

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису
УДК 004.852

До захисту допущено
Завідувач кафедри ММСА
О.Л.Тимошук

«___»_____2021 р.

Магістерська дисертація
на здобуття ступеня магістра
за спеціальністю 124 Системний аналіз
на тему: «Система підтримки прийняття рішень для оцінювання і
прогнозування стану страхової компанії»

Виконав:
студент II курсу, групи КА-02 мп
Панібратов Роман Сергійович

Керівник:
професор кафедри ММСА,
д.т.н, проф. Бідюк П.І.

Рецензент:
декан ФІОТ КПІ ім. Ігоря Сікорського
д.т.н., проф.,
Теленик С.Ф.

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань

Студент _____

Київ
2021

**Національний технічний університет України
«Київський політехнічний інститут імені
Ігоря Сікорського»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

Рівень вищої освіти – другий (магістерський)

Спеціальність – 124 «Системний аналіз»

ЗАТВЕРДЖУЮ
Завідувач кафедри

(підпис) (ініціали, прізвище)
«___» _____ 2021 р.

**ЗАВДАННЯ
на магістерську дисертацію студенту
Панібратову Роману Сергійовичу**

- 1. Тема дисертації:** «Система підтримки прийняття рішень для оцінювання та прогнозування стану страхової компанії», науковий керівник дисертації Бідюк Петро Іванович, доктор технічних наук, професор, затверджені наказом по університету від «02» листопада 2021 р. № 3651-с.
- 2. Термін подання студентом дисертації** 14.12.2021
- 3. Об'єкт дослідження:** страховий ринок.
- 4. Предмет дослідження:** фінансово-економічні показники страхових компаній, методи машинного навчання, статистичні критерії якості.
- 5. Перелік завдань, які потрібно зробити:**
 - 1) Здійснити огляд технічної літератури за темою роботи;
 - 2) Дослідити актуальність обраної теми;
 - 3) Ознайомитись із існуючими методами оцінювання ризиків у страхуванні;
 - 4) Здійснити порівняльний аналіз наявних методів, виявити їх переваги та недоліки;
 - 5) Дослідити сучасний стан страхового ринку України та його основні проблеми;
 - 6) Запропонувати методи оцінювання стану страхових компаній;
 - 7) Розробити та реалізувати систему, що дозволяє оцінювати та прогнозувати стан страхових компаній за фінансовими показниками;
 - 8) Провести аналіз результатів;
 - 9) Провести аналіз ринкових можливостей запуску стартап проекту;
 - 10) Розробити концептуальні висновки;

- 11) Підготувати ілюстративний матеріал;
- 12) Оформити пояснювальну записку.

6. Орієнтовний перелік ілюстративного матеріалу:

1. Аналіз стану страхового ринку України;
2. Методи, що використовувалися для оцінки стану страхових компаній;
3. Структура системи підтримки прийняття рішень, що була створена в ході дослідження;
4. Приклад результатів використання системи;
5. Розрахункові таблиці з результатами.

7. Орієнтовний перелік публікацій:

- 1) Панібратов Р.С., Бідюк П.І. Система підтримки прийняття рішень для оцінювання та прогнозування стану страхової компанії. Університетський науковий збірник «Системні науки та кібернетика» – К: NTUU «КРІ», 2021. С. 7-33 (прийнято до друку).
- 2) Панібратов Р. С. Система підтримки прийняття рішень для оцінювання та прогнозування стану страхової компанії, *VIII Міжнародна науково-технічна Internet-конференція «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами»*, 26 листопада 2021 року, Київ. С. 116-117.

8. Дата видачі завдання: 1 вересня 2020 року.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Отримання завдання на магістерську дисертацію	01.09.2021 – 05.09.2021	Виконано
2.	Огляд технічної літератури за темою	06.09.2021 – 12.09.2021	Виконано
3.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів. Дослідження актуальності обраної теми	13.09.2021 – 19.09.2021	Виконано
4.	Перший розділ. Аналіз існуючих методів оцінки ризику у страхуванні та сучасного стану страхового ринку України	20.09.2021 – 26.09.2021	Виконано
5.	Другий розділ. Математична постановка задачі оцінювання стану страхової компанії. Вибір методів оцінки стану компаній	27.09.2021 – 03.10.2021	Виконано
6.	Третій розділ. Розробка системи підтримки прийняття рішень	04.10.2021 – 10.10.2021	Виконано
7.	Розширення функціоналу системи	11.10.2021 – 17.10.2021	Виконано
8.	Аналіз результатів	18.10.2021 – 24.10.2021	Виконано
9.	Проведення аналізу ринкових можливостей стартап – проекту	25.10.2021 – 31.10.2021	Виконано
10.	Підготовка ілюстративного матеріалу	01.11.2021 – 14.11.2021	Виконано
11.	Оформлення пояснювальної записки	15.11.2021 – 26.11.2021	Виконано

Студент _____

Р.С. Панібратов

Науковий керівник дисертації _____

П.І. Бідюк

РЕФЕРАТ

Магістерська дисертація: 134 с., 39 рис., 23 табл., 1 додаток, 16 джерел.

СТРАХОВА КОМПАНІЯ, СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, БІНАРНА КЛАСИФІКАЦІЯ, МЕТОД *K*-НАЙБЛИЖЧИХ СУСІДІВ, МЕТОД ОПОРНИХ ВЕКТОРІВ, НАЇВНИЙ БАЙЄСІВСЬКИЙ КЛАСИФІКАТОР, ВИПАДКОВИЙ ЛІС, ГРАДІЄНТНИЙ БУСТИНГ, НЕЙРОННІ МЕРЕЖІ.

Об'єктом дослідження є страховий ринок.

Предметом дослідження є фінансово-економічні показники страхових компаній, методи машинного навчання.

Мета дослідження:

1. Розробити систему підтримки прийняття рішень щоб базувалась на методах машинного навчання.
2. Відтворити засоби прийняття рішень з точки зору страховика за певними моделями.

Актуальність теми: формування ринкової економіки країни включає надійний страховий захист суб'єктів господарювання. Через це постає потреба у пошуку ефективних засобів мобілізації фінансових ресурсів страхових компаній та їх використання.

Результат роботи: розроблено та реалізовано систему, що дозволяє оцінити майбутній стан страхової компанії за її поточними фінансово-економічними показниками.

Новизна роботи: спроектовано і реалізовано нову (авторську) систему підтримки прийняття рішень для розв'язання задач оцінювання і прогнозування стану страхової компанії, яка відрізняється функціональною повнотою і можливістю врахування невизначеностей, характерних для процесу оцінювання стану страхової компанії.

ABSTRACT

The topic: Decision support system for estimation and forecasting state of insurance company.

Master's Thesis 134 pp., 39 Fig., 23 tables, 1 supplement, 16 sources.

INSURANCE COMPANY, DECISION MAKING SUPPORT SYSTEM, BINARY CLASSIFICATION, K-NEAREST NEIGHBORS, SUPPORT VECTOR MACHINE, NAIVE BAYES CLASSIFIER, RANDOM FOREST, GRADIENT BOOSTING, NEURAL NETWORKS.

The object of the study is the insurance market.

The subject of the study is the financial and economic indicators of insurance companies, methods of machine learning, statistical quality criteria.

The aim of the study:

1. Develop a decision support system based on machine learning methods.
2. Reproduce decision-making tools from the point of view of the insurer according to certain models.

Relevance of the topic: the formation of a market economy of the country includes reliable insurance protection of business entities. Because of this, there is a need to find effective means of mobilizing the financial resources of insurance companies and their use.

The result of the work: a decision support system has been implemented that allows to assess the future condition of the insurance company by its financial and economic indicators.

Novelty of work: designed and implemented a new (author's) decision support system for solving problems of assessment and forecasting of the insurance company, which differs in functional completeness and the ability to take into account the uncertainties inherent in the process of assessing the insurance company.

ЗМІСТ

ВСТУП	9
РОЗДІЛ 1 АНАЛІЗ ЗАДАЧІ СТРАХУВАННЯ ВІД РИЗИКІВ ТА СТРАХОВОГО РИНКУ УКРАЇНИ	11
1.1 Актуальність страхування від ризику та його роль в ринковій економіці	11
1.2 Поняття оцінювання фінансових ризиків у страхуванні	15
1.3. Функції страхування	16
1.4. Наукове поняття класифікації страхування	17
1.5 Ризик-менеджмент у страхуванні	18
1.6 Методи оцінювання ризиків у страхуванні	19
1.6.1 Методи, які орієнтовані на аналіз наслідків	21
1.6.2 Методи, які орієнтовані на аналіз факторів ризику	25
1.6.3 Класифікація методів оцінювання ризику за рівнем визначеності	33
1.7 Поняття страхового ринку	35
1.8 Фінансовий аналіз страхової компанії	36
1.9 Сучасний стан страхового ринку України	38
1.10 Основні проблеми ринку страхування України	45
1.10.1 Низька платоспроможність та ліквідність	45
1.10.2 Недосконалість бізнес-моделей страховиків	46
1.10.3 Недостатня прозорість ринку та проблеми в корпоративному управлінні	47
1.10.4 Проблема захисту прав споживачів послуг страхування	47
1.11 Висновки до розділу 1	48
РОЗДІЛ 2 МАТЕМАТИЧНА ПОСТАНОВКА ЗАДАЧІ	49
2.1 Формалізація задачі оцінювання страхових компаній	49
2.2 Обробка даних та метрики якості класифікації	49
2.2.1 Попередня обробка даних	49
2.2.2 Вибір даних-ознак для навчання моделі	50
2.2.3 Метрики моделей класифікації	51
2.3 Математичні моделі класифікації	57
2.3.1 Метод k -найближчих сусідів	57
2.3.2 Метод опорних векторів	58
2.3.3 Наївний Байєсівський класифікатор	60
2.3.4 Випадковий ліс	63
2.3.5 XGBoost	65
2.3.6 Багатошаровий перцептрон	66
2.4 Висновки до розділу 2	76
РОЗДІЛ 3 РОЗРОБКА І ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ СТАНУ СТРАХОВИХ КОМПАНІЙ	77
3.1 Поняття системи підтримки прийняття рішень та його вимоги	77

3.2 Архітектура СППР	78
3.3 Вибір засобу реалізації СППР	80
3.4 Дані для дослідження	81
3.5 Інструкція використання СППР для користувача та аналіз результатів ..	82
3.5.1 Інструкція використання СППР	82
3.5.2 Обчислювальні дослідження та аналіз результатів	92
3.6 Висновки до розділу 3	97
РОЗДІЛ 4. СТАРТАП АНАЛІЗ ПРОЕКТУ	98
4.1 Вступ та постановка задачі стартап проекту	98
4.2 Карта стартап проекту	99
4.3 Технологічний аудит ідеї проекту	101
4.4 Аналіз ринкових можливостей запуску стартап-проекту	105
4.5 Розроблення ринкової стратегії проекту	114
4.6 Висновки до розділу 4	116
ВИСНОВКИ	117
ПЕРЕЛІК ПОСИЛАНЬ	119
ДОДАТОК А. КОД ОСНОВНОГО МОДУЛЮ ПРОГРАМНОГО ПРОДУКТУ	121

ВСТУП

Процес формування ринкової економіки будь-якої країни включає в себе надійний страховий захист суб'єктів господарювання. Саме це спричиняє необхідність пошуку ефективних засобів мобілізації фінансових ресурсів страхових компаній та їх цілеспрямованого використання. Для забезпечення надійності виконання зобов'язань, даний тип установи має ефективно формувати та використовувати фінансові ресурси.

Фінансовий аналіз діяльності страхових компаній виділяють в окрему гілку на класифікаторі дослідження грошових потоків, що впливає з економічної природи страхування, яке в ринковій економіці поступає, з одного боку, засобом захисту майнових інтересів, а з іншого — видом діяльності, що приносить дохід. Крім того, страхування — важливий чинник розвитку інвестиційного потенціалу країни. Таким чином, діяльність страховика складається із страхових, інвестиційних і фінансових операцій.

Процес формування ринкової економіки в Україні вимагає надійного страхового захисту суб'єктів господарювання.

Це спричиняє необхідність пошуків ефективних механізмів мобілізації фінансових ресурсів страхових компаній та їх цілеспрямованого використання.

Але якщо провести аналіз стану страхового ринку України, то можна виявити наступне. У першу чергу є розбіжність у вітчизняній законодавчій базі з приводу нормативів макроекономічних показників розвитку страхового ринку. Відсутність відповідного досконалого регулювання діяльності та поведінки страхових компаній є причиною накопичення системних проблем, які наявні в страховому ринку України.

Сьогоднішні теорії інформаційних технологій та бізнес-менеджменту і принципи побудови інтелектуальних систем вимагають того, щоб мета створення системи базувалася на цілеспрямованій та усвідомленій людській

діяльності. До нинішніх напрямів розробки людино-машинних систем відносять системи автоматичного керування, експертні системи та системи підтримки прийняття рішень (СППР). Остання являє собою найефективнішим засобом вирішення багатьох задач [1]. Завдяки їй, особа, що приймає рішення, може прямо за рахунок обчислювальних засобів проектувати, порівнювати та робити вибір альтернативних варіантів рішень найбільш різноманітними засобами[2]. СППР об'єднує в собі такі властивості, що робить їх не тільки доволі корисними для виконання системних задач управління та прийняття рішень, але й невід'ємними засобами аналізу даних в нинішніх умовах економічного розвитку [3].

У зв'язку з цим виникає потреба у створенні системи, що дозволила б оцінити стан будь-якої страхової компанії і спрогнозувати її стан у майбутньому, вказавши на можливі шляхи вирішення проблем.

У першому розділі сформульована задача оцінювання фінансових ризиків у страхуванні та роль страхування в ринковій економіці. Було розглянуто функції страхування та наукове поняття класифікації страхування. Проведено аналіз методів, які використовуються для оцінки ризиків у страхуванні та виявлено їх переваги та недоліки. Розглянуто поняття страхового ринку та було оцінено його сучасний стан в Україні та його актуальні проблеми.

У другому розділі було проведено дослідження основних економічних параметрів, що дозволяють оцінити стан страхової компанії. Також було створено математичні моделі для системи підтримки прийняття рішень для прогнозування стану страхової компанії.

У третьому розділі описано архітектуру та схему функціонування системи підтримки прийняття рішень. Також було проведено аналіз застосування математичних моделей та результати обчислюваних експериментів

Четвертий розділ присвячено формулювання стартап-проекту на основі побудованої системи.

РОЗДІЛ 1 АНАЛІЗ ЗАДАЧІ СТРАХУВАННЯ ВІД РИЗИКІВ ТА СТРАХОВОГО РИНКУ УКРАЇНИ

1.1 Актуальність страхування від ризику та його роль в ринковій економіці

Будь-яка діяльність суб'єкта ринкових відносин нерозривно пов'язана з ризиком, який потрібно спрогнозувати та оцінити, у чому і полягає задача страхування.

Ризик — небезпека непередбачуваних втрат, що виникають внаслідок природних явищ (стихійні лиха), непередбачуваними подіями техногенного характеру (вибухи, аварії, пожежі тощо), кримінальними явищами (крадіжки, шахрайство, тощо) та явища економічного характеру, різноманіття яких швидко зростає внаслідок поглиблення суспільного поділу праці та розвитку кредитної системи.

Страхування — система економічних відносин, що дає можливість відшкодовувати витрати і захистити майнові інтереси підприємців та фізичних осіб від стихійних лих та випадкових подій. Воно є перспективним сектором ринкової інфраструктури та є важливою складовою суспільства в умовах розвинутої економіки.

Ризиковий характер будь-якої сфери діяльності зумовлює створення спеціальної системи відносин, приклад якої можна побачити на рисунку 1.1.



Рисунок 1.1 — Система відносин при ризиковому характері суспільно виробничих відносин

Для даних відносин характерна специфічність і вони носять об'єктивний характер і утворюють категорію “страховий захист”

Страховий захист — економічна категорія, що відображає сукупність розподільчих та перерозподільчих відносин, які пов'язані з подоланням та відшкодуванням втрат, котрі можуть бути спричинені різними особливими обставинами.

Мета страхового захисту — накопичення і витрачання грошових та інших ресурсів для реалізації дій для попередження та мінімізації негативного впливу ризиків та відшкодування втрат, що були спричинені несприятливими ситуаціями.

Для страхового захисту характерні наступні властивості:

- випадковий характер появи надзвичайних подій;
- нерівномірність нанесення збитку різним суб'єктам;
- необхідність попередження наслідків вказаних подій;
- необхідність відшкодування завданих збитків у натуральній або грошовій формі.

Для забезпечення страхового захисту створюються страхові фонди, задача яких полягає у відшкодуванні збитків, що виникли внаслідок стихійними лихам, техногенними аваріями та непередбачуваними подіями.

Розрізняють три основні форми організації страхового фонду, де суб'єктами власності на ресурси є:

- держава;
- окремий суб'єкт;
- страхова організація або страховик.

Страховик — фінансова установа, яку створено у формі акціонерного, повного, командитного товариства або товариства з обмеженою відповідальністю відповідно до Закону України “Про господарські товариства” з урахуванням особливостей, передбачених Законом України “Про страхування”, а також яка одержала у встановленому порядку ліцензію на здійснення страхової діяльності.

Страхова діяльність в Україні здійснюється виключно страховиками, які є резидентами України [4].

Як правило, страхова діяльність орієнтована на захист майнових інтересів фізичних та юридичних осіб у випадку виникнення певного виду ситуацій за рахунок грошових фондів, що утворюються за рахунок страхових внесків, що сплачуються страховиками. Крах страховика у фінансовому сенсі стосується і великої групи осіб. Так, представник наукової думки про страхування К. Г. Воблий під страхуванням визначає вид господарської діяльності за рахунок солідарності та відплатності, що ставить перед собою в якості задачі покриття майбутніх потреб або нестатків, що є результатом несподіваної події, яку можна зафіксувати статистично [5]. К. Турбіна дає означення поняттю “страхова діяльність” як вид підприємництва, яка призначена для прийняття ризиків великої кількості суб'єктів страхових відносин і забезпеченні страхових виплат [6]. З точки зору Р. Юлдашева, страхова діяльність — діяльність страхових організацій та товариств страхування, що пов'язана з формуванням спеціальних грошових фондів або страхових резервів, необхідних для подальших виплат [7]. Звідси можна зробити висновок, що погляди стосовно страхування та страхової діяльності є доволі неоднозначними, оскільки вони розглядаються розглядаються з точки зору формування грошових фондів для наступного забезпечення грошових виплат.

Протягом періоду 2002-2008 року спостерігався стабільний розвиток українського страхового ринку. Позитивні тенденції реального сектору економіки та споживчого сектору гарантували постійний притік фінансових ресурсів до страхових компаній. Але внаслідок глобальної фінансово-економічної кризи 2008 року відбулися доволі різкі зміни. Відтік грошових коштів та зменшення платоспроможності призвели до зменшення об'єму фінансових ресурсів страхових компаній та надали актуального статусу питання їх ефективного формування за кризових умов. Внаслідок останнього спостерігається негативна динаміка розвитку страхових компаній та відсутності стабільності страхових компаній і страховиків.

Розмір і структура страхових фондів залежать від різноманітних факторів: стан економічної ситуації, співвідношення форм власності, засобів залучення коштів для реалізації соціальних програм, міжнародних відносин тощо.

Страхові резерви являють собою грошові ресурси і утворюються завдяки внескам страхувальників. Ще одним засобом формування страхового фонду є доходи від інвестиційної діяльності страховика, в оперативному управлінні якого знаходяться фінансові ресурси.

За основними задачами фінансового менеджменту можна зробити висновок, що фінансові ресурси є головним об'єктом фінансового управління. Правильне налаштування тактик та стратегій управління є головним фактором фінансової стійкості страхових компаній [8]. Як можна побачити, засоби забезпечення страхового захисту суб'єктів є доволі обмеженими. Тому навантаження повинно йти на страхові фонди, що є доволі перспективним щодо страхового захисту.

1.2 Поняття оцінювання фінансових ризиків у страхуванні

Страхова діяльність нерозривно пов'язана зі страховим ризиком. Але головним завданням страхової діяльності є розподіл ризику. Ризик є об'єктивною передумовою утворення страхових відносин. Якщо ризик відсутній, то відсутня і необхідність страхування. Але варто звернути увагу на те, що далеко не кожний ризик може бути передумовою створення страхових відносин. Ризик може бути застрахованим лише у тому випадку, якщо можна оцінити ймовірність настання даної події, оцінити розмір можливих збитків та втрат та забезпечити еквівалентну страхову премію. Термін «страховий ризик» має на увазі прогнозований збиток об'єкта страхування внаслідок виникнення страхової події.

Оцінювання фінансових ризиків у страхуванні є головним завданням фінансового аналізу, оскільки для управління ризиком у першу чергу потрібно провести його аналіз, і тоді його оцінити.

Під поняттям «оцінювання» мають на увазі систематичний процес виявлення факторів та типів ризиків, їх кількісне оцінювання, тобто сукупність методів аналізу ризиків включає в себе взаємодоповнюючі кількісні та якісні підходи.

В основному теоретичні дослідження проблем невизначеності присвячені аналізу причин появи, класифікації та методам кількісного оцінювання ризиків, але менше уваги звернено щодо якісних методів з врахуванням впливу неекономічних чинників на фінансово-господарську діяльність підприємств, які при практичному застосуванні могли б бути більш адекватними до реальних можливостей підприємців, а результати були б основою для їх використання у наступних етапах аналізу.

Оскільки пошук ймовірності розорення страхової компанії є окремим головним завданням страхової математики, на основі якої створюються актуарні концепції відносно фінансової оцінки стабільності з врахуванням не

тільки відсутності банкрутства, а його запобігання, то виникає необхідність використання економіко-математичних методів моделювання та аналізу фінансових ризиків, а також побудові актуарних моделей для пошуку раціональної величини страхової премії за знайденою ймовірністю розорення.

1.3. Функції страхування

Згідно Закону України «Про страхування», страхування — вид цивільно правових відносин щодо захисту майнових інтересів громадян та юридичних осіб у разі настання страхових випадків, що визначені договором страхування або чинним законодавством, за рахунок грошових коштів, що формуються за рахунок сплати фізичними та юридичними особами страхових платежів, та доходів від розміщення коштів цих фондів.

Це економічні відносини, в яких учасниками, як правило, виступають двоє суб'єктів відносин: страхова організація, яка є страховиком, і страхувальник — юридична особа або дієздатний громадянин, який уклав зі страховиком договір страхування або і є страхувальником за законодавством України.

Суть страхування розкривається в його функціях. До них відносять ризикову, превентивну та заощаджувальну функцію.

- Ризикова функція: передача страховикові за певну плату матеріальної відповідальності за наслідок ризику, зумовлену подіями, які перебачені договором або діючим законодавством.

- Превентивна функція: страхова компанія об'єктивно зацікавлена в більш тривалому використанні коштів страхового фонду і намагається мінімізувати ймовірність виплат за рахунок зменшення ймовірності страхового випадку. Для цього страховик має можливість застосовувати фінансову та правову превенцію.

– Заощаджувальна функція: договори довгострокового страхування життя мають накопичувальний характер. Сприяючи розвитку заощаджувальних видів страхування, держава дає можливість впливати на вирішення соціальних проблем, прискорювати грошовий обіг і збільшувати інвестиційний потенціал країни.

1.4. Наукове поняття класифікації страхування

Для збору та групування даних про об'єкти та явища для розробки оптимальної та ефективної стратегії розвитку потрібна єдина взаємопов'язана система — класифікація.

У теорії страхування розрізняють декілька класифікаційних ознак: історичні, де є можливість відстежити протягом тривалого історичного періоду динаміку страхування та його складових; економічні, на які впливають економічні умови здійснення страхування в певній країні; юридичні, які визначають умови та форми проведення страхування, спираючись на законодавчу базу та міжнародні угоди, які підписані державою.

За економічною ознакою страховика розрізняють спеціалізацію страховика, що веде до розгляду двох сфер діяльності.

1. «Life insurance» — страхування життя.
2. «Non-life insurance» — загальні види страхування, що включають в себе ті випадки, що не містять ознаки договорів страхування життя.

За статусом страхувальника або за їх інтересами, страхові послуги класифікуються за інтересами:

- громадян, які пов'язані з соціальними та майновими потребами однієї або декількох осіб;
- юридичних осіб.

За статусом страховика виокремлюють:

- Державне страхування або державні страхові організації;
- Комерційне страхування або страхові компанії різних форм власності за Законом України «Про страхування».

1.5 Ризик-менеджмент у страхуванні

Під менеджментом у страхуванні розуміють науково-практичний напрямок, що орієнтований на забезпечення ефективного функціонування страхових організацій у ринкових умовах господарювання. Нинішній менеджмент включає в себе керування інтелектуальними, фінансовими та матеріальними ресурсами для забезпечення ефективної діяльності страховика.

Але варто звернути увагу на те, що менеджмент, як складова ефективного управління будь-якою організацією, пов'язана не лише з управлінською діяльністю. Під цим поняттям розуміють і множину заходів для вивчення, аналізу, планування, прогнозування і формування страхового ринку. Головна мета в умовах невизначеності — не уникати ризику, а спрогнозувати його і мінімізувати його негативні наслідки.

Задача менеджменту — створити ефективну страхову організацію і при можливості коригувати її таким чином, щоб вона відповідала умовам і задачам ринку, що змінюються, з врахуванням непередбачуваних результатів.

Управління ризиком завжди постає як перед страхувальником так і перед страховиками. Першим кроком є усвідомлення ситуації ризику, вивчення та аналіз інформації, яка дозволяє керівниками прийняти рішення та розробити методи зменшення або максимальне обмеження негативних результатів ризику та його збитків.

Загальний підхід, на якому базується страхування, — створення резервів за стабільних умов, що дають можливість здійснювати діяльність з ризиком.

Безпосередні дії з метою мінімізації ризику визначають сутність поняття ризик-менеджменту, що містить в собі: виявлення наслідків діяльності економічних суб'єктів у ситуації з ризиком; здатність реагувати на можливі негативні наслідки; розробку й реалізацію заходів з метою компенсації можливих негативних результатів або їх повного усунення.

