ограничения на индивида. Растящото им приложение (Battaglini & Rasmussen, 2019) показва, че автоматизацията има значителен потенциал да спомогне дейността на съвременните организации и да доведе до по-добри резултати.
- **Постоянен и реалновремеви процес** – наличието на автоматичен поток от данни<sup>3</sup> към алгоритъм за управление на риска и взимане на решения означава, че наличната информация, с която разполага този модул се обновява в реално време. Това налага и наблюдение, анализ и преценка в реално време, която да обезпечи включването на най-актуалните данни във взимането на решение за управление на риска на всяко конкретно действие или транзакция (Davernport et al., 2012). Това важи с особена сила за широката група от рискове в областта на информационната сигурност (пр. кибер

---

<sup>3</sup> Т. нар. „стриймिंग“, от англ. – streaming.

атака), при която секунди забавяне могат да предопределят успеха или неуспеха на превантивните и корективни действия.



Фигура 2: Общ алгоритъм за автоматизирано управление на операционните рискове

В този смисъл процесът на управление на операционния риск вече не може да бъде периодичен спрямо предварително определен в политиките на организацията график, а следва да бъде постоянен и в реално време. Взимането на реалновремени решения също така се налага като активна и обещаваща сфера на научни и научно-приложни изследвания за риска в широк набор от различни области (вж. пр. Sanders et al., 2018).

Изведените изисквания за новите подходи за управление на операционния риск могат да бъдат използвани за разширяване на класическите подходи за управлението му (пр. Vose, 2008), като за целта се автоматизират стъпките, при които е необходима човешка намеса и се интегрира формална методология за извеждане на познания от данни, като CRISP-DM (Wirth & Hipp, 2000; Azevedo & Santos, 2008; Bosnjak et al., 2009; Schaefer et al., 2018; Huber et al., 2019). Въвеждайки подчертано аналитични компоненти и осъвременявайки подхода за управление на операционния риск чрез синтез на научните резултати от последното десетилетие,

достигаем до нов подход за управление на операционния риск (вж.



Фигура 2: Общ алгоритъм за автоматизирано управление на операционните *рискове*). Този иновативен подход надгражда настоящите научни достижения и най-добри практики и ги адаптира за използване в рамките на автоматизирани системи за анализ и автономно взимане на решения.

Общият алгоритъм за автоматизирано управление на операционните рискове се състои от 8 основни стъпки (процеса), който включват пълния управленски цикъл за оперативните рискове – от дефинирането на проблема и набирането на данни, през анализа и мерките, до тяхното приложение. Стъпките са разпределени в пет основни групи дейности, които до голяма степен следват обичайния управленски цикъл (Haimes, 2015; Vose, 2008; Crouhy, 2006; Sadgrove, 2016; Norkin, 2018; Зафирова, 2016; Цаневска, 2017):

- Дефиниция на проблема
- Информационно обезпечаване
- Обучение на модел за управление на операционния риск
- Приложение на модела
- Действия за управление

Всяка от изброените фази съдържа в себе си една или повече стъпки, представляващите конкретните дейности, които са необходими за нейното успешно провеждане. Тези стъпки са описани в детайли по-долу.

### Стъпка 1: Дефиниране на целева променлива и желани стойности

Класическите алгоритми за управление обичайно започват с дефиниране на контекста и разбиране на управленския проблем, който следва да се реши. Често това става чрез неформално или полу-формално представяне на структурата на проблема, като този процес се води от експерти в областта. От една страна това има значителен смисъл, тъй като позволява включването на широка група заинтересовани страни и позволява ясна комуникация между тях. От друга страна, подобен подход с вербално дефиниране на проблема значително затруднява автоматизацията на неговото решение, тъй като представянето под формата на естествена реч не изключва логическите пропуски и не гарантира пълен обхват на пространството на решаваната задача. Това затруднява

директното му програмиране в автоматизирана информационна система и в този смисъл не дава окончателен критерий за определяне на проблема или индикатор за решаването му. Вместо това, предлагаме въз основа на бизнес контекста и организационните нужди да се дефинира една или повече целеви променливи и техните желани стойности.

### **Стъпка 2: Захранване с данни в реално време чрез интеграция с приложни програмни интерфейси (API)**

Втората стъпка е и началото на информационното обезпечаване на процеса по управление на операционния риск. Основната задача тук е да се намерят необходимите данни за процеса, като те включват както дефинираните целеви променливи в първата стъпка, така и всички налични свързани с тях данни. Идеята тук е, че целевите променливи зависят от често много на брой различни двигатели, които определят тяхната динамика и дефинират очакванията ни за реализацията на техните стойности. Ако прогнозираната реализация се разминава значително от реалната, това би означавало материализиране на операционен риск. За целта следва да се дефинира сравнително прецизен прогностичен модел, нуждаещ се от съответната информация. В този смисъл е важно модулът за управление на риска да получава автоматичен поток от съответстващи на разрешавания проблем данни.

### **Стъпка 3: Изследване на структурата на данните и при необходимост намаляване на тяхната размерност**

Наличието на голям масив от данни има както своите преимущества, така и някои недостатъци при практическото им използване. Сред последните си струва да отбележим необходимостта от значително пространство за тяхното съхранение и високи изискванията към изчислителната мощ на системата при оценка на комплексни алгоритми върху големи масиви от информация. Поради това е възможно и желателно преди оценката на алгоритъма да се осъществи опознавателна фаза, която да отчете основната структура на данните. На този етап е удачно да се изведат и визуализации за данните и връзките между тях. Макар това да не е стриктно необходимо за автоматизираната система, то е изключително полезно за осъществяване на наблюдение и контрол от стана на експерти. Визуализациите ще спомогнат както за по-добро разбиране на самия процес и организационните рискове, така и за по-прецизен контрол върху резултатите от автоматизираната система.

### **Стъпка 4: Итеративно търсене на оптимален алгоритъм за моделиране на риск предвид данни и ограничения**

На следващата стъпка от приложението на общия алгоритъм се обучават различни модели, които да опишат оптимално зададената в началото целева променлива. Типовете статистически модели (алгоритми) които се тестват зависят от широк набор характеристики, но от особено значение са типа на целевата променлива, особеностите на анализирания данни и очакваните резултати. Тук могат да се приложат два основни подхода. На първо място, можем да изведем набор от популярни алгоритми с добри резултати в предишни изследвания и сходни приложения и да оставим алгоритъма да избере оптималният сред тях. На второ място можем да се освободим от това изкуствено ограничение и вместо това да

осъществим търсене сред широк набор от десетки (стотици) потенциални алгоритми, като накрая изведем оптималният сред тях.

### **Стъпка 5: Оценка на параметри на модела**

След избора на оптимален модел идва стъпката за изчисление на неговите параметри. Тук отново може да се подходи по два различни начина. Първият е да се използват вече изчислените и тествани коефициенти на оптималния модел от предходната стъпка. Вторият е да се оцени моделът върху различна извадка от данни. Има известен резон да се подходи по втория начин чрез разширяване на обема анализирана информация, тъй като е вероятно по-големият размер на извадката да доведе до по-точни оценки на параметрите на модела. Допълнително, възможно е на четвъртата стъпка да са изтеглени сравнително по-малки извадки от наличните данни за изчисляването на голям брой алтернативи от гледна точка на това да се спести време и изчислителни ресурси. Тук това вече не е така необходимо, тъй като в петата стъпка се оценява само един модел.

### **Стъпка 6: Тестване и класификация на нови наблюдения спрямо обучен модел**

Приложението на самия модел върху реални инстанции на даден процес се случва в шестата стъпка. Тук вече оцененият модел се използва като новите наблюдения, с които се захранва управленския модул през приложни програмни интерфейси се тестват спрямо него. Всяко едно от тези наблюдения се класифицира, като се оценява до каква степен числените стойности на неговите цели променливи се отклоняват от очакваните (прогнозираните от модела) обичайни стойности. Ако отклонението надхвърли зададените от критерия за аномалност граници, то се класифицира като необичайно и подлежащо на предприемане на допълнителни действия. Това по същество е стъпката за взимане на решения за рисковата експозиция. Подчертаваме, че за разлика от традиционните подходи, предложеният алгоритъм позволява да се извършва управление на риска на много високо ниво на детайл (грануларност), като се работи на ниво атомистично действие или транзакции, вместо на ниво бизнес процес или сфера на дейност.

### **Стъпка 7: Предприемане на действия**

След класифицирането на дадено наблюдение като легитимно или аномално, следва да се предприемат конкретни действия за управление на рисковете, свързани с него. В традиционните методологии експерти в областта извеждат стратегии и дейностите по управлението на рисковете и се назначава отговорник за всеки тип риск, който ги осъществява. При автоматизацията на процеса и преминаване към висока грануларност на управлението това вече не е реалистично – едва ли има експерт или дори отдел, който да може да адресира стотици хиляди или дори милиони транзакции всеки ден. Това и налага предприемането на автоматични действия в рамките на информационната система и интегрираните с нея други системи. Тук се явява и основният въпрос, който доминира повечето дискусии за обхвата на дигиталната трансформация – до каква степен действията следва да се предприемат автоматично и до каква степен за тях следва да са отговорни човешка работна сила. Постигането на оптимален баланс между двете не е тривиална задача и е удачно да се подходи формално като се изведе оценка на въздействието от прилагането

на изкуствен интелект в икономически контекст. Тук предлагаме два основни подхода, които са приложими в съвременните организации.

#### *Известяване на отговорник за риска*

При първия подход се залага предимно на ролята на човешката работна сила, която извършва или поне експлицитно оторизира необходимите действия. Тук се следва обичаен бизнес процес за управление на рисковете, при който е определен отговорник за риска, от когото се очаква извършване на всички дейности по управлението му.

#### *Интеграция с други системи за автоматични действия*

При втория подход се залага пълна автоматизация на действията по управление на операционния риск. Това се постига чрез интеграция с транзакционните системи на организацията и предполага ясно изведени политики за целта. В този случай модулът за управление на операционния риск е свързан със съответните информационни системи чрез определен приложен програмен интерфейс и в резултат от направената рискова класификация може да изпрати заявки за действия към тези системи.

### **Стъпка 8: Итеративен процес на постоянно усъвършенстване**

Общият алгоритъм за автоматизирано управление на операционните рискове е по същество гъвкав и итеративен, тъй като е възможно взаимодействието и преплитането на различни стъпки (пр. подготовката на данните за специфичен алгоритъм може да се извърши и при самото изчисляване на модела, а не в предходната стъпка). За да се засили тази итеративност и да се обезпечи устойчивото подобряване на алгоритъма, той включва и процес за постоянно усъвършенстване. Процесът е формален, като се осъществява както автоматично на определен от анализатора или потребителя интервал от време (пр. веднъж месечно), така и в резултат от човешка намеса за стартирането му (при неочаквани промени на обстоятелствата). Този процес предполага преразглеждане на целевите променливи, актуалната структура на данните, генериране на алтернативни модели, избор и изчисляване на нов оптимален такъв, като не е необходимо всички тези стъпки да бъдат засегнати. В този смисъл могат да се обновят само определени стъпки, като това най-често включва преизчисляване на параметрите на използвания модел с актуални данни.

Общият алгоритъм за автоматизирано управление на операционните рискове цели да зададе рамката, в която този процес да може да бъде успешно изпълнен, като огромната част от дейностите по него се извършват от автоматизирани информационни системи.

### **Глава 3: Автоматизирано прилагане на класификационни алгоритми за управление на операционния риск**

Класификацията на различни наблюдения като рискови или нерискови такива е най-интуитивната и често срещана, но и същевременно най-добре проучена задача от сферата на управлението на риска. От гледна точка на използваните данни тя предполага, че

разполагаме с маркирани данни, където наблюденията са разделени на добре дефинирани дискретни класове (обичайни/аномални). Оттам и целевата променлива естествено е променливата, която описва тези класове. От друга страна в много приложения изходите, които представляват икономически интерес не са дискретни, а продължителни променливи. Сред примерите си струва да споменем променливи като икономически растеж, вероятност за фалит, възвръщаемост на даден актив, приходи от определени дейности или бизнес линии, производителност на труда и други подобни. Управлението на процесите с продължителни изходи предполага и използването на друг тип алгоритми от сферата на статистиката и машинното самообучение, които са известни като регресионни алгоритми (Hastie et al., 2005).

Главата има за цел да изследва автоматизираното прилагане на оптимални алгоритми за управление на риска в случаите на продължителни стойности на целевата променлива, да очертае критерии за класификацията на определени наблюдения като рискови и да демонстрира възможността за трансфер на тези дейности от експерти към автоматизиран агент. За постигане на тези цели са разгледани пет основни ситуации на операционните рискове, като към тях е приложен общия алгоритъм за автоматизирано управление. Тези ситуации включват моделиране на отсъствията от работното място, на популярността на онлайн комуникацията на организацията, за управление на инцидентите в процеса на поддръжка, за оценка на недвижимата собственост на организацията, както и за оценка на неблагоприятните промени във външната среда. За идентифицирането и оценката на рисковете във всяка от ситуацията използваме широк набор от регресионни алгоритми, като на базата на степента на аномалност идентифицираме и потенциалните рискови наблюдения. Тук показваме и как на базата на широко възприети статистически критерии за качество на моделите и прогнозите, както и за нивото на аномалност на даденото наблюдение можем да проведем цялостния процес по идентификация и оценка на рисковете по абсолютно автономен начин и без човешка намеса. Това е ключово, за да може той да бъде автоматично изпълнен от информационна система. В същото време отчитаме и важността от човешки контрол върху процеса, поради което като важни стъпки включваме визуализацията и докладването на основни връзки и резултати, генерирани в рамките на управленския процес.

