

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління

Кафедра Комп'ютерних інтелектуальних технологій та систем

## **КВАЛІФІКАЦІЙНА РОБОТА**

### **Пояснювальна записка**

рівень вищої освіти

другий (магістерський)

Оцінювання фінансових ризиків з використанням  
інтелектуального аналізу даних

Виконав:

студент 2 курсу, групи КІТм-21-1

Артур КОНОВАЛОВ

Спеціальність 123 Комп'ютерна інженерія

Тип програми освітньо-професійна

Освітня програма Комп'ютерні інтелектуальні  
технології

Керівник проф. каф. КІТС\_Микола КОРАБЛЬОВ

Допускається до захисту

Зав. кафедри

Олег РУДЕНКО

(підпис)

2023 р.

Харківський національний університет радіоелектроніки

Факультет	Комп'ютерної інженерії та управління
Кафедра	Комп'ютерних інтелектуальних технологій та систем
Рівень вищої освіти	другий (магістерський)
Спеціальність	123 Комп'ютерна інженерія
Тип програми	освітньо-професійна
Освітня програма	Комп'ютерні інтелектуальні технології

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 202\_ р.

## ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові \_\_\_\_\_ Коновалову Артуру Євгеновичу  
(прізвище, ім'я, по батькові)

1. Тема роботи Оцінювання фінансових ризиків з використанням інтелектуального аналізу даних

затверджена наказом по університету від “ 03 ” листопада 2023 р. № 1290Ст

2. Термін подання студентом роботи до екзаменаційної комісії 13.01.2024

3. Вхідні дані до роботи \_\_\_\_\_

- 1) методи оцінювання фінансових ризиків
- 2) моделі кредитного ризику
- 3) інструментальні засоби оцінювання кредитних ризиків
- 4) мова програмування Python
- 5) мова програмування R-package

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

- 1) Аналіз предметної області і постановка задачі дослідження
- 2) Аналіз методів і моделей оцінювання фінансових ризиків
- 3) Аналіз типів даних і невизначеностей в моделюванні фінансових ризиків
- 4) Моделювання кредитних ризиків
- 5) Оцінювання кредитного ризику з використанням адаптивної мережі Байєса
- 6) Прведення експериментальних досліджень

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій Плакати – 16 арк. ф. А4

---

---

---

---

---

---

---

---

6. Консультанти розділів роботи и (п.6 включається до завдання за наявності консультантів згідно до наказу, зазначеному у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

### КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Видача та узгодження теми проєкту	06.11.2023	Виконано
2	Огляд стану проблеми та постановка задачі	06.11 -17.11	Виконано
3	Аналіз літератури за напрямком роботи	18.11 - 01.12	Виконано
4	Аналіз методів і моделей оцінювання фінансових ризиків	02.12 - 09.12	Виконано
5	Побудова математичних моделей для аналізу та прогнозування кредитних ризиків	10.12 – .17.12	Виконано
6	Проведення експериментальних досліджень	18.12.23 - 07.01.24	Виконано
7	Підготовка пояснювальної записки	08.01 - 15.01	Виконано
8	Підготовка графічного матеріалу	16.01 – 23.01	Виконано
9	Захист проєкту	24.01.2024	Виконано

Дата видачі завдання 06 листопада 2023 р.

Студент \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_ професор Микола КОРАБЛЬОВ  
(підпис) (посада, ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 89 с., 3 рис., 1 табл., 2 дод., 36 джерел.

### ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ФІНАНСОВІ РИЗИКИ, КРЕДИТУВАННЯ, КРЕДИТОСПРОМОЖНІСТЬ, МАТЕМАТИЧНІ МОДЕЛІ, МОДЕЛЮВАННЯ, МЕРЕЖА БАЙЄСА

Метою кваліфікаційної роботи є побудова математичних моделей для прогнозування фінансових процесів та оцінювання кредитних ризиків можливих втрат на основі застосування методів інтелектуального аналізу даних.

Об'єктом дослідження є процеси аналізу та моделювання фінансових ризиків.

Предметом дослідження є методи, моделі та критерії аналізу процесів кредитування з використанням інтелектуального аналізу даних.

У роботі пропонується побудова математичних моделей нелінійних нестационарних процесів, якими є фінансові процеси, для оцінювання і моделювання кредитоспроможності позичальників, що дозволяє отримати високу якість проміжних та кінцевих результатів обчислювальних експериментів та уникнути неадекватності моделей.

З метою підвищення ефективності прийняття рішень при аналізі та моделюванні фінансових ризиків доцільно використовувати мережі Байєса, які дозволяють враховувати невизначеності різних типів, характеризуються швидкими алгоритмами навчання, підвищують адекватність ймовірнісної моделі та забезпечують зменшення розмірів кредитного ризику. Проведені експериментальні дослідження показали, що оцінка кредитоспроможності позичальників з використанням байєсівського підходу дозволила отримати кращі результати і покращити загальну точність результатів даних на 2.5%.

## ABSTRACT

Master's thesis:: 89 pages, 3 figures, 1 tables, 2 appendices, 36 sources.