1.6 Методи оцінювання ризиків у страхуванні

Міра ризику — ймовірність настання випадку втрат і розмір можливих збитків. Крім цього варто звертати увагу на випадковий характер можливих втрат. Ймовірність виникнення даної події можна розглядати як об'єктивно, так і суб'єктивно. У об'єктивному випадку для ймовірності застосовують обчислення частоти, з якою відбувається подія. Суб'єктивний варіант використовує суб'єктивні критерії, які базуються на різних припущеннях. Тому суть задачі оцінювання фінансових ризиків полягає у пошуку залежності між визначеними розмірами втрат і ймовірністю їх виникнення. Дана залежність відображається в кривій ймовірностей виникнення визначеного рівня втрат.

На даний момент налічують ефективні математичні методи оцінювання фінансових ризиків у страхуванні. Так, стандарти Basel II та Solvency I заохочують розвиток моделей управління та оцінювання ризиків фінансової сфери.

У страховій практиці для оцінювання ризику застосовують різні моделі, які постійно вдосконалюються та покращуються. Розрізняють два засоби аналізу ризиків — за позицією факторів ризику та за позицією наслідків ризикової події. За таким підходом, математичні методи оцінювання ризиків поділяють на дві групи: методи, що базуються на аналізі наслідків («високорівневі моделі»), та методи, що базуються на аналізі факторів ризику («низькорівневі моделі»). В якості властивості для класифікації моделей

застосовують набір задач, для вирішення яких ця модель може застосовуватися. Так, моделі, що можуть відображати функцію розподілу втрат у явному вигляді, можуть оцінювати середні втрати страхової компанії та максимальні втрати за встановленим рівнем значущості (таблиця 1.1).

Таблиця 1.1 — Класифікація методів аналізу та управління фінансовими ризиками

Клас моделей	Застосування для розподілу ризиків за рівнем критичності	Застосування для оцінки середніх втрат	Застосування для оцінки максимальних втрат
<i>Методи, які орієнтовані на аналіз наслідків (високорівневі моделі)</i>			
BIA	немає	немає	допустимо
LDA	допустимо	допустимо	допустимо
IMA	немає	немає	допустимо
<i>Методи, які орієнтовані на фактори ризику (низькорівневі моделі)</i>			
Метод функціональних кореляцій	допустимо	допустимо	допустимо
Регресійний аналіз	допустимо	допустимо	допустимо
Байєсівські мережі	допустимо	допустимо	допустимо
Методи нечіткої логіки	допустимо	немає	немає

1.6.1 Методи, які орієнтовані на аналіз наслідків

Дана група методів базується на засобах математичної статистики. Більшість моделей включає в себе збір та аналіз даних втрат, що є наслідком фінансового ризику протягом попередніх періодів з подальшим прогнозом втрат на майбутній період. До даного класу відносять наступні найбільш відомі моделі:

A. BIA (Basic Indicator Approach — метод базових показників).

Модель BIA описує вимоги відносно забезпечення достатності капіталу для покриття ризиків. Дана модель була запропонована для розв'язання задачі оцінювання максимально можливих втрат від операційного ризику за встановленим рівнем значущості (99 %). Метод базується на гіпотезі про те, що операційні втрати є залишковими від загальних втрат, які пов'язані з ризиками. Квантиль розподілу випадкової величини X , що характеризує величину втрат, задається такою формулою:

$$\widehat{Q}_{99\%} = \overline{G}_I \alpha,$$

де $\widehat{Q}_{99\%}$ — оцінка 99%-го квантиля випадкової величини, що характеризує втрати;

$\overline{G}_I$ — середній валовий дохід за три попередні роки, протягом яких валовий річний дохід був додатнім;

$\alpha = 15\%$ — коефіцієнт, визначений за дослідженням галузевої практики (стосується фінансових організацій та банків) Базельським комітетом.

Значення останньої величини як правило уточнюють з використанням результатів оцінювання втрат, що стосуються фінансових ризиків українських страхових компаній. Недолік моделі BIA полягає в тому, що обчислена оцінка об'єму можливих фінансових втрат залежить тільки від розміру бізнесу і ніяк не залежить від специфіки страхового портфелю.

B. LDA (Loss Distribution Approach — метод розподілу збитків).

Згідно за моделями класу LDA висувається гіпотеза, що випадкова величина, що вказує на розмір втрат x , які відбулися протягом часу t , задається таким чином:

$$x = \sum_{i=1}^{n(t)} L_i ,$$

де $n(t)$ — випадкова величина, що характеризує кількість випадків втрат певного типу за конкретний проміжок часу t ;

L_i — множина випадкових величин, що характеризують величини можливих окремих втрат.

Крім того, висувається додаткова гіпотеза щодо величин множин L_i : вони незалежні та мають однаковий закон розподілу для конкретного типу збитків. Для побудови даних моделей аналізуються втрати за обраний проміжок часу для пари вигляду «лінія бізнесу»/«тип втрат». Для кожної пари даного формату за даними про частоту втрат і величину збитків, що були зафіксовані у минулому, обчислюється вибіркове середнє значення частоти появи ризикових подій $E(n(t))$ і вибіркове середнє значення величини втрату у разі нстання ризикової події $E(L)$. Після визначення функцій розподілу випадкових величин $n(t)$ і L , будується функція розподілу випадкової величини x (використовується метод Монте-Карло). Функція розподілу випадкової величини x , що характеризує загальний обсяг втрат за певним видом ризиків, дозволяє знайти точкову оцінку математичного сподівання втрат і обчислити квантиль встановленого рівня, тобто значення OpVaR (Operational Value at Risk).

Варто також звернути увагу на те, що варіанти методу LDA класифікують за двома групами. Перша множина моделей ґрунтується на прямому аналізі ризикових подій, і пов'язаних з ними втрат не враховуючи факторів ризику та причинно-наслідкових зв'язків у межах складнішої моделі. До цієї групи

відносять як пряму оцінку характеристик розподілу частоти та дії ризикових подій, так і методики, що орієнтуються на теорію екстремальних значень. Друга множина моделей базується на розширеній множині змінних, до якої відносять і фактори ризику.

C. IMA (Internal Measurements Approach — метод внутрішніх вимірів).

Метод IMA дозволяє обчислити оцінку максимально можливих втрат для певного типу ризику без побудови розподілу випадкової величини X , що характеризує об'єм втрат. У даній моделі висувається наступна гіпотеза: якщо розділити усі збитки на очікувані (їх сума близька до математичного сподівання суми збитків за період) та непередбачувані (перевищують середнє і належать до “хвоста” статистичного розподілу), то існує функціональна залежність між величинами очікуваних та непередбачуваних збитків. Найбільш типовим випадком є лінійна залежність, тобто:

$$\widehat{Q}_{99\%} = E_I P_E L_{GE} = \gamma E_L \alpha,$$

де $\widehat{Q}_{99\%}$ — оцінка 99%-го квантиля розподілу можливих втрат (величини капіталу, необхідного для покриття ризику) за даним видом події;

P_E — ймовірність настання негативної події даного типу за цей період;

L_{GE} — середня величина збитку певного типу за умови того, що негативна подія відбулася;

E_I — коефіцієнт масштабу;

γ — коефіцієнт для оцінювання вимог до капіталу через оцінку очікуваних збитків E_L .

Загальна оцінка максимально можливих втрат обчислюється як сума всіх оцінок, які розраховані для різних типів ризикових подій та ліній бізнесу.

D. Статистичний метод оцінювання ризику.

Суть даного методу полягає у дослідженні статистики негативних наслідків внаслідок реалізації рішень, які спостерігалися в схожих видах підприємницької діяльності. Можливе використання різних засобів оцінювання,

включаючи дисперсійний аналіз. Головним показником, який обчислюється за допомогою статистичного методу, є частота втрат, що має відношення до конкретного виду діяльності:

$$f = \frac{n'}{n_{\text{заг}}},$$

де n' — кількість зафіксованих випадків втрат за статистичною вибіркою;

$n_{\text{заг}}$ — загальна кількість випадків за статистичною вибіркою.

У момент прийняття рішень, орієнтуючись на даний метод, показник частоти витрат переносять на прогнозовані дані і він розглядається в якості ймовірності настання втрат певного рівня.

Зараз статистичний метод використовується в різних модифікаціях і найбільшого поширення отримав метод статистичного моделювання — метод Монте-Карло. До плюсів даного методу відносять можливість аналізу та оцінки різних сценаріїв розвитку проекту, що досліджується, з можливим включенням різних чинників впливу в межах одного підходу. Мінусом даного підходу є серйозний рівень використання ймовірнісних характеристик, що не завжди є задовільним для менеджерів проекту.

Середні величини дозволяють виразити звичайні розміри ознак, що змінюються, якісно однорідних явищ та простежити їх динаміку відносно середнього рівня розвитку. Одним із статистичних методів дослідження є метод середніх величин. При оцінюванні ризику відбувається розбиття окремих ризикових груп на менші підгрупи для створення аналітичної бази для визначення ризику за певними ризиковими параметрами.

Ще одним методом статистичного аналізу є метод процентів. За термінологією оцінювання ризиків він являє собою певну скупність додатних та від'ємних відхилень від середнього ризикового типу наявної аналітичної бази. Ці відхилення вимірюються у відсотках або проміле від середнього ризикового типу.

На практиці також використовують економетричні, статистичні методи оцінювання та аналізу ризиків, тощо. Доречно звернути увагу на те, що зараз вітчизняні та зарубіжні вчені мають доволі сильний інструментарій оцінювання та відстеження ризику. Також вони мають можливість використовувати актуарні розрахунки, комплексне моделювання страхових процесів, методи експертних оцінок, асоціацій та аналогій тощо.

1.6.2 Методи, які орієнтовані на аналіз факторів ризику

Методи, що відносяться до даної групи, включають в себе властивість поглибленого дослідження процесів організації і дозволяють застосовувати інформацію про причинно-наслідкові зв'язки. Прикладами цієї множини моделей є Sb-AMA, метод функціональних кореляцій, регресійні моделі, байєсівські мережі, методи нечіткої логіки та інші.

A. Sb-AMA (Scenario-bases Advanced Measurement Approach — сценарний аналіз).

Цей метод базується на визначенні факторів ризику (можливих передумов ризиків), за якими створюються можливі сценарії, використовуючи формат правила “що буде, якщо”. Тобто у протизвагу до розглянутих попередніх методів, що мають можливість аналізувати збитки, які відбулися, дана модель оцінює втрати, які можливі у майбутньому, коли відбуваються певна подія. Для кожного сценарію за історичними даними або експертними оцінками знаходять частоту та розмір збитків. Після перевірки даних та корекції неточних оцінок сценарії розбивають на множини за факторами ризику і для кожної групи сценаріїв оцінюють параметри статистичних розподілів частоти та обсягу збитків. Кінцевим кроком є застосування імітаційного моделювання з використанням методу Монте-Карло, що дозволяє оцінити повний розподіл збитків. Оцінка надається з використанням гіпотези, що є інформація щодо

сімейства, до якого належить розподіл випадкових величин кількості збитків (наприклад, розподіл Пуассона) і величини самих збитків. Для останнього зазвичай використовують закони розподілу з “важкими хвостами”, до яких належить логнормальний розподіл та розподіл Парето.

Постійне відстежування фінансового стану за цим способом дає можливість моніторингу негативних тенденцій і дозволяє мінімізувати ймовірність ризику банкрутства будь-якої компанії.

В. Метод функціональних кореляцій.

Даний метод базується на побудові структурної моделі організації. Перехід до математичної моделі здійснюється за рахунок побудови орієнтованого графа, де вершини виступають в ролі співробітників та підрозділів, а ребра — в ролі потоку інформації. Кожній вершині відповідає випадковий процес, що відображає її функціональність. Особливість даної моделі полягає у встановленні залежностей між функціональністю пов’язаних вершин. Аналіз ризиків проводиться не з позиції наслідків або окремих факторів ризику, а відносно співробітників, систем бізнес-процесів, яким відповідають вершини графа.

Подібно методу LDA задаються розподілом величини втрат (для вершини i реалізується розподіл за допомогою випадкової величини L_i), яка має зв’язок з окремими випадками реалізації ризиків. Моменти реалізації ризиків описуються випадковим процесом $n_i(t)$ для вершини i . Для даного процесу характерні наступні стани:

- 1) $n_i(t) = 0$ — стан нормального функціонування.
- 2) $n_i(t) = 1$ — стан збою системи.

Загальні втрати $X_i(t + \Delta t)$, понесені на момент $(t + \Delta t)$ внаслідок збоїв на i -ій вершині графа можна записати наступним чином:

$$X_i(t + \Delta t) = X_i(t) + n_i(t + \Delta t)L_{i,t+\Delta t},$$

де $L_{i,t+\Delta t}$ — реалізація випадкової величини L_i .

З метою встановлення зв'язку між різними функціональними елементами задаються поняттям “підтримки”, що отримує функціональний елемент від зв'язаних з ними елементів. Коли розмір “підтримки” є меншим, ніж вказана величина, то даний функціональний елемент перебуватиме у стані відмови, тобто для вказаної його i -ої вершини графа встановлюється $n_i(t) = 0$. Для відображення даної структури в модель вводять випадковий процес $h_i(t)$ наступного вигляду:

$$h_i(t) = \vartheta_i - \sum_j w_{ij} n_j(t) + \varepsilon_i(t),$$

де ϑ_i — середній об'єм “підтримки”, яку i -й процес отримує за умови повної функціональності системи (всі $n_i(t) = 0$);

w_{ij} — об'єм “підтримки”, яку i -й процес отримує від j -го процесу і відповідно втрачає, якщо j -ий процес перейшов у стан відмови;

$\varepsilon_i(t)$ — гаусівський білий шум.

Даний метод використовується для аналізу функціональності процесів та фіксації збоїв. Фінансові ризики мають інше походження і мають зв'язок з моментом відсутності коштів на відшкодування страхового випадку, у разі його виникнення згідно укладеного страхового договору. Тому цей метод є недоцільним у використанні для оцінювання фінансових ризиків.

С. Регресійні моделі.

Цей клас моделей орієнтований на пошук причинно-наслідкових зв'язків між індикаторами, що спостерігаються, та рівнем ризику. Виділяють наступні два види показників, які можливо застосовувати в якості пояснювальних змінних:

1. Змінні оточення — кількісні показники, які описують бізнес-процеси певного підприємства.
2. Фактори ризику — кількісні показники, що відображають характеристику спостережуваних випадків реалізації ризиків.

Дану математичну модель можна записати таким чином:

$$x = Af + b + \varepsilon ,$$

де x — величина втрат, яка пов'язана з фінансовим ризиком;

f — вектор змінних, що спостерігається;

ε — випадкова величина, що задає рівень похибки;

A та b — параметри, які оцінюються, та описують залежність між змінною x та факторними змінними f .

За умови використання великого об'єму даних, цей метод гарантує високу точність оцінок.

D. Методи нечіткої логіки.

Моделі нечіткої логіки дозволяють якнайкраще застосовувати експертне оцінювання для аналізу ризиків за умови часткової або повної відсутності точних даних. Метод найближує модель до людських міркувань під час прийняття та обґрунтування рішень. Метод нечіткої логіки можна використовувати для оцінки об'єму втрат за виявлення ризику банкрутства страхової компанії. На зарубіжних ринках є автоматизовані системи для оцінювання фінансових ризиків, що використовують інструментарій нечіткої логіки, як RiskShield.

E. Метод аналогів.

Для аналізу ризику, яким може бути обтяжений проект, може бути корисною інформація про вплив чинників ризику схожих за сутністю проектів, які були виконані раніше. З цією метою створюється інформаційна база на базуючись на її дослідженнях роблять узагальнення та приймають рішення з приводу виконання проекту. Мінусом цього методу є описовий характер. Крім того, можлива ситуація зміни впливу чинників ризику на схожі проекти з плином часу. Також може змінитися і сама множина чинників ризику.

Г. Метод експертного оцінювання.

Завичай при дослідженні складних систем, до яких наліжать і фінансові, зустрічаються проблеми, що виходять за межі формальних математичних задач. Суть методу експертного оцінювання полягає у застосуванні людського інтелекту та їх здібності здатності вирішувати задач, які є слабо формалізованими.

Процес експертного оцінювання складається з наступних кроків:

1. Формування мети оцінювання.
2. Постановка задачі.
3. Формування групи управління процесом оцінювання.
4. Опис форми отримання необхідних результатів.
5. Вибір експертів та оцінка їх компетентності.

Вибір експертів проводиться таким чином, щоб спеціалісти, які входять в склад створеної групи, з одного боку володіли релевантною інформацією щодо специфіки роботи підприємства, а з іншого, не були зацікавлені щодо результатів оцінювання. Тому зазвичай до складу команди включають 2–3-х спеціалістів підприємства та 2–3-х зовнішніх експертів. Компетентність експертів об'єктивно оцінюється ступенем їхньої кваліфікації у певній сфері знань за допомогою аналізу професійної, наукової та іншої діяльності. Суб'єктивний метод оцінювання компетентності включає в себе взаємне оцінювання кожним експертом кваліфікації колег та своєї за певною шкалою. Після обробки результатів опитування визначається компетентність експертної групи, що визначає певну похибку оцінювання.

6. Створення анкет опитування.
7. Вибір методу отримання інформації.
8. Пряме опитування експертів.
9. Обробка результатів та створення звіту з метою прийняття рішення.

До засобів обробки результатів відносять наступні методи:

1. Метод надання переваг.

У даному методі експерти нумерують об'єкти або показники ризику в порядку їх значущості, причому найменш значущий елемент отримує № 1.

При обробці результатів обчислюють коефіцієнт відносної важливості j -го елемента за наступною формулою:

$$K_j = \frac{\sum_{i=1}^n K_{ij}}{\sum_{j=1}^s \sum_{i=1}^n ij},$$

де K_{ij} – місце, на яку поставлена значущість j -го елемента у i -го експерта;

n – кількість експертів;

s – кількість розглянутих елементів.

Показник, який має найбільший коефіцієнт і є тим самим показником ризику, що з точки зору експертів, найкраще характеризує об'єкт.

2. Метод рангів.

Експерти оцінюють важливість кожного елемента за шкалою відносної значущості в апіорі встановленому діапазоні: 0-1, 0-10, 0-100. Даний метод у порівнянні з попереднім дозволяє оцінити не лише важливість j -го елемента, а й степінь його вагомості.

Експертні методи оцінювання рівня фінансового ризику використовують за умови відсутності необхідних інформативних даних підприємства для обчислень економіко-статистичними методами. Вони орієнтовані на опитування кваліфікованих спеціалістів з подальшою математичною обробкою результатів опитування.

На даний момент на практиці стали часто використовувати групові методи експертизи: консиліуми, наради та бізнес-тренінги. Результати досліджень вказують на те, що методи колективного генерування ідей дають ідей в більшій кількості, ніж за умови індивідуальної експертизи кожного спеціаліста. Не дивлячись на формат проведення експертних процедур, всі

експертні методи ґрунтуються на бальній оцінці окремих чинників ризику і визначенні їх частки. Головним мінусом експертного підходу є суб'єктивність окремих оцінок.

Г. Аналітичний метод.

У даному підході застосовується класичне правило економіки або ефект левериджу, що полягає у наступному: чим більший ризик, тим більший дохід. Через це застосування менеджером будь-якого аналітичного методу зводиться до оцінки приросту доходу проекту і приросту ризику проекту, тобто граничної корисності.

В рамках аналітичного підходу використовуються різноманітні методи, до яких найчастіше відносять:

- метод аналізу абсолютних і відносних показників;
- метод аналізу чутливості.

Для наглядної демонстрації розглянемо метод аналізу чутливості, що складається з наступних кроків:

1. Вибір головного ключового показника, по відношенню до якого виконується оцінка чутливості.
2. Вибір можливих факторів ризику, що впливають на ключовий показник шляхом відхилення від очікуваного значення.
3. Обчислення конкретних значень ключового показника на різних етапах реалізації проекту і за умов різних значень обраних факторів ризику.

Створені потоки надходжень і витрат дають можливість у будь-який момент часу обчислювати поканики ефективності. Також будують діаграми залежності між результуючим показником та змінними вхідними параметрами. На основі порівнянь між побудованими діаграмами можна визначити ті фактори, які мають найбільший ступінь вплив на кінцеве значення ключового параметра.

Завдяки розрахункам критичних лоя проекту значення факторів ризику менеджер знаходить слабкі місця і створює програму дій для посилення цих слабких місць. Наприклад, якщо в якості критичного фактору виступає ціна на

продукцію, то доцільним буде посилення програми маркетингу або зниження вартості продукту.

Головний мінус даного методу в тому, що він не може охопити абсолютно всі можливі обставини реалізації. Також він не уточнює ймовірності здійснення альтернативних проектів.

Н. Байєсівські мережі.

Цей клас методів дозволяє продемонструвати причинно-наслідкові зв'язки між різними впливовими факторами ризику та змінами середовища. У порівнянні з регресійними моделями, байєсівські мережі дозволяють враховувати не тільки прямі залежності рівня ризиків від чинників ризику, а й залежності між чинниками. Також ці моделі дають більше варіантів для формування висновку за непоним обсягом даних. У рамках математичної термінології байєсівська мережа виступає в ролі орієнтованого графа, де вершини являють собою фактори ризику та зміни середовища, а ребра — зафіксовані взаємозв'язки або такі, що прогнозуються. Крім того мережа описується множиною умовних розподілів випадкових величин, які характеризують фактори ризику та змінні середовища.

Нехай існує n випадкових величин $X_1, \dots, X_n$. Спільну ймовірність розподілу значень цих величин можна записати у вигляді добутку n умовних ймовірностей:

$$P(x_1, \dots, x_n) = P(x_1) \prod_{j=2}^n P(x_j | x_1, \dots, x_{j-1}).$$

До переваг байєсівських мереж відносять можливість одночасного застосування експертного оцінювання (наприклад, оцінювання структури мережі за допомогою визначення залежностей між змінними, вибір типів початкових розподілів змінних), і математичних методів для обчислення висновку мережі.

Висновок, сформований мережею, може здійснюватися через поширення інформації у будь-якому напрямку. Метод використовується для формування

ймовірнісного висновку — обчислення умовної ймовірності отримання значень для частини випадкових величин, коли відомі значення інших величин. З математичної точки зору виконується розрахунок ймовірності виду $P(y, x)$, де X — множина значень, що спотергіаються, а Y — множина значень, що оцінюються. Тобто задача полягає у пошуку умовних ймовірностей за наступною формулою:

$$P(y|x) = \frac{\sum_S P(y, x, s)}{\sum_{y, S} P(y, x, s)},$$

де S — множина всіх змінних, крім X та Y .

Такі обчислення є доволі трудомісткими, а задача формування висновку є NP -повною. Для забезпечення високої точності та спрощення механічних обчислень створено було проведено випробовування ряду алгоритмів побудови висновку у байєсівській мережі.

1.6.3 Класифікація методів оцінювання ризику за рівнем визначеності

За рівнем повноти інформації, яку має суб'єкт підприємницької діяльності, методи оцінювання ризиків поділяють на певні групи. за наступних умов (рисунок. 1.2):

- визначеності, коли інформація представлена у вигляді повного звіту про прибутки та збитки, тощо;
- часткової невизначеності, коли інформацію про ситуацію ризику являє собою частоту виникнення ризикової події;
- повної невизначеності, коли повністю відсутня інформація щодо ситуації ризику, але можна скористатися допомогою фахівців та експертів з метою часткового подолання невизначеності.

Зменшення обсягу вихідної інформації про сферу діяльності веде до ускладнення розв'язку задачі оцінки ступеню ризику у зв'язку зі скороченням можливостей та зменшення достовірності отриманих результатів. Розглянемо смисловий зміст кожної групи методів оцінювання ризику.

За рисунком 1.2 в умовах визначеності використовують розрахунково-аналітичні методи, які використовуються для обчислення показників за даними звітності. Показники оцінки ризику можуть виражатися в абсолютних, відносних та середніх величинах.

Абсолютні величини характеризують результати ризикових подій:

- у вигляді грошового чи фізичного вираження, якщо відповідні втрати піддаються одиницям виміру;
- у складі балансових відношень, що показують результати фінансово-господарської діяльності: ліквідність балансу організації, здатність фінансових джерел до формування матеріально-обробних засобів, тощо.

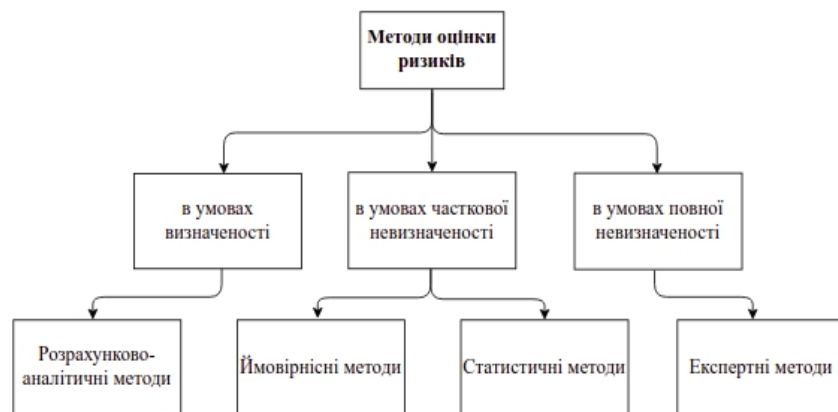


Рисунок 1.2 — Класифікація методів оцінки ризиків за ступенем визначеності

Доречно звернути увагу на те, що показники розглядаються як абсолютні величини.

У відносному представленні ризик розглядають як величину можливих втрат, що відносять до певної бази, яку можна розглядати як затрати ресурсів на певний виді підприємницької діяльності, або очікуваний прибуток від діяльності. Відносні показники є похідними від абсолютних.

Показники середніх величин розглядаються як узагальнені показники, в яких грають роль діючі причини, чинники ризику, закономірності підприємницької діяльності. Водночас проводиться згладжування різниць результатів діяльності окремих підприємств та відображає загальне або властиве лише усій групі підприємців у певній сфері діяльності за однієї ознакою.