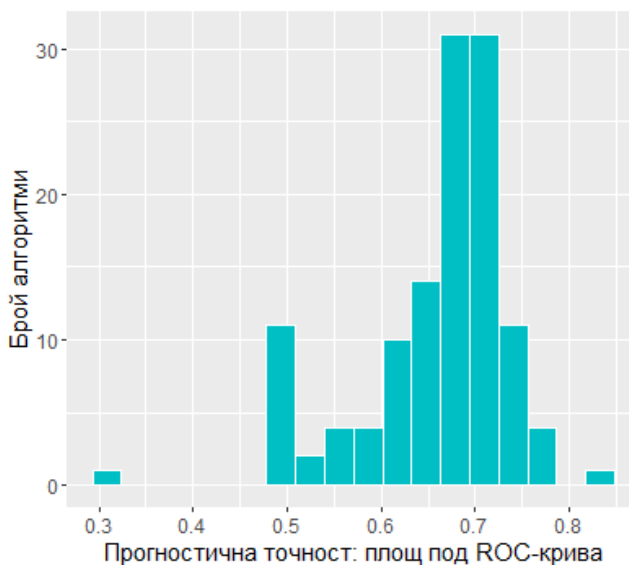
Значителният брой актуални изследвания в областта на статистическите методи и машинното обучение предполагат и бурно развитие на методите, подходите и алгоритмите, които са достъпни за изследователски цели. С цел изчерпателен обзор и оптимален избор от широк набор от алтернативни алгоритми, използваме данните за оценка и тестване на 136 от най-популярните методи. Броят наблюдения за обучение и тестване на всеки от алгоритмите е представен в приложение, но общият принцип е, при изчерпателното тестване на класификационни алгоритми избираме случайна подизвадка от дадения информационен масив, от която 80% е обучаваща извадка, а другите 20% - тестова такава. При тестването изчисляваме площта под ROC-кривата на всеки алгоритъм, но и мярка за неговата комплексност.



Като приближение на мярката за комплексност използваме времето, необходимо за изчисление, като стандартизираме най-дългото необходимо време на 100% и представяме останалите времена като част (пропорция) от него. В този смисъл мярката за комплексност се изменя в интервала от 0% до 100%. Изчислителното време на даден алгоритъм е силно зависимо от използваната инфраструктура и метод на изчисление, като от особена важност е дали изчислението е разпределено или не. Подвеждащо е да се докладва „чистото“ необходимо време, тъй като то ще зависи от машините или клъстерите от машини, които се използват. Мярката за комплексност разрешава отчасти този проблем, като извежда отношения на времена, а не абсолютни стойности. Макар да остават някои проблеми – пр. изчислителното време би се променило от типа процесор, архитектура, управление на натоварването и др., то мярката за комплексност представлява едно задоволително приближение към това доколко ресурсоемък е даден алгоритъм.

### Провеждане на кампания по директен маркетинг

Прогностичната точност на разгледаните методи е обобщена в хистограмата на Фигура 3. На първо място прави впечатление значителната вариация в точността между различните използвани алгоритми. Дори да пренебрегнем екстремните стойности (в положителен и отрицателен план), основната част на разпределението се изменя в интервала от 0.5 до 0.78. Това навежда на мисълта, че някои алгоритми са значително по-подходящи за определен тип задачи отколкото други. В този смисъл изборът на подходящ алгоритъм за класификация може да доведе до много значима разлика в генерираните резултати и оттам – в създадената стойност на бизнеса. На второ място отбелязваме, че показаното разпределение е близко до нормалното, като наблюдаваме пик около  $AUROC = 0.7$ . Това би била очакваната прогностична точност на „средния“ алгоритъм, решаващ тази задача. На трето място отчитаме сравнително високия брой алгоритми, които не добавят стойност (площ под кривата от 0.5), като подчертаваме, че те следва да се избягват.



Фигура 3: Разпределение на класификационните методи спрямо тяхната прогностична точност в задача за управление на риска при директен маркетинг

Първите десет класификационните методи с показани най-добри резултати са представени в Таблица 3: Топ 10 класификационни метода с най-висока прогностична точност. Прави впечатление, че различни модели на дискриминантен анализ се представят изключително добре при тази задача, като най-добрият класификатор е именно робастния регуляризиран дискриминантен анализ (rrlda) с площ под кривата на работната хараткеристика от 0.82.

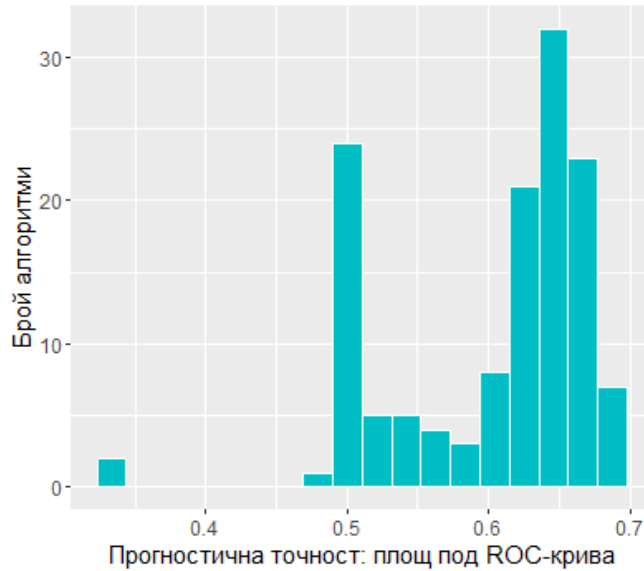
Дърветата на решенията и различни типове ансамблови алгоритми също попадат сред десетте най-добри класификатора, като площта под ROC-кривата за алгоритми на места от 2 до 10 се изменя в интервала от 0.73 до 0.77. Подчертаваме и че най-точната класификация не се постига от най-ресурсоемките алгоритми, като в Топ 10 попадат оптимизирани и сравнително бързи алгоритми.

Таблица 3: Топ 10 класификационни метода с най-висока прогностична точност

Тип алгоритъм	Метод	Площ под ROC-крива	Мярка за комплексност
<i>Робастен регуляризиран линеен дискриминантен анализ</i>	rrlda	0.824	1.8%
<i>Меко независимо моделиране на класови аналогии, SIMCA-модел</i>	CSimca	0.773	0.3%
<i>Класификатор, базиран на правила</i>	JRip	0.767	7.8%
<i>Смесен дискриминантен анализ</i>	mda	0.764	0.1%
<i>Условно дърво за изводи</i>	ctree	0.762	0.3%
<i>Дърво на решенията от тип C4.5</i>	J48	0.754	3.0%
<i>Модел на осреднена невронна мрежа</i>	avNNet	0.740	3.9%
<i>Класификатор, базиран на ROC</i>	rocc	0.738	0.7%
<i>Ансамблово адаптивно засилване</i>	AdaBag	0.734	2.3%
<i>Сборен модел, базиран на дървета на решенията</i>	nodeHarvest	0.734	56.7%

### Провеждане на операции с кредитни карти

Обслужването на дължимите баланси по кредитни карти е ключов проблем във финансовия сектор, като невъзможността това да се случи и потенциалните измами могат да имат значим ефект върху финансовите потоци и платежоспособността на зависещите от тях организации. В този контекст е особено важно да се избере оптимален алгоритъм, като дори малки подобрения в прогностичната точност могат да доведат до отключване на значителна стойност за кредиторите. За целта извършваме изчерпателно тестване на 136 основни алгоритъма от сферата на машинното обучение и анализираме тяхната точност при класификация.



Фигура 4: Разпределение на класификационните методи спрямо тяхната прогностична точност в задача за управление на риска при операции с кредитни карти

Фигура 4 обобщава данните за прогностичната точност на алтернативните алгоритми, измерена като площ под кривата на работната характеристика (ROC-крива). Голям брой от разгледаните алгоритми имат AUROC около 0.50, което е резултат, равен на този на шанса – следователно този първи връх от разпределението показва неплодотворните алгоритми за тази задача. В хистограмата наблюдаваме и втори връх със стойности около 0.65, като огромната част от разгледаните алгоритми са концентрирани именно в интервала от 0.62 до 0.68. Най-добрите алгоритми клонят към площ от 0.7, но реално нито един не преминава тази граница. Отчитаме класификационната задача върху данните от кредитни карти като сравнително трудна такава, което обяснява и получените резултати.

Десетте най-добри метода за класификация са представени в Таблица 3. Прави впечатление, че групата се доминира от два основни типа модели – този на машини с подкрепящи вектори и на класификационни и регресионни дървета (CART<sup>4</sup>). Най-висок резултат има машината с подкрепящи вектори с полиномно ядро, изчислена по метода на най-малките квадрати с AUROC = 0.684, последвана от дървета на решенията от тип C5.0 и три други вариации на машини с подкрепящи вектори (всички с AUROC – 0.682). В първата десетица се нареждат още два метода, базирани на дървета, един ROC-базиран класификатор и модел на ансамблово адаптивно засилване.

Отбелязваме, че точността на всички тези алгоритми е много сходна и на практика ще се наблюдават сравнително малки различия, които биха имали значение само при обработка на значителни масиви от данни. Колкото до времето необходимо за изчисление, най-добрите алгоритми отново не са най-ресурсоемките. Оптималният метод е близо сто пъти по-бърз от най-бавният такъв, вторият най-добър е хиляда пъти по-бърз, а третият – 167 пъти. Това

<sup>4</sup> От англ. – Classification and Regression Trees.

показва, че и в тази задача отново виждаме възможност за балансиране между изчислителното натоварване и точността на получените резултати.

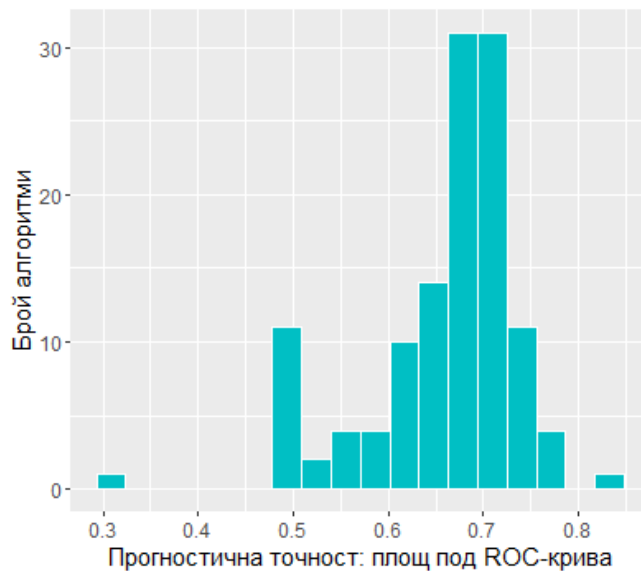
Таблица 4: Топ 10 класификационни метода с най-висока прогностична точност

Тип алгоритъм	Метод	Площ под ROC-крива	Мярка за комплексност
Машина с подкрепящи вектори по метода на най-малките квадрати и полиномно ядро	svmPoly	0.684	1.1%
Дърво на решенията от тип C5.0 за извеждане на набор от правила	C5.0Rules	0.682	0.1%
Линейна машина с претеглени подкрепящи вектори	svmLinearWeights	0.682	0.6%
Линейна машина с подкрепящи вектори	svmLinear	0.682	0.2%
Линейна машина с подкрепящи вектори	svmLinear2	0.682	0.2%
Класификатор, базиран на ROC	rocc	0.680	0.2%
Класификационни и регресионни дървета	rpart1SE	0.679	0.1%
Ансамблов адаптивно засилване	AdaBag	0.677	37.4%
Засилено дърво на решенията	bstTree	0.677	1.9%
Засилени класификационни дървета	ada	0.673	3.2%

### Отпускане на кредити

Моделирането на риска при отпускане на кредит е класическа класификационна задача и към нея могат да се приложат стандартните алгоритми за машинно обучение. За целта оценяваме 136 от тях, като обобщените резултати за тяхната прогностична точност са представени на Фигура 5.

Разпределението, което наблюдаваме в този случай се различава в значителна степен от нормалното такова. Наблюдаваме пик на алгоритмите с прогностична точност около шанса (AUROC = 0.5), последвано от сравнително равномерно разпределение на алгоритми с прогностична точност в интервала 0.52 до 0.64. Голяма част от разгледаните алгоритми има точност в порядъка от 0.68 до 0.70, което може да се каже, че са очакванията ни за „средния“ алгоритъм, подходящ за тази конкретно задача. Наблюдава се малък брой алгоритми с AUROC > 0.70, които са и най-добре представящите се класификатори при целевата променлива на разглеждания масив от данни. Първите десет класификатора с най-висока прогностична точност са представени в Таблица 5.



Фигура 5: Разпределение на класификационните методи спрямо тяхната прогностична точност в задача за управление на риска при отпускане на кредити

Всички те имат площ под кривата на работната характеристика от поне 0.7, като най-добрият от тях – регуляризираната случайна гора – достига до стойност от 0.73. Прави впечатление, че тази група се доминира от различни типове и методи за изчисление на случайни гори, като те съставляват половината от десетте най-добри алгоритъма. Извън това се наблюдават методи за засилване на градиента, един специфичен тип невронна мрежа (многопластов перцептрон), както и вариант на дискриминантния анализ – локализиращия линеен дискриминантен анализ. Отново отбелязваме, че най-тежките от изчислителна гледна точка алгоритми не дават най-добри резултати. Най-точният класификатор се изчислява 2.6 пъти по-бързо от най-ресурсоемкия, а вторият по точност – 13.5 пъти по-бързо.

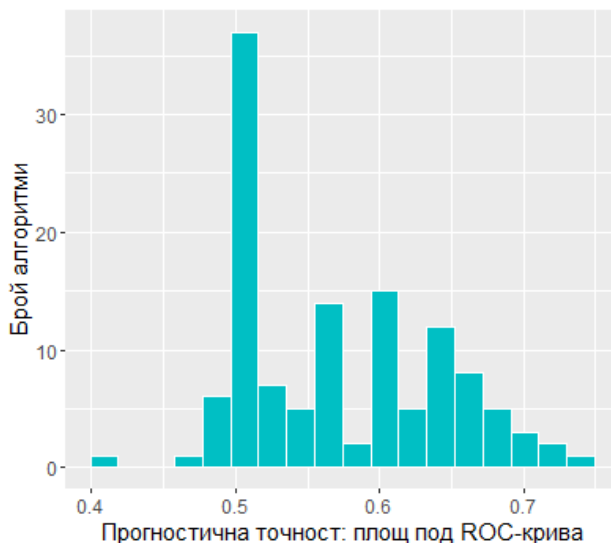
Таблица 5: Топ 10 класификационни метода с най-висока прогностична точност

Тип алгоритъм	Метод	Площ под ROC-крива	Мярка за комплексност
Регуляризирана случайна гора	RRF	0.730	38.5%
Екстремно засилване на градиента	xgbLinear	0.718	7.4%
Регуляризирана случайна гора	RRFglobal	0.712	5.2%
Екстремно засилване на градиента	xgbDART	0.707	26.3%
Многостъпкова адаптивна МСР-мрежа	msaenet	0.702	5.1%
Случайни дървета	rFerns	0.701	2.0%
Локализиран линеен дискриминантен анализ	loclda	0.701	0.8%
Случайна гора	rf	0.700	1.7%
Случайна гора	ranger	0.698	1.6%
Машина със засилване на градиента	gbm	0.698	0.2%

### Управление на отношенията с външни партньори

Задачата за определяне на класова принадлежност на фирми към това дали ще фалират или не, се оказва със значителна трудност. Сред седемте най-популярни подхода само случайната гора показва задоволителни резултати. Това са именно случаите, в които търсенето на

оптимален алгоритъм следва да се разшири значително, като тук тествахме 136 различни алтернативи. Съответните модели са изчислени върху обучаваща извадка, а техните прогнози – тествани върху тестова такава. Разпределението на точността им, измерена чрез площта под кривата на работната характеристика е обобщено в хистограмата на Фигура 6 Фигура 7. Огромен брой (над 40) от оценените алгоритми имат прогностична точност около AUROC = 0.5, което е точно равно на класификация спрямо безусловната вероятност за принадлежност към даден клас. Трудността на задачата се подчертава и от това, че има и някои алгоритми с площ под ROC-кривата от под 0.5, което е резултат по-лош от случайно генерирания такъв. Наблюдаваме лек пик на прогностичната точност при стойности на площта от 0.6, като най-добрите класификационни алгоритми достигат до прогностична точност AUROC над 0.7. Отбелязваме значителните различия между резултатите на различните методи, като все пак се наблюдава и малцинство подходи с доста висока прогностична точност.