INTELLECTUAL DATA ANALYSIS, FINANCIAL RISKS, LENDING, CREDIT WRITING, MATHEMATICAL MODELS, SIMULATION, BAYESIAN NETWORK

The major goal of this thesis is the construction of mathematical models for forecasting financial processes and credit risk assessment of possible losses based on the application of intelligent data analysis methods.

The object of the research is the processes of analysis and modeling of financial risks.

The subject of the research is the methods, models and criteria of the analysis of lending processes using intelligent data analysis

The paper proposes the construction of mathematical models of non-linear non-stationary processes, which are financial processes, for assessing and modeling the creditworthiness of borrowers, which allows obtaining high-quality intermediate and final results of computational experiments and avoiding the inadequacy of models.

In order to increase the efficiency of decision-making in the analysis and modeling of financial risks, it is advisable to use Bayesian networks, which allow taking into account uncertainties of various types, are characterized by fast learning algorithms, increase the adequacy of the probabilistic model and ensure the reduction of credit risk. The conducted experimental studies showed that the assessment of borrowers' creditworthiness using the Bayesian approach allowed to obtain better results and improve the overall accuracy of data results by 2.5%.

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет                      Комп'ютерної інженерії та управління  
Кафедра                        Комп'ютерних інтелектуальних технологій та систем

## **АНОТАЦІЯ**

### **КВАЛІФІКАЦІЙНОЇ РОБОТИ**

рівень вищої освіти другий (магістерський)

---

Оцінювання фінансових ризиків з використанням  
інтелектуального аналізу даних

---

---

Виконав:

студент 2 курсу, групи КІТм-22-1

Артур КОНОВАЛОВ

Спеціальність 123 Комп'ютерна інженерія

Тип програми      освітньо-професійна

Освітня програма Комп'ютерні інтелектуальні  
технології

Керівник    проф. каф. КІТС Микола КОРАБЛЬОВ

2023 р.

## АНОТАЦІЯ

Коновалов А.Є. Оцінювання фінансових ризиків з використанням інтелектуального аналізу даних. – Магістерська кваліфікаційна робота.

У магістерській кваліфікаційній роботі вирішено актуальну задачу побудови математичних моделей для аналізу і прогнозування фінансових процесів та оцінювання кредитних ризиків можливих втрат на основі застосування методів інтелектуального аналізу даних, методів аналізу часових рядів і прикладної статистики.

Метою кваліфікаційної роботи є побудова математичних моделей для аналізу і прогнозування фінансових процесів та оцінювання кредитних ризиків можливих втрат на основі методів інтелектуального аналізу даних, які дозволяють покращити загальну точність результатів та підвищити якість оцінок прогнозів.

Об'єктом дослідження є процеси аналізу та моделювання фінансових ризиків.

Предметом дослідження є методи, моделі та критерії аналізу процесів кредитування з використанням інтелектуального аналізу даних.

Методи дослідження. Особливості об'єкту і мета досліджень обумовлюють необхідність застосування наступних методів: системний аналіз, інтелектуальний і статистичний аналіз даних; методи фільтрації та обробки невизначеностей; ймовірно-статистичний та регресійний аналіз даних; байєсівські мережі; методи прогнозування на основі часових рядів, різницевого рівняння, регресійного аналізу і байєсівських мереж.

Наукова новизна полягає у застосуванні адаптивної Байєсівської мережі для оцінювання кредитоспроможності позичальників, що забезпечує високу адекватність ймовірнісної моделі і зменшує величини кредитного ризику.

Практична цінність отриманих результатів полягає у тому, що розроблені методи і моделі інтелектуального аналізу даних для оцінювання кредитних ризиків можуть бути впроваджені у фінансових організаціях з метою автоматизованого розв'язання задач моделювання, оцінювання і прогнозування втрат.

У першому розділі розглянуто актуальність задачі побудова математичних моделей для аналізу і прогнозування фінансових процесів та оцінювання кредитних ризиків. Проведено аналіз і класифікація фінансових ризиків. Під фінансовим ризиком слід розуміти можливість виникнення фінансових наслідків, які можуть бути непередбачуваними, у формі втрат або одержання доходів за умов невизначеності при здійсненні фінансової діяльності юридичною або фізичною особою. Фінансові ризики класифікують як: кредитний ризик; ринковий ризик; операційний ризик та ризик ліквідності.

Проведено аналіз моделей кредитного ризику. Важливою задачею менеджменту ризиків фінансових організацій, що забезпечують кредитами клієнтів, являється оцінювання кредитних ризиків. Розв'язання цієї задачі може забезпечити своєчасне і можливо повне повернення кредитів, а також зменшення можливих втрат. Для аналізу кредитних ризиків найчастіше застосовують статистичні методи, які ґрунтуються на інформації, що взята впродовж великого проміжку часу, та на ймовірнісних розрахунках того, чи настануть ті чи інші події, що можуть мати позитивні або негативні наслідки.