Ризик з точки зору часткової невизначеності розглядається як ймовірнісна категорія. Через це доречно застосовувати ймовірнісні та статистичні методи оцінювання ризику, за допомоги яких обчислюються величини оцінки ризику. Ймовірнісні величини являють собою міру можливості настання або появи ризикового випадку та його наслідків. Обчислення виконуються за допомогою частоти ризикової події, яка потребує потрібної інформації. Наслідки ризикової події являють собою інтервальні та точкові оцінки.

В умовах повної невизначеності можливе використання експертних методів, які несуть суб'єктивний характер і дозволяють застосовувати способи зниження ступеня невизначеності та прийняття обґрунтоване рішення.

1.7 Поняття страхового ринку

Страховий ринок — система фінансово-економічних відносин, де об'єктом купівлі-продажу виступає страхова послуга, формується попит і пропозиція на неї. Потреба підтримки безперервності відстворювального процесу у випадку несподіваних і несприятливих обставин зумовлює розвиток страхового ринку.

Такий ринок є також формою організації фінансових відносин у формуванні та розподілі страхового фонду для забезпечення страхового захисту.

Невід'ємною складовою умовою функціонування страхового ринку є наявність суспільної потреби на страхові послуги і наявність страховиків, які можуть задовольнити потреби.

Страховий ринок являє собою складну та інтегровану систему, що включає в себе різні структурні ланки.

Базовою ланкою страхового ринку виступає страхова компанія.

Страхова компанія — історично визначена суспільна форма функціонування страхового фонду і являє собою відособленою структурою, що здійснює укладання договорів страхування та їх обслуговування. Для страхових компаній характерна техніко-організаційна єдність. Економічна відокремленість страхової компанії полягає в повній відокремленості її ресурсів та їх самостійному повному обсягу.

Страховий ринок виконує регулюючу функцію за наявності економічної конкуренції. Успіх на страховому ринку залежить від страховика, його мотивації співробітників страхового товариства до пошуку нових можливих клієнтів, удосконалення методів страхового обслуговування.

1.8 Фінансовий аналіз страхової компанії

Аналіз фінансової діяльності будь-якої страхової компанії відносять до окремої області класифікатора дослідження, що є результатом з точки зору економічного походження страхування, яке в ринковій економіці виступає з однієї сторони в ролі засобу захисту майнових інтересів, а з іншої — джерелом доходу. Також варто зазначити, що страхування є невід'ємним фактором розвитку інвестиційного потенціалу будь-якої країни. Діяльність страховика включає в себе страхові, інвестиційні та фінансові операції. Тут наявні і відповідно три види ризиків, які властиві для функціонування страхової компанії. Ризиковий характер страхової діяльності і є причиною специфіки фінансового аналізу страхової компанії.

Фінансовий аналіз страхової компанії — це в першу чергу аналіз фінансового стану установи та оцінка її надійності. Фінансова стійкість є

першочерговим показником фінансово-економічної діяльності. Якщо страхова компанія фінансово стійка, то вона є платоспроможною і вона має переваги в плані залучення інвестицій, клієнтів, тощо. Таке підприємство не конфліктує з державою та суспільством щодо перерахування податків і забезпечення виплат. Результати фінансового аналізу стану страхової компанії є об'єктом інтересу як для внутрішніх користувачів при ухваленні рішень, так і для зовнішніх, що взаємодіють безпосередньо з цією компанією. Постає питання у виборі оптимального методу фінансового аналізу даної установи.

Методика фінансового аналізу будь-якого підприємництва у вузькому сенсі — це сукупність показників фінансового стану та методів оцінювання. Дані методики поділяють за варіантами групування тих або інших показників фінансового стану підприємництва, але вони охоплюють показники, що дають оцінку платоспроможності, ліквідності, власного капіталу та зобов'язання. Основна частина методів фінансового аналізу створена для промислових підприємств. Для їх застосування відносно оцінки фінансового стану страховика, потрібно розуміти особливості страхової діяльності, розуміння базових понять і показників, що використовуються у фінансовому менеджменті страховика. Фінансовий аналіз будь-якої страхової компанії має бути спрямований на загальні показники фінансового стану і на специфічні показники, що властиві лише страховій діяльності. Дослідженнями даної задачі займаються українські та зарубіжні вчені: В. Куликов, В. В. Шахова, Л. А. Орланюк-Малицька. Вони виокремили головні напрями фінансового менеджменту страховика та запропонували показники та методи аналізу діяльності страхової компанії.

Фінансові ресурси страхової компанії складаються з двох головних частин — власного та залученого капіталу. Як правило, залучена частина капіталу в якості страхових премій за об'ємом є більшою, ніж власна. Вони не є власністю страховика і є тимчасовими протягом дії договорів страхування. Ці ресурси знаходяться під управлінням страхової компанії і використовуються для страхових виплат або стають прибутковою базою у випадку беззбиткового

проходження договорів. Дану частину частину капіталу не можна однозначно класифікувати як позиковий капітал, так як він має бути відторгнутим у повному обсязі у майбутньому. Перетворення даних засобів через певний інтервал часу в дохід страховика аргументує коректність поняття «залучений капітал». Вони можуть використовуватися страховиком в якості інвестиційного джерела лише на певний час.

Накопичуючи грошові кошти, страховики можуть здійснювати страховий захист страхувальників. Також, використовуючи значний об'єм тимчасово вільних коштів, страхова компанія може залучати інвестиційну діяльність. Саме через це страхові організації відносять до фінансових установ. У цьому секторі економіки циркулюють великі грошові потоки, а страхові організації грають роль активних учасників перерозподілу валового внутрішнього продукту, стаючи фінансовими корпораціями зі своєю галузевою специфікою.

1.9 Сучасний стан страхового ринку України

Станом на 30 вересня 2019 року в Державному реєстрі фінансових установ було зареєстровано 234 страхові компанії, серед яких 23 забезпечували послуги «life», а 211 — «non-life». Починаючи з 2000-х років кількість страхових компаній була постійною, але внаслідок кризи 2008-2009 років почала зменшуватися (рисунок 1.3).

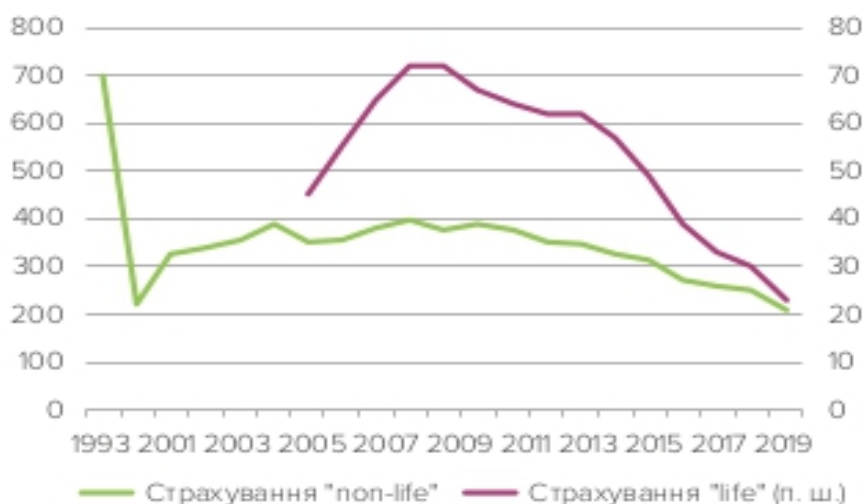


Рисунок 1.3 — Графік кількості страхових компаній України

Протягом дев'яти місяців 2019 року з ринку страхування пішла 51 компанія. Більшість з них, за вимогами законодавства, покинули ринок, так як вони вели свою діяльність більше ніж 6 місяців або вирішили анулювати власні ліцензії.

Кількість страхових компаній станом на 31.03.2020 становила 225, з яких 22 страхові компанії зі страхування життя та 203, що здійснювали страхування, що відмінні від життя. Станом на кінець березня 2020 року кількість страхових компаній у порівнянні зі схожим періодом у 2019 році кількість страхових компаній зменшилася на 40 одиниць (таблиця 1.2) [9].

Таблиця 1.2 — Кількість страхових компаній з 2018 до 2020 року

Кількість страхових компаній	Станом на 31.12.2018	Станом на 31.12.2019	На кінець кварталу	
			I кв. 2019	I кв. 2020
Загальна кількість	281	210	265	225
з них страхові компанії «non-life»	251	210	238	203
з них страхові компанії «life»	30	23	27	22

Обсяги валових страхових премій зростали, починаючи з 2015 року. Водночас, рівень страхових виплат залишався незмінним в інтервалі 24-27% (рисунок 1.4).



Рисунок 1.4 — Графік страхових премій та рівня виплат.

Валові страхові премії, що були отримані страховиками при страхуванні та перестрахуванні від страхувальників та перестрахувальників за I квартал 2020 року, становили 11548,8 млн. грн., з яких 46,8 % — надійшли від фізичних осіб, а 53,2 % — від юридичних осіб.

За I квартал 2020 року сума отриманих страховиками валових премій зі страхування «non-life» становила 10295,6 млн. грн., а зі страхування «life» — 1253,2 млн. грн.

Чисті страхові премії за II-й квартал 2020 року становили 83,7% від валових страхових премій.

Обсяги валютних страхових премій за I квартал 2020 року зменшились порівняно з I кварталом 2019 року на 13,4%, чисті страхові премії зменшились на 0,1% (рисунок 1.5).

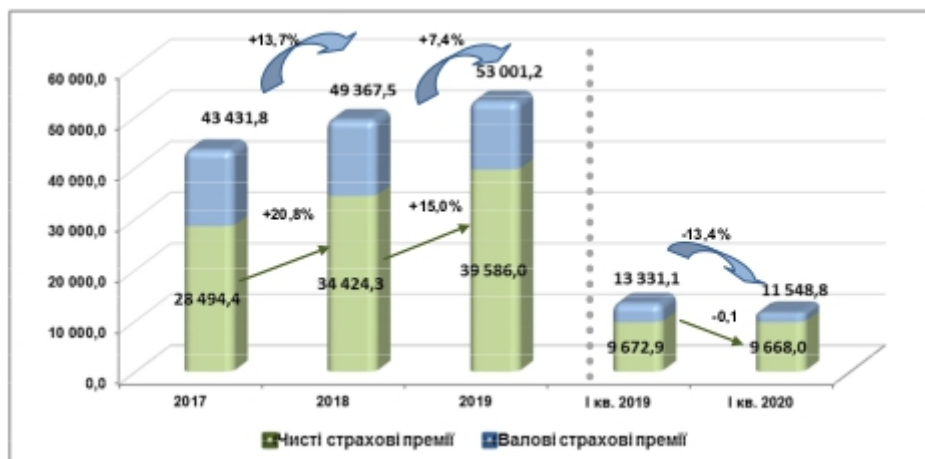


Рисунок 1.5 — Графік змін страхових премій за 2017-2019 рік, I квартал 2020 року

Концентрація страхового ринку за надходженням валових страхових премій станом на кінець березня 2020 року подана в таблиці 1.3.

Таблиця 1.3 – Концентрація страхового ринку за I квартал 2020 року

Перші (Тор)	страхування «life»		страхування «non-life»		
	Надходження премій (млн. грн.)	Частка на ринку, %	Надходження премій (млн. грн.)	Частка на ринку, %	Кількість страхових компаній, які більше 50% страхових премій отримали від перестраховувальників
Тор 3	712,1	56,8	1750,4	17,0	0
Тор 10	1206,0	96,2	4098,0	39,8	1
Тор 15	1253,0	100,0	5271,8	51,2	3
Тор 30	х	х	7859,5	76,3	8
Тор 50	х	х	9247,5	89,8	8
Тор 100	х	х	10174,8	98,8	9
Тор 150	х	х	10290,4	99,9	11
Всього по ринку	1253,2	100,0	10295,6	100,0	13

Звідси можна зробити висновок, що, не дивлячись на велику кількість компаній, основну частку валових страхових премій — 98,8% — акумулюють 100 страхових компаній «non-life» та 96,2% — 10 страхових компаній «life».

За ринком страхування життя Індекс Герфіндаля-Гіршмана (ННІ) становив 1529,92 за I квартал 2020 року (у I кварталі 2019 року — 1573,93), за ринком ризикових видів страхування за I квартал 2020 року показник становив 257,13, (у I кварталі 2019 року — 236,07). Загалом по страховому ринку ННІ дорівнював 222,37 (у I кварталі 2019 року — 210,52).

Дані вказують на те, що у I кварталі 2020 року на ринку видів страхування інших, ніж страхування життя, був високий рівень конкуренції, водночас на ринку страхування життя спостерігається помірна монополізація.

В Україні ринок страхування є незначним за обсягами активів та рівнем споживання страхових послуг. Показник проникнення (Insurance Penetration ratio), що визначається як співвідношення обсягу залучених страхових премій до ВВП, становить 1,4% (з них «non-life» — 1,3%). Для прикладу: у Європі цей показник становить 5,0%, а в середньому у світі — 6,1% [10]. Країни, які мають доволі високий показник проникнення — це Великобританія, Японія та Франція (рисунок 1.6).

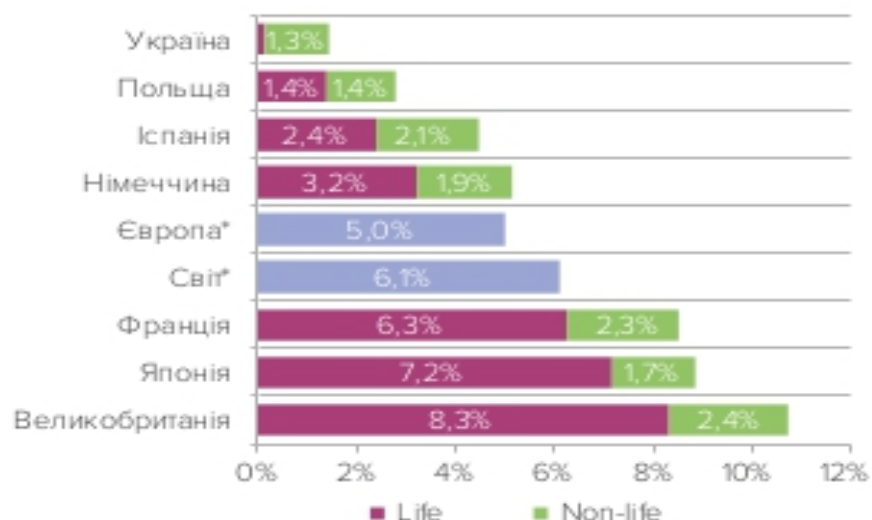


Рисунок 1.6 — Графік рівня проникнення послуг страхування в економіку у 2018 році (синім кольором відмічено сумарний показник «life» і «non-life»)

Показник щільності страхування (Insurance Density ratio) — це розмір страхової премії на одну особу. В Україні він перебуває на дуже низькому рівні і на кінець 2018 року становив 34 дол. США (рисунок 1.7).

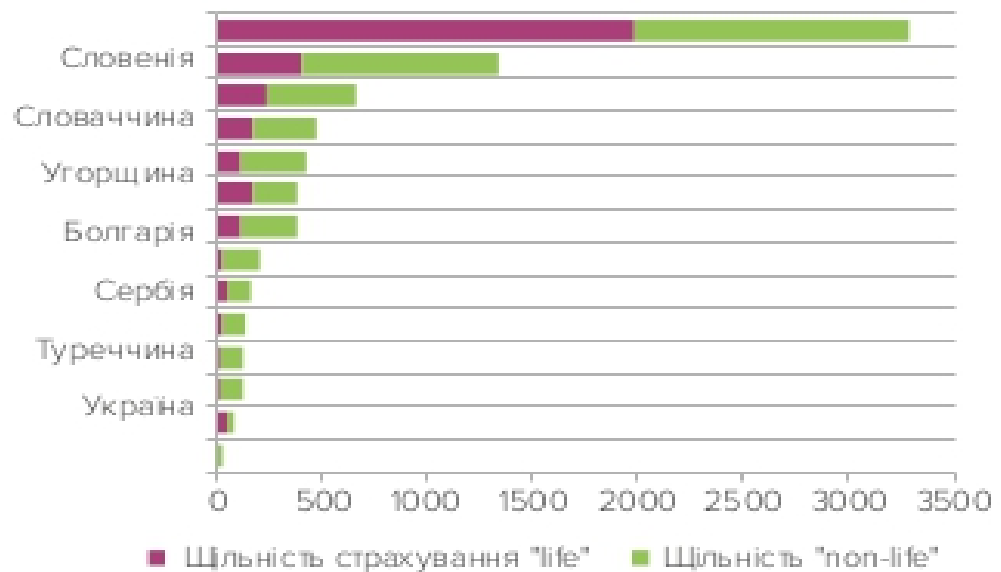


Рисунок 1.7 — Страхові премії на одну особу у 2018 році в дол. США

З-поміж 88 країн, які досліджував Swiss Re Institute, Україна за цими показниками споживання послуг «non-life» страхування посідає 73 місце та за послугами страхування життя — 82 місце.

Станом на кінець 2018 року частка п'яти найбільших страхових компаній за обсягом активів страхування «non-life» становила 25% ринку, а частка 20 найбільших за обсягом активів компаній становила 58% ринку. Частка двох найбільших компаній страхування «life» становила 74% активів. Частка валових премій страхування «non-life» у 2018 році становила 92,1% від загального обсягу валових премій, і лише 7,9% було отримано від страхування «life». Для прикладу: у ЄС частка премій за договорами страхування життя становить 58% та є джерелом довгострокового фінансування для фінансового сектору. У структурі страхування «non-life» найбільшу частку страхування займає автострахування — 26,3% від об'єму валових страхових премій у 2018 році. Більшість страхових компаній, які виходять на ринок, в основному орієнтуються на здійснення моторних видів страхування. Страхування «life» за

договорами накопичувального страхування є непопулярним в Україні і зазвичай проводиться в якості ініціативи з боку роботодавців.

Обсяг активів для покриття страхових резервів на кінець 2018 року дорівнював 40,7 млрд грн, з яких 26,0 — для покриття технічних резервів та резервів зі страхування життя (рисунок 1.8).



Рисунок 1.8 – Структура активів для технічних резервів та резервів із страхування життя на кінець 2018 року

1.10 Основні проблеми ринку страхування України

Недосконале законодавство, відсутність відповідного регулювання та нагляду за діяльністю та поведінкою страхових компаній стали причиною накопичення системних проблем, які вимагають невідкладних рішень. Усе це призводить до низького рівня довіри до компаній сектору та не дають йому розвиватися відповідним чином.

1.10.1 Низька платоспроможність та ліквідність

Платоспроможність — здатність страховика виконати свої зобов'язання згідно укладеному договору страхування. Зараз система оцінки платоспроможності страхових компаній є спрощено адаптованою формою європейської системи оцінки платоспроможності Solvency I.

Дана система оцінки фінансового стану страховиків не дає можливості отримати вчасно об'єктивні дані про діяльність компанії та вчасно реагувати на проблеми платоспроможності. Через це у регулятора відстуні необхідні механізми контролю за виконанням вимог щодо платоспроможності страховиків. За чинної регуляції страхові компанії можуть суттєво занижувати резерви та підвищувати вартість активів (завищена вартість копоративних цінних паперів, нерухомості, дебіторської заборгованості). Як результат, неможливо оцінити реальну платоспроможність, так як формально наявний капітал був насправді недостатнім.

Крім того, недостатній обсяг високоякісних ліквідних активів для потенційних виплат є причиною проблем ліквідності. Внаслідок цього страхові компанії починають занижувати виплати. У той же момент у регулятора

відсутня інформація щодо строковості за потенційними виплатами та активами, що робить неможливою оцінку щодо розривів ліквідності.

Через це рівень платоспроможності та ліквідності багатьох страхових компаній є доволі низьким та потребує підвищення.

1.10.2 Недосконалість бізнес-моделей страховиків

Більшість страхових компаній України працюють без чіткої стратегії розвитку та бізнес-моделі. Особливості ціноутворення веде до формування концентрованої структури страхового портфеля, недиверсифікованості продуктів та ризиків страхування. Внаслідок цього демпінг страхових компаній призводить до поглиблення проблеми на ринку. Зараз через недоліки діючого законодавства та регуляторної політики на ринку страхові послуги використовуються для заниження бази оподаткування. Аналіз діяльності страхових компаній дає можливість висунути гіпотезу про те, що головною бізнес-метою окремих страховиків є діяльність, що пов'язана з виведенням капіталу, легалізацією незаконно отриманих доходів, оптимізацією податків, включаючи конвертацію безготівкових коштів у готівку.

Діяльність компаній, що пов'язана зі страхуванням життя, на даний момент спрямована на залучення клієнтів за короткостроковими договорами страхування. Недостатня кількість засобів для залучення довоготрокових ресурсів обмежує можливості бізнесу розвиватися та ставати стійкішим.

1.10.3 Недостатня прозорість ринку та проблеми в корпоративному управлінні

Недосконалим є і відображення фінансових результатів страхових компаній. Наприклад, значну частку страхових премій насправді становлять комісійні винагороди агентам (30-50%). Страхові тарифи за багатьма напрямками страхування не є економічно обґрунтованими і не базуються на актуарних розрахунках, і тому не дають можливості покрити усі можливі страхові виплати. Проблеми з ліквідністю внаслідок поганої якості активів стимулює страхові компанії занижувати виплати. Валовий прибуток від операцій страхування напрямлений на покриття серйозних операційних втрат. Через це сектор у значній мірі операційно неефективний.

На даний момент немає дієвих механізмів контролю за розкриттям власності страхових компаній, не встановлено належних стандартів та вимог щодо ділової репутації власників та осіб, що виконують ключові функції у страхових компаніях, немає структурованого підходу до побудови корпоративного управління та системи внутрішнього контролю. Це веде до можливої непрозорості діяльності та звітування. Також на ринку працює велика кількість страховиків, кінцеві бенефіціарні власники яких невідомі, що унеможливорює встановити відповідальності у разі доведення компанії до неплатоспроможності.

1.10.4 Проблема захисту прав споживачів послуг страхування

В Україні немає механізму гарантування страхових виплат за довгосроковими договорами страхування життя, що є іншою причиною непопулярності даного виду страхування.

Дійсна процедура виходу з ринку не захищає прав споживачів страхових послуг належним чином. Вона є доволі тривалою та у зв'язку з її недосконалістю дозволяє недобросовісним страховикам вийти з ринку без виконання власних зобов'язань за укладеними договорами страхування.

Низький попит на страхові послуги в Україні зумовлений такими факторами:

- недовіра до непрозорого ринку страхування та фінансової системи;
- відсутність дієвої системи захисту прав споживачів страхових послуг;
- фінансова необізнаність громадян у страхових продуктах;
- низька платоспроможність населення;
- неринкова поведінка страховиків щодо споживачів.

1.11 Висновки до розділу 1

У першому розділі розглянуто поняття страхування від ризику, його роль в страховій економіці, його функції та наукову класифікацію. Визначено поняття страхового ринку та оцінено його поточний стан в Україні. У межах оцінювання та прогнозування стану страхових компаній, можна зробити висновок, що найкращими методами для оцінювання стану та прогнозування є моделі, що працюють з невеликим об'ємом даних.

Звідси з дослідженої інформації можна сформулювати задачу оцінювання стану страхової компанії, яку має вирішувати майбутня система підтримки прийняття рішень, як класифікацію вибірки статистично-економічних даних та пошуку залежності між даними та ймовірністю платоспроможності. Було прийнято рішення використати декілька математичних моделей для проектування системи підтримки прийняття рішень, обчислювальних дослідів та проведення їх якісного аналізу.

РОЗДІЛ 2 МАТЕМАТИЧНА ПОСТАНОВКА ЗАДАЧІ

2.1 Формалізація задачі оцінювання страхових компаній

Задача оцінювання та прогнозування стану страхової компанії є задачею бінарної класифікації: прогнозування, чи є діяльність компанії ефективною. Частина методів оцінювання ризиків, що були розглянуті у попередньому розділі, не є доступними для громадськості, і в загальному випадку не дають можливості дати повну оцінку стану страхових компаній.

Вхідними даними для системи та математичних моделей машинного навчання з вчителем є економічні дані страхових компаній України за 2020 рік. Потім обирається модель з найкращими показниками навчання та використовується для прогнозування подальшого стану компанії, використовуючи дані компаній за перше півріччя 2021 року.

2.2 Обробка даних та метрики якості класифікації

2.2.1 Попередня обробка даних

Як правило, дані в реальному часі характеризуються великою кількістю пропусків, неповнотою та непослідовністю. Попередня обробка даних дозволяє покращити якість даних, водночас допомагаючи покращити якість та ефективність навчання моделей класифікації з вчителем. Це є важливим кроком у пошуку корисних знань, оскільки якісні рішення повинні ґрунтуватися на якісних даних.

Обробка даних складається з наступних кроків. Спочатку видаляються стовпці, які не містять в собі релевантної інформації. Якщо стовпці числових даних містять в собі декілька пропусків і несуть в собі важливу інформацію, то пропуски заповнюються середнім значенням або медіаною. На останньому

кроці проводиться стандартизація даних, щоб дані мали стандартний нормальний розподіл. Стандартизація є важливим кроком, оскільки більшість алгоритмів машинного навчання (лінійна та логістична регресія, нейронна мережа, тощо) використовують градієнтний метод оптимізації і потребують масштабованих даних. Різниця в області значень ознак задає різний розмір кроку для кожної ознаки. Щоб бути впевненим у тому, що градієнтний спуск рухається плавно до точки мінімуму і розмір кроку оновлюється з однаковою швидкістю для всіх ознак, ми масштабуємо дані перед їх завантаженням у моделі. Крім того, алгоритми відстані (k-найближчих сусідів, k-середніх та метод опорних векторів) повинстю залежать від розмаху ознак. Це пов'язано з тим, що вони використовують відстані між точками даних для визначення їх подібності.

2.2.2 Вибір даних-ознак для навчання моделі

Для того щоб визначитися з ознаками, що беруть участь у навчанні моделей, будується кореляційна матриця вибірки даних, елементи якої, обчислюються за формулою співвідношення Пірсона:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} ,$$

де r — коефіцієнт кореляції Пірсона;

x_i — i -ий компонент ознаки x ;

y_i — i -ий компонент ознаки y ;

$\bar{x}$ — вибіркове середнє x ;

$\bar{y}$ — вибіркове середнє y .