Фигура 6: Разпределение на класификационните методи спрямо тяхната прогностична точност в задача за управление на риска при прогнозиране на риск от фалит

Най-точните десет алгоритъма са представени в Таблица 6. Най-оптималния сред тях е робастното меко независимо моделиране на класови аналогии, RSIMCA<sup>5</sup>. Този метод е сравнително по-слабо познат и използван в икономиката и бизнеса, но по същество включва надзираван анализ, при който се осъществява разделяне на данните на главни компоненти и конструиране на под-пространства въз основа на тези компоненти, които се използват впоследствие за класификация. За повече детайли насочваме читателя към оригиналната разработка на Brandon & Hubert (2005), както и към изследването на Fauziyah et al. (2018). RSIMCA-моделът има площ под ROC-кривата от 0.733 и е повече от триста пъти по-бърз от най-бавният алгоритъм – Метода за бавно извеждане на правила, който е на второ място по прогностична точност с 0.721. На трето място с много близка прогностична точност (AUROC – 0.714) се нареждат случайните дървета, последвани от шест други метода от семейството на

<sup>5</sup> От англ. – Robust Soft Independent Modeling of Class Analogies

дърветата на решенията или случайните гори. Всички те са относително бързи и в този смисъл изискват сравнително по-ограничени изчислителни ресурси.

Таблица 6: Топ 10 класификационни метода с най-висока прогностична точност

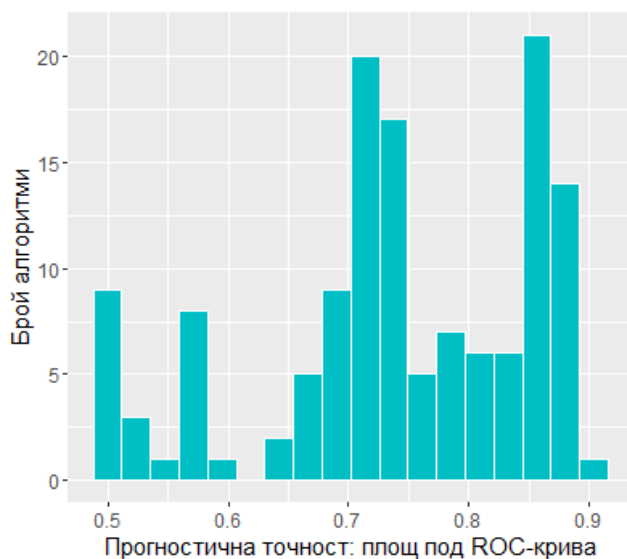
Тип алгоритъм	Метод	Площ под ROC-крива	Мярка за комплексност
<i>Робастно меко независимо моделиране на класови аналогии, RSIMCA-модел</i>	RSimca	0.733	0.3%
<i>Метод за бавно извеждане на правила, PRIM</i>	PRIM	0.721	100.0%
<i>Случайни дървета</i>	rFerns	0.714	1.7%
<i>Класификационни и регресионни дървета</i>	rpart1SE	0.693	0.1%
<i>Класификационни и регресионни дървета</i>	rpart2	0.693	0.0%
<i>Дърво на решенията от тип C5.0 за извеждане на набор от правила</i>	C5.0Rules	0.693	0.2%
<i>Класификатор, базиран на правила</i>	PART	0.679	0.3%
<i>Регуляризирана гора на решенията (случайна гора)</i>	RRF	0.676	35.6%
<i>Дискриминантен анализ със свиване</i>	sda	0.676	0.2%
<i>Ансамбलोво адаптивно засилване</i>	AdaBag	0.674	0.9%

На десето място по прогностична точност се класира специфичен метод за дискриминантен анализ, като отново регистрира сравнително висока точност – AUROC = 0.676. Отчитаме, че именно в задачи с по-висока трудност има особено голям смисъл да се тества широк набор от алтернативи и да се избере оптималната за дадената ситуация и тип данни. Това позволява дадения процес по управление на риска да се възползва от алгоритми, които са особено подходящи за конкретни нишови ситуации, но средно статистически имат по-слабо представяне в по-широк спектър от приложения<sup>6</sup> и поради това не са популярни и често използвани.

### Дейности по електронна търговия

За намирането на оптималния класификационен алгоритъм при изследването на онлайн потребителско поведение оценяваме 136 различни алгоритъма за машинно обучение. Обобщените резултати за тяхната прогностична точност, измерена чрез площта под кривата на работната характеристика са показани чрез хистограма на Фигура 7. Средната прогностична точност при класификацията на онлайн поведението е чувствително по-висока спрямо другите разгледани рискови ситуации. Прави впечатление, че разпределението на площите се характеризира с два върха – единият е около AUROC = 0.70-0.75, а другият – при стойности в порядъка 0.85-0.90. Най-добрият класификационен алгоритъм отбелязва стойност дори над 0.90. Отчитаме и значителната дисперсия при резултатите от изчислените методи. Немалка част от тях имат резултати близки до шанса, но и значително малцинство регистрират много висока прогностична точност. Това подчертава важноста от избора на оптимален алгоритъм поради много различното качество на генерираните прогнози.

<sup>6</sup> Това, разбира се, е следствие от Теоремата за липса на безплатен обяд в оптимизацията – поради спецификите на различните задачи не е възможно да се намери оптимален алгоритъм или стратегия, които са най-добри за всяка една задача. В този смисъл доброто представяне при един клас задачи се компенсира от по-слабо при друг клас задачи. За повече информация вж. Yo & Perune, 2002.



Фигура 7: Разпределение на класификационните методи спрямо тяхната прогностична точност в задача за управление на риска при електронна търговия

Най-точните десет алгоритъма според площта под ROC-кривата са представени в Таблица 7. Ротационната гора има най-добро представяне с AUROC = 0.902, следвана от случайната гора с претеглени пространства (0.889), метода за бавно извеждане на правила PRIM (0.886), засилено дърво на решенията (0.885), както и поредица от методи от семейството на класификационните и регресионни дървета (всички с 0.880). Отчитаме семейството от методи на случайна гора, както и съставляващите ги дървета, като най-оптимален подход за решаването на разгледаната задача на потребителско поведение в дигитална среда.

Таблица 7: Топ 10 класификационни метода с най-висока прогностична точност

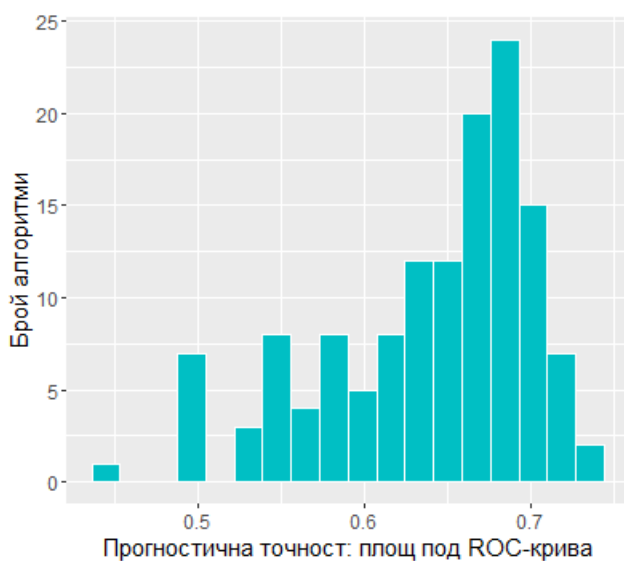
Тип алгоритъм	Метод	Площ под ROC-крива	Мярка за комплексност
Ротационна гора	rotationForestCp	0.902	1.0%
Случайна гора с претеглени пространства	wsrf	0.889	4.4%
Метод за бавно извеждане на правила, PRIM	PRIM	0.886	14.3%
Засилено дърво на решенията	bstTree	0.885	1.4%
Класификационни и регресионни дървета	rpart	0.880	0.0%
Класификационни и регресионни дървета за редни изходи	rpartScore	0.880	1.4%
Условно дърво за изводи	ctree2	0.880	0.1%
Дърво на решенията от тип C5.0	C5.0	0.879	0.3%
Дърво на решенията от тип C5.0 с корекция за разход	C5.0Cost	0.879	0.7%
Алгоритъм с дълбоко засилване	deepboost	0.876	4.9%

От гледна точка на необходимото време и ресурси за изчисляване на тези методи подчертаваме, че отново най-ресурсоемките методи не водят до най-добри прогностични резултати. Напротив, оптималният алгоритъм се изчислява 100 пъти по-бързо от най-бавният такъв, като наблюдаваме подобни и по-добри съотношения и в последващите



методи, попаднали в първата десетица. Този устойчив резултат при всички разглеждани задачи показва възможността за едновременна оптимизация както на точността, така и на използваните ИТ ресурси и изчислителна инфраструктура.

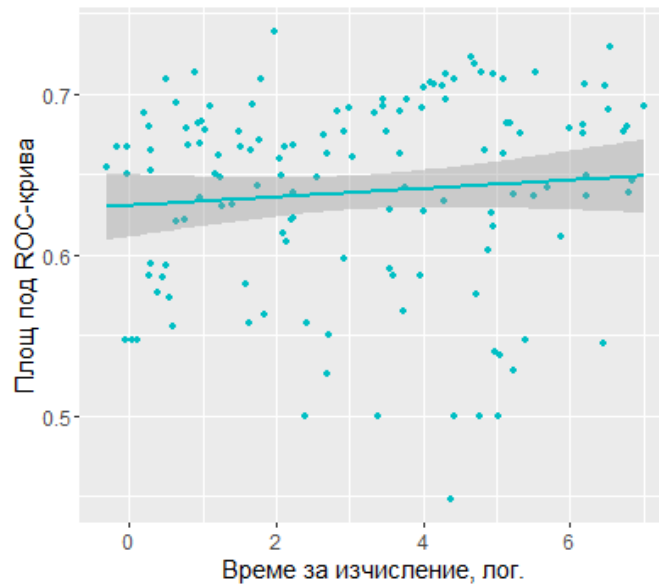
Прегледа на широк набор от популярни алгоритми от сферата на машинното обучение върху пет различни операционни ситуации позволява да се очертаят някои основни изводи. На първо място, не се припокриват най-добрите алгоритми за всяка от разглежданите задачи – при всяка от индивидуалните класификация наблюдаваме различия в най-добрите алгоритми за решаването ѝ. Това вероятно се дължи на факта, че различни семейства от алгоритми и техните конкретни методи се справят по-добре с определени типове данни, но пък генерират по-лоши резултати с други типове. Това е добре известната теорема за липса на безплатен обяд в оптимизацията (Branden & Hubert, 2005) и подчертава, че е теоретично и практически неудачен подход да се използват едни и същи методи за всяка една задача.



Фигура 8: Разпределение на класификационните методи спрямо тяхната осреднена прогностична точност

На второ място отбелязваме, че прогностичната точност между различните алгоритми може да се изменя в изключително широки интервали. Фигура 8 обобщава разпределението на осреднените стойности на площта под кривата на работната характеристика за всички разглеждани алгоритми. Средната прогностична точност варира в интервала от AUROC = 0.50 до 0.74. Това подчертава, че важността от избора на оптимален алгоритъм е не само теоретично обоснована, но може да има значителни практически последствия. Този резултат подчертава допълнително важност от изчерпателно търсене на най-добрите алгоритми за управление на операционния риск, тъй като при съответния мащаб на бизнес операциите подобренията в прогностичната точност имат потенциала да генерират огромна бизнес стойност.

На трето място отбелязваме, че определени семейства от алгоритми имат тенденцията да показват по-добри резултати спрямо други такива. По-специално, различните методи на случайна гора често се явяват сред най-добрите алгоритми за решаване на всяка една от разгледаните задачи. В анализа забелязахме обичайната за тях тенденция да се нагаждат към определените анализирани данни, но въпреки това те показват отлични резултати и при нови тестови извадки. От гледна точка на класическите класификационни алгоритми изглежда, че дискриминантния анализ в различни негови варианти има доста добри прогностични резултати. Тъй като тези методи са в повечето случаи високо оптимизирани, те биха могли да са разумен компромис в ситуации, при които е необходимо да се анализира значителен обем от данни при условия на ограничени изчислителни ресурси.



Фигура 9: Разпределение на класификационните методи спрямо тяхната осреднена прогностична точност

На четвърто място подчертаваме, че не е задължително най-ресурсоемките от изчислителна гледна точка алгоритми да достигат до най-точни прогностични резултати. Във всяка от разгледаните задачи най-добрият метод за класификация не е този, който отнема най-много ресурси за изчислението му. Сред първите десет метода според площта под кривата на работната характеристика за всяка от задачите се нареждат методи, които са десетки и дори стотици пъти по-бързи от най-бавният такъв. Фигура 9 представя графично тази връзка между средната площ под ROC кривата за всички методи върху всички задачи и логаритъма на времето, необходимо за тяхната оценка. Графично отчитаме слаба положителна връзка между двете, но при изследването ѝ в рамките на модел на линейна регресия тази връзка не достига статистическа значимост ( $p = 0.323$ ). В този смисъл е възможно да се избере оптимално съотношение между прогностична точност и необходимими изчислителни ресурси за даден метод, така че да се генерира максимална ползва за процеса по управление на риска.