Банки застосовують скорингову систему, яка являється системою оцінювання позичальників та використовує математично-статистичні методи. Розроблення інформаційних технологій, з використанням яких можна обробляти значну кількість кредитних заявок, сприяло розвитку скорингу та його різновидів. Правильним і коректним може бути підхід, що ґрунтується на статистиці дефолтів та враховує всі попередні періоди, тобто скорингову модель. Такий підхід висуває до статистики дефолтів високі вимоги.

Розглянуто та проведено регресійний аналіз для оцінювання фінансових ризиків. Серед регресійних моделей розглянута лінійна ймовірнісна модель, логіт та пробіт моделі, моделі на основі експертного оцінювання та Байєсівські моделі на основі мереж довіри. З розглянутих моделей для оцінювання і прогнозування фінансових ризиків пропонується використовувати моделі Байєса, які дозволяють підвищити адекватність ймовірнісної моделі та забезпечити зменшення розмірів кредитного ризику.

Розглянуті моделі ринкового ризику, під яким розуміється ризик втрати позицій, які виникають під час руху ринкових цін. На сучасному етапі для менеджменту та аналізу ринкового ризику банки застосовують різні досить складні математичні та статистичні методи, за допомогою яких забезпечується створення адекватних математичних моделей та обчислення високоякісних оцінок. Розглянутий метод Value-at-Risk, який оцінюється у грошовому вираженні як кількісна міра потенційного збитку згідно позицій фактичної вартості активів внаслідок руху ринку, та метод Stressed Value-at-Risk, який використовується для розрахунку вартості під ризиком за умов наявності на ринку значного стресу протягом одного року.

У другому розділі виконано моделювання кредитних ризиків. Однією з проблем при побудові моделей є проблема відбору вхідних ознак, значущих для моделі. Для зменшення розмірності моделі необхідне зменшення числа незалежних змінних з метою видалення з неї всі незначущих ознак, які не несуть в собі інформації, корисної для аналізу, і таким чином, спростити модель та видалити надлишкові ознаки. Згідно цього виконано оцінювання інформативності змінних та здійснений вибір підмножини ознак.

Можна використати різні підходи для відбору ознак, застосування яких для відбору ознак базується на тому, що в даних містяться деякі ознаки, які при формуванні результату являються зайвими або значущими, тому їх можна видалити без великих втрат інформації. Виділення ознак дозволяє створювати нові ознаки у вигляді функцій від оригінальних ознак, а відбір ознак дозволяє повертати їх підмножину. Найпростішим алгоритмом являється проведення перевірки кожної можливої підмножини ознак та знаходження серед них такої, яка дозволяє мінімізувати величину похибки.

Для оцінювання інформативності моделей в якості показника було обрано критерій  $\chi^2$  та загальний статистичний тест. Використовуючи цей критерій виконується порівняння вибірок, які можуть мати альтернативні ознаки та виконується оцінювання ймовірностей кореляції між альтернативними ознаками. Виконане рекурсивне виключення змінних, для чого застосовано метод рекурсивного

усунення ознак, який може використовуватися для задач з невеликими вибірками.

Оцінка значущості результатів при розв'язанні задачі кредитоспроможності платників може виконуватися із застосуванням великої кількості алгоритмів. Під значимістю атрибута розуміється показник, який характеризує те, наскільки вихідний результат сильно залежить від вхідних показників. В цілому, для атрибутів обчислення показників значущості можливе лише після побудови дерева класифікаційних правил.

Для остаточного включення змінних до моделі необхідно за результатами оцінювання провести скоринг за кожним з методів. Для кожної змінної обчислюється загальна оцінка, згідно якої визначається кількість методів, що обрали відповідну змінну як значущу для моделі.

Для аналізу даних застосована Байєсівської мереж (БМ), використання якої потребує вирішення математичних задач побудови структури БМ і формування імовірнісного висновку. Задача побудови байєсівської мережі, тобто оцінювання її структури і параметрів, складається з таких кроків: побудова за два етапи БМ евристичним методом та формування (обчислення) імовірнісного висновку в БМ, яке реалізується також двома кроками. Використання БМ дозволяє проводити аналіз причинно-наслідкових зв'язків між окремими змінними, подіями та даними, а також на цій основі формулювати обґрунтований ймовірнісний висновок або прогноз. Розглянуто реалізацію алгоритма прогнозування кредитоспроможності індивідуального позичальника за допомогою БМ.

У третьому розділі проведені експериментальні дослідження. Була розглянута критеріальна база для аналізу якості результатів, згідно чого досліджені вихідні дані для статистичного аналізу і прогнозування та виділені основні особливості його виконання. Вказані вимоги, що пред'являються щодо вихідних статистичних даних, якими є: релевантність, надійність і точність, порівнянність та репрезентативність. Обчислення прогнозу, побудова та експериментальна перевірка (верифікація) ймовірнісно-статистичної моделі базуються на використанні одночасно інформації двох типів: апіорна інформація про змістовну сутність і природу аналізованого явища та вихідні статистичні дані, які характеризують як процес, так і результати

функціонування системи.