Після кореляційного аналізу видаляються ті стовпці які мають близьку до нульової кореляцію, що веде до зменшення розмірності вхідних ознак та зменшення шуму.

2.2.3 Метрики моделей класифікації

Для оцінювання якості навчання моделей класифікації використовувалися наступні метрики.

– Матриця похибки

Це вид таблиці, що дозволяє показати інформацію щодо якості класифікаційної моделі на валідаційному наборі даних, де дійсні значення є відомими. Кожен рядок матриці містить екземпляри реальних класів, в той же час як стовпець містить екземпляри спрогнозованих класів. Назва матриці походить від факту, що вона дозволяє легко показати, чи допускає модель класифікації помилки у випадку двох класів. Загальний вид матриці зображено на рисунку 2.1.

		Predicted classes	
		Negative 0	Positive 1
Actual classes	Negative 0	TN	FP
	Positive 1	FN	TP

Рисунок 2.1 — Загальний вигляд матриці похибок

Таблиця складається з чотирьох елементів:

1. TN (True Negative) — правильно спрогнозовані значення як дійсні негативні значення.
2. FP (False Positive) — прогнозовані значення неправильно оцінили дійсне позитивне значення, тобто негативні значення оцінювалися, як позитивні (помилка першого роду).
3. FN (False Negative) — позитивні значення були спрогнозовані як негативні (помилка другого роду).
4. TP (True Positive) — правильно спрогнозовані значення як дійсні позитивні значення.

— Точність (Accuracy)

Це відношення числа правильних прогнозів до загальної кількості вхідних значень. Вона обчислюється за допомогою елементів матриці похибок за такою формулою:

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP}.$$

Дана метрика приймає значення від 0 до 1. Для найкращої моделі метрика приймає максимальне значення. Але у випадку незбалансованого набору даних точність може бути хибним показником якості моделі. Для цих умов використовують збалансовану точність, яка обчислюється за формулою:

$$Balanced Accuracy = \frac{TPR + TNR}{2},$$

де $TPR = \frac{TP}{TP + FN}$ — чутливість; частка істинно позитивних значень, що були спрогнозовані правильно;

$TNR = \frac{TN}{TN + FP}$ — специфічність; частка істинно негативних значень, що були спрогнозовані правильно.

– Загальна точність (Precision)

Дана метрика показує, наскільки точна модель класифікації з прогнозованих позитивних значень і скільки з них є істинно позитивними. Обчислюється за формулою:

$$Precision = \frac{TP}{FP + TP}.$$

Влучність є гарною величиною для визначення, коли ціна хибно позитивних значень є доволі високою.

– Повнота (Recall)

Цей показник обчислює скільки дійсно позитивних значень модель правильно оцінила серед усіх прогнозів. Формула має такий вигляд:

$$Recall = \frac{TP}{FN + TP} .$$

– F1-міра

Легко зрозуміти, що високий показник влучності та повноти свідчать про те, що модель класифікації навчається дуже добре. Але в реальних ситуаціях як правило цього не відбувається одночасно, тому часто доводиться шукати деякий компроміс. Тому виникає потреба у такій метриці, що дозволяє об'єднувати інформацію про влучність та повноту моделі. За рахунок цього стає легшим прийняття рішення про те, яка модель буде використовуватися для прогнозу. У цьому випадку допомагає F1-міра.

F1-міра являє собою гармонійне середнє між влучністю та повнотою. Метрика обчислюється за такою формулою:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Дана величина представляє собою гармонійне середнє між влучністю та повнотою. Вона прямує до нуля у випадку, коли влучність або повнота прямує до нуля. F1-міра дає однакові ваги як для точності так і для повноти, тому метрика буде зменшуватися одночасно зі зменшенням влучності або повноти. За наявності даної оцінки якості класифікатора легко прийняти рішення, чи спостерігаються зміни в моделі чи ні.

– ROC-AUC

При перетворенні дійсної відповіді моделі (ймовірності того, що спрогнозована зміна належить до одного з класів) у бінарну змінну, ми маємо обрати поріг, при якому негативна мітка стає позитивною. Зазвичай в якості

цього порогу виступає значення 0.5, але воно не завжди є оптимальним, у випадку незбалансованості даних.

Одним з методів оцінки моделі без прив'язки до конкретного порогу, є метрика ROC-AUC — площа (Area Under Curve) під кривою помилок (Receiver Operating Characteristic curve). Крива представляє собою лінію від (0,0) до (1,1) у координатах True Positive Rate (TPR) та False Positive Rate (FPR). Остання обчислюється за формулою:

$$FPR = \frac{FP}{FP + TN}.$$

FPR показує частку елементів негативного класу, що були оцінені алгоритмом неправильно. Графік кривої для випадку звичайного та ідеального класифікатора зображено на рисунку 2.2.

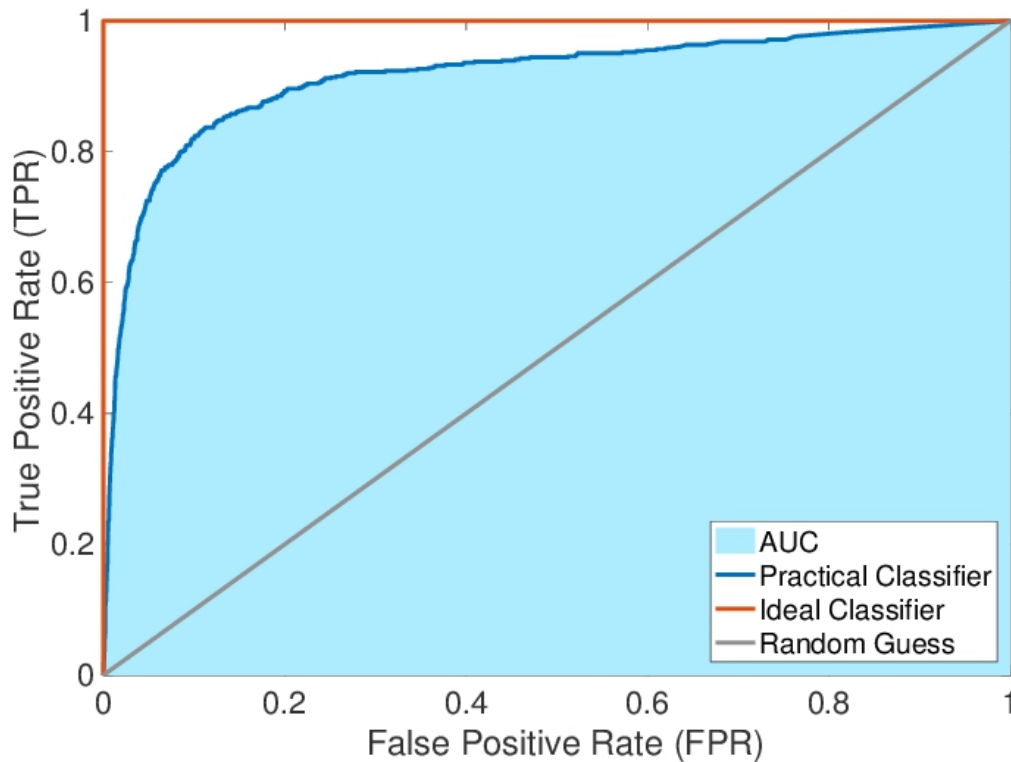


Рисунок 2.2 — Графік кривої похибки; сірим кольором відображено випадкові значення; синім кольором показано роботу звичайного класифікатора: чим вище позиція графіку над діагоналлю, тим краща якість роботи класифікатора; так само чим більша площа кривої під графіком, тим краще; оранжевим кольором зображено роботу ідеального класифікатора

В ідеальному випадку, коли класифікатор працює безпомилково ($TPR = 1$, $FPR = 0$), ми отримуємо площу кривої, що дорівнює одиниці; інакше, коли класифікатор випадково оцінює ймовірності класів, ROC-AUC прямуватиме до 0.5, оскільки класифікатор буде давати однакову кількість TP та FP. Площа під кривою відображає якість алгоритму, крім того, важливою складовою є крутизна цієї кривої — ми максимізуємо TPR, мінімізуючи FPR, тому крива має прямувати до точки (0,1).

Метрика ROC-AUC стійка до незбалансованих даних і може інтерпертуватися як ймовірність того, що випадково позитивне значення буде мати більш високу ймовірність бути позитивним, ніж випадково обране негативне значення.

2.3 Математичні моделі класифікації

Для оцінювання та прогнозування було обрано шість моделей класифікації: k -найближчих сусідів, метод опорних векторів, наївний Байєсівський класифікатор, випадковий ліс, XGBoost та багат шаровий перцептрон.

2.3.1 Метод k -найближчих сусідів

Метод k -найближчих сусідів — метричний алгоритм для задачі автоматичної класифікації об'єктів або регресії. У випадку задачі класифікації результатом є належність об'єкта до відповідного класу. Об'єкт класифікується множиною голосів його сусідів, причому об'єкт відносять до класу, який є найбільш поширеним серед k найближчих сусідів. Для задачі регресії обчислюється середнє значення цільової змінної сусідів.

Параметр k є константою і задається користувачем. Вектор ознак без мітки класифікується за допомогою присвоєння мітки, яка найчастіше зустрічається серед k тренувальних зразків, які найближче розташовані до цієї точки. Схема навчання алгоритму наведено на рисунку 2.3.

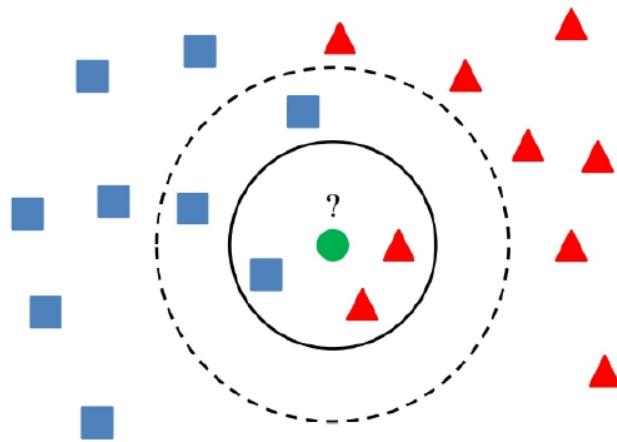


Рисунок 2.3 – Графічне зображення алгоритму k -найближчих сусідів

Метод k -найближчих сусідів є типом класифікації, де функція апроксимується тільки локально, а всі обчислення відкладаються до оцінки функції. Так як алгоритм ґрунтується на відстані для класифікації, якщо ознаки являють собою фізичні одиниці чи мають доволі різні масштаби, то стандартизація або нормалізація навчальної вибірки дозволяє підвищити точність алгоритму.

Для оцінювання стану страхових компаній значення k дорівнювало трьом.

2.3.2 Метод опорних векторів

Метод опорних векторів — метод навчання з учителем, що використовується як для розв’язування задачі класифікації так і для регресії. Дається набір навчальних зразків з позначенням належності до одного з двох класів. Алгоритм будує модель, що призначає кожен новий зразок до однієї чи іншої категорії, створюючи її як наймовірніший бінарний лінійний класифікатор. Метод опорних векторів відображає навчальні зразки у точки в просторі, максимізуючи при цьому ширину проміжку між двома категоріями. Нові зразки у цьому випадку відображаються у той же простір і прогнозується

їх належність до певного класу в залежності від того, на яку сторону проміжку вони потрапляють (рисунок 2.4).

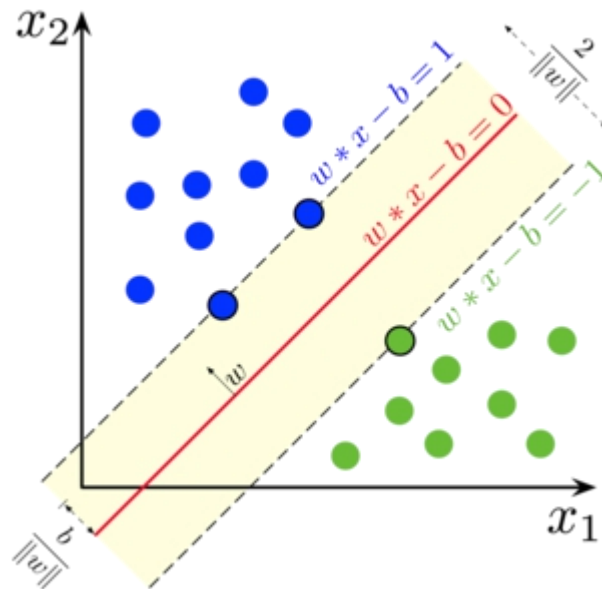


Рисунок 2.4. — Схема алгоритму методу опорних векторів

У формальному сенсі, метод опорних векторів будує гіперплощину чи множину гіперплощин у багатовимірному чи нескінченновимірному просторі. Гарне ділення даних досягається за рахунок гіперплощини, що має найбільшу відстань найближчих тренувальних зразків будь-якого класу (функціональна маржа), оскільки чим більша маржа, тим нижча узагальнена помилка класифікатора.

Під час навчання моделі розв'язується така задача оптимізації [11]:

$$\begin{cases} \min_{w,b} \frac{1}{2} w^T w \\ y_i((w^T x_i) + b) \geq 1, i = 1, \dots, l. \end{cases},$$

де w – ваговий вектор;

b – вектор констант.

Для класифікації використовувалося лінійне ядро.

2.3.3 Наївний Байєсівський класифікатор

Наївний Байєсівський класифікатор — простий ймовірнісний класифікатор, який ґрунтується на застосуванні теореми Байєса з наївним припущенням щодо незалежності ознак. Для оцінки параметрів моделей використовується метод максимальної правдоподібності, тобто працюючи з наївною Байєсівською моделлю, можна і не знати теорії байєсівської ймовірності і не використовувати байєсівські методи.

Ймовірнісна модель класифікатора — це модель виду

$$p(L | F_1, \dots, F_n)$$

залежної змінної класу L з невеликою кількістю класів, що залежить від змінних $F_1, \dots, F_n$.

Використовуючи теорему Байєса, запишемо ймовірність [12]:

$$p(L | F_1, \dots, F_n) = \frac{p(L) p(F_1, \dots, F_n | L)}{p(F_1, \dots, F_n)}.$$

У цій формулі нас не цікавить знаменник, оскільки він не залежить від L , а значення властивостей F_i апріорі відомі.

А чисельник представляє собою еквівалентність формули спільної ймовірності моделі виду:

$$p(L, F_1, \dots, F_n),$$

яку можна переписати таким чином, використовуючи повторне застосування визначення умовної ймовірності:

$$\begin{aligned} p(L, F_1, \dots, F_n) &= p(L) p(F_1 | L) p(F_2 | L, F_1) \dots \\ &\dots p(F_n | L, F_1, F_2, \dots, F_{n-1}). \end{aligned}$$

Застосуємо наївні припущення щодо умовної незалежності: нехай кожна ознака F_i умовно незалежно від будь-якої іншої ознаки F_j при $i \neq j$, тобто

$$p(F_i | L, F_j) = p(F_i | L).$$

Звідси спільну модель можна записати таким чином:

$$\begin{aligned} p(L, F_1, \dots, F_n) &= p(L) p(F_1 | L) p(F_2 | L) \dots p(F_n | L) = \\ &= p(C) \prod_{i=1}^n p(F_i | L). \end{aligned}$$

Виходячи з припущення про незалежності ознак, ймовірність за міткою класу дорівнює:

$$p(L | F_1, \dots, F_n) = \frac{1}{p(F_1, \dots, F_n)} p(L) \prod_{i=1}^n p(F_i | L).$$

У роботі використовувався гаусівський класифікатор.

При роботі з неперервними даними, звичайним припущенням є те, що неперервні значення, які пов'язані з кожним класом, розподілені за нормальним законом. Припустимо, що навчальна вибірка містить неперервну ознаку x . Спочатку дані сегментуються за класом і потім обчислюється середнє значення і дисперсія для кожного класу. Позначимо μ_k — середнє значення x , яка пов'язана з класом C_k , а σ_k^2 — незміщена дисперсія значень x , яка пов'язана з класом C_k . Нехай було обрано спостереження v . Тоді щільність імовірності v даного класу C_k , $p(x = v | C_k)$ може бути обчислена введенням v у рівняння нормального розподілу з параметрами μ_k і σ_k^2 .

Тобто:

$$p(x = v | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v - \mu_k)^2}{2\sigma_k^2}}.$$

Іншим загальним методом при роботі з неперервними значеннями є використання інтервалів для дискретизації значень ознак для отримання нової вибірки, розподіленої за законом Бернуллі. У деякій літературі висувається припущення, що це необхідно для застосування наївного байєсівського класифікатора, але це не так, і дискретизація може відкинути дискретну інформацію.

2.3.4 Випадковий ліс

Випадковий ліс – алгоритм машинного навчання, який застосовується для задач регресії та класифікації. Він використовує ансамблеве навчання – метод, що комбінує велику кількість класифікаторів для вирішення складних проблем. Алгоритм випадкового лісу складається з множини дерев рішень. “Ліс”, згенерований даним алгоритмом, навчається за допомогою беггінгу — ансамблевого мета-алгоритму, що покращує точність алгоритмів машинного навчання за рахунок усереднення результатів [13]. Метод встановлює результати на основі прогнозів дерев рішень.

Головна відмінність алгоритму випадкового лісу від алгоритму дерева рішення полягає в тому, що встановлення кореневих вузлів та відокремлення вузлів відбувається випадково у першому випадку. Метод використовує беггінг для генерації необхідних прогнозів. Беггінг включає використання різних наборів даних з навчальної вибірки частіше ніж один. Навчальна вибірка складається зі спостережень та ознак, що використовуються для прогнозів. Дерева рішень дають різні результати в залежності від навчальних даних, що подаються до алгоритму. Результати ранжуються, прогноз з найвищою оцінкою обирається як кінцевий результат (рисунок 2.5).

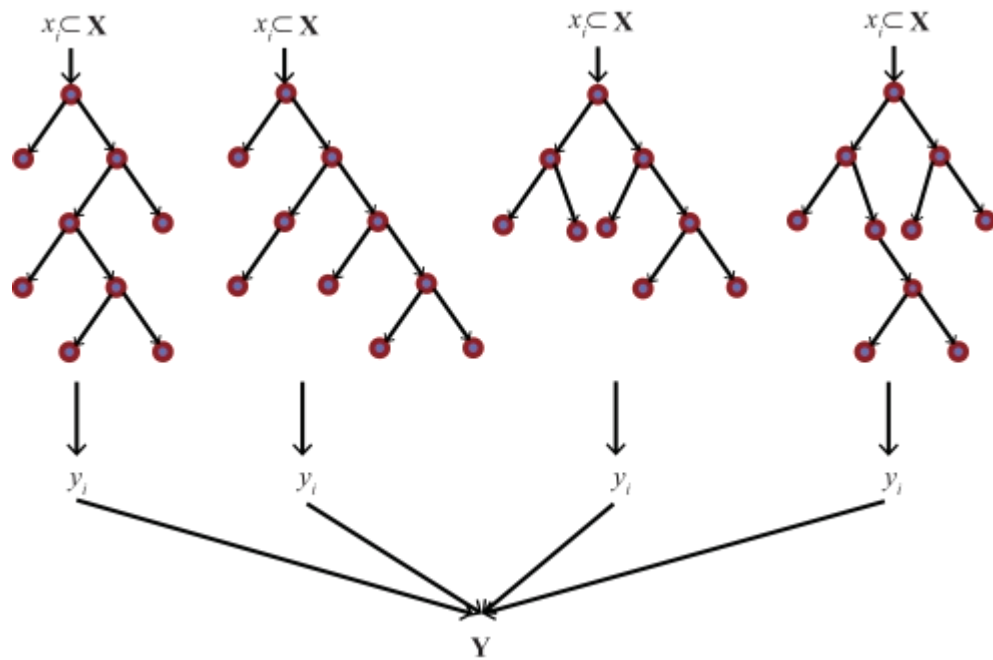


Рисунок 2.5 – Алгоритм роботи випадкового лісу

Для задачі класифікації вибір остаточного результату відбувається за допомогою системи мажоритарного голосування. Вихід, що був обраний більшістю дерев рішень, стає кінцевим результатом випадкового лісу.

2.3.5 XGBoost

Бустинг — це підхід, що комбінує множину простих моделей у єдину композиційну модель. З додаванням більш простих моделей, загальна модель перетворюється у сильний предиктор. За термінологією бустингу, прості моделі називаються слабкими моделями.

Градiєнтний бустинг перевизначає бустинг як проблему чисельної оптимізації, у якій задача полягає у мінімізації функції втрат моделі за рахунок слабких моделей, використовуючи алгоритм градієнтного спуску. Оскільки градієнтний бустинг ґрунтується на мінімізації функції втрат, можуть використовуватися різні функції втрат, що можна застосовувати для задач регресії, класифікації, тощо.

Градiєнтний бустинг — покрокова аддитивна модель, що генерує навчальні моделі протягом процесу навчання (додається одне дерево за один крок, дерева, що існують, в моделі не змінюються). Внесок слабкої моделі навчання до ансамблю ґрунтується на оптимізаційному процесі градієнтного спуску. Розрахований внесок кожного дерева базується на мінімізації загальної помилки сильної моделі. Градієнтний бустинг не модифікує розподіл вибірки, оскільки слабкі моделі тренуються на залишкових помилках сильної моделі. За рахунок тренування на залишках моделі, під цією альтернативою розуміють надання більшої важливості спостереженням, які були класифіковані неправильно. Нова слабка модель додається для зосередження на тих областях, де існуючі моделі працюють погано.

XGBoost — один з найбільш популярних варіантів градієнтного бустингу. Це ансамблевий алгоритм машинного навчання, який базується на деревах рішень. Він був створений для покращення продуктивності та швидкості моделей машинного навчання. Процедура XGBoost використовує попередньо-відсортований алгоритм та алгоритм на основі гістограми для обчислення найкращого поділу. Алгоритм на основі гістограм розбиває точки

даних для ознаки на дискретні інтервали і використовує ці інтервали, щоб знайти значення поділу гістограми (рисунок 2.6).

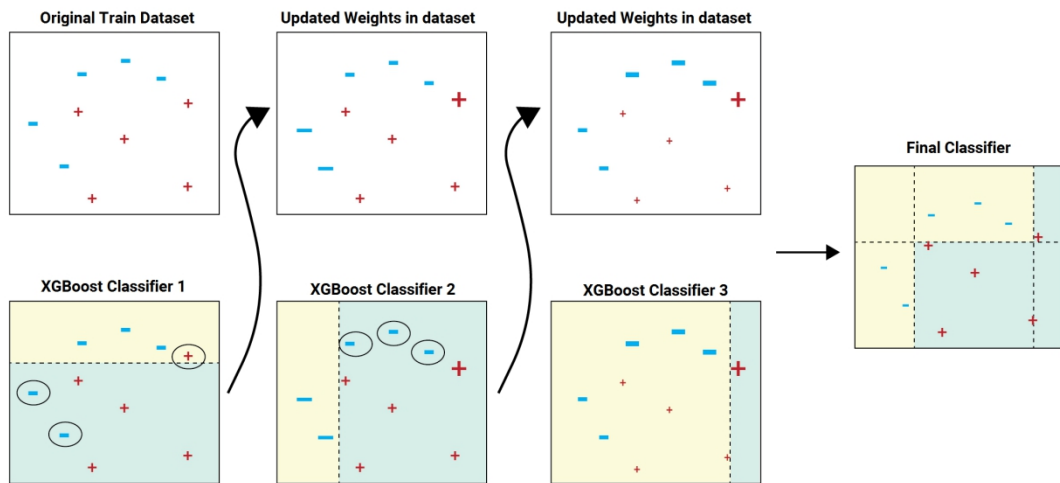


Рисунок 2.6 — Алгоритм роботи XGBoost

2.3.6 Багатошаровий перцептрон

Багатошаровий перцептрон — це глибока, штучна нейронна мережа. Вона складається з вхідного шару для отримання сигналу, вихідного шару для прийняття рішень або передбачення щодо вхідного сигналу, а між ними — довільна кількість шарів, які являють собою обчислювальну основу багатошарового перцептрону. Дана мережа з одним прохованим шаром здатна апроксимувати будь-яку неперервну функцію.

Багатошарові перцептрони часто застосовуються для задач навчання з учителем: вони тренуються на вибірках даних пари «вхід-вихід» і навчаються моделювати кореляцію або залежності між входами та виходами. Навчання включає в себе налаштування параметрів, ваг та зміщення, моделі з метою мінімізації помилок. Зворотнє поширення використовується для коригування ваг та зміщення щодо похибки, а саму помилку можна виміряти різними способами, як середньоквадратичною помилкою для випадку регресії чи перехресною ентропією для випадку класифікації.

При прямому шляху сигнал проходить через приховані шари до вихідного шару, і рішення останнього вимірюється за цільовими значеннями.

При зворотньому шляху, використовуючи зворотнє поширення та ланцюгове правило обчислення, часткові похідні функції помилок з посиленням до ваг та зміщень поширюються у зворотньому напрямку через багатошаровий персептрон. Даний процес диференціювання дає градієнт, за допомогою якого можна налаштовувати параметри мережі, коли вони переміщують нейронну мережу на крок ближче до мінімуму помилки. Для розробки системи підтримки прийняття рішень в якості оптимізації для багатошарового персептрону використовувався метод Adam.

Повну схему архітектури нейронної мережі подано на рисунку 2.7.