Показаните резултати в настоящата глава демонстрират възможността за автоматизация на важни стъпки от процеса по идентификация и оценка на операционните рискове. Автоматичният избор на алгоритъм за оценка на риска на база на единен индикатор (площ под крива на работната характеристика) и неговото приложение позволяват управленския процес да протича на ниво индивидуално наблюдение или събитие и освобождават служителите от непосилната отговорност да следят и преценяват в реално време всички събития, които имат отношение към процеса. Нещо повече, при използването на алгоритми за надзиравано обучение при човешките експерти остават важните задачи по определяне на цели променливи, техните желани стойности и наблюдение и контрол върху действието на автоматизираната информационна система. Това позволява едновременна оптимизация на огромна част от необходимите дейности и запазване на ценния човешки експертен елемент по качествена калибрация на използваните модели и системи.

#### **Глава 4: Автоматизирано прилагане на регресионни алгоритми за управление на операционния риск**

Класификацията на различни наблюдения като рискови или нерискови такива е най-интуитивната и често срещана, но и същевременно най-добре проучена задача от сферата на управлението на риска. От гледна точка на използваните данни тя предполага, че разполагаме с маркирани данни, където наблюденията са разделени на добре дефинирани дискретни класове (обичайни/аномални). Оттам и целевата променлива естествено е променливата, която описва тези класове. От друга страна в много приложения изходите, които представляват икономически интерес не са дискретни, а продължителни променливи. Сред примерите си струва да споменем променливи като икономически растеж, вероятност за фалит, възвръщаемост на даден актив, приходи от определени дейности или бизнес линии, производителност на труда и други подобни. Управлението на процесите с продължителни изходи предполага и използването на друг тип алгоритми от сферата на статистиката и машинното самообучение, които са известни като регресионни алгоритми (Hastie et al., 2005).

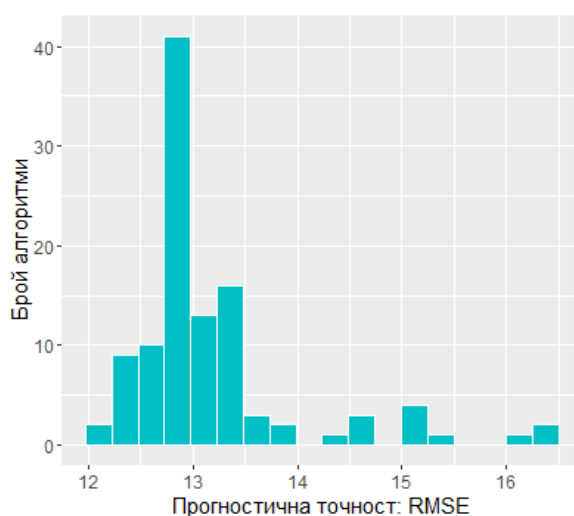
Целта на тяхното използване е да се определи с максимална точност очакването за реализация на целевата променлива и при вероятност за излизането ѝ от желаните или допустими граници (събитие на операционен риск) да се предприемат мерки за управление, които да намалят негативния ефект.

Настоящата глава има за цел да изследва автоматизираното прилагане на оптимални алгоритми за управление на риска в случаите на продължителни стойности на целевата променлива, да очертае критерии за класификацията на определени наблюдения като рискови и да демонстрира възможността за трансфер на тези дейности от експерти към автоматизиран агент. За постигане на тези цели са разгледани пет основни ситуации на операционните рискове, като към тях е приложен общия алгоритъм за автоматизирано управление. Тези ситуации включват моделиране на отсъствията от работното място, на популярността на онлайн комуникацията на организацията, за управление на инцидентите в

процеса на поддръжка, за оценка на недвижимата собственост на организацията, както и за оценка на неблагоприятните промени във външната среда. За идентифицирането и оценката на рисковете във всяка от ситуациите използваме широк набор от регресионни алгоритми, като на базата на степента на аномалност идентифицираме и потенциалните рискови наблюдения.

### Прекомерни отсъствия от работното място

Предвид значителният избор от различни регресионни алгоритми, следва да отчетем възможността някои от тях да има значително по-добри прогностични възможности спрямо други такива. За да проверим това оценяваме 106 алтернативни алгоритъма върху данни за прекомерно отсъствие от работното място представяме прогностичната им точност, измерена чрез корена от средната квадратична грешка, на *Фигура 10*.



*Фигура 10: Хистограма на прогностичната точност на регресионните методи за анализ на прекомерно отсъствие от работното място*

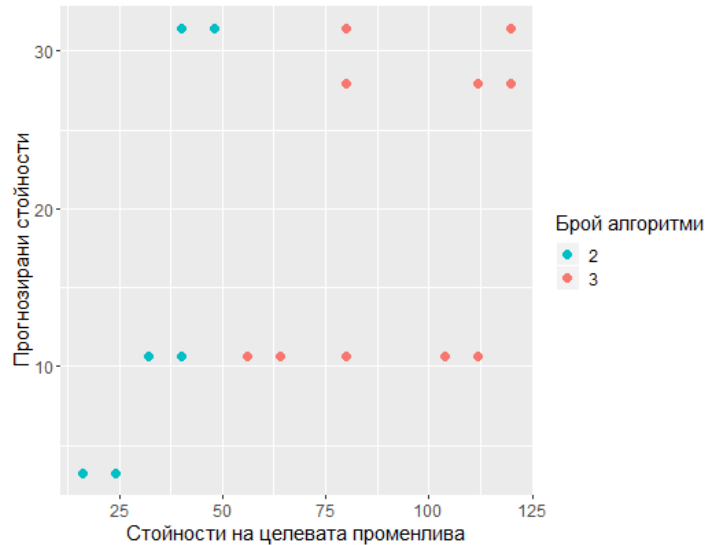
На нея отбелязваме значителна група алгоритми с прогностична точност в околността на  $RMSE = 13$ , като почти всичките разгледани алтернативи имат корен от средната квадратична грешка за тази задача в интервала от  $RMSE = 12$  до  $RMSE = 14$ . Най-добрите алгоритми регистрират  $RMSE \sim 12$ , но съществуват и няколко такива с особено слабо представяне, при които  $RMSE > 15$ . Разглеждайки резултатите от тази задача отбелязваме, че при нея почти всички алгоритми дават задоволителни резултати, като разликата между най-добрите и средните такива е сравнително малка. Ключовото в този случай е да се избегне използването на някои от особено слабите методи, тъй като техните резултати регистрират чувствително по-високи нива на грешката спрямо средната такава. Методите с най-ниски прогностични грешки са представени в *Таблица 8*. Веднага забелязваме, че шест от най-добрите десет метода са различни варианти от семейството на случайната гора. Най-добра прогностична точност има един от приложните методи на класификационните и регресионни дървета (gpart) със корен от средната квадратична грешка 12.17 и средна грешка от 0.62, последван от алгоритъма за екстремно засилване на градиента със съответно  $RMSE = 12.17$  и  $ME = 0.04$ .

Таблица 8: Прогностична точност на регресионни алгоритми за прогнозиране на часовете отсъствие от работа

Тип алгоритъм	Метод	Средна грешка ME	Корен от средна квадратична грешка, RMSE	Средна абсолютна грешка, MAE	Мярка за комплексност
Класификационни и регресионни дървета	rpart2	0.621	12.172	5.376	0.1%
Екстремно засилване на градиента	xgbDART	0.043	12.174	5.459	1.8%
Рандомизирана случайна гора	extraTrees	-0.314	12.229	5.418	2.9%
Гаусов процес с радиална ядрена функция	gausspr Radial	0.054	12.287	5.573	0.7%
Асемблирана класификационни и регресионни дървета	treebag	0.205	12.325	5.459	0.4%
Регуляризирана случайна гора	RRFglobal	-0.603	12.330	5.724	0.4%
Случайна гора с условни изводи	cforest	-0.141	12.371	5.581	14.5%
Условно дърво за изводи	ctree	0.798	12.381	5.387	0.3%
Регуляризирана случайна гора	RRF	-0.587	12.424	5.792	2.9%
Случайна гора	rf	-0.500	12.430	5.688	8.5%

Отчитаме изключително малката разлика в прогностичната точност между десетте най-добри метода – тя варира от  $RMSE = 12.17$  до  $RMSE = 12.43$ , което подчертава възможността изборът на метод сред това вече тясно множество да бъде извършен на база и на други съображения – необходими изчислителни ресурси, организационен опит и съображение, възможност за интерпретация на резултатите и интеграция с други системи и т.н. От гледна точка на използваните ресурси, отчитаме, че нито един от топ десетте алгоритъма не е особено високотелен от тази гледна точка. Най-добрият метод е 1,000 пъти по-бърз спрямо най-бавната алтернатива и има време на изчисление съизмеримо с това на многомерната линейна регресия. Тези резултати отчасти се дължат и на популярността на случайните гори, които водят до редица приложни методи в рамките на това семейство, като те често се стремят да оптимизират алгоритъма от гледна точка на използваните ресурси и оттам – времето за изчисление.

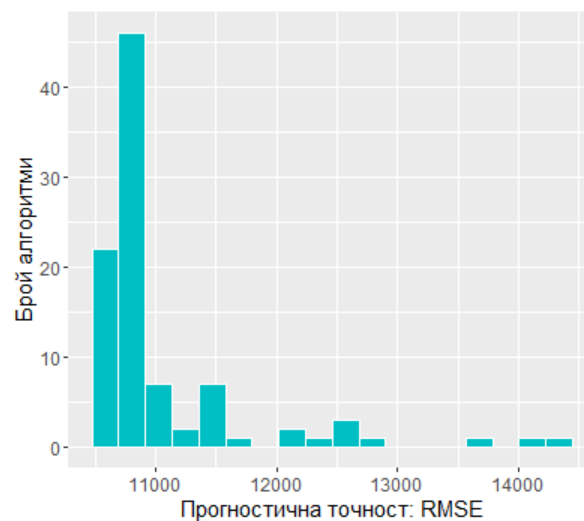
Разглеждайки часовете отсъствие от работното място спрямо разстоянието до дома на служителите не забелязваме ясно изразена тенденция сред аномалните наблюдения. Тук отново отбелязваме, че екстремните наблюдения са успешно идентифицирани с помощта на предложените критерии, като колкото по-екстремно е дадено наблюдение, толкова е по-голяма вероятността да бъде определено като аномално според повече от един критерий. *Фигура 11* фокусирано представя най-рисковите наблюдения от извадката според два или повече критерия. Тя позволява да се степенуват по приоритет различните събития, представляващи потенциални рискове, като адресирането им започва от най-високо приоритетните (аномалии според поне 3 алгоритъма) и продължи към по-ниско приоритетните такива. Този подход обезпечава рационалното използване на организационните ресурси и насочването им към дейностите, които биха отключили потенциално най-голяма бизнес стойност.



Фигура 11: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск

### Онлайн комуникация

В търсенето на оптимален прогностичен метод за рисковете при онлайн комуникация, изчисляваме прогнозата за целевата променлива с помощта на над 100 алтернативни алгоритъма, като разпределението на тяхната прогностична точност (RMSE) е представено графично на *Фигура 12*. Огромната част от тестваните алгоритми имат много добра прогностична точност от  $RMSE < 11,000$ . Наблюдаваме силно изместено разпределение, при което голям брой подходи генерират близки до оптималните резултати, а има и по-малък брой алгоритми с много по-слабо представяне.



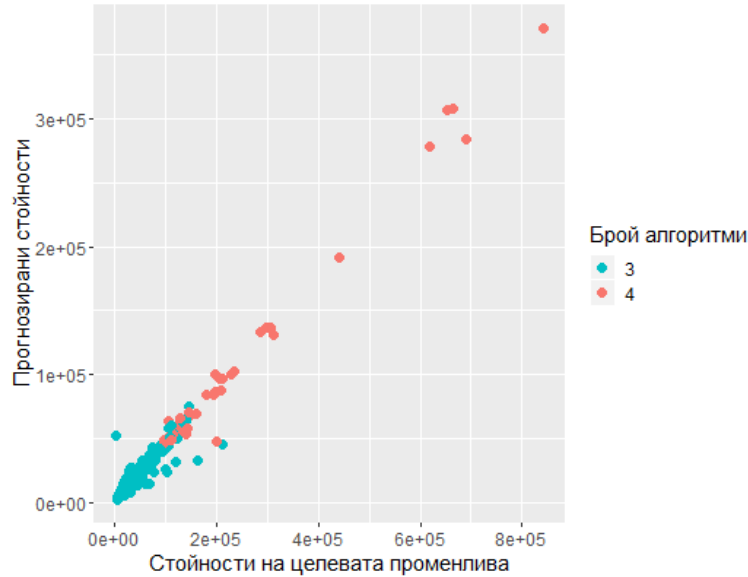
Фигура 12: Хистограма на прогностичната точност на регресионните методи при прогнозиране на онлайн комуникация

Десетте подхода с най-висока прогностична точност са представени в Таблица 9. Прави впечатление, че шест от тях представляват различни методи за изчисляване на случайна гора. Най-добрият алгоритъм в този случай е *ranger*, който представлява високо оптимизирано приложение на случайната гора, което постига съизмерима прогностична точност при много нисък разход на изчислителни ресурси. Неговият корен от средната квадратична грешка е едва  $RMSE = 10514$ , като и останалите десет алгоритъма са в тази околност. Извън методите за изчисляване на случайна гора, забелязваме и алгоритми на гаусов процес, бейсова и ласо регресия, както и обобщен линеен модел. От гледна точка на всички изчислени мерки за прогностична точност представените топ десет алгоритъма са трудно различими и изборът за оптималният такъв можем да бъде направен на база подходящ ресурсен или организационен критерий.