Виділені основні етапи прогнозування, такі як постановка завдання; аналіз сутності процесу, що досліджується; збір необхідної статистичної інформації; специфікація моделі; ідентифікація моделі та верифікація моделі. Розглянуті основні вимоги до математичної моделі і проведено аналіз якості моделі. Аналіз адекватності моделі виконано з використанням відповідних параметрів адекватності.

Точність і обґрунтованість прогнозів є обов'язковим етапом прогнозування, на якому використана певна сукупність підходів, критеріїв та процедур, що дозволили отримати оцінку якості прогнозу, для якої обрані відповідні показники.

Отримані результати моделювання кредитних ризиків банку з використанням байєсівських мереж для управління кредитним ризиком. Процес аналізу являє собою виконання декількох етапів. Відбір значущих змінних для прогнозування відбувався за 3 кроки. На основі отриманих статистичними методами розрахунків було обчислене фінальне значення для кожної зі змінних, а потім було отримано результат: чи варто до фінальної моделі обирати дану змінну. Згідно розглянутих основних кроків за початковими даними була побудована структура БМ.

Отримані результати обчислювальних експериментів свідчать, що використання моделі у формі БМ дозволяє отримати високі значення точності моделі, що дорівнюють 0,779 або 77,9%. Наведені результати вказують на доцільність використання БМ при оцінюванні кредитоспроможності позичальників кредитів.

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ФІНАНСОВІ РИЗИКИ,  
КРЕДИТУВАННЯ, КРЕДИТОСПРОМОЖНІСТЬ, МАТЕМАТИЧНІ МОДЕЛІ,  
МОДЕЛЮВАННЯ, МЕРЕЖА БАЙЄСА

Публікації здобувача за темою роботи:

1. Коновалов А.Є. Оцінювання кредитних ризиків на основі адаптивної мережі Байєса // Матеріали IV Всеукраїнської студентської наукової конференції «Розвиток сучасної науки: актуальні питання теорії та практики», 17 листопада 2023, м. Львів, Україна.

## ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ	14
ВСТУП	15
1 АНАЛІЗ МЕТОДІВ ОЦІНЮВАННЯ ФІНАНСОВИХ РИЗИКІВ	17
1.1 Аналіз фінансових ризиків	17
1.2 Аналіз моделей кредитного ризику	21
1.3 Регресійні моделі та регресійний аналіз	24
1.3.1 Регресійні моделі.....	25
1.3.2 Лінійна ймовірнісна модель.....	25
1.3.3 Логіт та пробіт моделі.....	26
1.3.4 Моделі на основі експертного оцінювання.....	27
1.3.5 Байєсівські моделі на основі мереж довіри.....	29
1.4 Моделі ринкового ризику	29
1.4.1 Value-at-Risk.....	30
1.4.2 Stressed Value-at-Risk.....	34
2 МОДЕЛЮВАННЯ КРЕДИТНИХ РИЗИКІВ	38
2.1 Оцінювання інформативності змінних	38
2.1.1 Вибір підмножини ознак.....	39
2.1.2 Виконання загального статистичного тесту.....	41
2.1.3 Рекурсивне виключення змінних.....	42
2.1.4 Оцінка значущості результатів.....	45
2.1.5 Остаточне включення змінних до моделі.....	47
2.2 Оцінювання кредитного ризику з використанням адаптивної мережі Байєса	48
3 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ	55
3.1 Критеріальна база для аналізу якості результатів	55
3.1.1 Вихідні дані для статистичного аналізу та прогнозування.....	55
3.1.2 Вимоги до математичної моделі.....	58
3.1.3 Аналіз якості моделі.....	60

3.1.4 Аналіз якості прогнозу .....	61
3.2 Результати моделювання кредитних ризиків	62
ВИСНОВКИ	69
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	71
ДОДАТОК	А
	О

**шибка! Закладка не определена.**

Графічний	матеріал	атестаційної	роботи
			О

**шибка! Закладка не определена.**

ДОДАТОК	Б
	О

**шибка! Закладка не определена.**

Наукова	публікація
	О

**шибка! Закладка не определена.**

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

БМ – Байєсівська мережа

МКМЛ – метод Монте-Карло для марковських ланцюгів

ОМД – оцінка мінімальної довжини

VaR – Value-at-Risk (вартість під ризиком)

SVaR – Stressed Value-at-Risk (стресована вартість під ризиком)

LASSO – Least Absolute Shrinkage and Selection Operator (оператор найменшого абсолютного скорочення та відбору)

SVM – Support Vector Machine (Опорна машина векторів)

RFE – Recursive Feature Elimination (рекурсивного усунення ознак)

MSE – Mean Squared Error (середня квадратична похибка)

RMSE – Root Mean Squared Error (середньоквадратична помилка)

## ВСТУП

Одним із ключових завдань сучасного розвитку методів інтелектуального аналізу даних, методів аналізу часових рядів та прикладної статистики є побудова адекватних математичних моделей, а також розробка ефективних методів розв'язання задач моделювання і прогнозування фінансових процесів, які дозволяють оцінювати ризики можливих втрат.