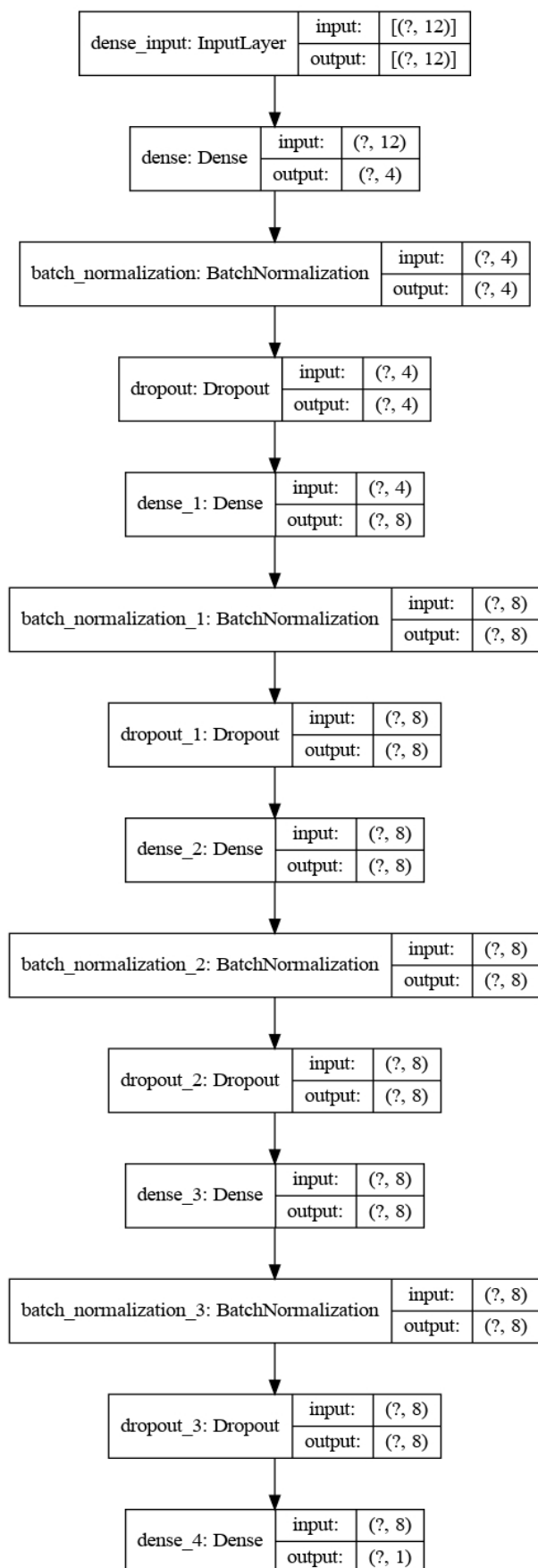


Рисунок 2.7 – Архітектура багатошарового перцептрону

Архітектура нейронної мережі складалася з чотирьох прихованих шарів, де перший шар містив чотири нейрони, решта — вісім. Така кількість була обрана через те, що цього було достатньо для вирішення задачі оцінювання. В якості функції активації для прихованих шарів використовувалася функція гіперболічного тангенсу:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Оскільки розглядалася задача бінарної класифікації, то для вихідного шару нейронної мережі використовувалася функція “сигмоїд”:

$$f(x) = \frac{1}{1 + e^{-x}}.$$

При навчанні башатошарового персептрону використовувалися наступні методи.

— Batch normalization (Пакетна нормалізація)

Це метод, що дозволяє підвищити продуктивність та стабілізувати роботу штучних нейронних мереж. Суть методу полягає у тому, що деяким шарам нейронної мережі на вхід подаються дані, що були попередньо оброблені та мають нульове математичне сподівання та одиничну дисперсію.

На практиці можливі два підходи реалізації алгоритму градієнтного спуску для навчання нейронних мереж: стохастичний і пакетний.

1. Стохастичний градієнтний спуск — реалізація, в якій на кожній ітерації алгоритму з навчальної вибірки випадковим чином обирається тільки один об’єкт.

2. Пакетний — реалізація градієнтного спуску, коли на кожній ітерації навчальна вибірка переглядається повністю, і тільки після цього змінюються ваги моделі.

Існує “золота середина” між цими двома підходами — коли переглядається тільки деяка підмножина навчальної вибірки фіксованого розміру. Так підмножини називають міні-пакетами (mini-batch).

Пакетна нормалізація зменшує величину, на яку зміщуються значення вузлів в прихованих шарах (коваріантне зміщення).

Коваріантне зміщення — ситуація, коли розподіл значення ознак в навчальній та тестовій вибірці мають різні параметри (математичне сподівання, дисперсія, тощо).

Найпростіший спосіб вирішення проблеми коваріантного зміщення для вхідного шару — це випадковим чином перемішувати дані перед створенням пакетів. Але для прихованих шарів нейронної мережі даний метод не підходить, оскільки розподіл вхідних даних для кожного вузла прихованих шарів змінюється кожний раз, коли відбувається оновлення параметрів в попередньому шарі. Це проблема внутрішнього коваріантного зміщення. Для вирішення цієї проблеми часто використовують низьку швидкість навчання та методи регуляризації при навчанні моделі. Іншим способом усунення внутрішнього коваріантного зміщення є метод пакетної нормалізації.

До корисних властивостей пакетної нормалізації відносять наступне. За рахунок цього підходу досягається більш швидка збіжність моделей, незважаючи на виконання додаткових обчислень. Пакетна нормалізація дозволяє кожному шару мережі навчатися більш незалежно від інших шарів. Також виникає можливість використання більш високої швидкості навчання, оскільки пакетна нормалізація гарантує, що виходи вузлів нейронної мережі не матимуть дуже великих чи малих значень. Крім того, метод є в деякому сенсі окремим механізмом регуляризації: даний метод привносить у виходи прихованих шарів деякий шум. Моделі стають менш чутливими до початкової ініціалізації ваг.

Опишемо метод пакетної нормалізації. Нехай на вхід деякого шару нейронної мережі подається вектор розмірності l : $x = (x^{(1)}, \dots, x^{(l)})$. Нормалізуємо даний вектор по кожній розмірності k :

$$\widehat{x}^{(k)} = \frac{x^{(k)} - E(x^{(k)})}{\sqrt{D(x^{(k)})}},$$

де $E(x^{(k)})$ — математичне сподівання;

$D(x^{(k)})$ — дисперсія.

Останні дві величини обчислюють по всій навчальній вибірці. Дана нормалізація вхідного шару може змінити представлення даних в шарі. Для уникнення цієї ситуації використовують два параметри стиску та зсуву нормалізованої величини для кожного $x^{(k)}$: $\gamma^{(k)}$, $\beta^{(k)}$, які застосовуються наступним чином:

$$y^{(k)} = \gamma^{(k)} \widehat{x}^{(k)} + \beta^{(k)}$$

Дані параметри налаштовуються у процесі навчання разом з іншими параметрами моделі.

Нехай навчання моделі проводиться за допомогою пакетів B розмірністю n [14]:

$$B = \{x_1, \dots, x_n\}.$$

Нормування застосовується до кожного елементу входу з індексом k окремо, тому у подальшому $x^{(k)}$ ми пропускаємо індекс для зрозумілого пояснення. Припустимо, що ми отримали нормовані значення пакету $\widehat{x}_1, \dots, \widehat{x}_n$. Після застосування операції стискування та зсуву було отримано $y_1, \dots, y_n$. Позначимо дану функцію пакетної нормалізації наступним чином:

$$BN_{\gamma, \beta}: \{x_1, \dots, x_n\} \rightarrow \{y_1, \dots, y_n\}.$$

Алгоритм пакетної нормалізації складається з таких кроків:

1. Обчислюється математичне сподівання і дисперсію пакету за такими формулами:

$$\mu_B = \frac{1}{n} \sum_{i=1}^n x_i;$$

$$\sigma_B^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_B)^2$$

2. Нормалізуються дані пакету:

$$\widehat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}},$$

де ε — константа, яка використовується для обчислювальної стійкості (налаштовується разом з іншими параметрами).

3. Застосовується операція стиску та зсуву:

$$y_i = \gamma \widehat{x}_i + \beta$$

Варто звернути увагу на те, що якщо

$$\beta = \mu_B$$

та

$$\gamma = \sqrt{\sigma_B^2 + \varepsilon},$$

то y_i дорівнює x_i , тобто $BN_{\gamma, \beta}$ є тотожнім відображенням. Використання пакетної нормалізації не може призвести до зниження точності, так як оптимізатор може лише використовувати нормалізацію як тотожне відображення.

– Dropout (виключення)

Однією з проблем, яка виникає при побудові нейронної мережі, є явище перенавчання. Метод виключення є одним найбільш поширеним засобів запобігання цій проблемі, завдяки простому алгоритму та непоганим результатам [15].

Для випадку нейронних мереж явище перенавчання виникає через випадок сумісної адаптації, тобто при поновленні ваг нейрона під час навчання метод зворотнього розповсюдження помилки, враховується діяльність інших нейронів з метою мінімізації функції похибки. Через це ваги нейронів можуть змінюватися, коригуючи похибки інших нейронів. Метод виключення дозволяє запобігти цій адаптації. Схему роботи подано на рисунку 2.8.

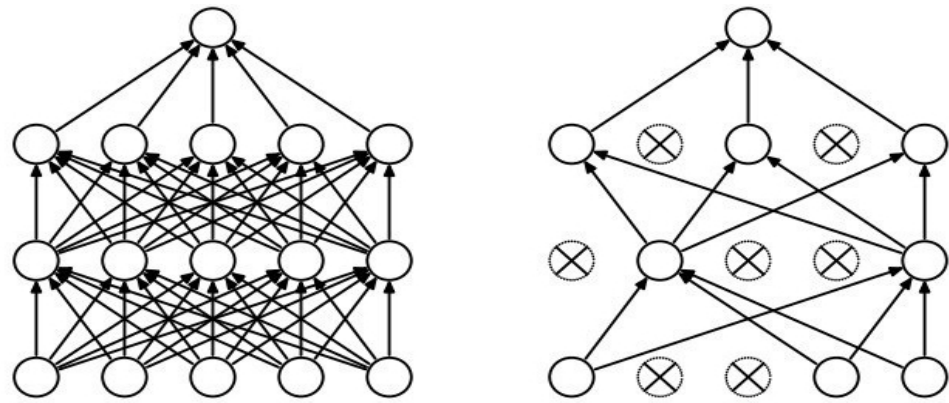


Рисунок 2.8. – Схема роботи виключення: зліва зображення нейронної мережі до застосування методу, справа – ця ж сама мережа після використання виключення ; мережа зліва використовується при тестуванні після навчання параметрів

Розглянемо шар нейронної мережі, що складається з N нейронів. Метод виключення вимикає нейрон з ймовірністю p , відповідно, залишає увімкненим з ймовірністю $q = 1 - p$, причому ймовірність вимкнення будь-якого нейрона є однаковою.

Нехай $f(x)$ – функція активації, тоді застосування дропауту для i -го нейрону записується таким чином:

$$U_i = \Theta_i f\left(\sum_{j=1}^N w_j x_j + b\right),$$

де w_j — ваги нейрону,

b — зміщення,

x_j — вхідні дані,

а ймовірність $P(\Theta_i = 0) = p$.

Ця формула використовується під час навчання моделі. Але оскільки на даному стані нейрон залишається в мережі з ймовірністю q , на етапі тестування потрібно емулювати поведінки нейронної мережі, що була використана під час навчання.

Для цього результат вихідного значення функції активації множать на коефіцієнт q :

$$U_i = q f(\sum_{j=1}^N w_j x_j + b).$$

На практиці також використовують техніку зворотнього виключення. На відміну від звичайного методу, множення на коефіцієнт відбувається під час навчання, причому цей коефіцієнт дорівнює оберненій ймовірності того, що нейрон залишиться в мережі $\frac{1}{q}$. А під час тестування значення нейрону залишається таким же, як і в методі зворотнього поширення помилки.

Тобто вихідне значення i -го нейрону на етапі навчання таке:

$$U_i = \frac{1}{q} \Theta_i f(\sum_{j=1}^N w_j x_j + b),$$

а на етапі тестування:

$$U_i = f(\sum_{j=1}^N w_j x_j + b) .$$

Зворотня модифікація виключення на практиці застосовується частіше, ніж звичайна, тому що в ній не потрібно змінювати кожного разу модель для проведення етапу тестування. Для розробки системи бралася ймовірність 30%, так як для нейронної мережі результати були найкращими.

2.4 Висновки до розділу 2

Формалізовано математичну постановку задачі оцінювання та прогнозування стану страхової компанії, яка вирішується створеною системою підтримки прийняття рішень. Обрано метрики оцінювання якості моделей класифікації.

Описано процес попередньої обробки даних та їх підготовка до навчання моделей. У подальшому було прийнято рішення стосовно проведення обчислювальних дослідів для оцінювання якості роботи алгоритмів класифікації.

Також було обрано шість методів машинного навчання з учителем для задачі класифікації, використання яких дасть можливість оцінити стан страхової компанії та спрогнозувати стан компанії у майбутньому та порівняти якість моделей.

РОЗДІЛ 3 РОЗРОБКА І ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ СТАНУ СТРАХОВИХ КОМПАНІЙ

3.1 Поняття системи підтримки прийняття рішень та його вимоги

Система підтримки прийняття рішень (СППР) – інтерактивна комп'ютерна система, що використовується з метою допомогти у прийнятті рішень в бізнес-діяльності. Такі системи дозволяють одержати важливу інформацію з вихідних даних, забезпечують проведення аналізу та надають відповідну модель з метою забезпечення вирішення поставленої задачі.

Для досягнення високої точності оцінки прогнозу, потрібно використовувати засоби, які ґрунтуються на використанні структуризації поставленої проблеми, способів попередньої обробки даних, їх візуалізації та експлораторного аналізу, моделей та оцінки їх якостей.

Але сьогоднішня робота зі значним обсягом даних потребує легкого, сучасного та ефективного інструментарію. До головних вимог до створення сучасних СППР відносять наступне:

- мова програмування, що має інтуїтивний синтаксис та використовується для різноманітних задач, та забезпечує використання ресурсів обчислювальних машин для виконання процесів, що гарантує малий час виконання запитів користувача.
- легкий та інтуїтивний інтерфейс, що забезпечує завантаження даних, їх візуалізацію, відображення обробки даних та фінальних результатів.

СППР розроблена мовою програмування Python і включає в себе модулі, які забезпечують створення вікно інтерфейсу користувача, завантаження даних та їх візуалізацію, розбиття даних на навчальну, валідаційну та тестувальну вибірку, тренування моделей класифікації та обчислення метрик якості, безпосередній прогноз.

3.2 Архітектура СППР

Архітектура СППР, що була розроблена для задачі оціювання стану страхових компаній наведена на рисунку 3.1:

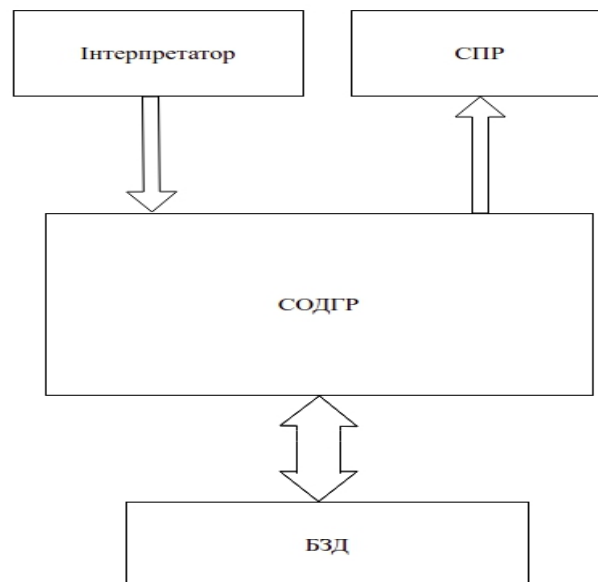


Рисунок 3.1 — Архітектура СППР

Компоненти архітектури виконують такі функції:

- Інтерпретатор відповідає за безпосереднє завантаження даних джерела інформації. В якості даних завантажуються файли формату *.XLSX.
- СОДГР або система обробки даних та генерації результатів відповідає за візуалізацію даних за допомогою графіків.
- БЗД або база даних і знань відповідає за забезпечення оцінювання стану страхової компанії. Програма відображає значення метрик якості навчання моделей класифікації.
- СПР або система прийняття рішень дозволяє зобразити результати. Програма забезпечує відображення прогнозу майбутнього стану страхової компанії та можливі причини в разі негативної оцінки стану.

Функціональна схема СППР наведена на рисунку 3.2.

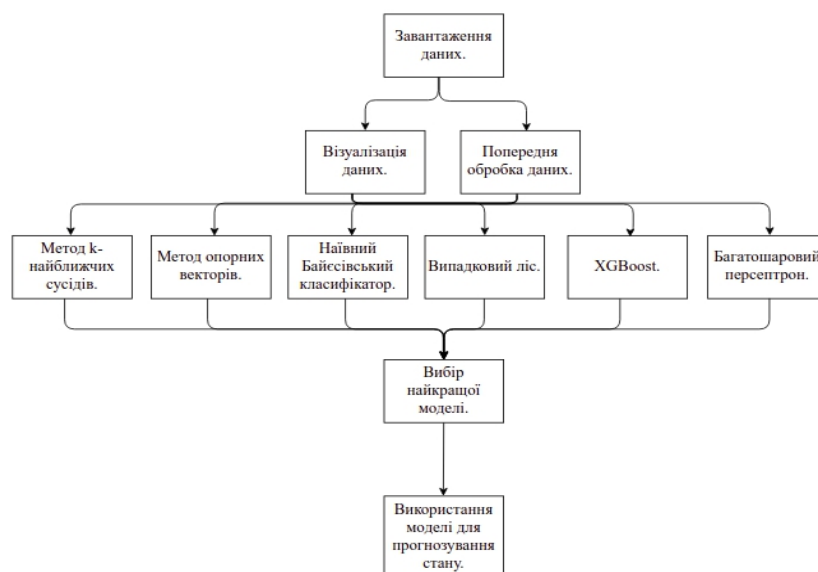


Рисунок 3.2 – Схема функціонування СППР

3.3 Вибір засобу реалізації СППР

Для реалізації СППР була використана мова програмування Python. Це одна з найбільш сучасних мов програмування, яка займає провідні позиції серед інших мов програмування, як C++, Java, тощо. Для Python характерний інтуїтивний синтаксис та динамічна типізація.

Просте імпортування модулів та бібліотек дає можливість спрощення поставленої задачі та дозволяє сфокусуватися на реалізації алгоритму в залежності від поставлених задач.

Для реалізації СППР були створені наступні модулі:

- Design.py — модуль, що створює вікно інтерфейсу користувача.
- Utils.py — модуль, що містить допоміжні функції для системи;
- Main.py — модуль, що забезпечує функціонал роботи СППР.

При розробці використовувалися такі бібліотеки:

1. Numpy.
2. Matplotlib.
3. Seaborn.
4. Pandas.
5. Scikit-learn.
6. XGBoost.
7. Keras.
8. PyQt5.

3.4 Дані для дослідження

Для дослідження роботи програми в якості даних було взято основні показники діяльності страхових компаній (у розрізі установ), відповідно до постановки Кабінету Міністрів України № 835 з сайту Національного банку України. В якості навчально-валідаційної вибірки бралися показники за 2020 рік (185 компаній), а в якості тестової вибірки — показники за перше півріччя 2021 року (166 компаній).

Вибірка складалася з таких стовпців:

1. Код установи.
2. Назва установи.
3. Активи.
4. Власний капітал.
5. Зобов'язання.
6. Страхові резерви.
7. Чисті зароблені страхові премії.
8. Премії підписані.
9. Премії, передані у перестраховуванні.
10. Страхові виплати та страхові відшкодування.
11. Виплати на оплату праці.
12. Фінансовий результат до оподаткування (прибуток).
13. Фінансовий результат до оподаткування (збиток).
14. Витрати з податку на прибуток.
15. Чистий фінансовий результат (прибуток).
16. Чистий фінансовий результат (збиток).

Для навчально-валідаційної вибірки перші дві колонки видаляються, оскільки вони не містять в собі релевантної інформації. Також у цій же вибірці створювалася колонка “Оцінка діяльності”, яка заповнювалася наступним чином. СППР обчислює рентабельність капіталу кожної компанії: відношення

чистого прибутку до власного капіталу, яке вимірюється у відсотках. Якщо цей показник був більше 10% [16], то компанії присвоювалася оцінка “ефективна”, інакше — “неефективна”.

3.5 Інструкція використання СППР для користувача та аналіз результатів

3.5.1 Інструкція використання СППР

Після запуску СППР перед користувачем з’явиться вікно, зображене на рисунку 3.3, де буде відбуватися основна взаємодія. Користувач може завантажити потрібні файли з даними по страховим компаніям; побудувати кореляційну матрицю даних та графік компаній за оцінкою діяльності; змінити структуру даних, видаливши некорельовані дані або відновивши вибірку; обрати математичну модель класифікації, попередньо встановивши розмірність валідаційної вибірки; побудувати матрицю похибки.

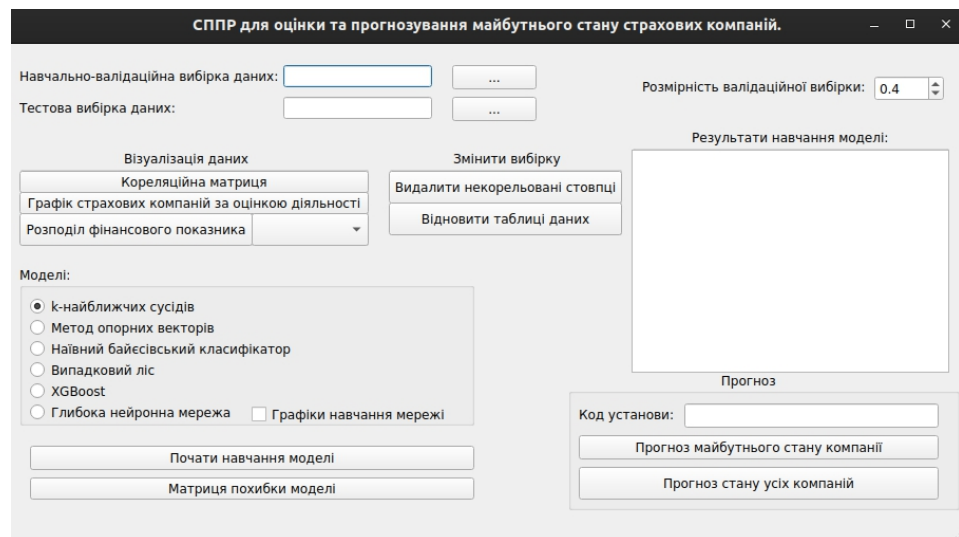


Рисунок 3.3 — Основне вікно програми

Натиснувши на кнопку “Кореляційна матриця”, користувач побачить графічне зображення матриці кореляції вибірки даних (рисунок 3.4).

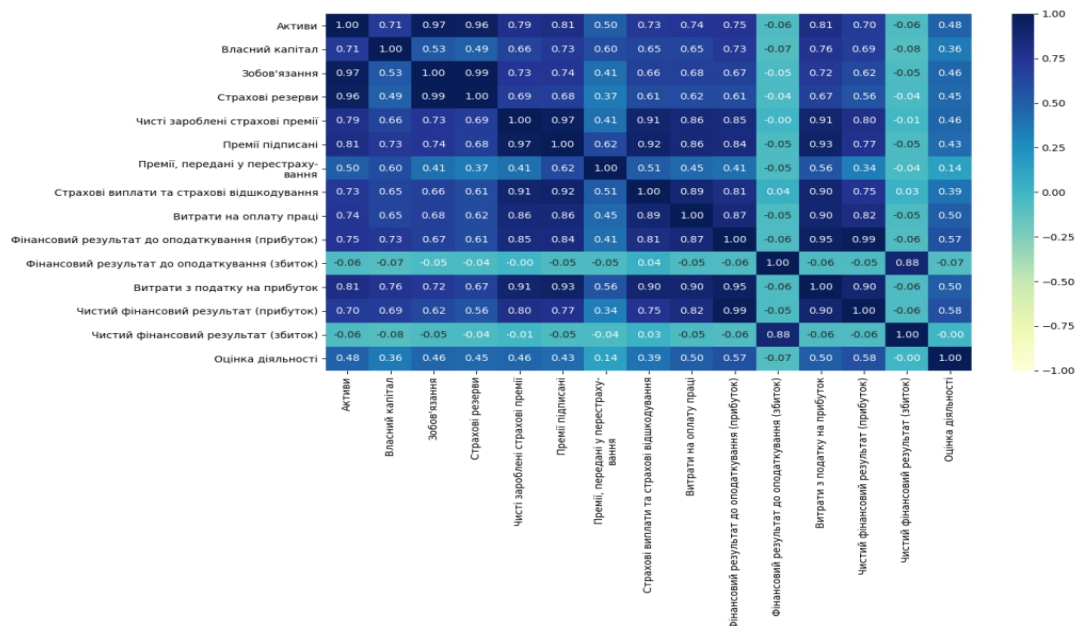


Рисунок 3.4 — Кореляційна матриця даних

Якщо користувач обере команду “Графік страхових компаній за оцінкою діяльності”, то він побачить кругову діаграму страхових компаній за оцінкою діяльності. Приклад даного графіку наведено на рисунку 3.5.

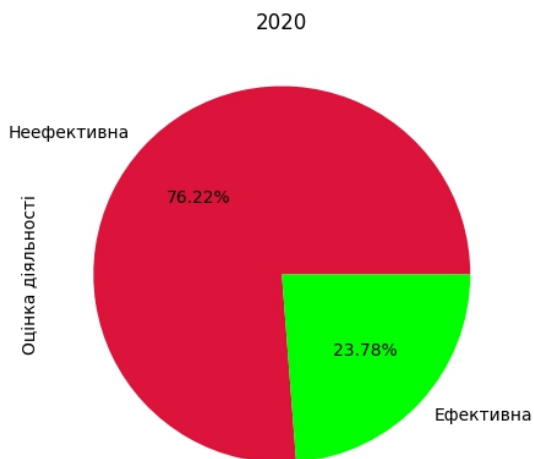


Рисунок 3.5 — Графік страхових компаній за оцінкою діяльності за 2020 рік

Можна побачити за графіком, що дані є незбалансованими, оскільки кількість страхових компаній з ефективною діяльністю складає лише 24%.

Також користувач може переглянути графік розподілу фінансового показника за оцінкою діяльності. Для цього він має обрати фінансовий показник, який його цікавить, після завантаження файлів даних та обрати команду “Розподіл фінансового показника”. Після цього система будує графік розподілу формату “ящик з вусами”. Приклад графіку наведено на рисунку 3.6.