Таблица 9: Прогностична точност на регресионни алгоритми за определяне на споделяния на онлайн комуникация

Тип алгоритъм	Метод	Средна грешка ME	Корен от средна квадратична грешка, RMSE	Средна абсолютна грешка, MAE	Мярка за комплексност
Случайна гора	ranger	384.52	10514.60	3566.87	0.35%
Регуляризирана случайна гора	RRF	247.73	10532.25	3630.94	9.92%
Регуляризирана случайна гора	RRFglobal	231.90	10533.87	3633.60	1.46%
Случайна гора	rf	213.69	10573.80	3662.55	0.52%
Рандомизирана случайна гора	extraTrees	209.17	10584.73	3639.87	3.33%
Гаусов процес с полиномна ядрена функция	gaussprPoly	311.21	10636.78	3703.16	1.42%
Бейсова прагова регресия	bridge	349.77	10651.67	3637.69	0.30%
Условно дърво за изводи	cforest	366.84	10656.54	3646.23	0.68%
Обобщен линеен модел	glmnet	370.03	10657.97	3607.08	0.00%
Регресия от тип ласо	lasso	369.31	10659.49	3608.71	0.01%

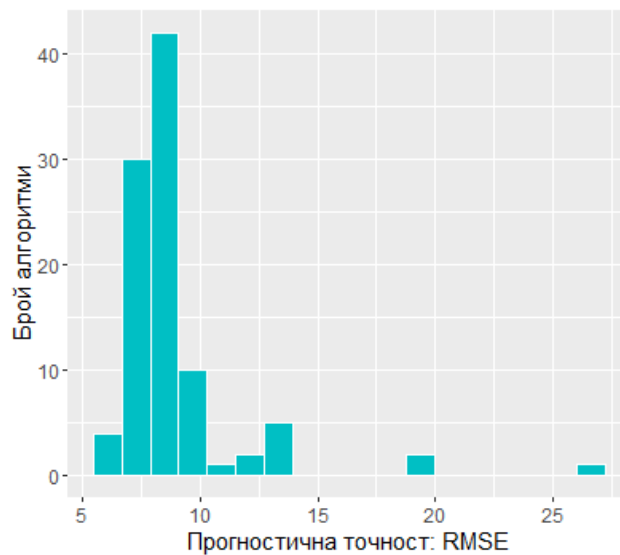
За целта може да бъде използвана мярката на комплексност като приближение на ресурсните нужди на дадения алгоритъм. С изключение на регуляризираната случайна гора всички останали алгоритми са сравнително бързи, като включително *ranger* е около 300 пъти по-бърз от най-бавния разгледан вариант. В този смисъл той е подходящ избор за регресионен алгоритъм за решаване на подобни задачи. За целите на управлението на операционните рискове предлагаме да се анализират наблюденията, които са идентифицирани като аномални от поне половината алгоритми. Прилагайки този критерий значително стесняваме множество казуси, които подлежат на активно управление – едва 1,827 (4.6%) наблюдения са класифицирани като аномални според поне 3 критерия и 43 (0.1%) – според четири. Тези наблюдения са представени визуално на Фигура 13 и подлежат на последващи действия за управление на риска.



Фигура 13: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск

### Оценка на цените на активи (недвижими имоти)

На база на данните за недвижими имоти са оценени и 106 алтернативни модела от сферата на машинното самообучение, като тяхната прогностична точност е детайлно изследвана спрямо генерираните от тях грешки при прогнозата.



Фигура 14: Хистограма на прогностичната точност на регресионните методи при оценки на цени на активи

Хистограма на тяхната прогностична точност спрямо коренът от средните квадратични грешки е представен на Фигура 14. Прави впечатление, че огромната част от методите имат отклонение RMSE в интервала от 7 до около 9. Най-добрите алгоритми сред изследваните



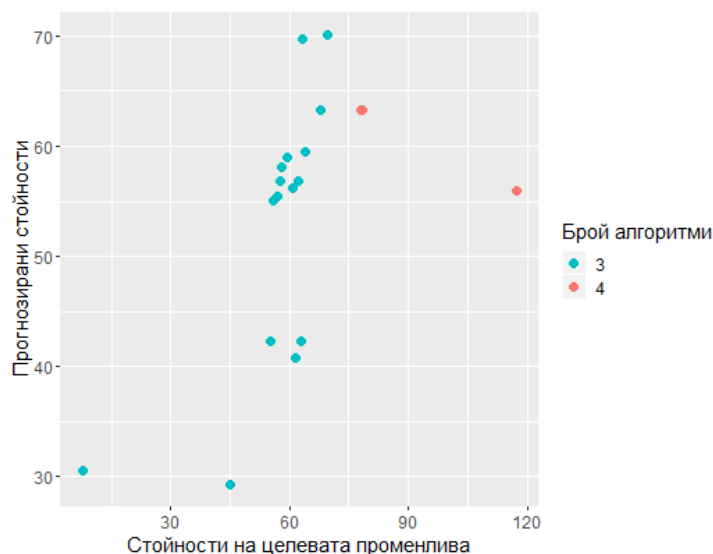
имат прогностична точност от  $RMSE < 6.5$ , а тези с най-лошите резултати могат да достигнат до стойност от  $RMSE$  над 25. Прави впечатление, че макар като цяло при решаването на тази задача да забелязваме значително групиране около една стойност, то има и екстремно слаби резултати. Малък брой от алгоритмите имат и малко по-добро представяне спрямо средната точност. В този смисъл има потенциална бизнес стойност в тестването и избирането на най-добрите сред тях. Десетте подхода с най-нисък корен от средната квадратична грешка са представени в Таблица 10. Веднага прави впечатление, че седем от тях са различни приложения в рамките на семейството на случайните гори. Всички те регистрират изключително добра прогностична точност, като коренът от средната квадратична грешка на прогнозите спрямо тестовата извадка е в интервала от  $RMSE = 6.46$  до  $RMSE = 7.10$ .

Таблица 10: Прогностична точност на регресионни алгоритми за определяне на цена на единица площ

Тип алгоритъм	Метод	Средна грешка ME	Корен от средна квадратична грешка, RMSE	Средна абсолютна грешка, MAE	Мярка за комплексност
Регуляризирана случайна гора	RRF	-0.750	6.459	4.831	64.8%
Квантилна случайна гора	qrf	0.001	6.470	4.695	5.2%
Регуляризирана случайна гора	RRFglobal	-0.832	6.568	4.890	9.9%
Случайна гора	ranger	-0.878	6.600	4.884	7.0%
Паралелна случайна гора	parRF	-0.936	6.689	4.965	3.8%
Случайна гора	ranger	-0.943	6.689	4.908	4.0%
Регуляризирана най-малки квадрати с радиално ядрена функция	krlsRadial	-0.435	7.068	5.388	14.8%
Бейсови адитивни регресионни дървета	bartMachine	-0.795	7.076	5.353	11.1%
Рандомизирана случайна гора	extraTrees	-0.948	7.081	5.137	7.8%
Гаусов процес с полиномно ядро	gaussprPoly	-0.501	7.082	5.443	2.1%

Останалите три алгоритъма, които не са от типа случайни гора са 2 базирана на ядрена функция и един – на гаусов процес, като прогностичната им точност е около  $RMSE = 7.10$ . Мярката за комплексност, отчитаща пропорционално време за изчисление спрямо най-ресурсоемният алгоритъм, също варира в много широки граници. Най-добрият метод – този на регуляризираната случайна гора е едва 35% по-бърз от най-бавният в извадката. От друга страна, вторият най-добър – квантилната случайна гора – е близо 20 пъти по-бърз от най-ресурсоемният, а разликата в прогностичната точност между двата е почти незабележима. Това подчертава, че и при този тип задача е възможно да се намери оптимална точка между ползите и разходите при изчислението на дадените алгоритми. Нещо повече, скоростта на изчисление на алгоритъма показва и възможността за преминаване от асинхронни към синхронни операции, т.е. от изчисление на модели и последващото им използване и обновяване в бъдещ момент към реално-времева аналитика, което се използва и обучава едновременно. Фигура 15 представя наблюденията с потенциално най-висок риск – те са класифицирани като екстремни спрямо три или четири от използваните четири подхода. Графиката позволява да проследим, че тези наблюдения не следват очакваната тенденция и

не показват връзка между прогноза и реализация. В този смисъл те са съвсем неочаквани и представляват потенциална реализация на операционни рискове.



Фигура 15: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск

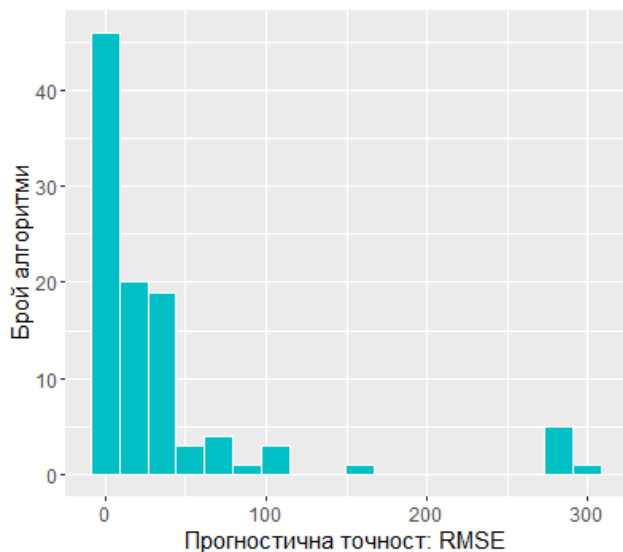
### Резки промени на пазарното търсене

Върху разглежданите данни изчисляваме 109 алтернативни прогностични модела с различни регресионни алгоритми от областта на машинното самообучение. Тяхната прогностична точност, измерена чрез корена от средната квадратична грешка, е представена на Фигура 16. Видимо от хистограмата, малкият брой наблюдения води до неустойчиви прогнозни резултати. Има голям брой алгоритми, които чрез свръхнагаждане достигат изключително ниски стойности на RMSE, близки до нулата. Забелязваме и малък брой алгоритми с изключително лошо представяне – RMSE в околността на 300. Този резултат е значително по-слаб от наивна прогноза за всяка стойност (пр. средната за извадката). Като потенциален начин да избегнем уповаването на модели с несигурно представяне извън обучителната извадка е удачно да не се разглеждат моделите с  $RMSE < 1$ , като подчертаваме, че при огромната част от тях то е близко до нула.

Сред останалите алгоритми можем да открием тези с най-висока прогностична точност, като десетте най-добри са представени в Таблица 11. Най-оптималните модели са на ансамбловите многомерни адаптивни регресионни сплайнове (MARS)<sup>7</sup>, като те достигат  $RMSE = 2.51$  и  $RMSE = 2.61$ . В тази задача ясно се открояват и редица представители на семейството на обобщените линейни модели, както и модели на машина с подкрепящи вектори. Отново отчитаме, че най-добре представящите се алгоритми не са най-ресурсоемките, което позволява на базата на прогностичната точност и мярката за комплексност да се вземе оптимизационно решение за оптимален модел за решаване на дадената задача. Предвид

<sup>7</sup> От англ. – Multivariate Adaptive Regression Splines, MARS.

точността и мярката за комплексност на модела на ансамблови многомерни адаптивни регресионни сплайнове (MARS) без ограничения, избираме него като алгоритъм за последващото тестване. Подчертаваме, че при подобен размер на извадката всички резултати следва да се интерпретират с повишено внимание и че при практическото приложение на предложения метод следва да се търсят начини по възможност за избягване на извадки с подобен размер.

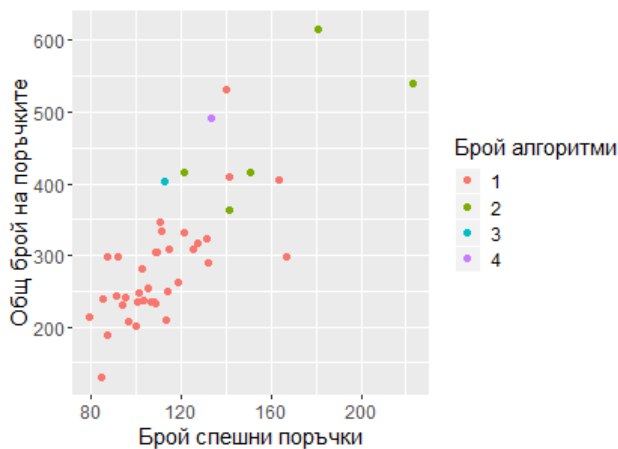


Фигура 16: Хистограма на прогностичната точност на регресионните методи при резки промени на пазарното търсене

Таблица 11: Прогностична точност на регресионни алгоритми за прогнозиране на логистични поръчки

Тип алгоритъм	Метод	Средна грешка ME	Корен от средна квадратична грешка, RMSE	Средна абсолютна грешка, MAE	Мярка за комплексност
Ансамблови многомерни адаптивни регресионни сплайнове	bagEarth	-0.056	2.509	2.076	1.12%
Ансамблови многомерни адаптивни регресионни сплайнове с ограничения	bagEarthGCV	0.107	2.605	2.185	0.48%
Мрежа с обобщен линеен модел	glmnet	-1.188	2.780	2.384	0.14%
Гаусов процес с линейно ядро	gaussprLinear	-0.966	2.843	2.357	0.28%
Регуляризирана бейсови невронни мрежи	brnn	-0.910	2.853	2.374	0.08%
Машина с подкрепящи вектори с линейно ядро 2	svmLinear2	-3.153	5.865	5.129	0.07%
Машина с подкрепящи вектори с линейно ядро	svmLinear	-3.153	5.865	5.129	0.28%
Гаусов процес с полиномно ядро	gaussprPoly	-3.527	7.345	5.660	0.16%
Машина с подкрепящи вектори по метода на най-малките квадрати и полиномно ядро	svmPoly	-3.146	7.660	6.936	0.29%
Засилен обобщен линеен модел	glmboost	-2.669	7.863	6.994	0.10%

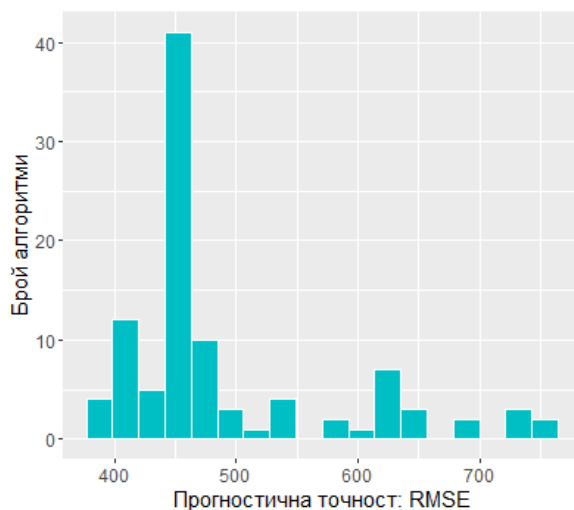
На *Фигура 17* са разгледани в повече детайли само наблюденията, които са класифицирани като аномални от поне един алгоритъм (критерий). Огромната част от тях са необичайно според само един подход и това най-често е локалният фактор на екстремна стойност, докато има само едно наблюдение, което е аномално според 3 алгоритъма и едно – спрямо четири. Подчертаваме, че е удачно в ситуациите на ненадзиравано самообучение като потенциално рискови да разглеждаме наблюденията, които са маркирани като такива от повече от един критерий.