В теперішній час підвищення якості та розширення можливостей сучасних систем оптимального оцінювання станів, систем математичного моделювання, систем прогнозування розвитку стохастичних процесів в технічних системах і технологіях, в економіці та фінансах, а також систем оцінювання фінансових ризиків різних типів неможливе без застосування інформаційних технологій та сучасних методів системного аналізу, без коректної обробки великої кількості статистичних даних, а також без врахування невизначеностей статистичного, структурного та параметричного характеру.

Як показує досвід різних країн, економічний успіх будь-якої країни у будь-якій галузі практичної діяльності на всіх рівнях управління процесами різної природи неможливий без застосування наукового підходу. Тому актуальною задачею є використання методів інтелектуального аналізу даних та сучасних методів математичного моделювання для розв'язання практичних задач моделювання і прогнозування розвитку фінансових процесів різної природи та виникаючих фінансових ризиків, оцінювання станів ризиків за умов наявності невизначеностей і випадкових впливів. Розв'язання цих задач є важливим за умов відсутності ефективних методів поглибленого аналізу даних, особливо при необхідності подальшого розвитку нелінійних нестационарних процесів.

Використання цих методів для розв'язання складних задач у різних видах соціальної діяльності дозволяє досягати результатів, які неможливо отримати в іншій спосіб. Особливо це важливо для таких задач, при розв'язанні яких необхідно враховувати багато різних чинників, або у постановці яких може мати місце

невизначеність того чи іншого типу.

Слід зазначити, що фінансові ризики являються специфічною сферою кредитно-фінансових взаємовідношень банку з юридичними та фізичними особами, метою яких є отримання прибутку. Указані відносини можуть піддаватися великій кількості різноманітних ризиків, які в більшості випадків являються невизначеними. Тому вони повинні бути мінімізовані у повсякденній роботі банку.

# 1 АНАЛІЗ МЕТОДІВ ОЦІНЮВАННЯ ФІНАНСОВИХ РИЗИКІВ

## 1.1 Аналіз фінансових ризиків

Під фінансовим ризиком слід розуміти можливість виникнення фінансових наслідків, які можуть бути непередбачуваними, у формі втрат або одержання доходів за умов невизначеності при здійсненні фінансової діяльності юридичною або фізичною особою [1, 2]. Зазначимо, що ризики можуть мати місце на будь-якому підприємстві або у фінансовій установі, тому важливим моментом для ведення успішного бізнесу є належне управління ризиками. Керівництво підприємств або фінансових установ має різний рівень менеджменту ризиків, тому деякими з ризиків можна керувати безпосередньо, а деякі з них можуть значною мірою знаходитися поза межами контролю управлінців. Що можна зробити, коли виникають умови настання ризику, – це передбачити можливі виникнення ризиків, оцінити їх можливий вплив і бути готовим до реагування та реалізації плану дій на несприятливі події. Фінансові ризики можна класифікувати, використовуючи ті чи інші підходи. Один із підходів заснований на поділі фінансового ризику на чотири категорії [3]:

- кредитний ризик;
- ринковий ризик;
- операційний ризик;
- ризик ліквідності.

Ринковим ризиком передбачається зміна умов перебування на конкретному ринку, де фірми і підприємства конкурують за свої сегменти ринку. Прикладом ринкового ризику може бути зростаюча тенденція виконувати банківські операції споживачами в Інтернеті. Цей аспект ринкового ризику спричинив для традиційного роздрібного бізнесу значні виклики. Так, підприємства і організації, які виконували необхідні дії для швидкої адаптації до обслуговування, були процвітаючими та мали зростання доходу, в той час як організації, які не так швидко адаптувались або

робили неправильний вибір у своїй реакції на зміну ринка, відходили в сторону.

Іншим елементом ринкового ризику є ризик перевищення конкурентами рівня функціонування компанії. На конкурентоспроможному світовому ринку компанії, які найбільш фінансово успішні, іноді пропонують унікальну цінову пропозицію, яка часто може бути із зменшенням норм прибутку, але дозволяє їм виділитися з множини інших, що надає їм надійну ринкову ідентичність.

Підприємства стикаються з кредитним ризиком внаслідок надання кредитів клієнтам. Це може також стосуватися власного кредитного ризику компанії з постачальниками. Коли забезпечується фінансування закупівель для своїх клієнтів за умов того, що клієнт може замовчуватись в оплаті, бізнес також приймає фінансовий ризик [4]. У компанії завжди повинен бути достатній грошовий потік для своєчасного сплачування своїх кредиторських рахунків, забезпечуючи виконання своїх власних кредитних зобов'язань. В іншому випадку кредитори можуть припинити надавати компанії кредит, або припинити взагалі ведення бізнесу з цією компанією.

Згідно з Базельським комітетом з банківського нагляду кредитний ризик визначається як можливість того що позичальник або агент банку не може у погоджений термін виконати свої зобов'язання [5]. Метою управління кредитним ризиком являється максимізація прибутку банку, підтримуючи кредитний ризик в межах допустимих значень параметрів. Банки управляють кредитним ризиком, а також ризиком окремих операцій і кредитів, що властиві всьому портфелю. У Базельському комітеті відрізняли ринковий ризик від кредитного, а також при розрахунку ризикового капіталу враховували незалежно обидві ці категорії.