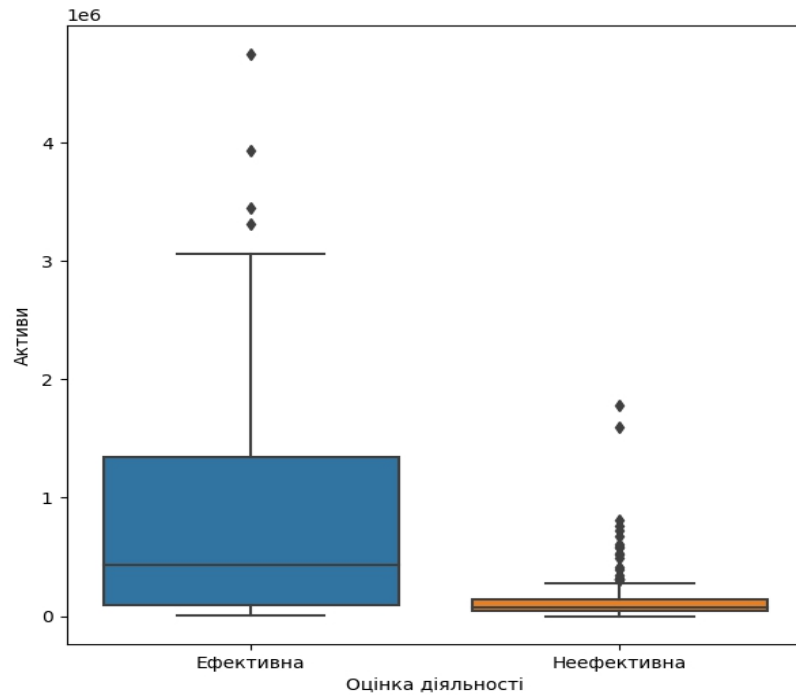


Рисунок 3.6 — Приклад графіку розподілу фінансового показника за оцінкою діяльності

Графік усіх фінансових показників за оцінкою діяльності наведено на рисунку 3.7.

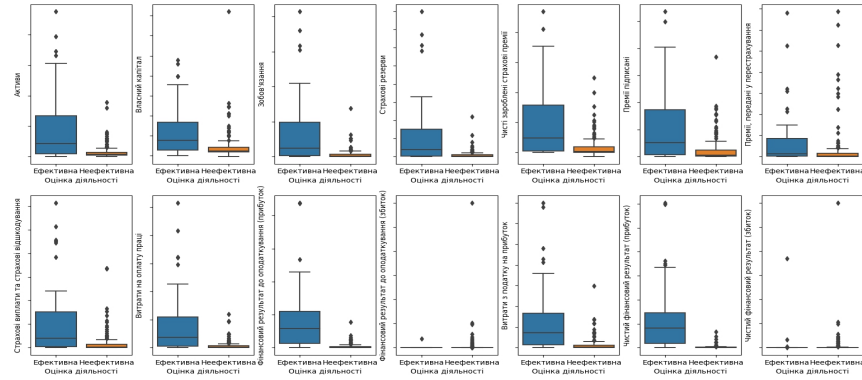


Рисунок 3.7 — Графіки розподілу всіх фінансових показників за оцінкою діяльності

У випадку, коли користувач не хоче використовувати некорельовані показники при побудові моделі, то він може обрати команду “Видалити некорельовані стовпці”. Внаслідок цього програма видалить ті стовпці даних, які мають слабку кореляцію. Приклад роботи наведено на рисунку 3.8.

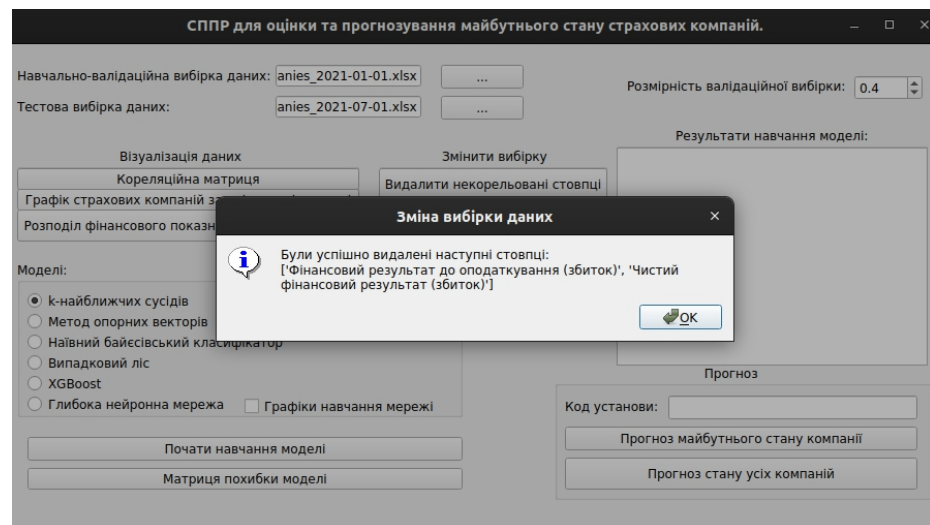


Рисунок 3.8 — Приклад результату видалення некорельованих стовпців

Якщо користувач хоче відновити стартову вибірку та працювати з нею, то він може натиснути на кнопку “Відновити таблиці даних”. В результаті навчально-валідаційна та тестова вибірка повернуться до початкового стану. Результат роботи команди наведено на рисунку 3.9.

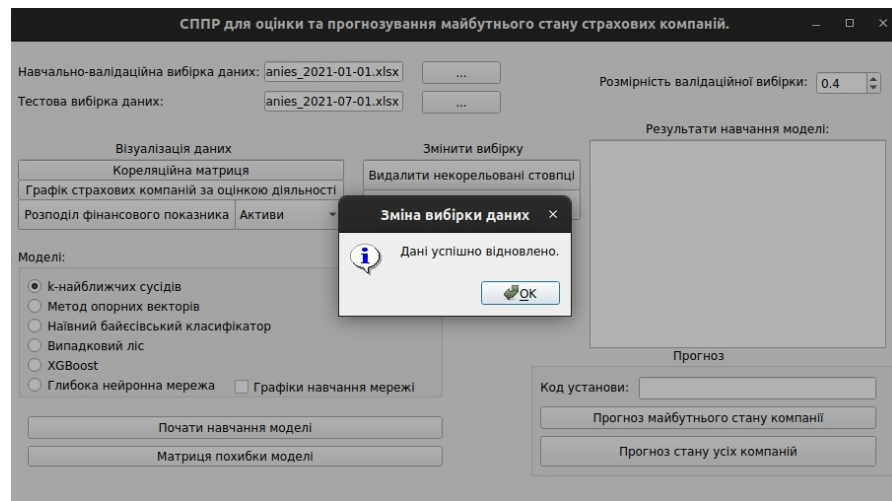


Рисунок 3.9 — Приклад результату відновлення таблиці даних

Також користувач може задати розмірність валідаційної вибірки від 0.1 до 0.9, що буде використовуватися при розбитті даних. Обравши потрібну модель та натиснувши кнопку “Почати навчання моделі”, користувач побачить значення метрик якості класифікації. Приклад результату наведено на рисунку 3.10.

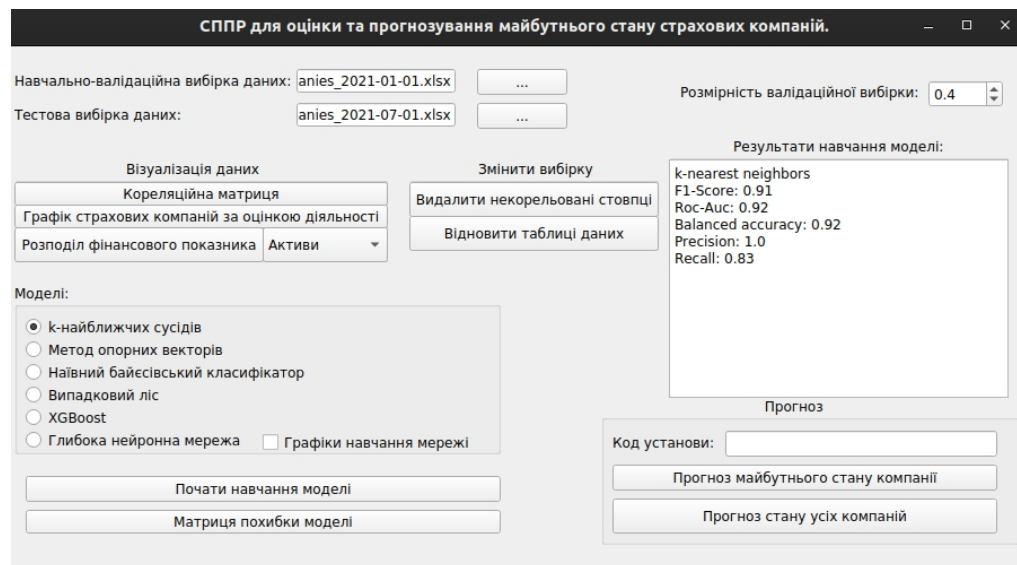


Рисунок 3.10 — Результат навчання моделі

Якщо користувач обере модель “Багатошаровий персептрон”, то йому потрібно зачекати, оскільки навчання мережі займає деякий час. Після завершення навчання СППР сповістить про те, що модель готова (рисунок 3.11).

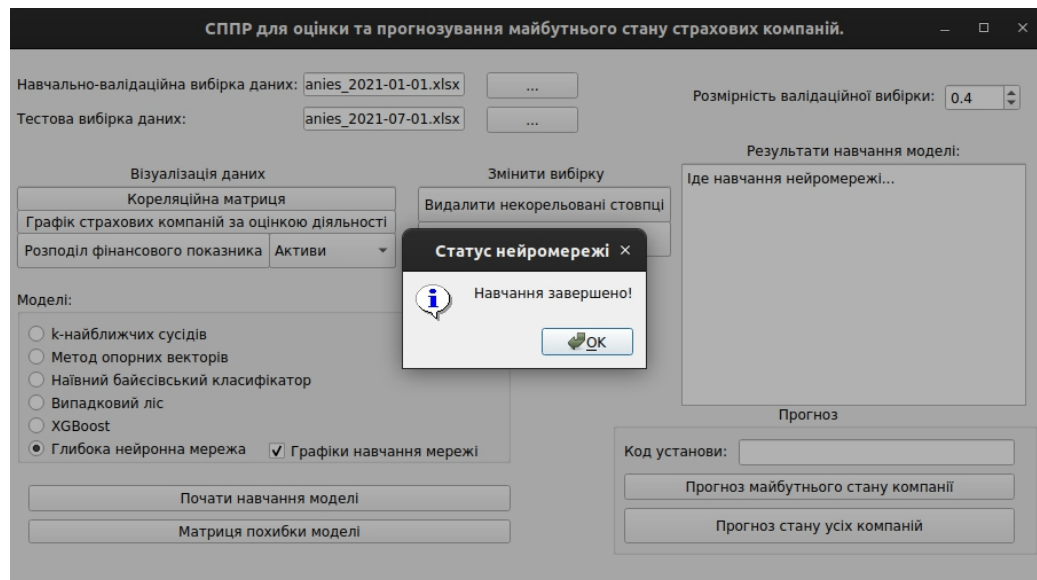


Рисунок 3.11 — Сповіщення системи про завершення навчання нейронної мережі

Якщо користувач при виборі багатошарового перцептрону також увімкнув прапорець “Графіки навчання мережі”, то система також продемонструє графіки втрат та точності для навчальної та валідаційної вибірки (рисунок 3.12 та рисунок 3.13).

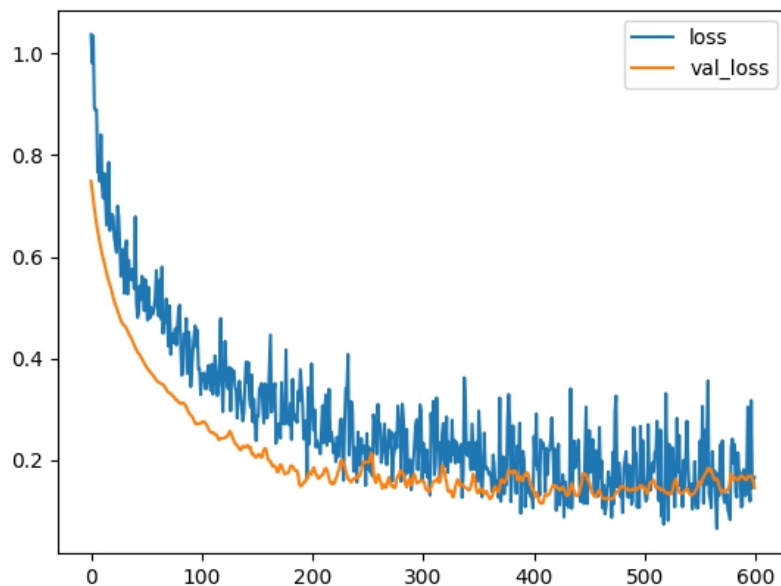


Рисунок 3.12 — Приклад графіку функції втрат для тренувальної та валідаційної вибірки

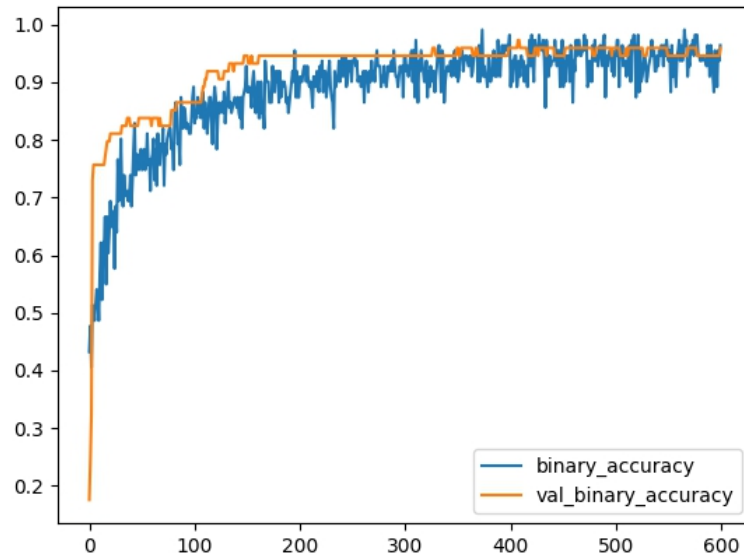


Рисунок 3.13 — Приклад графіку точності для тренувальної та валідаційної вибірки

Користувач також може подивитися матрицю похибки обраної ним моделі. Для цього він має натиснути на кнопку “Матриця похибки моделі”. Після цього система відображає графічне зображення матриці похибки моделі. Приклад зображення наведено на рисунку 3.14.

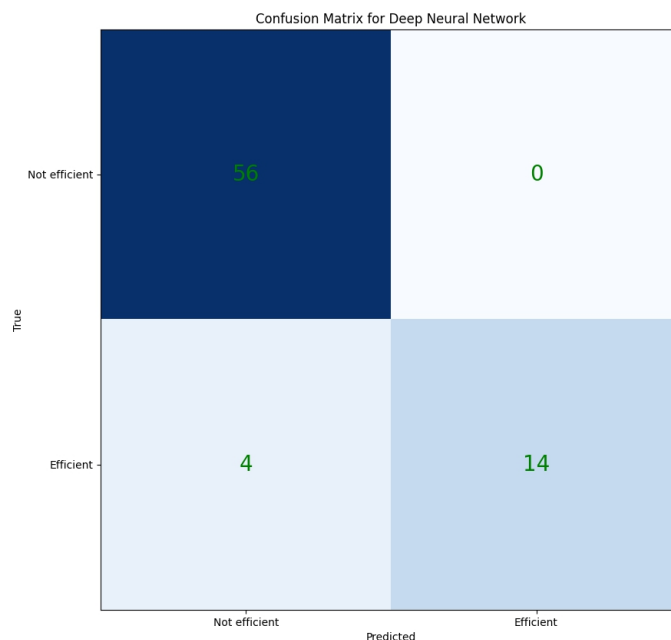


Рисунок 3.14 – Приклад матриці похибки

Після того, як користувач обрав найкращу модель за метриками якості

класифікації, він може дізнатися щодо прогнозу стану будь-якої страхової компанії. Для цього він має ввести код установи, що складається з 8 цифр. Якщо користувач ввів правильно код та якщо страхова компанія існує в тестовій вибірці, то система надає результат прогнозу (рисунк 3.15).

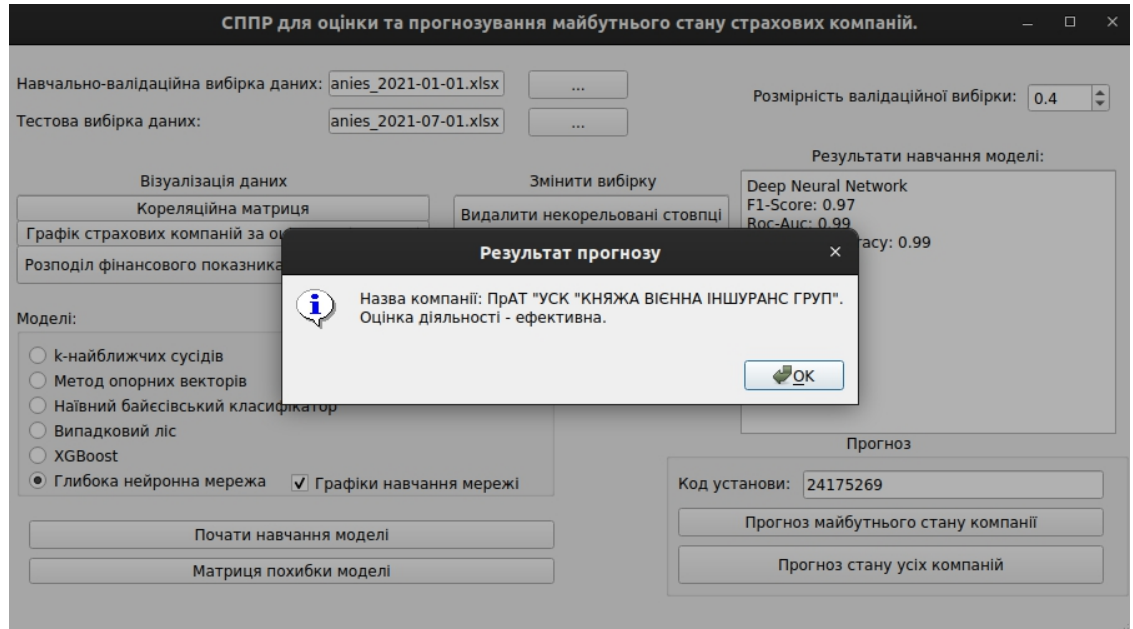


Рисунок 3.15 – Приклад позитивного прогнозу стану страхової компанії

Якщо прогноз страхової компанії є негативним, то система також вказує на те, на що варто звернути увагу (рисунк 3.16).

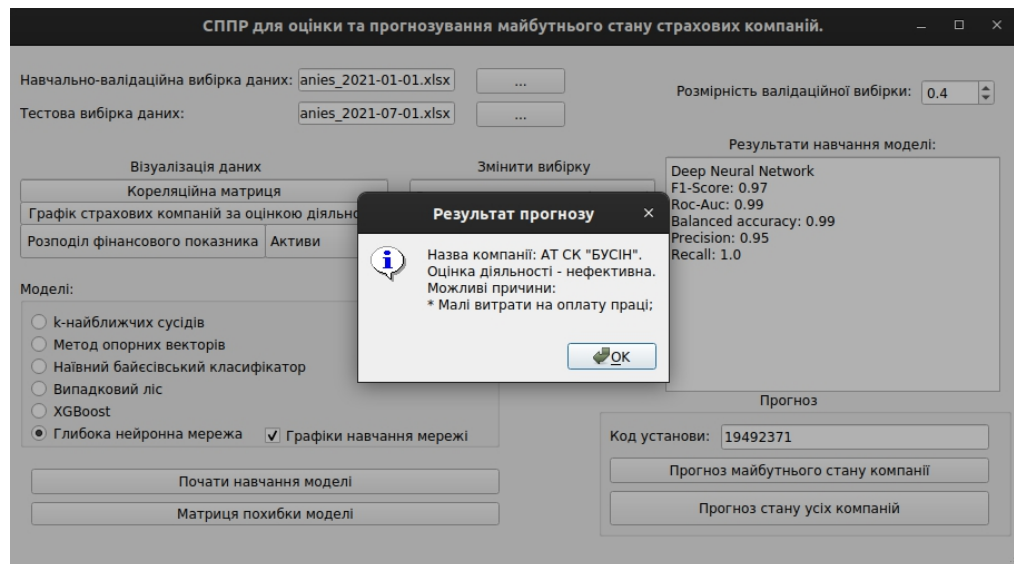


Рисунок 3.16 — Приклад негативного прогнозу стану страхової компанії

При пошуку причин неефективної діяльності система перевіряє показники активів, витрат на оплату праці, страхові резерви та премії. Якщо хоча б один показник менше середнього значення, то система включає його у список можливих причин. У разі їх відсутності система надає пораду щодо повного перегляду менеджменту компанії (рисунок 3.17).

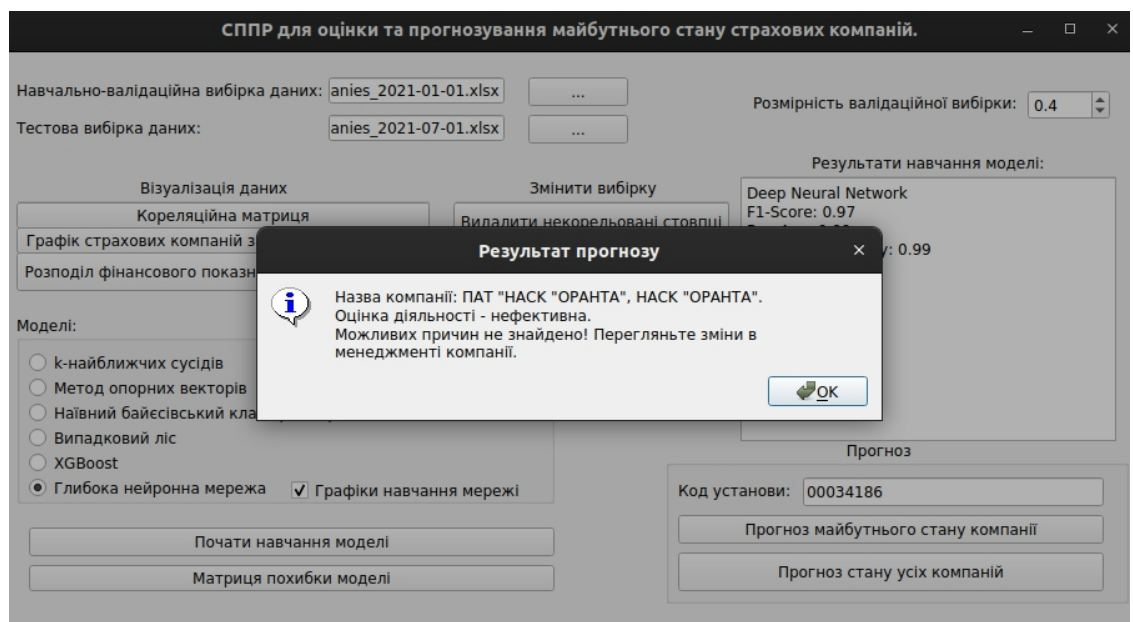


Рисунок 3.17 — Приклад негативного прогнозу стану страхової компанії з відсутністю причин негативної оцінки

Якщо користувача цікавить оцінка прогнозу абсолютно всіх страхових

компаній у тестовій вибірці, то він має обрати команду “Прогноз стану усіх компаній”. Після цього система спрогнозує майбутню оцінку всіх страхових компаній і збереже результати прогнозу обраної користувачем моделі у текстовий файл формату *.CSV. Програма сповістить про успішність прогнозу (рисунок 3.18) та виведе графік страхових компаній протягом прогнозного періоду за оцінкою діяльності (рисунок 3.19).

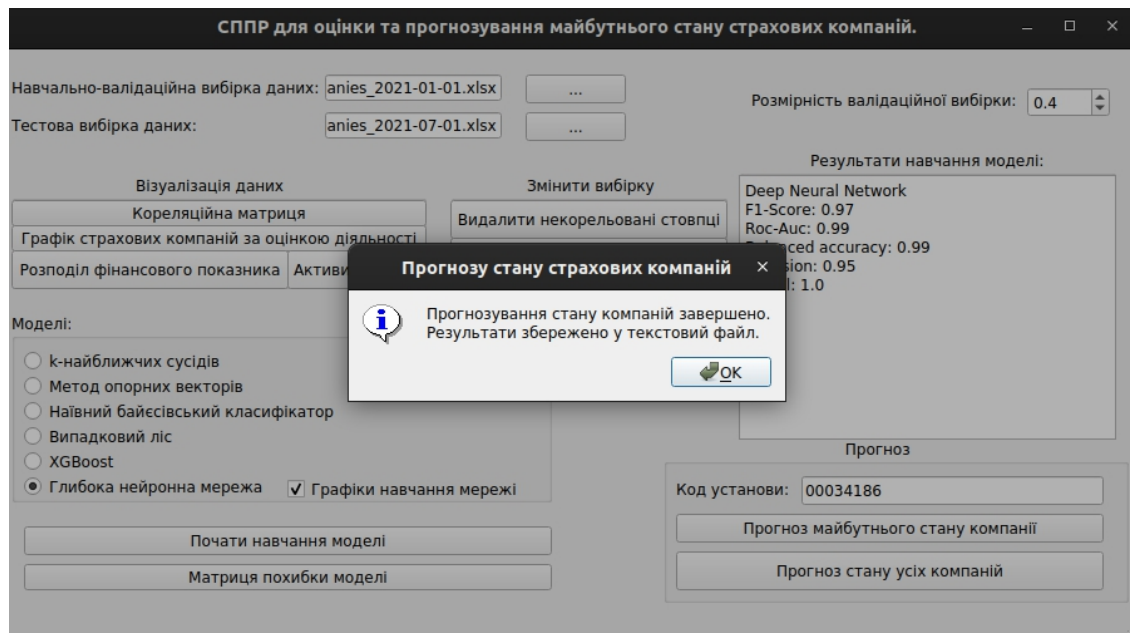


Рисунок 3.18 — Сповіщення системи про завершення прогнозу стану страхових компаній

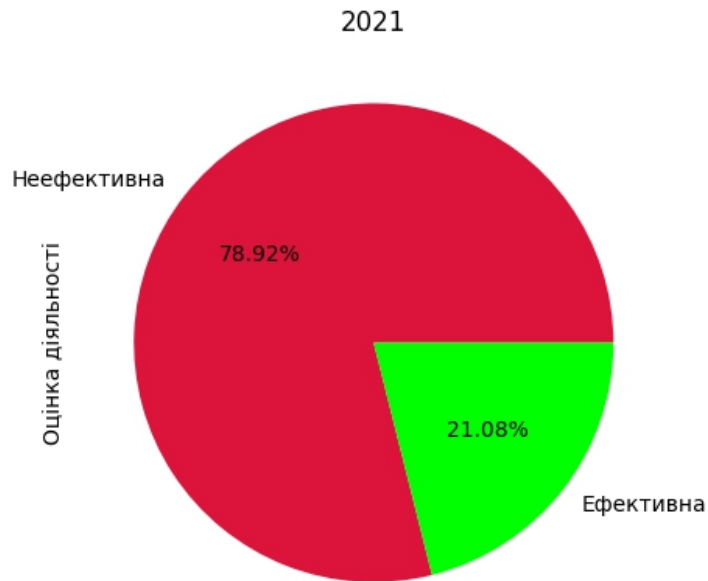


Рисунок 3.19 — Графік страхових компаній за прогнозною оцінкою діяльності за 2021 рік

3.5.2 Обчислювальні досліді та аналіз результатів

Для проведення обчислювальних дослідів навчально-валідаційна вибірка розбивалася за співвідношенням 60:40. Тренувальна вибірка мала розмірність 111, валідаційна — 74. Розглядався випадок повної вибірки, що складалася з 14 змінних, та скороченої, що складалася з 12 змінних (випадок видалення некорельованих стовпців).