*Фигура 17: Брой алгоритми, класифициращи всяко наблюдение като аномално спрямо потенциално рисковите наблюдения*

### Обработка на сигнали от клиенти

Използвайки автоматизиран подход за изчисление на прогнозните стойности за време за обработка на сигнал, можем значително да разширим обхвата на апробираните алгоритми за машинно самообучение. Върху настоящите данни оценяваме 109 различни алгоритъма, от които 102 достигат конвергенция и могат да бъдат използвани за генериране на прогнози.



*Фигура 18: Хистограма на прогностичната точност на регресионните методи при сигнали*

Разпределението на тяхната точност, измерена чрез корена от средната квадратична грешка (RMSE) е показано графично в хистограмата на *Фигура 18*. Огромната част от алгоритмите регистрират RMSE около 450, като прави впечатление дългата опашка на разпределението – тя показва наличието на немалък брой алгоритми с особено лошо представяне. Поради спецификата на задачите и данните е възможно да наблюдаваме тенденцията определени алгоритми да се справят особено добре с някои задачи, което да се компенсира от особено лошо представяне на други. Затова е неудачно да се използва един и същи алгоритъм при всяка една задача въз основа само на минали добри резултати от решаването на други задачи с друг набор от данни. Това е и една от причините в предложения тук общ алгоритъм за автоматизирана оценка и управление на операционните рискове да се включва компонент на тестване и апробация на широк набор от алтернативни алгоритми.

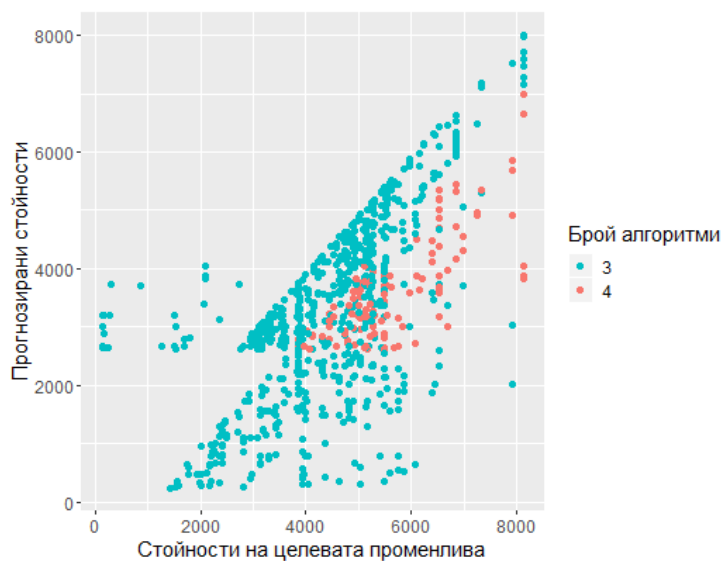
Десетте подхода с най-висока прогностична точност са обобщени в *Таблица 12*. Осем от десетте метода са разновидности и различни приложения на алгоритъма на случайната гора. Най-висока прогностична точност достига метода RRFglobal, като неговата прогностична точност е  $RMSE = 388.8$ , следван от още пет различни приложения на случайната гора със съизмерими нива на грешката. В списъка на десетте най-точни алгоритъма се нареждат и две приложения на екстремното засилване на градиента. От гледна точка на мярката за комплексност, най-точният метод има сравнително добри нива на ресурсоемкост, като той е 50 пъти по-бърз от най-бавната алтернатива. При необходимост от изчислителна оптимизация е възможно да се избере и метод със съизмеримо висока точност и по-ниски изчислителни нужди. За нуждите за настоящия анализ отчитаме представянето на най-точният алгоритъм като задоволително и можем да го използваме за нуждите на оценка и управление на риска.

*Таблица 12: Прогностична точност на регресионни алгоритми за определяне на време за обработка на сигнали*

Тип алгоритъм	Метод	Средна грешка ME	Корен от средна квадратична грешка, RMSE	Средна абсолютна грешка, MAE	Мярка за комплексност
<i>Регуляризирана случайна гора</i>	RRFglobal	-23.95	388.82	183.81	2.0%
<i>Регуляризирана случайна гора</i>	RRF	-24.22	389.64	184.59	13.2%
<i>Случайна гора</i>	rf	-24.30	389.75	184.61	1.1%
<i>Паралелна случайна гора</i>	parRF	-28.04	393.42	188.02	0.7%
<i>Рандомизирана случайна гора</i>	extraTrees	-23.79	405.33	187.67	5.6%
<i>Случайна гора</i>	ranger	-24.27	406.17	188.66	0.3%
<i>Екстремно засилване на градиента</i>	xgbLinear	-17.56	407.67	188.13	0.3%
<i>Екстремно засилване на градиента</i>	xgbTree	-10.43	411.34	191.64	0.2%
<i>Засилено дърво на решенията</i>	bstTree	-8.33	412.05	185.00	0.1%
<i>Бейсови адитивни регресионни дървета</i>	bartMachine	-7.16	414.76	196.90	55.5%

В крайна сметка е удачно организациите да фокусират своето внимание само върху високорисковите транзакции в рамките на своите оперативни процеси. Автоматизираното

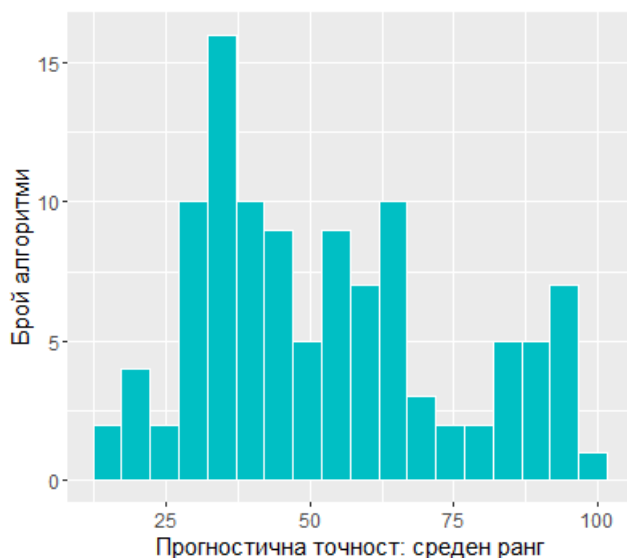
извеждане на наблюдения (транзакции или агенти) с високо ниво на риск по необходимост следва да бъде резултат от приложението на повече от един критерий, като различните критерии и методи в крайна сметка автоматично достигат до относителен консенсус за аномалността на даден случай. Паралелът с процес по управление на риска, воден от човешки агенти, би бил достигането до съгласие на група от експерти или ръководители. В този смисъл предлагаме критерият за наличие на потенциален риск да бъде повече от половината автоматични алгоритми да го класифицират като такъв. *Фигура 19* графично показва рисковите наблюдения (определени от поне три критерия), както и високо-рисковите наблюдения (определени от всички четири критерия). Това са именно наблюденията, които подлежат на активно управление след тяхната автоматизирана идентификация.



*Фигура 19: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск*

В настоящата глава е осъществен преглед на пет основни ситуации, в които могат да възникнат потенциални операционни рискове – при отсъствие на служители от работното място, при оценка на недвижими активи, при управление на комуникацията със заинтересовани страни, при промени в пазарното търсене, както и при обработка на сигнали от клиенти. Основната задача на управлението на тези рискове е да се оптимизира баланса между потенциално негативни и потенциално позитивни последствия, като за тази цел прогнозираме очакваните изходи и търсим значителни отклонения от тях. Като потенциален подход за автоматизация тук предлагаме първоначално да се приложи общия алгоритъм за управление на операционните рискове и на база прогнозните стойности да се дефинират значителните отклонения сред тях. За тази цел предлагаме да се използва консенсусен подход, при който за потенциална аномалия се маркира всяко наблюдения, определено за аномално от повече от половината използвани критерии. Представеното тук изследване позволява да очертаем и някои основни изводи при този тип задачи.

Първо, важноста при избор на подходящ регресионен алгоритъм при управлението на операционните рискове е още по-ясно изразена. Това е ясно видимо както в представените по-горе резултатите, така и в ранговото разпределение на прогностичната точност. Всички разгледани алгоритми са ранкирани по такъв начин, че този с най-висока точност, измерена чрез корена от средната квадратична грешка, получава ранг от 1. Алгоритъмът с втората най-висока точност – от 2 и така до най-неточният алгоритъм, който получава ранг от 109. *Фигура 8* представя хистограма на осреднените рангове на разгледаните алгоритми. Съществува малък брой алгоритми, които устойчиво генерират прогнози с високо ниво на точност спрямо всички разгледани задачи. От друга страна, това не е така при повечето алгоритми, като повечето от тях имат среден ранг между 25 и 75, но наблюдаваме и малък брой особено неуспешни алгоритми със среден ранг над 80. Това подчертава ползата от прилагането на етапа на селекция на статистическите алгоритми преди използването им в конкретни научни или приложни задачи.

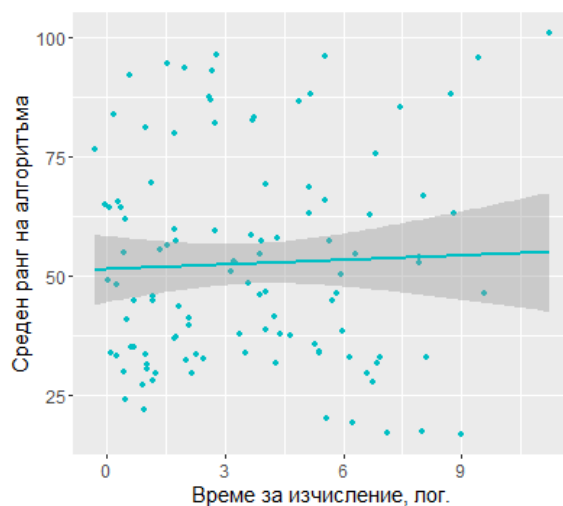


*Фигура 20: Разпределение на регресионните методи спрямо тяхната осреднена прогностична точност*

На второ място отчитаме, че и при регресионните алгоритми съществуват групи такива, които водят до устойчиво добри резултати. Това често са различните приложения на метода на случайна гора, като прогнозните стойности на различните алтернативни варианти на случайни гори са много близки една до друга. Това дава основание да се използват високо-оптимизирани варианти на метода (пр. имплементацията *ranger* в езика R), тъй като те позволяват значителна икономия на изчислителни ресурси на практика без загуба на прогностична точност. Класическите статистически методи като линейната регресия в общия случай регистрират много по-ниска прогностична стойност спрямо методите от сферата на машинното самообучение и е удачно те да бъдат заместени или поне допълнени от по-съвременни подходи. Изключение в този смисъл са ситуациите, в които се анализират изключително малки извадки. При тях не наблюдаваме разлика между линейният регресионен модел и алтернативните алгоритми от сферата на машинното самообучение. От

друга страна, мерките за точност при такива извадка са неустойчиви и с по-ниска надеждност и този резултат следва да се интерпретира с повишено внимание. При всички случаи, предложените методи имат по-добри резултати когато са запазени с по-голям обем от данни.

На трето място отбелязваме, че на практика няма връзка между прогностичната точност и необходимите изчислителни ресурси за даден алгоритъм. *Фигура 9* представя графично връзката между среден ранг на алгоритъма и логаритъм на времето, необходимо за изчисление му. Визуално не се наблюдава връзка между тези две характеристики на разглежданите алгоритми. Формално тестваме връзката като оценяваме регресия на средния ранг спрямо необходимото време за изчисление. Макар коефициентът пред времето да достига статистическа значимост на нива от 5%, то размерът му е изключително малък:  $\beta = 0.0006$ , а обяснената дисперсия е едва 4%, което показва, че липсва висока практическа значимост. Това показва, че е удачно при автоматизацията на процеса по управление на операционния риск като допълнителен второстепенен критерий на етапа на селекция на оптимален прогностичен алгоритъм да се добави и неговата ефективност от гледна точка на необходими изчислителни ресурси. Почти незабележимата връзка между време за изчисление и прогностична точност отново подчертава възможността за избор на алгоритъм с висока точност и сравнително ниски изчислителни нужди.



*Фигура 21: Разпределение на регресионните методи спрямо тяхната осреднена прогностична точност*

Четвърто, ситуациите на операционен риск, при които целевата променлива е продължителна, позволяват да се апробират набори от критерии за аномалност. Изложените тук резултатите показват, че предложеният подход за консенсус между четири основни критерия е удачен и работещ. Разгледахме едновременното приложение на два статистически критерия (формулата на Тъки и критерият за 4 стандартни отклонения разлика), както и два алгоритъма за клъстериране, базирани на плътността на дадена околност (алгоритмите DBSCAN и LOF). Анализът показва, че колкото по-далеч от основни



кълъстери в данните е дадено наблюдение, толкова е по-вероятно то да бъде идентифицирано като аномално от повече от един критерий. От друга страна, приложението на всеки от четирите критерия води до много различен брой потенциално необичайни наблюдения, като някои от критериите генерират нереалистично голям брой аномални маркировки. С цел избягване на този проблем и минимизиране на броя на неверните позитивни класификации предлагаме като рискови да се класифицират само наблюдения, които са такива според поне 3 критерия (или над 50% от използваните критерии), а като високо-рискови – според поне 4. Този подход намалява броя на маркираните като рискови наблюдения и позволява предприемане на фокусирани действия за управление спрямо тях.

## **Глава 5: Автоматизирана система за управление на операционния риск**

Автоматизацията на процеса по управление на операционните рискове в условията на дигитална трансформация неизбежно преминава през изграждане на информационни системи, които да изпълняват автоматично задачите по набиране, обработка и анализ на информация, както и да оказват подкрепа при взимането на решения и предприемането на конкретни управленски действия за адресиране на потенциалните рискове. Качеството на автоматизираните аналитични процеси, с които е натоварена управленската информационна система за оценка и управление на операционния риск е функция на обема и качеството на данните, с които тя е захранена. В този смисъл въпросът за информационното обезпечаване на системата е ключов за реалното осъществяване на автоматизирано управление на риска. Като първи най-важен критерий за използване на даден масив от данни за нуждите на общия алгоритъм за автоматизирано управление се налага нивото на качество на разглежданите данни. Независимо от източника на данните, те следва във висока степен да отговарят на изискванията за качество както от технологична, така и от организационна и бизнес гледна точка.