Як правило, позиції в портфелі компанії залежать на практиці одночасно від ринкових і кредитних факторів ризику. В такому випадку оцінювання функції вартості портфеля, яка поділяє зміну його вартості на чистий ринковий ризик та компонент чистого кредитного ризику, може призвести до недооцінювання ризику. Тому з точки зору оцінювання ризиків існуючий підхід до керування ризиками буде завжди консервативним. Цей факт часто має місце при наданні позик в іноземній валюті, тому при використанні традиційного підходу до регулювання буде значно

недооцінений реальний ризик портфеля позик в іноземній валюті.

У банківському регулюванні існує різниця між ринковим і кредитним ризиками та їх незалежним аналізом. Кредитний ризик використовується в основному для банківської діяльності, а ринковий ризик в основному використовується для торгової діяльності. Згідно цього нормативна категоризація банків передбачає організацію в них в кредитного відділу та відділу ринкових інвестицій. Це вимагає для кредитного та ринкового ризику наявності окремого регуляторного капіталу:

$$RC_c + RC_m. \quad (1.1)$$

Нормативний капітал для кредитного ризику  $RC_c$  розраховується окремо для кожної позики. Слід зазначити, що в теперішній час моделі кредитного ризику портфеля для розрахунку регулятивного капіталу майже не використовуються. Разом з тим, якщо вони приймають як детерміновані фактори ринкового ризику, то вони підходять для цієї схеми.

Для запобігання виникненню несприятливих змін ринкових цін використовується регуляторний капітал для ринкового ризику  $RC_m$ , але він не враховує можливість дефолту контрагента. Але потрібно мати в торговій книзі регуляторний капітал для ризику контрагента для деяких позицій [5]. Ось чому зв'язок ринкового ризику з торговою книгою та кредитного ризику з банківською книгою слугував поточному регулюванню, що подано виразом (1.1), яке являється консервативним та потребує аналізу відповідних передумов.

Передумова 1. При вимірюванні суб-адитивного ризику всього портфеля «диверсифікація» буде меншою або дорівнюватиме максимально сумі ризику по банківській і торговельній книгах.

Передумова 2. Кредитний ризик має відношення тільки до банківської книги, а ринковий ризик має відношення тільки до торгової книги.

Загальний ризик за всіма показниками суб-адитивного ризику буде максимально дорівнювати сумі ринкового ризику і кредитного ризику, або буде меншим. Це буде вірно в тому випадку, коли передумови виконуються і висновок

повинен обов'язково бути вірним.

Зазвичай передумова 1 виконується згідно визначення суб-адитивності, а передумова 2 не сприймається буквально, але вона вважається гарним наближенням до реальності. Крім того, регулювання вважають широко консервативним, так як воно вимагає для ринкового і кредитного ризику окремого ризикового капіталу.

В науковій літературі по інтеграції ринкового та кредитного ризиків проблема інтеграції ризиків розглядається по-різному. Існує напрямок досліджень, в якому традиційна категоризація розглядається критично.

У роботі [6] за скороченою формою розглядається модель для включення в традиційні моделі кредитного ризику стохастичних процентних ставок. В [7] розроблена система кредитного ризику, що заснована на структурній моделі кредитного ризику, яка включає стохастичні процентні ставки. Для інтегрованого аналізу ринкових і кредитних ризиків в [8] запропонована система моделювання для портфелів з фіксованою прибутковістю.

В [7] запропоновано підхід до інтегрованого моделювання ризиків, який не заперечує традиційну класифікацію, а скоріше за все вказує на те, що портфелі, які аналізуються в різних категоріях ринку і кредитного ризику, можуть бути розглянуті для всього банківського портфеля як ризики підпортфелей. Після того, коли будуть побудовані підпортфелі, то виникає проблема, яку ще треба вирішити, – це кількісна оцінка ефекту диверсифікації при виконанні умов, що підпортфелі будуть об'єднані в загальний портфель.

Також стверджується, що проблема аналізу ринкових та кредитних ризиків не є комплексною проблемою диверсифікації. Крім того, ця проблема полягає ще і в тому, що за ринковими та кредитними факторами побудова підпортфеля ризику неможлива. Цей факт в першу чергу повинен бути проаналізований, якщо це так. Якщо в такій ситуації замість цього вартість портфеля формалізується та апроксимується суб-позиторіями кредитного і ринкового ризику, то помилка в аналізі та оцінюванні складових може призвести зазвичай до помилки оцінювання ризику в цілому, що може призвести до значного заниження справжнього ризику.

## 1.2 Аналіз моделей кредитного ризику

Важливою задачею менеджменту ризиків фінансових організацій, що забезпечують кредитами клієнтів, являється оцінювання кредитних ризиків. Розв'язання цієї задачі може забезпечити своєчасне і можливо повне повернення кредитів, а також зменшення можливих втрат. Слід зазначити, що для надходжень капіталу кредитний ризик являє собою потенційний ризик, що може виникати через неспроможність сторін, які взяли на себе зобов'язання щодо виконання умов фінансової угоди з банком, або виконати взяті зобов'язання в інший спосіб [6].