Результати навчання моделей наведено в таблиці 3.1 та 3.2. Вибір найкращої моделі виконувався за показниками F1-міри та збалансованої точності.

Таблиця 3.1 — Результати навчання моделей класифікації для повного набору даних

Модель	F1-Score	Roc-Auc	Balanced Accuracy	Precision	Recall
k-найближчих сусідів	0.91	0.92	0.92	1.0	0.83
Метод опорних векторів	0.76	0.81	0.94	1.0	0.61
Наївний байєсівський класифікатор	0.69	0.78	0.83	0.79	0.61
Випадковий ліс	0.94	0.94	0.94	1.0	0.89
XGBoost	0.91	0.94	0.94	0.94	0.89
Багатошаровий перцептрон	0.91	0.92	0.92	1.0	0.83

Таблиця 3.2 – Результати навчання моделей класифікації для скороченого набору даних

Модель	F1-Score	Roc-Auc	Balanced Accuracy	Precision	Recall
k-найближчих сусідів	0.91	0.94	0.94	0.94	0.89
Метод опорних векторів	0.86	0.9	0.9	0.88	0.83
Наївний байєсівський класифікатор	0.77	0.86	0.86	0.71	0.83

Продовження таблиці 3.2

Випадковий ліс	0.86	0.95	0.95	0.75	1.0
XGBoost	0.88	0.96	0.96	0.78	1.0
Багатошаровий перцептрон	0.95	0.98	0.99	0.9	1.0

За результатами таблиць можна побачити, що модель випадкового лісу показала найкращі результати серед усіх моделей. При скороченому наборі даних метрики якості усіх моделей покращилися. Водночас багатошаровий перцептрон продемонстрував майже ідеальні результати — його точність становила 99%. Графік втрат та точності для нейронної мережі подано на рисунку 3.20 та 3.21.

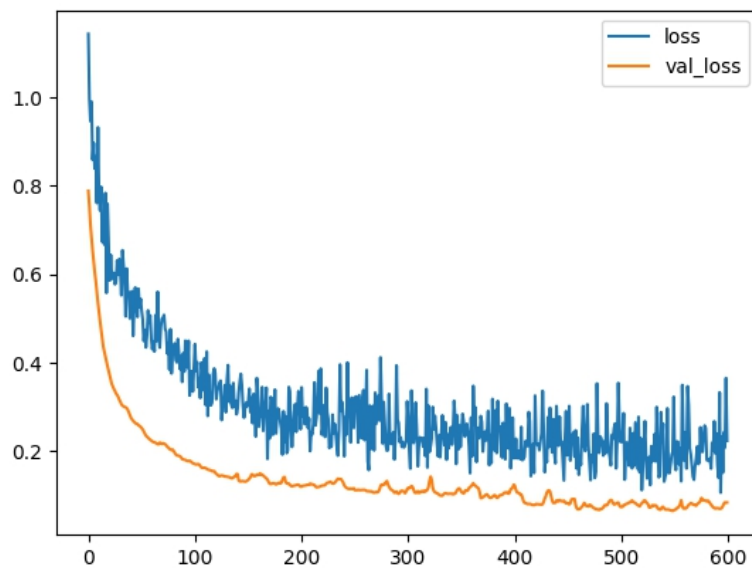


Рис 3.20 — Графік функції втрат для навчальної та валідаційної вибірки для скороченого набору даних

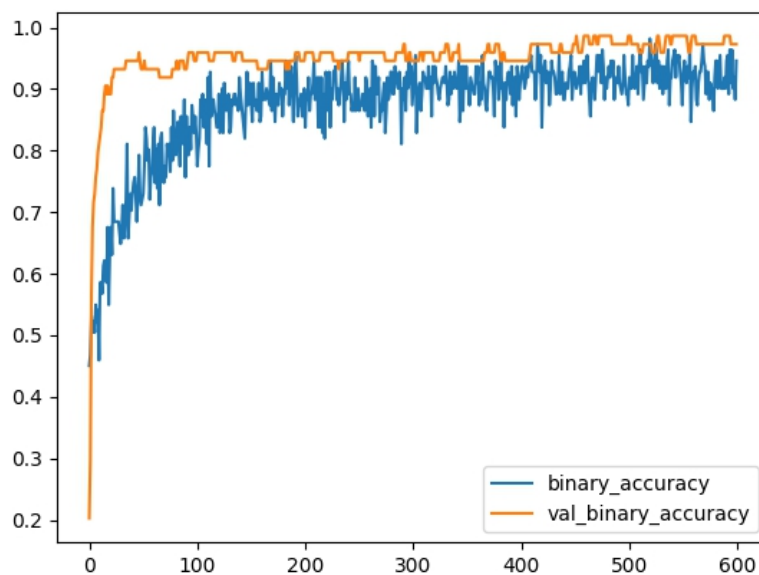


Рис 3.21 — Графік точності для навчальної та валідаційної вибірки для скороченого набору даних

Також наведено матриці похибок для моделей класифікації для випадків повного та скороченого обсягу даних (рисунок 3.22 та рисунок 3.23).

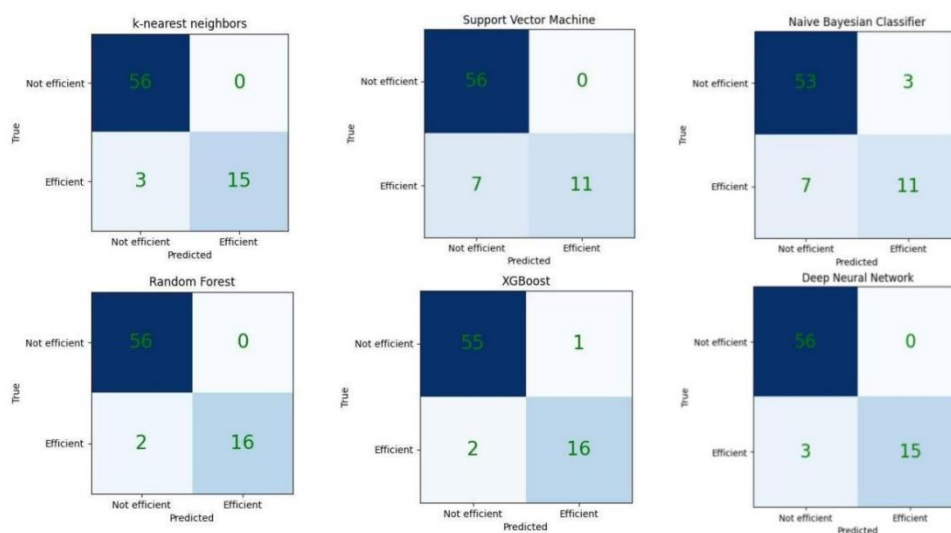


Рисунок 3.22 – Матриці похибок моделі при повному наборі даних

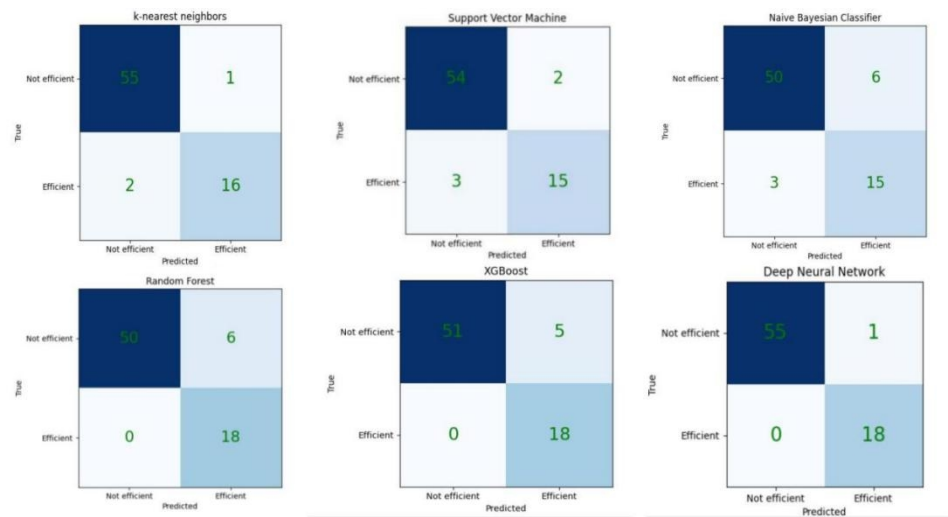


Рисунок 3.23 – Матриці похибок моделі при скороченому наборі даних

За матрицями наочно видно, що при повному наборі даних, найкращі результати має модель випадкового лісу. При скороченому наборі даних для моделі k-найближчих сусідів, опорних векторів та наївного Байєсівського класифікатора зменшилася кількість помилково-негативних оцінок, але при цьому збільшилася кількість помилково-позитивних оцінок. При цьому модель випадкового лісу, XGBoost та нейронної мережі взагалі не містили помилково-негативних оцінок. Для скороченого набору даних багатoshаровий перцептрон показав майже безпомилкові результати. Тому при виборі найкращої моделі при оцінюванні та прогнозуванні страхових компаній користувач має вирішити чи включати показники збитків під час навчання моделей чи ні.

3.6 Висновки до розділу 3

В розділі розглянуто розробку та використання авторської системи підтримки прийняття рішень для оцінювання та прогнозування стану страхових компаній.

Розроблена архітектура та схема функціонування системи підтримки прийняття рішень. Обрано інструментарій для реалізації інтерфейсу користувача та математичних моделей класифікації, що були обрані в попередньому розділі.

Наведена інструкція для експлуатації СППР для користувача та проведено обчислювальний дослід для порівняння ефективності моделей класифікації. Відносно повної вибірки даних найкращі результати продемонструвала модель випадкового лісу. Моделі k -найближчих сусідів, методу опорних векторів та наївного Байєсівського класифікатора продемонстрували набагато кращі результати, навчаючись за скороченою вибіркою. Водночас випадковий ліс показав гірші результати. Серед усіх моделей за навчанням на скороченій вибірці ідеальні результати продемонстрував багатошаровий персептрон – його точність становила 99%. У випадку вибору найкращої моделі при оцінюванні та прогнозуванні страхових компаній користувач має вирішити чи включати показники збитків під час навчання моделей чи ні.

РОЗДІЛ 4. СТАРТАП АНАЛІЗ ПРОЕКТУ

4.1 Вступ та постановка задачі стартап проекту

На сьогоднішній день є актуальним питання у страхуванні життя та власності в Україні. Але в Україні нема досконалої системи оцінювання, що надавала би повну інформацію про страхову компанію та рішення в залежності від платоспроможності. Робота дозволяє створити засіб оцінювання будь-якої компанії, взявши за основу описану систему підтримки прийняття рішень.

Для розробки стартап проекту та виведення його на ринок необхідно провести детальне дослідження, яке передбачає виконання наведених нижче чотирьох кроків.

Здійснити маркетинговий аналіз стартап-проекту, в рамках якого:

- 1) розробити опис ідеї проекту, визначити основні напрямки використання товару чи послуги та сформулювати основні відмінності від товарів/послуг конкурентів;
- 2) проаналізувати ринкові можливості для його реалізації;
- 3) розробити стратегію виведення товару на ринок, базуючись на аналізі ринкового середовища.

Організація стартап-проекту, яка включає такі кроки:

- 1) скласти календарний план реалізації та запуску стартап-проекту;
- 2) визначити плановий обсяг виробництва потенційного товару та на його основі розрахувати потребу у матеріальних ресурсах і персоналі;
- 3) розрахувати витрати, необхідні для реалізації проекту, та витрати на запуск проекту.

Виконати фінансово-економічний аналіз та оцінити ризики стартап-проекту, в межах якого:

- 1) визначити обсяг інвестиційних витрат;
- 2) розрахувати основні фінансово-економічні показники проекту (собівартість, ціну продукту/послуги, податковий збір та чистий прибуток) та

визначити показники інвестиційної привабливості проекту (рентабельність продажів, період окупності проекту);

3) визначити основні ризики проекту та способи для їх запобігання.

Розробити заходи з комерціалізації проекту. Цей етап націлений на пошук фінансування проекту та просування інвестиційної пропозиції. Для його досягнення необхідно:

1) визначити цільову групу інвесторів та описати їх бізнес інтереси;

2) скласти інвестиційну пропозицію: стислий опис проекту для ознайомлення інвестора із стартап-проектом;

3) визначити основні канали та заходи для просування оферти інвесторам.

В розділах нижче наведено результати проведених кроків.

4.2 Карта стартап проекту

Стартап проект полягає у створенні системи підтримки прийняття рішень для оцінювання та прогнозування стану страхових компаній. Дана програма аналізує масиви даних, після чого надає оцінку діяльності страхової компанії та можливі причини у випадку незадовільної оцінки (обґрунтований історичними даними). На основі результатів, отриманих при практичному використанні запропонованої системи підтримки прийняття рішень пропонується створити центр моніторингу стану страхових компаній України.

В таблиці 4.1 представлено основна інформація проекту, розкрито основну ідею та рішення поставленої проблеми.

Таблиця 4.1 – Інформаційна карта проекту

1. Назва проекту	Система оцінювання та прогнозування стану страхових компаній на основі методів машинного навчання
2. Автори проекту	Панібратов Роман
3. Коротка анотація	Система дозволяє оцінювати стан страхової компанії за її фінансово-економічними показниками .
4. Термін реалізації проекту	6 місяців
5. Необхідні ресурси	Обладнання – комп'ютер. Програмне забезпечення, операційна система, антивірусні програми. Електрика, газ, водопостачання, Інтернет. Фінансові ресурси – заробітна плата працівниками на 6 місяців роботи, гроші на оплату комунальних послуг, оренди, реклами тощо. Приміщення з усім необхідним обладнанням для комунікації.
6. Опис проблеми, які вирішує проект	Дана система дозволяє забезпечити прийняття рішень, пов'язаних зі страхуванням. Система не є суб'єктивною та забезпечує спосіб прийняття рішень, який близький до страховиків. Система працює безпосередньо з акаунтом користувача і не має доступу до фінансів, що гарантує захист даних користувача.
7. Головні цілі та завдання проекту	Основна мета проекту – доведення ефективності обраного методу дослідження. Додаткові завдання – новий досвід, створення структурної системи, робота з реальними даними та створення комерційно успішного продукту.
8. Очікувані результати	Повністю автоматична система підтримки прийняття рішень, що може працювати з даними будь-якої розмірності.

4.3 Технологічний аудит ідеї проекту

Далі було виділено основну ідею стартапу та зведено її в наступній таблиці 4.2.

Таблиця 4.2 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Створена структурована система дозволяє запропонувати правильне рішення в сфері страхування та вирішити питання платоспроможності.	1. Робота в якості страховика	Система дозволяє оцінити стан конкретної компанії-страховика і вказує на можливі причини у випадку неефективної оцінки.
	2. Робота в якості страхувальника	Система дозволяє оцінити стан усіх страхових компаній. На основі цих результатів страхувальник може прийняти рішення щодо вибору компанії для страхування життя чи власності.

Далі було визначено конкурентів на ринку і зроблено порівняльний аналіз програмних продуктів конкурентів, виявлено їх переваги та недоліки. Крім того, було наведено перелік переваг над існуючими програмними рішеннями (таблиця 4.3, 4.4).

Таблиця 4.3 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні товари/концепції конкурентів)				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Sofvie	Riskturn	Optial SmartStart			
1	Точність прогнозування	Застосування моделей, що безпосередньо оцінюють ймовірність платоспроможності компанії	Власний алгоритм	Власний алгоритм	Власний алгоритм			+

Продовження таблиці 4.3.

2.	Процедура оцінювання діяльності компанії	Процедура проста – лише використання фінансово-економічних показників компанії	Проєктура складна	Спрощена процедура оцінювання	Проєктура складна			+
3.	Ризики невірної оцінки	Є можливість заявленої неінформативних показників	Невідомо (комерційна таємниця, але вони точно існують)	Невідомо (комерційна таємниця, але вони точно існують)	Невідомо (комерційна таємниця, але вони точно існують)		+	

Кінець таблиці 4.3.

4.	Доступність/Зручність	Інтерфейс, що не вимагає спеціальних знань	Власний інтерфейс	Власний інтерфейс	Власний інтерфейс		+	
----	-----------------------	--	-------------------	-------------------	-------------------	--	---	--

Таблиця 4.4 – Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
1	Створення структурованої системи оцінювання та прогнозування стану страхової компанії з використанням методів машинного навчання	Використання мови програмування C++	Потрібні доопрацювання	Не доступні
2		Використання мови програмування Java	Потрібні доопрацювання	Доступні
3		Використання мови програмування Python	Наявна	Доступні
Обрана технологія реалізації ідеї проекту: мова програмування Python				

4.4 Аналіз ринкових можливостей запуску стартап-проекту

Характеристика потенційного ринку ринку майбутнього стартап проекту наведено в таблиці 4.5.

Таблиця 4.5 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	0
2	Загальний обсяг продаж	2 млн \$
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає
6	Середня норма рентабельності в галузі (або по ринку), %	10%

Після аналізу ринку приходимо до висновку, що він є сприятливим для створення програмного продукту, так як динаміка ринку є позитивною та відсутні конкуренти.

Наступним кроком потрібно охарактеризувати головні групи потенційних користувачів продукту та скласти опис вимог для кожної групи (таблиця 4.6).

Таблиця 4.6 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Необхідність у страхуванні життя або власності	Громадяни України в віці від 18 до 50 років	Малий об'єм початкового капіталу	Легкість використання та відстеження
2	Збереження бюджету установи та його примноження	Акціонерне, повне, командитне товариство або товариство з додатковою відповідальністю	Цікавить відсутність ризику.	Високі вимоги до керування капіталом.

Далі потрібно провести аналіз можливих загроз, що можуть виникнути під час виведення продукту на споживацький ринок і перешкодити успішному запуску проекту. Результати наведено в таблиці 4.7.

Таблиця 4.7. – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренція	Дуже малий рівень виходу на ринок, можливий момент виходу нових компаній протягом 5 наступних років.	Прискорити вихід нових програмних продуктів
2	Збут	Ускладнення збуту у зв'язку з недовірою користувачів	Розміщення додаткової реклами в інтернеті, розширена наукова база.

Крім того потрібно проаналізувати фактори, що мають позитивний вплив на запуск проекту. Результати наведено у таблиці 4.8.

Таблиця 4.8 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Відсутність конкуренції	Лідуюча позиція	Розширення діапазону можливостей продукту, фокус на якості
2	Створення позитивного іміджу компанії	Надання послуг на найвищому рівні та забезпечення задоволення клієнтів.	Створення ефективної рекламної кампанії, техпідтримка існуючих клієнтів.

Наступним кроком є характеристика конкурентного середовища, а саме визначення типу та рівня конкуренції. Результати представлені у таблиці 4.9.

Таблиця 4.9 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції: - Досконала конкуренція	Багато систем/команд аналітиків	Розробити продукт, який легко розпізнається та має якість, що дозволяє вирізняти нас від конкурентів
2. За рівнем конкурентної боротьби: - міжнародний	На ринку присутні системи, що розроблені за кордоном.	Розширення аудиторії та списку мов, які підтримуються системою.
3. За галузевою ознакою: - внутрішньогалузева	—	—
4. Конкуренція за видами товарів: - товарно-родова	Конкуренція між прогнозами інформаційних систем та команд аналітиків.	Збільшення точності та швидкості відправки результатів оцінки.
5. За характером конкурентних переваг: - нецінова	Різні засоби прогнозування дають різну точність	Розробка більш точних алгоритмів
6. За інтенсивністю: - марочна	Бренд, що легко впізнається, дає великі переваги	Звернення великої уваги на розвиток бренду

Далі потрібно виконати детальний аналіз конкуренції за моделю 5 сил конкуренції Майкла Портера, яка застосовується для розуміння структури галузі, аналізу її привабливості з точки хору отримання прибутку, оцінки конкуренції і розробки стратегії бізнесу. Результати аналізу представлені в таблиці 4.10.

Таблиця 4.10 – Аналіз конкуренції в галузі за М. Портером.

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Інші структуровані системи	Гнучкі ціни, розмір капіталовкладень	Фактори сили постачальників	Контроль якості, система інформації	Ціна, лояльність споживачів
Висновки:	Інтенсивна конкуренція можлива в майбутньому	Є як можливість входження на ринок, так і нові потенційні конкуренти	Постачальники відсутні	Клієнти не диктують умови роботи на ринку	Товарозамінники відсутні

Результати аналізу конкурентного середовища свідчать про те, що на ринку сприятлива ситуація для створення і запуску даного стартап-проекту. Ґрунтуючись на проведеному аналізі конкуренції (таблиця 4.10), а також враховуючи характеристики ідеї стартап-проекту (таблиця 4.5), характеристики потенційних клієнтів і їх вимоги до продукту (таблиця 4.6) та фактори ринкового середовища (таблиці 4.7 та 4.8) було сформульовано та обґрунтовано перелік факторів конкурентоспроможності. Результати аналізу подано в таблиці 4.11.

Таблиця 4.11 – Обґрунтування факторів конкурентоспроможності

Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
Низька конкуренція	В нашій країні при розробці стартапу не було знайдено конкурентів
Доступність програмного продукту	Розроблений продукт є загальнодоступним і кросплатформним. Для доступу потрібна наявність файлу даних страхових компаній.
Зручність використання	Інтуїтивно-зрозумілий інтерфейс, який не вимагає попередніх знань.

Після проведення аналізу можна виділити сильні та слабкі (які потребують вдосконалення) сторони продукту (таблиця 4.12).

Таблиця 4.12 – Порівняльний аналіз сильних та слабких сторін системи

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів						
			-3	-2	-1	0	+1	+2	+3
1	Низька конкуренція	20				+			
2	Доступність програмного продукту	20				+			
3	Зручність використання	19			+				

Кінцевим кроком аналізу ринкових можливостей для запуску проекту є складання SWOT-аналізу. Він дає можливість надавати оцінку можливостям та загрозам бізнесу, а також сильним і слабким сторонам продукту. Результати відображено в таблиці 4.13.

Таблиця 4.13 – SWOT-аналіз стартап-проекту

Сильні сторони: відсутність конкурентів; дружній інтерфейс; не потрібні спеціальні знання та навички для використання.	Слабкі сторони: відсутність налагодженої клієнтської бази.
Можливості: розширення списку мов для аналізу; додавання нових предметних областей; інтеграція з іншими програмними системами.	Загрози: поява конкурентів; збут.

За результатами SWOT-аналізу спроектовано альтернативну ризикову поведінку для інтеграції стартап-проекту на ринок та приблизний час реалізації системного комплексу, з урахуванням потенційних проектів, що можуть бути виведені на ринок (таблиця 4.14).

Таблиця 4.14 – Альтернативи ринкового впровадження стартап проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Швидкий вихід на ринок із «сирим» продуктом, можливі проблеми з точністю прогнозу та універсальністю.	5%	2 місяці
2	Поступовий вихід з готовим продуктом. Висока якість та кокурентоспромо жна ціна.	30%	4 місяці

За детальним аналізом ринкового та конкурентного середовища, факторів загроз та можливостей, сильних та слабких сторін продукту ми приходимо до висновку, що на ринку склалися сприятливі умови для впровадження товару і даний товар відповідає вимогам користувачів.

4.5 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії перш за все передбачає визначення стратегії охоплення ринку. Для цього було охарактеризовано цільові групи потенційних клієнтів (таблиця 4.15).

Таблиця 4.15 – Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Громадяни, що хочуть застархувати своє життя або свою власність	Низька готовність	5%	Низька	Середня
2	Страхові компанії	Висока	20%	Низька	Висока
3	Великі підприємства	Висока	35%	Висока	Висока
Які цільові групи обрано: 1,2					

Стратегія охоплення ринку було обрано недиференційований маркетинг — компанія фокусує зусилля одночасно на всіх сегментах споживачів. Для роботи в обраних сегментах ринку сформовано базову стратегію розвитку (таблиці 4.16, 4.17, 4.18).

Таблиця 4.16 – Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1	1 та 2	Стратегія недиференційованого маркетингу	Висока універсальність, багатогалузевість, висока ціна та якість.	Масовий маркетинг.