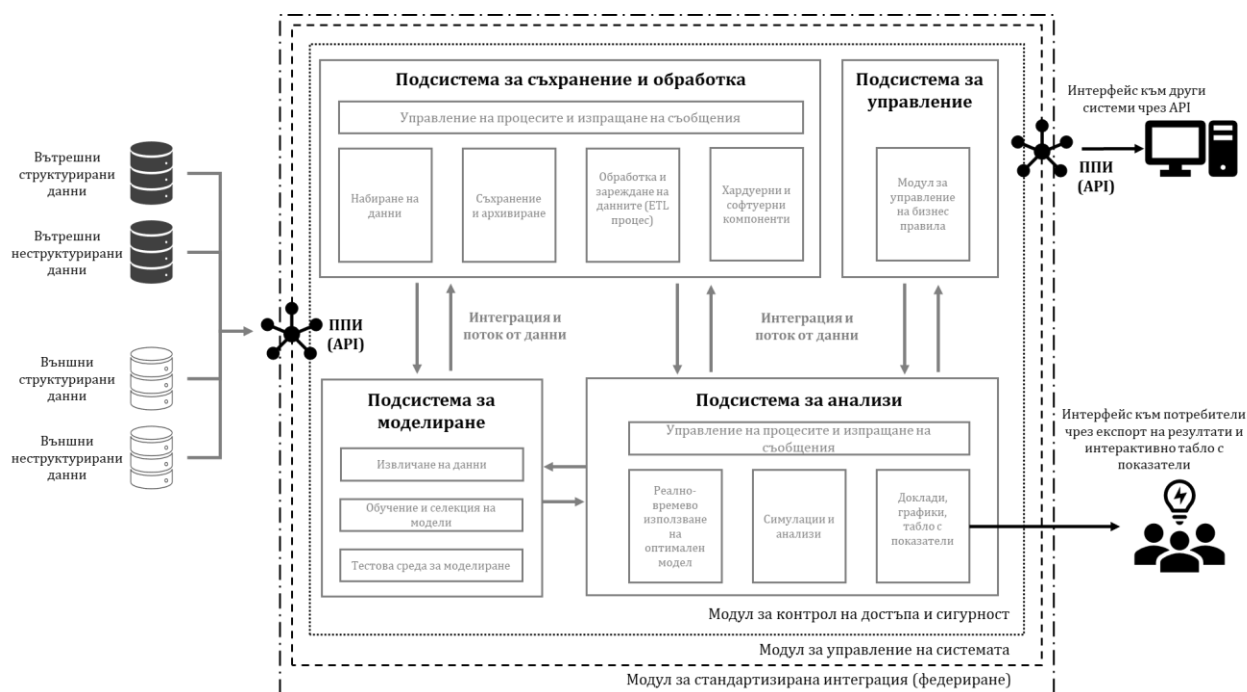
Предвид високата информационна-интензивност на предложения алгоритъм за автоматизирано управление, то като най-важни измерения на качеството в този случай отчитаме следните:

- Точност (липса на грешки);
- Обективност;
- Приложимост;
- Подходящо количество;
- Добавена стойност.

Единствено набори от данни, имащи задоволителни резултати спрямо тези измерения подлежат на включване в УИС за управление на операционните рискове. Основните източници на тези данни са два – вътрешни за организацията бази от структурирани и неструктурирани данни, както и публично достъпни данни.

Естественят завършек на процеса по автоматизация на операционните рискове е успешната му интеграция и изпълнение в рамките на единна информационна система. На базата на

общата рамка за ИТ архитектура на предприятието на Отворената Група TOGAF<sup>8</sup> (вж. Weismann, 2011; Harrison, 2018), както и предложените модели от Klein et al. (2016) и други автори (Helu et al., 2017; Hashem et al., 2016; Clement et al., 2017; Theorin et al., 2017; Mayer et al., 2019; Raghupathi & Raghupathi (2014)) можем да разширим тези референтни архитектури по такъв начин, че те да могат да обслужват процеса по управление на операционните рискове на база на общия алгоритъм за автоматизирано управление. За целта следва да се добавят специални модули, които да обезпечат предварителната селекция на оптималния модел, да се разширят аналитичните модули, така че да използват в реално време оптимален модел, но и да могат да осъществят по-сложни задачи като анализа на чувствителността и симулации, както и да се разшири обхват на модулите за сигурност и управление на системата. Допълнително системата следва да включва модул, който да позволи управлението на риска веднъж след като той е идентифициран. Този модул поддържа информация за набора от действия, които следва да се предприемат при различни нива на риска и позволява интеграция към други системи, които да извършат желаните действия (пр. прекратяване на транзакция). Схематично разширената система е показана на *Фигура 22*.



*Фигура 22: Референтна архитектура на управленска информационна система за автоматизирано управление на операционния риск*

Като основен принос на настоящата глава се явява предложената тук референтна архитектура на специализирана управленска информационна система. Архитектурата е така изградена, че отделните подсистеми съответстват на стъпките на общия алгоритъм, предложен в Глава 2. Подсистемата за съхранение и обработка е натоварена с функциите по

<sup>8</sup> От англ. – The Open Group Architecture Framework, TOGAF.

информационно обезпечаване (стъпки 2 и 3 от алгоритъма). Подсистемата за моделиране отговаря за избора и оценката на параметрите на оптималния модел за управление на риска във всяка една от ситуациите или задачите, които се решават. Целта ѝ е да автоматизира дейностите по обучение на модел за управление на операционния риск (стъпки 4 и 5 от алгоритъма). Подсистемата за анализи автоматизира процеса по приложение на модела (стъпка 6), а интеграцията към други системи дава възможност за автоматизация и на процеса по управление (стъпка 7 от алгоритъма). Процесът за постоянно усъвършенстване (стъпка 8) се автоматизира чрез програмирана периодична преоценка на оптималните модели и техните параметри. Извън тези подсистеми, архитектурата включва и три хоризонтално модула, които са необходими за безпроблемната експлоатация на УИС.

На практика огромната част от транзакционните данни в съвременните информационни системи не са маркирани данни. Те съдържат подробна информация за транзакцията (пр. идентификатори на извършител и сесия, времеви печат, същност и детайли на самата транзакция, система на извършване и т.н.), но индивидуалните транзакции са отбелязани като рискови или не. Това се дължи в голяма степен на факта, че повечето информационни системи на този етап записват транзакциите, но рядко ги класифицират автоматично (Beunon-Davies, 2016; Dumas et al., 2018, стр. 257-277). Затова и често входните данни в алгоритъма за автоматизация са немаркирани данни, при които следва да се извърши ненадзиравано обучение за търсене на аномалии. От една страна този подход е по-неточен спрямо надзираваното обучение (Chandola et al., 2009), но от друга позволява още по-високо ниво на автоматизация и интеграция в рамките на подходяща информационна система. Това е така, тъй като при ненадзираваното обучение извършваните от експерти дейности за определяне на целевата променлива и нейните желани стойности отпадат и процесът на оценка на риска може да протече изцяло без човешка намеса.

Можем да осъществим автоматична идентификация и оценка на потенциално аномалните (рискови) наблюдения с помощта на вече описания набор от алгоритми. На първо място използваме статистически алгоритъм за разстояние на Махаланобис. Основната му идея е да се изчисли разстоянието между дадено наблюдение и останалите такива в многомерното пространство и на тази база да се идентифицират аномалиите (De Maesschalck et al., 2000). Добре известен факт е, че тези разстояния на Махаланобис приблизително следват  $\chi^2$ -разпределение (Hardin & Rocke, 2005). Затова и изчисляваме точните нива на значимост на разстоянието на Махаланобис на всяко наблюдение, като го сравним с теоретично  $\chi^2$ -разпределение. Отчитаме тези наблюдения, които имат статистически значима разлика спрямо референтното  $\chi^2$ -разпределение като аномални. Като критерий за аномалност приемаме точни нива на значимост по-ниски от 5%.

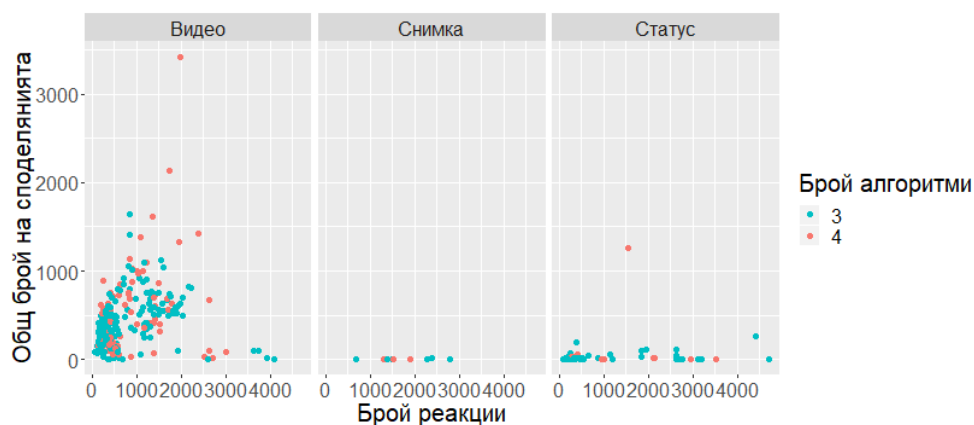
Втората група подходи, които използваме за търсене на аномални наблюдения в рамките на процеса по ненадзиравано самообучение са алгоритми за клъстериране, базирана на плътността на дадена околност. Тъй като тук се фокусираме върху контекстуалните аномалии е удачно да използваме многомерните версии на тези подходи. По-конкретно използваме многомерната версия на алгоритъма DBSCAN, който цели да идентифицира

околности с различни плътности (Soni & Ganatra, 2016; Hahsler et al., 2019). Основната идея е, че аномалните поведения се намират в околности с ниска плътност, а нормалните – с висока такава. По подобен начин действа и алгоритъмът за локален фактор на екстремна стойност (LOF), анализирайки малки околности, на чиято база изчислява мярка за аномалност – LOF (Ma et al, 2013).

Като последен потенциален подход за идентифициране на рискови поведения можем да използваме анализа на главни компоненти (Zenati et al., 2018). При него размерността на данните се намалява, като от тях се извеждат определен брой главни компоненти. Обичайните наблюдения следва да имат високи натоварвания на първите по значение главни компоненти, докато аномалните такива – на по-ниски по значение главни компоненти. Разстоянията между наблюдения и компоненти могат отново да бъдат измерени с помощта на разстояния на Махаланобис и по този начин да се изчисли мярка за аномалността на всеки отделен случай (Filzmoser & Todorov, 2013).

### Маркетингова комуникация чрез социалните мрежи

Предвид значителните различия между алтернативните алгоритми предлагаме за аномални да бъдат класифицирани наблюдения, при които поне три алгоритъма имат консенсус за това. Това предложение се базира на основният резултат, че сборът от различни алгоритми като цяло има по-добра точност отколкото всеки един индивидуален от тях (вж. Makridakis et al., 2020). На *Фигура 23* са представени графично наблюденията, които са класифицирани като аномални от съответно три или четири алгоритъма. Аномалните случаи според три алгоритъма са 256 на брой (3.63% от общата извадка), а тези, които са аномални според четири алгоритъма са 92 на брой (1.3%). Този брой е достатъчно малък, за да позволи съответната по-задълбочена проверка или при невъзможност за провеждане на такава – дори автоматично предприемане на действия от страна на информационната система.



*Фигура 23: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск спрямо различните типове Фейсбук Лайв сесии*

Можем да разгледаме и профила на аномалиите спрямо типа статус, който е постнат във Фейсбук Лайв сесията (вж. *Фигура 11*). Преди всичко отбелязваме, че при хипер-връзките (линковете) не наблюдаваме аномални поведения, а само при другите три типа статуси. Те са

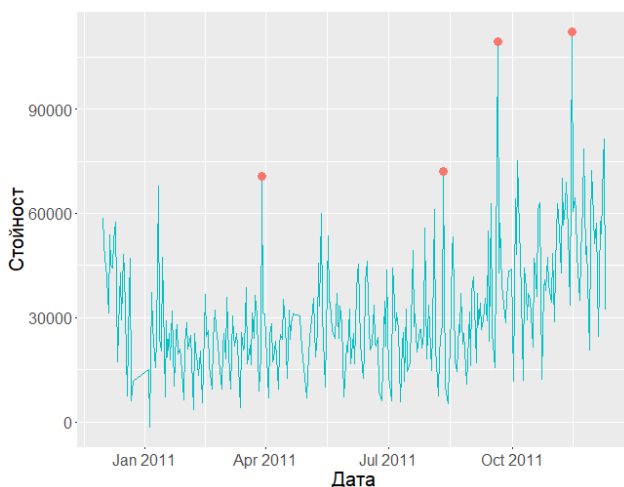
най-ясно изразени при видео материалите и при текстовите такива (статусите), като алгоритмите успяват да маркират както аномално високи наблюдения – с много споделяния, така и аномално ниски такива – с прекалено малко такива. Освен при линковете и при снимките имаме сравнително обичайно поведение, което показва тяхната устойчивост, но и липса на потенциал за генериране на неочаквано висока потребителска ангажираност. Множеството аномалии при видео материалите отразяват техният голям брой, но и значителна дисперсия. Получените дотук резултати ясно показват, че статусите с видео се характеризират с най-голям потенциал за генериране на положителни рискове, но и с най-високо ниво на ангажимент. В този смисъл е препоръчително за организацията да проучи допълнително видео материалите с неочаквано добри резултати и да ги репликира, като в процеса отключи както бизнес стойност за себе си, така и подобри потребителското преживяване за клиентите си.

### **Управление на търсенето в електронен магазин**

Търсенето на потенциални рискове на ниво индивидуална транзакция носи известни предизвикателства. Най-забележително, тук проблемът е, че целим да идентифицираме контекстуални аномалии, но значителна част от контекста липсва. Важна част от него е извършителят на действието, тъй като дадени наблюдения биха били аномални ако са извършени от един клас агенти (пр. клиенти), но напълно обичайни ако са извършени от друг клас агенти (пр. служители). В този смисъл при задачите за идентификация на риска е винаги удачно да се работи с данни и за агентите, извършили действието, като те могат да са човешки агенти или автоматизирани такива (системи, интерфейси, скриптове и др.). В случаите когато идентификацията на извършителя е невъзможна поради липсващи данни, то може да се използва подход на търсене на аномалии на ниво транзакции. В останалите случаи е удачно да се използват подходи, които включват както особености на транзакциите, така и особености на агентите.