Слід розрізняти індивідуальний та портфельний ризики при оцінюванні кредитного ризику. Джерелом індивідуального ризику може бути позичальник, емітент цінних паперів, боржник, кожен з яких може бути окремим, конкретним контрагентом банку. Для оцінювання індивідуального ризику необхідне оцінювання кредитоспроможності кожного окремого контрагента, тобто його індивідуальної спроможності в повному обсязі і своєчасно розрахуватися за взятими зобов'язаннями. Тому для таких ризиків можливе поверненням кредиту із запізненням, або взагалі можливе його неповернення.

Для аналізу масових явищ найчастіше застосовують статистичні методи оцінювання ризиків, які ґрунтуються на інформації, що взята впродовж великого проміжку часу про події, що можуть вплинути негативно на результат. Визначають, з якою періодичністю ці події відбувалися, якщо вони мали місце раніше. Статистичні методи оцінювання ризиків також базуються на ймовірностних розрахунках того, чи настануть ті чи інші події, що можуть мати позитивні або негативні наслідки.

Якщо випадки ризиків зустрічаються не так часто і не можуть бути статистично визначені, то застосовують якісні методи оцінювання ризиків. Їх оцінювання здійснюються на підставі досвіду і знань експертів. Тому такі методи оцінювання ризиків називають ще методами експертних оцінок. Слід зазначити, що рівень ризику залежить від ймовірності настання небажаних подій, тобто від його значення, яке від нього очікують, а також від варіантів можливого результату.

Банки намагаються відкинути найменш ризикованих і найменш привабливих позичальників у великому потоці кредитних заявок з тим, щоб приділити час найбільш цікавим позичальникам, використовуючи автоматизацію системи управління ризиками. Це можна зробити із застосуванням скорингової системи і моделі, яка дозволяє нараховувати бали в анкеті позичальника, а також розраховувати кредитоспроможність клієнта.

Скоринг являється системою оцінювання позичальників, яка використовує математично-статистичні методи. Його метою є оцінювання рівня платоспроможності клієнта на основі факторів, які стосуються відібраних потенційних позичальників. Використання скорингу є актуальним в банківській системі, оскільки його модель являється основним індикатором оцінювання ризику щодо прийняття кредитного рішення. Так як система може прийняти рішення, яке може не відповідати реальному стану справ, то такі рішення не завжди ідеальні, тому що банк може видати неплатіжеспроможному клієнту кредит, або навпаки не видати кредит хорошому позичальнику [17].

Розроблення інформаційних технологій, з використанням яких можна обробляти значну кількість кредитних заявок, сприяло розвитку скорингу та його різновидів. Відомо, що підхід, який базується на скорингу, передбачає на основі кредитних історій банку побудову математичних моделей для оцінювання кредитоспроможності позичальників, а також виходячи з соціально-демографічних показників потенційного позичальника оцінювання рівня ймовірності настання його дефолту. Банк може визначити безпосередньо набір факторів, які прямо створюють передумови для повернення або неповернення кредиту, спираючись на проаналізовані та існуючі статистичні дані «негативних» та «позитивних» кредитів за встановлений або обраний період, а також на основі цих характеристик для кожного нового клієнта визначити його спроможність до повернення кредиту [19].

Так як першочергово необхідно визначити та зібрати характеристики, які безпосередньо відносяться до клієнта, початкові етапи скорингового підходу базуються на експертному висновку. При цьому важливо перевірити дані, які були надані самим клієнтом. Під час співбесіди з позичальником кредитні менеджери збирають якомога

більше інформації про нього та нараховують бали за кожен сприятливий фактор. А вже після цього скорингова система обробляє інформацію.

Слід зазначити, що у процесі використання скорингової моделі (кожні 2-3 роки) її необхідно корегувати, так як ситуація змінюється динамічно, можуть з'явитися нові зловмисники, які можуть підрахувати, які саме потрібно фіктивні дані надати банку з тим, щоб скорингова модель визначила, що можна йому надати кредит та ін. Спільне використання цих підходів дозволяє полегшити процес прийняття рішення щодо видачі кредиту клієнтам банків [20].

Більш правильним і коректним може бути підхід, що ґрунтується на статистиці дефолтів та враховує всі попередні періоди, тобто скорингову модель. [21]. Щоб реалізувати цей підхід, слід визначити і передбачити ймовірність реалізації відповідних подій з наступними вимогами:

- для аналізу еквівалентними є ті об'єкти, які використовують статистичні дані та для яких збираються ці дані;
- відповідні умови є еквівалентними, для яких пропонується збирати та використовувати відповідні статистичні дані;
- є достатніми обсяги вибірок статистичних даних, методи їх обробки є коректними, а також джерело інформації є надійним.

Такий підхід висуває до статистики дефолтів високі вимоги, що наведені нижче.