Таблиця 4.17 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристик товару конкурента, і які?	Стратегія конкурентної поведінки*
Так	Так	Ні	Стратегія виклику лідера

Таблиця 4.18 – Визначення стратегії позиціонування

Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
Легкість у використанні, висока точність аналізу	Позиціонування як продукту, що використовує передові технології для аналізу	Якість прогнозу, універсальність	По іміджу. Позиціонування на низькій ціні Легкість у використанні

4.6 Висновки до розділу 4

В даному розділі був проведений перший етап розроблення стартап проекту — виконано маркетинговий аналіз стартап проекту. За результатами можна сказати, що є можливість ринкової комерціалізації проекту, так як на ринку немає конкурентів і рентабельність роботи є високою. Говорячи про потенційну групу клієнтів: громадян, малий бізнес, та інноваційність технології, виявлено перспективи впровадження даного програмного продукту у вигляді центру моніторингу стану страхових компаній.

Звідси можна зробити висновок, що роботу можна розглядати як стартап, що орієнтується на стратегії блакитних океанів, тобто не зайнятих ринків. На це вказує відсутність конкурентів та велика потенційна аудиторія. У зв'язку з цим вважаю необхідним продовжувати досліджувати страховий ринок з позиції платоспроможності компаній.

ВИСНОВКИ

У даній магістерській дисертації розглянуто поняття ризиків у страхуванні; подано відомі методи оцінювання ризиків, вказано їх переваги та недоліки. Також був проведений аналіз нинішнього стану страхового ринку України та його проблеми. Описано математичну постановку задачі оцінювання та прогнозування стану страхової компанії.

Описано моделі класифікації, які були застосовані у роботі. Створена авторська система підтримки прийняття рішень, що дозволяє оцінювати та прогнозувати стан страхових компаній за фінансово-економічними показниками за допомогою методів машинного навчання. Крім того, були використані методи попередньої обробки даних, для перевірки вибірки на корельованість та інформативність щодо оцінки діяльності страхової компанії. Після проведення обчислювальних досліджень була оцінена ефективність та точність роботи моделей.

За результатами дослідження можна зробити висновок, що у випадку повної вибірки модель випадкового лісу показала найкращі результати. Водночас багат шаровий персептрон показав майже ідеальні результати, навчаючись за скороченою вибіркою. Тому при виборі найкращої моделі при оцінюванні та прогнозуванні страхових компаній потрібно прийняти рішення чи включати показники збитків під час навчання моделей чи ні, оскільки вони є некорельованими.

В останньому розділі було виконано роботу над стартап-проектом. В якості стартап-проекту було обрано організацію у вигляді центру моніторингу діяльності страхових компаній України. Дана стратегія має властивості блакитного океану, тобто вона є повністю відкритою для розробленого продукту. Незважаючи на невелику рентабельність, коло можливих користувачів є доволі великим, через це даний стартап є інвестиційно привабливим.

При подальшому дослідженні даної теми необхідно виділити наступні завдання:

1. Детальний аналіз економічних показників, що впливають на оцінку діяльності страхових компаній (аналіз звітності компаній, перевірка хибних значень тощо).
2. Застосування більш сучасних методів машинного навчання для задачі класифікації.

ПЕРЕЛІК ПОСИЛАНЬ

1. Олексюк О.С. Системи підтримки прийняття фінансових рішень. Київ: Наукова думка, 1998. 508 с.
2. Бідюк П.І., Тимощук О.Л., Коваленко А.Є., Коршевнік Л.О. Системи і методи підтримки прийняття рішень. Київ: НТУУ «КПІ імені Ігоря Сікорського», 2020. 600 с.
3. Робсон М., Уллах Ф. Практическое руководство по реинжинирингу бизнес-процессов / пер. с англ. под ред. Н. Д. Эриашвили. Москва: Аудит, ЮНИТИ, 1997. 224 с.
4. Страхування: теорія та практика: Навчально-методичний посібник / Внукова Н.М., Успенко В.І., Временко Л.В. та ін.; За загальною редакцією проф. Внукової Н.М. Харків: Бурун Книга, 2004. 376 с
5. Вовчак О.Д. Страхування: навч. посіб. 3-тє видання, стереотипне. Львів: «Новий світ – 2000», 2006. 480 с.
6. Теория и практика страхования: учеб. пособие / под ред. К. Е. Турбиной. Москва: Анкил, 2003. 704 с.
7. Юлдашев Р.Т. Организационно-экономические основы страхования бизнеса. Москва: Анкил, 2002. 247 с.
8. Фінанси: підручник. / С. І. Юрій та ін.; ред. С. І. Юрій, В. М. Федосов. Київ: Знання, 2008. 611 с.
9. Підсумки діяльності страхових компаній за I квартал 2020 року. URL: <https://forinsurer.com/files/file00676.pdf> (дата звернення: 24.10.2021).
10. Майбутнє регулювання ринку страхування в Україні: Біла книга. НБУ 2020. URL: https://bank.gov.ua/admin_uploads/article/White_paper_insurance_fin_pr_2020-04-22.pdf?v=4 (дата звернення: 21.10.2021).
11. Support-vector machine. URL: https://en.wikipedia.org/wiki/Support-vector_machine (дата звернення: 21.10.2021)

12. Naive Bayes classifier. URL: https://en.wikipedia.org/wiki/Naive_Bayes_classifier (дата звернення: 21.10.2021)
13. Ali, J., Khan, R., Ahmad, N., Maqsood, I. Random Forests and Decision Trees. *International Journal of Computer Science Issues (IJCSI)*. 2012. Vol. 9. P. 272-278.
14. Ioffe, S., Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ICML*. 2015. Vol. 37. P. 448-456.
15. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. / Srivatsava N. et.. al. *Journal of Machine Learning Research*. 2014. Vol 15. P. 1929-1958.
16. Шірінян, А.А., Шірінян, Л.В. Показники ефективності страхового ринку України з позицій власників страхових компаній. *Ефективна економіка*. 2019. №10. С. 1-7.

ДОДАТОК А. КОД ОСНОВНОГО МОДУЛЮ ПРОГРАМНОГО ПРОДУКТУ

Main.py

```
import sys
from PyQt5 import QtGui

import Design
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')

from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.naive_bayes import GaussianNB
from tensorflow import keras
from keras import layers

from PyQt5 import QtWidgets
from PyQt5.QtWidgets import QFileDialog
from PyQt5.QtWidgets import QMessageBox

from Utils import set_id, evaluate
```

```

class App(QtWidgets.QMainWindow, Design.Ui_MainWindow):
    def __init__(self):
        super().__init__()
        self.setupUi(self)
        self.setWindowTitle("СППР для оцінки та прогнозування майбутнього стану
страхових компаній.")
        self.model = None
        self.model_name = ""
        self.Train_Valid_Data = None
        self.Test_Data = None
        self.X = None
        self.Y = None
        self.X_train = None
        self.X_valid = None
        self.Y_train = None
        self.Y_valid = None
        self.Y_pred = None
        self.scaler = StandardScaler()
        self.reduced_scaler = StandardScaler()

self.Load_Train_Valid_Data_pushButton.clicked.connect(self.load_train_valid_data)
self.Load_Test_Data_pushButton.clicked.connect(self.load_test_data)
self.Heatmap_pushButton.clicked.connect(self.plot_heatmap)
self.Pieplot_pushButton.clicked.connect(self.plot_pie)
self.Start_Learning_pushButton.clicked.connect(self.train)

self.Confusion_Matrix_pushButton.clicked.connect(self.visualize_confusion_matrix)
self.pushButton.clicked.connect(self.forecast)
self.Reduce_Train_Valid_Data_pushButton.clicked.connect(self.reduce_data)
self.Restore_Data_pushButton.clicked.connect(self.restore_data)

self.forecast_all_companies_pushButton.clicked.connect(self.forecast_all_companies)
self.box_and_whiskers_pushButton.clicked.connect(self.box_and_whiskers_plot)
self.Reduced = False
self.feature_cols = None

```

```

def load_train_valid_data(self):
    options = QFileDialog.Options()
    options |= QFileDialog.DontUseNativeDialog
    fileName, _ = QFileDialog.getOpenFileName(self, "Завантажити
тренувально-валідаційну вибірку даних", "", "XLSX(*.xlsx)", options=options)
    if fileName and ("Train-Valid Data" in fileName):
        self.Train_Valid_Data_lineEdit.setText(fileName)
        self.Train_Valid_Data = pd.read_excel(fileName, engine='openpyxl',
skiprows = 5)

        self.Train_Valid_Data.drop(["№ з/п", "Код установи", "Назва установи"],
axis = 1, inplace = True)

        self.Train_Valid_Data["Оцінка діяльності"] =
np.where(round((self.Train_Valid_Data["Чистий фінансовий результат (прибуток)"] /
self.Train_Valid_Data["Власний капітал"])*100,2) > 10, "Ефективна", "Неефективна")
        self.Train_Valid_Data.dropna(inplace = True)
        self.feature_cols = list(self.Train_Valid_Data.columns[:-1])
        self.Train_Valid_Data[self.feature_cols] =
np.round(self.Train_Valid_Data[self.feature_cols],2)
        self.X = self.Train_Valid_Data.drop("Оцінка діяльності", axis = 1)
        self.Y = self.Train_Valid_Data["Оцінка діяльності"]
        self.Y = self.Y.map({'Ефективна':1, 'Неефективна':0})
        self.features_comboBox.addItem(self.feature_cols)
    else:
        QMessageBox.critical(self, "Помилка", "Не вдалося завантажити
файл", QMessageBox.Ok)

def load_test_data(self):
    options = QFileDialog.Options()
    options |= QFileDialog.DontUseNativeDialog
    fileName, _ = QFileDialog.getOpenFileName(self, "Завантажити тестову
вибірку даних", "Excel file (.xlsx)", "XLSX(*.xlsx)", options=options)
    if (len(self.Train_Valid_Data_lineEdit.text()) == 0):
        QMessageBox.critical(self, "Помилка", "Спочатку завантажте
навчально-валідаційну вибірку даних.", QMessageBox.Ok)

```

```

elif fileName and ("Test Data" in fileName):
    self.Test_Data_lineEdit.setText(fileName)
    self.Test_Data = pd.read_excel(fileName, engine = 'openpyxl', skiprows = 4)
    self.Test_Data.drop(["№ з/п"], axis = 1, inplace = True)
    self.Test_Data.dropna(inplace = True)
    self.Test_Data["Код установки"] = self.Test_Data["Код
установи"].astype(int)
    self.Test_Data["Код установки"] = self.Test_Data["Код
установи"].apply(str)
    self.Test_Data["Код установки"] = self.Test_Data["Код
установи"].apply(set_id)
    self.Test_Data[self.feature_cols] =
np.round(self.Test_Data[self.feature_cols],2)
else:
    QMessageBox.critical(self,"Помилка","Не вдалося завантажити
файл",QMessageBox.Ok)

def reduce_data(self):
    if (self.Train_Valid_Data is None) or (self.Test_Data is None):
        QMessageBox.critical(self,"Помилка!","Дані не завантажено.",
QMessageBox.Ok)
    else:
        corr_mean = self.Train_Valid_Data.corr().mean()
        cols_to_remove = []
        for feature_col in corr_mean.index:
            if np.abs(corr_mean[feature_col]) <= 0.1:
                cols_to_remove.append(feature_col)
                self.feature_cols.remove(feature_col)
        self.Train_Valid_Data.drop(cols_to_remove, axis = 1, inplace = True)
        self.Test_Data.drop(cols_to_remove, axis = 1, inplace = True)
        self.Reduced = True
        self.features_comboBox.clear()
        self.features_comboBox.addItem(self.feature_cols)
        if len(cols_to_remove) != 0:
            QMessageBox.information(self, "Зміна вибірки даних", "Були

```

успішно видалені наступні стовпці:\n" + str(cols_to_remove), QMessageBox.Ok)

else:

QMessageBox.information(self, "Зміна вибірки даних", "Змін не відбулося.", QMessageBox.Ok)

def restore_data(self):

if (len(self.Train_Valid_Data_lineEdit.text()) == 0) or (len(self.Test_Data_lineEdit.text()) == 0):

QMessageBox.critical(self, "Помилка", "Дані не завантажено", QMessageBox.Ok)

elif self.Reduced:

self.Train_Valid_Data = pd.read_excel(self.Train_Valid_Data_lineEdit.text(), engine = 'openpyxl', skiprows=5)

self.Train_Valid_Data.drop(["№ з/п", "Код установи", "Назва установи"], axis = 1, inplace = True)

self.Train_Valid_Data["Оцінка діяльності"] = np.where(round((self.Train_Valid_Data["Чистий фінансовий результат (прибуток)"] / self.Train_Valid_Data["Власний капітал"])*100,2) > 10, "Ефективна", "Неефективна")

self.Train_Valid_Data.dropna(inplace = True)

self.feature_cols = list(self.Train_Valid_Data.columns[:-1])

self.X = self.Train_Valid_Data.drop("Оцінка діяльності", axis = 1)

self.Y = self.Train_Valid_Data["Оцінка діяльності"]

self.Y = self.Y.map({'Ефективна':1, 'Неефективна':0})

self.features_comboBox.clear()

self.features_comboBox.addItem(self.feature_cols)

self.Test_Data = pd.read_excel(self.Test_Data_lineEdit.text(), engine = 'openpyxl', skiprows = 4)

self.Test_Data.drop(["№ з/п"], axis = 1, inplace = True)

self.Test_Data.dropna(inplace = True)

self.Test_Data["Код установи"] = self.Test_Data["Код установи"].astype(int)

self.Test_Data["Код установи"] = self.Test_Data["Код установи"].apply(str)

self.Test_Data["Код установи"] = self.Test_Data["Код установи"].apply(set_id)

```

        self.Train_Valid_Data[self.feature_cols]
np.round(self.Train_Valid_Data[self.feature_cols],2)
        self.Test_Data[self.feature_cols]
np.round(self.Test_Data[self.feature_cols],2)
        self.Reduced = False
        QMessageBox.information(self, "Зміна вибірки даних","Дані успішно
відновлено.",QMessageBox.Ok)
    else:
        QMessageBox.information(self, "Зміна вибірки даних", "Змін не
відбулося.", QMessageBox.Ok)

def plot_heatmap(self):
    if len(self.Train_Valid_Data_lineEdit.text()) == 0:
        QMessageBox.critical(self, "Помилка", "Не завантажено файл даних!",
QMessageBox.Ok)
    else:
        data_for_heatmap = self.Train_Valid_Data.copy()
        data_for_heatmap["Оцінка діяльності"] = data_for_heatmap["Оцінка
діяльності"].map({"Ефективна":1,"Неефективна":0})
        plt.figure(figsize = (15,15))
        sns.heatmap(data_for_heatmap.corr(), annot = True, fmt = '.2f',
cmap='YlGnBu', vmin = -1, vmax = 1)
        plt.tight_layout()
        plt.show()

def plot_pie(self):
    if len(self.Train_Valid_Data_lineEdit.text()) == 0:
        QMessageBox.critical(self, "Помилка", "Не завантажено файл даних!",
QMessageBox.Ok)
    else:
        plt.title('2020')
        self.Train_Valid_Data["Оцінка діяльності"].value_counts().plot(kind = 'pie',
figsize = (5,5), colors = ['crimson','lime'], autopct = '%1.2f%%')
        plt.show()

```

```

def box_and_whiskers_plot(self):
    if len(self.Train_Valid_Data_lineEdit.text()) == 0:
        QMessageBox.critical(self, "Помилка", "Не завантажено файл даних!",
QMessageBox.Ok)
    else:
        data_for_plot = self.Train_Valid_Data.copy()
        feature_for_boxplot = self.features_comboBox.currentText()
        plt.figure(figsize=(7,7))
        sns.boxplot(x      =      data_for_plot["Оцінка      діяльності"],y      =
data_for_plot[feature_for_boxplot])
        plt.show()

def train(self):
    if len(self.Train_Valid_Data_lineEdit.text()) == 0:
        QMessageBox.critical(self, "Помилка", "Не завантажено файл даних!",
QMessageBox.Ok)
    else:
        size = round(self.Valid_Size_doubleSpinBox.value(),1)
        if self.Reduced:
            self.X = self.X[self.feature_cols]
            self.X_train,      self.X_valid,      self.Y_train,      self.Y_valid      =
train_test_split(self.X, self.Y, test_size=size, random_state=42, stratify=self.Y)
            print(self.X_train.shape, self.X_valid.shape)
            self.X_train = self.reduced_scaler.fit_transform(self.X_train)
            self.X_valid = self.reduced_scaler.transform(self.X_valid)
        else:
            self.X = self.X[self.feature_cols]
            self.X_train,      self.X_valid,      self.Y_train,      self.Y_valid      =
train_test_split(self.X, self.Y, test_size=size, random_state=7, stratify=self.Y)
            print(self.X_train.shape, self.X_valid.shape)
            self.X_train = self.scaler.fit_transform(self.X_train)
            self.X_valid = self.scaler.transform(self.X_valid)
        if (self.KNN_radioButton.isChecked() == True):
            self.model = KNeighborsClassifier(n_neighbors=3)
            self.model_name = "k-nearest neighbors"

```

```

self.model.fit(self.X_train, self.Y_train)
self.Y_pred = self.model.predict(self.X_valid)
evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
elif (self.SVM_radioButton.isChecked() == True):
    self.model = SVC(kernel='linear', C = 1, gamma='scale')
    self.model_name = "Support Vector Machine"
    self.model.fit(self.X_train, self.Y_train)
    self.Y_pred = self.model.predict(self.X_valid)
    evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
elif (self.RF_radioButton.isChecked() == True):
    self.model = RandomForestClassifier(criterion='entropy', max_depth=30,
min_samples_leaf=7, n_estimators=4, random_state = 7)
    self.model_name = "Random Forest"
    self.model.fit(self.X_train, self.Y_train)
    self.Y_pred = self.model.predict(self.X_valid)
    evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
elif (self.NB_radioButton.isChecked() == True):
    self.model = GaussianNB()
    self.model_name = 'Naive Bayesian Classifier'
    self.model.fit(self.X_train, self.Y_train)
    self.Y_pred = self.model.predict(self.X_valid)
    evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
elif (self.XGB_radioButton.isChecked() == True):
    self.model = XGBClassifier(eval_metric = 'logloss',
use_label_encoder=False, n_estimators = 10, learning_rate = 0.05)
    self.model_name = 'XGBoost'
    self.model.fit(self.X_train, self.Y_train)
    self.Y_pred = self.model.predict(self.X_valid)
    evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
elif (self.DNN_radioButton.isChecked() == True):
    self.textEdit.setText("Іде навчання нейромережі...")
    self.model_name = "Deep Neural Network"
    self.model = keras.Sequential([
        layers.Dense(units = 4, input_shape = [self.X.shape[1]],
activation='tanh'),

```

```

layers.BatchNormalization(),
layers.Dropout(0.3),
layers.Dense(units = 8, activation = 'tanh'),
layers.BatchNormalization(),
layers.Dropout(0.3),
layers.Dense(units = 8, activation = 'tanh'),
layers.BatchNormalization(),
layers.Dropout(0.3),
layers.Dense(units = 8, activation = 'tanh'),
layers.BatchNormalization(),
layers.Dropout(0.3),
layers.Dense(units = 1, activation='sigmoid'))))
self.model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['binary_accuracy'])
history = self.model.fit(
    self.X_train, self.Y_train,
    validation_data = (self.X_valid, self.Y_valid),
    batch_size = 16,
    epochs = 600,
    verbose = 1
)
history_df = pd.DataFrame(history.history)
QMessageBox.information(self, "Статус нейромережі", "Навчання
завершено!", QMessageBox.Ok)
self.textEdit.clear()
self.Y_pred = [1 if pred > 0.5 else 0 for pred in
self.model.predict(self.X_valid)]
evaluate(self.model_name, self.Y_valid, self.Y_pred, self.textEdit)
if (self.plotting_checkBox.isChecked() == True):
    fig, axes = plt.subplots(1, 2, figsize = (8,4))
    cols =
np.array(['loss', 'val_loss'], ['binary_accuracy', 'val_binary_accuracy']))
titles = ['Loss', 'Accuracy']

for ax, col_row, title in zip (axes, cols, titles):

```

```

        history_df[col_row].plot(ax = ax)
        ax.set_title(title)
    plt.show()

def visualize_confusion_matrix(self):
    if (self.Y is None):
        QMessageBox.critical(self, "Помилка", "Не завантажена
трениувально-валідаційна вибірка даних!", QMessageBox.Ok)
    elif (self.Y_pred is None):
        QMessageBox.critical(self, "Помилка", "Не обрана модель для навчання!",
QMessageBox.Ok)
    else:
        cm = confusion_matrix(self.Y_valid, self.Y_pred)
        fig, ax = plt.subplots(figsize = (10,10))
        ax.imshow(cm, cmap = 'Blues')
        ax.grid(False)
        ax.set_title(self.model_name)
        ax.set_xlabel('Predicted')
        ax.set_ylabel('True')
        ax.xaxis.set(ticks = (0,1), ticklabels = ('Not efficient', 'Efficient'))
        ax.yaxis.set(ticks = (0,1), ticklabels = ('Not efficient', 'Efficient'))
        ax.set_ylim(1.5,-0.5)
        for i in range(len(np.unique(self.Y_pred))):
            for j in range(len(np.unique(self.Y_pred))):
                ax.text(j, i, cm[i,j], ha = 'center', va = 'center', color = 'green',
fontsize = 20)
        plt.show()

def forecast(self):
    search_id = self.Insurance_Company_ID_lineEdit.text()
    if (self.Test_Data is None):
        QMessageBox.critical(self, "Помилка", "Не завантажено тестувальну
вибірку даних!", QMessageBox.Ok)
    elif len(search_id) == 0:
        QMessageBox.critical(self, "Помилка", "Не введено код установи!",

```

```

QMessageBox.Ok)
        elif (len(search_id) != 8) or (not search_id.isdigit()):
            QMessageBox.critical(self, "Помилка", "Неправильно введений код
установи!\nКод має складатися з 8 чисел.", QMessageBox.Ok)
        elif (self.model is None):
            QMessageBox.critical(self, "Помилка", "Не обрана модель!",
QMessageBox.Ok)
        else:
            data_for_forecast = self.Test_Data.loc[self.Test_Data["Код установи"] ==
search_id, self.feature_cols]
            if data_for_forecast.empty:
                QMessageBox.critical(self, "Результати пошуку", "Не знайдено даних
страхової компанії за даним кодом.\nМожливо цієї компанії не існує або неправильно
введений код.", QMessageBox.Ok)
            else:
                if self.Reduced:
                    data_for_forecast_scaled =
self.reduced_scaler.transform(data_for_forecast)
                else:
                    data_for_forecast_scaled = self.scaler.transform(data_for_forecast)
                prediction = 1 if self.model.predict(data_for_forecast_scaled) > 0.5 else
0
                insurance_company_name = self.Test_Data.loc[self.Test_Data["Код
установи"] == search_id, "Назва установи"].values
                insurance_company_name = ",".join(insurance_company_name)
                info = "Назва компанії: " + str(insurance_company_name) + ".\n"
                cases = ""
                num_cases = 0
                if prediction == 0:
                    info += "Оцінка діяльності - нефективна.\n"
                    if data_for_forecast["Активи"].values <
self.Test_Data["Активи"].mean():
                        if num_cases == 0:
                            cases += "Можливі причини:\n"
                            cases += "* Низький рівень активів;\n"

```

```

        num_cases += 1
    else:
        num_cases += 1
        cases += "* Низький рівень активів;\n"
    if data_for_forecast["Витрати на оплату праці"].values <
self.Test_Data["Витрати на оплату праці"].mean():
        if num_cases == 0:
            cases += "Можливі причини:\n"
            cases += "* Малі витрати на оплату праці;\n"
            num_cases += 1
        else:
            num_cases += 1
            cases += "* Малі витрати на оплату праці;\n"
    if data_for_forecast["Страхові резерви"].values <
self.Test_Data["Страхові резерви"].mean():
        if num_cases == 0:
            cases += "Можливі причини:\n"
            cases += "* Малий об'єм страхових резервів;\n"
            num_cases += 1
        else:
            num_cases += 1
            cases += "* Малий об'єм страхових резервів;\n"
    if data_for_forecast["Премії підписані"].values <
self.Test_Data["Премії підписані"].mean():
        if num_cases == 0:
            cases += "Можливі причини:\n"
            cases += "* Низькі премії;\n"
            num_cases += 1
        else:
            num_cases += 1
            cases += "* Низькі премії;\n"
    if num_cases == 0:
        info += "Можливих причин не знайдено! Перегляньте
зміни в менеджменті компанії."
    info += cases

```

```

        QMessageBox.information(self, "Результат прогнозу", info,
        QMessageBox.Ok)

    else:
        info += "Оцінка діяльності - ефективна.\n"
        QMessageBox.information(self, "Результат прогнозу", info,
        QMessageBox.Ok)

def forecast_all_companies(self):
    if self.Test_Data is None:
        QMessageBox.critical(self, "Помилка", "Не завантажено тестувальну
        вибірку даних!", QMessageBox.Ok)
    elif self.model is None:
        QMessageBox.critical(self, "Помилка", "Не обрана
        модель", QMessageBox.Ok)
    else:
        Resulted_df = self.Test_Data[["Назва установи"]].copy()
        data_for_forecast = self.Test_Data[self.feature_cols].copy()
        if self.Reduced:
            data_for_forecast_scaled = self.reduced_scaler.transform(data_for_forecast)
        else:
            data_for_forecast_scaled = self.scaler.transform(data_for_forecast)
        Resulted_df["Оцінка діяльності"] = np.where(self.model.predict(data_for_forecast_scaled) > 0.5, "Ефективна", "Неефективна")
        Resulted_df.to_csv("Результати прогнозу.csv", index = None)
        QMessageBox.information(self, "Прогнозу стану страхових
        компаній", "Прогнозування стану компаній завершено.\nРезультати збережено у текстовий
        файл.", QMessageBox.Ok)

        plt.title("2021")
        Resulted_df["Оцінка діяльності"].value_counts().plot(kind = 'pie', colors =
        ['crimson', 'lime'], figsize = (5,5), autopct = '%1.2f%%')
        plt.show()

def main():

```

```
app = QtWidgets.QApplication(sys.argv)
window = App()
window.show()
app.exec_()

if __name__ == '__main__':
    main()
```