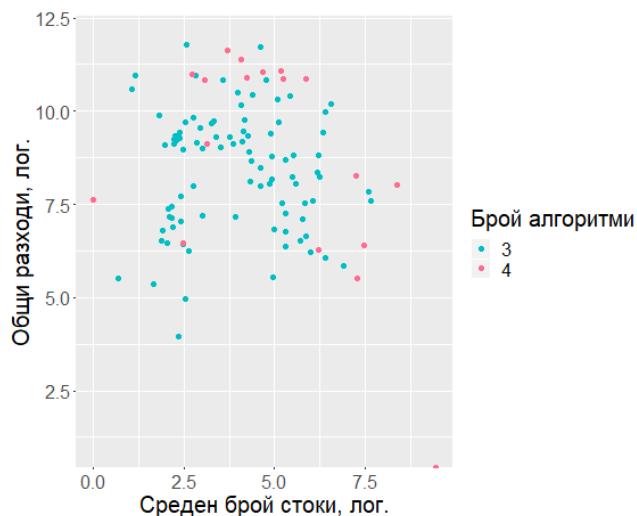
Анализът на данните за търговия на стоки или услуги и идентифицирането на екстремно високи или ниски стойности на потребителското търсене е класическа задача при управлението на операционния риск. Това е важно, тъй като позволява оптимално зареждане, което да обезпечи успешната търговия, както и позволява да се отделят необходимите човешки и организационни ресурси за посрещане на пиковите и да се оптимизират разходите при спадовете. В епохата на сравнително малките масиви от данни тази задача се решава на агрегирано (обобщено) ниво, като се разглеждат общи количества разменени стоки или услуги и се взимат решения за необходимите наличности и служители. В ерата на големите масиви от данни, ниските изчислителни разходи и гъвкавите статистически алгоритми вече се налага като практика задачата да се решава на ниво индивидуална поръчка или индивидуален потребител (агент). Това позволява и рискът от екстремно поведение да бъде преценен на много по-ниско ниво и на това ниво да бъдат предприети и съответните действия за адресирането му. Като първа стъпка от анализа на аномалиите на ниво индивидуален агент е удачно да се разгледа общия контекст на наблюдаваните действия. В случая разполагаме с детайлни времеви редове на транзакциите и можем да отчетем доколко тяхното поведение на агрегирано ниво е обичайно и дали наблюдаваме рискове на това по-

високо ниво. За целта е удачно да използваме методи за търсене на аномалии в рамките на времеви редове. Припомняме, че даден времеви ред може да бъде разложен на основните си компоненти – дългосрочна тенденция (тренд), сезонни компоненти и случайни остатъци (вж. Cleveland et al, 1990). За приложение на този подход и избор на оптимален модел, насочваме читателя към Gerunov (2016). Можем да извършим идентификация на рисковите наблюдения като използваме подходящ алгоритъм за идентификация на необичайните отклонения в реда от случайни остатъци при декомпозицията.



*Фигура 24: Аномални наблюдения на ниво обща стойност на агрегирани транзакции в данни за онлайн магазин*

На *Фигура 24* можем да видим аномалните наблюдения спрямо общият дневен оборот. Автоматизираният алгоритъм отчита четири такива (или 1.32% от извадката), като всички те са неочаквано високи стойности на оборота. В този смисъл това са рискове с положителни последствия. Макар организацията да има възможност да предприеме мерки за увеличаване на благоприятния им ефект, то не е необходимо да се предприемат мерки за предотвратяването или минимизирането им. Доколкото би съществувал проблем, той е свързан с възможностите за доставка на отправените поръчки. Предвид съвпадението между тези аномалии и аномалните наблюдения на ниво транзакции, организацията може да разглежда и двата подхода като удачни за нуждите на управлението.



Фигура 25: Брой алгоритми, класифициращи всяко наблюдение като аномално при наблюденията с най-високо ниво на риск

Прилагайки консенсусния подход за съгласие между различните алгоритми маркираме като потенциално аномални наблюденията, отбелязани от поне три алгоритъма. Тези наблюдения, определени за високо-рискови, са графично представени на Фигура 25. Аномалиите, идентифицирани като такива от поне три алгоритъма са 84 на брой или 1.92% от извадката, а маркираните от четири алгоритъма – 20 на брой или 0.46% от извадката. Тези пропорции показват, че методът на консенсус е успешен и при тази задача, като успешно успява да идентифицира като високо-рискови достатъчно малък брой наблюдения, които впоследствие да подлежат на дейности по управление на операционния риск.

Допълнителното предимство, което се получава при този анализ е, че тези наблюдения са конкретни индивиди, които могат да бъдат допълнително проучени или подходящо таргетирани, така че отрицателните рискове, към които организацията има експозиция да бъдат минимизирани, а положителните – максимизирани. Отбелязвайки, че предложената информационна система има някои характеристики на ограничен изкуствен интелект е удачно да измерим степента на съответствие на предложената система с основни етични критерии за нейната полезна и безопасна експлоатация. Обобщените параметри на съответствието са представени в Таблица 13, където оценяваме степента на съответствие с дадените критерии по тристепенна качествена скала (ниско-средно-високо). Оценката е експертна и е обоснована в последната колона от таблицата. Спряно шестте разгледани критерия на Bostrom & Yudkowski (2011) идентифицираме три със средно ниво на съответствие и три с високо ниво на съответствие на системата. Предвид, че предложената автоматизирана управленска информационна система за управление на операционните рискове не е критична система, която управлява процеси, обезпечавачи човешки живот или с функционалности, които го застрашават, то това ниво на съответствие е достатъчно за въвеждане на системата в продуктивен режим.

Таблица 13: Таблица на съответствието между етичните критерии на Bostrom & Yudkowski (2011) и автоматизираната УИС за управление на операционните рискове

№	Критерий	Степен на съответствие	Обяснение за степента на съответствие
1	<i>Отговорност</i>	Средно	Предложената система е така конструирана, че по дизайн целта ѝ е да намали отрицателните ефекти (компонент на чист риск) и да увеличи положителните такива (положителни отклонения от очакването). Макар това да е направено в организационен контекст, то подобряването на ефективността на определени организации води и до подобряване на общата производителност и бизнес среда и оттам – до увеличение на общественото благоденствие.
2	<i>Прозрачност</i>	Високо	Предложената система генерира голяма количество визуална информация и доклади, които могат да се използват за междинен, краен и последващ контрол от човешките експерти. Крайните решения се взимат по прозрачен начин (консенсус между алгоритми) и подлежат както на интуитивни обяснение, така и на визуализация.
3	<i>Проследимост</i>	Високо	Хоризонталният модул за управление на системата съдържа богати функционалности за записване на всяко действие и събитие под формата на журнални записи, както и за техния анализ. Наличието на подобни възможности позволява пълна проследимост на действията на системата както в реално време, така и под формата на последващ одит.
4	<i>Устойчивост</i>	Средно	Хоризонталният модул за информационна сигурност и управление на достъпа гарантира възможност за внимателно разпределяне на нивата на достъп според индивидуални агенти, групи потребители и функции на потребителите. Допълнителните функционалности за сигурност също допринасят за по-високата устойчивост на системата, особено към външни атаки. За да се постигне най-високо ниво тук е удачно системата да се внедри в архитектурата на информационна сигурност на организацията и да
5	<i>Предвидимост</i>	Високо	Нивото на предвидимост е високо, тъй като системата ползва набори от познати алгоритми с известно действие, които могат да генерират надеждни решения. В случаите когато се използват недетерминистични алгоритми (пр. невронни мрежи или случайни гори), дизайнът е такъв, че предполага обучение на моделите върху голям брой извадки, което да обезпечи елиминирането на случайните стохастични грешки. От гледна точка на критериите за аномалност обезпечаваме предвидимост като поддържаеме устойчивост на моделите и техните параметри.
6	<i>Избягване на щети за невинни хора</i>	Средно	Предложената автоматизирана УИС не предполага интеграция със системи, които имат функционалност



			за нанасяне на физически наранявания върху човешки същества. Управлението на операционните рискове се занимава предимно с разпределение и оптимизация на ресурси или процеси. Потенциални щети могат да възникнат при неправилно решение на системата, което лишава индивид от ресурси, които иначе би следвало да получи (заем, предимство при ред, плащане и др.), но не получава. Това може да доведе до потенциални материални и нематериални щети, но целта на процеса по постоянно усъвършенстване на системата (стъпка 8 от алгоритъма) е постепенно да намалява и минимизира тези погрешни решения.
--	--	--	---

#### IV. Справка за научните и научно-приложните приноси

Проведените изследвания в настоящата дисертация създават приноси в три основни групи: научни, научно-приложни и методологични, като основните сфери в които те са позиционирани са икономиката и наука за управление на риска, иконометрия и управленския информационни системи.

Сред **научните приноси** се нареждат следните:

- Изведен е общ алгоритъм за автоматизирано управление на операционните рискове чрез анализ и синтез на съществуващата литература и авторски разширения по метода на проектиращата наука (design science);
- Въз основа на тестване на 136 класификационни алгоритъма от сферата на машинното самообучение за изведени тези с най-висока прогностична точност при решаване на задачи за управление на операционните рискове с характер на дискретен избор;
- Въз основа на тестване на 109 регресионни алгоритъма от сферата на машинното самообучение за изведени тези с най-висока прогностична точност при решаване на задачи за управление на операционните рискове с продължителна целева променлива;
- Предложена е референтна архитектура на управленска информационна система, която автоматизира напълно дейностите от изведения общ алгоритъм за управление на информационните рискове и може да интегрира апробираните статистически алгоритми.

**Научни-приложните приноси** са следните:

- Операционализирана е концепцията за операционен риск по такъв начин, че да бъде директно приложима към бази данни;
- Предложен е консенсусен критерий за определяне на дадено наблюдение като потенциално рисково, който разчита на съгласието на четири различни индивидуални критерия от сферата на статистиката и машинното самообучение;

- Към предложената управленска информационна система са приложени етични критерии, разработени за анализ на системи или дигитални агенти с автономно взимане на решения.

Основните **методологически приноси** на дисертацията са както следва:

- Оценени и тествани са 136 алгоритъма с класификационен и 109 алгоритъма с регресионен характер, като е показано как те могат да бъдат приложени към задачи от сферата на управление на рисковете. Мнозинството от тези алгоритми не са прилагани към подобен тип задачи.
- Използвани са елементи и подходи от сферата на проектиращата наука за извеждане както на обобщения управленски алгоритъм, така и на референтната архитектура на информационната система, което показва как проектиращата наука може да бъде приложена към икономически задачи с интердисциплинарен характер<sup>9</sup>.

Комбинацията от тези приноси води както до нови резултати в областта на управлението на операционния риск, така и до създаване на конкретен дигитален артефакт (архитектура на управленска информационна система), който може да бъде използван в практиката.

## V. Публикации по темата на дисертационния труд

Представеният дисертационен труд е апробиран чрез публикации в страната и чужбина, изнасяне на доклади на специализирани научни конференции, както и обсъждане в неформален порядък с редица колеги, работещи в областта на управлението на риска, взимане на икономически решения, икономическото и бизнес моделиране. Публикациите по темата на дисертационния труд се изразяват в един монографичен учебник, една глава от колективна монография, три студии и пет статии, както следва:

### Монографичен учебник:

1. Gerunov, A. (2017). *Notes on Risk Management*. Sofia University "St. Kliment Ohridski", Faculty of Economics and Business Administration. ISBN: 978-954-9399-45-5.

### Глава от колективна монография:

2. Герунов, А. (2020). Финансови и бизнес аспекти на инвестицията в защита на данните. Глава 7 в Герунов и др. (ред.), *Privacy by Design: Принципи, практики и технологии*. СУ „Св. Климент Охридски“, Стопански факултет. ISBN: 978-954-9399-59-2.

### Студии:

3. Герунов, А. (2020). Приложение на класификационни алгоритми за моделиране на икономически избори. *Икономическа мисъл*, 2, 45-67.

---

<sup>9</sup> Използването на проектиращата наука в икономически задача е сравнително рядък феномен. Макар някои автори да го отчитат като потенциален метод в сферата на икономиката (Gonzalez, 2008), то приложенията са все още силно ограничени.

4. Gerunov, A. (2020). Binary Classification Problems in Economics and 136 Different Ways to Solve Them. *Bulgarian Economic Papers*, 2/2020, 1-31.
5. Герунов, А. (2019). Управление на риска: типологии, принципи и подходи. *Предприемачество*, 7(2), 205-244.

**Статии:**

6. Gerunov, A. (2020). Machine Learning Algorithms for Forecasting Asset Prices: An Application to the Housing Market. *Economics and Management*, 1, 27-42.
7. Герунов, А. (2020). Количествени подходи за управление на операционния риск във финансовия сектор. *Годишник на Стопанския факултет на СУ „Св. Климент Охридски“*. (приет за печат)
8. Герунов, А. (2020). Анализ и оценка на операционните рискове. *Икономически и социални алтернативи*, 2, 24-42.
9. Gerunov, A. (2019). Modelling economic choice under radical uncertainty: machine learning approaches. *International Journal of Business Intelligence and Data Mining*, 14(1-2), 238-253.
10. Gerunov, A. (2016). Automating Analytics: Forecasting Time Series in Economics and Business. *Journal of Economics and Political Economy*, 3(2), 340-349.

Представените публикации отговарят на изискванията на чл. 12 от Закона за развитие на академичния състав в Република България (ЗРАСРБ), чл. 35 от Правилника за прилагане на ЗРАСРБ (ППЗРАСРБ) и изпълняват националните наукометрични изисквания по чл. 1а, ал. 1 от ППЗРАСРБ за Област 3: Социални, стопански и правни науки, Професионално направление 3.8 Икономика.

## **VI. Заключение**

Представените в настоящата работа резултати имат потенциала да задълбочат познанията за процесите по управление на операционния риск, но и да ги трансформират по начин, отговарящ на изискванията на цифровата трансформация. Представените тук резултати, в допълнение към наличната литература и насоките за бъдеща работа, имат амбицията в научен и приложен план да спомогнат дигитализирането на съвременните организации в дейностите по управление на риска. Целта на този процес е с помощта на нови подходи, методи и технологии да се отключи ръст на производителността, подобряване на разходната структура и редица нови бизнес възможности, които да има ефект върху цялостната организационна среда.