1. Однорідність вибірки, тобто позичальники повинні бути досить схожими.
2. Вибірка повинна включати необхідну кількість випадків, виконуючи такі умови: чим більше буде у вибірки дефолтів, тим для результату краще; як правило, за експертними оцінками та рекомендаціями достатній обсяг вибірки може становити приблизно 2000 випадків, або більше.

3. Побудову моделі необхідно виконувати на вибірці, яка за обмежений термін була накопичена; до цієї вимоги призводить саме зміна макроекономічного середовища. Так позичальник з певним набором параметрів для одного макроекономічного середовища буде вчасно і в повному обсязі виплачувати кредит, а в іншому середовищі це може стати причиною розглядати його як дефолт. Тому

вважається, що для країн, що розвиваються постійно, необхідно кожні 2-3 роки змінювати модель оцінювання ймовірності дефолту.

4. Для отримання правильних результатів для кредитування фізичних осіб необхідно від позичальників просити надати не тільки особисту кредитну історію, тобто оплачені або неоплачені кредити, а також зберігати інформацію про вік, стать, місце роботи, посаду, сімейний стан тощо. Цю процедуру слід виконувати постійно з метою створення бази знань про клієнта та, щоб зрозуміти те, які саме характеристики при прийнятті рішення будуть мати вагу, а які з них для побудованої моделі будуть значущими. Тому спочатку необхідно по кожному клієнту накопичити якомога більше параметрів і характеристик.

5. У вибірку слід включати інформацію по тих кредитах, для яких цикл кредитування вже закінчився. Ця вимога є необхідною, тому що обов'язково необхідна інформація про те, чи був кредит повернений, чи не був повернутий [7].

6. Зазвичай відносять до кредитної історії наступні види продуктів: іпотечний кредит, споживчий кредит, кредит на авто. Для кожного з таких кредитних продуктів необхідно будувати скорингову модель [8].

Таким чином, дані, що стосуються процесу кредитування фізичних осіб, повинні практично задовольняти всім цим вимогам. Параметрами скорингової моделі для фізичних осіб можуть бути: вік, сімейний стан, освіта, робота, посада, кількість дітей, місце проживання, власність, наявність і розмір поточних боргів, кредитна історія, співвідношення кількості поданих заявок і виданих кредитів, тривалість відносин з кредиторами та ін. [22].

### 1.3 Регресійні моделі та регресійний аналіз

Регресійний аналіз являє собою відповідний розділ математичної статистики, в якому розглядаються методи аналізу залежності однієї величини від іншої. Такий підхід застосовується у випадку, коли залежності між змінними можна виразити кількісно як деякі комбінації цих змінних. Комбінації, що отримані, можна використовувати для передбачення значень, які може прийняти цільова (залежна)

змінна за результатами обчислень по заданому набору значень вхідних (незалежних) змінних. Для цього у найпростішому випадку можна використовувати стандартні статистичні методи, наприклад, лінійну регресію [23].

### 1.3.1 Регресійні моделі

Регресійні моделі орієнтовані на виявлення причинно-наслідкових зв'язків між індикаторами, що спостережуються, та рівнем ризику. Виділяють дві групи показників, які можна використовувати як індикатори, що спостережуються, які називають пояснювальними змінними:

- 1) змінні оточення, які є кількісними показниками;
- 2) фактори ризику, які також є кількісними показниками, що характеризують випадки реалізації ризиків, що спостережуються [7].

Така математична модель може бути подана у вигляді:

$$x = A \times f + b + \varepsilon, \quad (1.2)$$

де  $x$  – величина втрат, що пов'язані з кредитним ризиком;

$f$  – вектор значень змінних що спостережуються;

$\varepsilon$  – випадкова величина, яка визначає рівень похибки моделі;

$A$  і  $b$  – параметри, що оцінюються, які характеризують залежність між змінною  $x$  та змінними факторів  $f$ .

Зазначимо, що необхідно мати достатній обсяг даних, щоб застосувати цей метод для забезпечення високої точності оцінок.

### 1.3.2 Лінійна ймовірнісна модель

Лінійна ймовірнісна модель по суті являється регресійною моделлю, в якій залежна змінна може приймати значення 0 або 1, де 0 означає, що кредит затверджено, 1 – не затверджено. Лінійна ймовірнісна модель з математичної точки зору описується наступним правилом прийняття рішення:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad (1.3)$$

де  $y$  – залежна (пояснювана) змінна;

$(x_1, x_2, \dots, x_k)$  – незалежні (екзогенні) змінні;

$(b_1, b_2, \dots, b_k)$  – вагові коефіцієнти, які присвоєні незалежним змінним;

$\varepsilon$  – випадкова похибка, розподіл якої залежить від значень незалежних змінних, враховуючи те, що її математичне очікування дорівнює нулю.

Векторне представлення модель описується у вигляді:

$$y = b' \times x + \varepsilon, \quad (1.4)$$

де  $x$  – вектор незалежних змінних;

$b'$  – транспонований вектор параметрів.

Після того, як коефіцієнти моделі оцінені, її можна застосовувати для оцінювання кредитоспроможності майбутніх запитів [21]. Для прийняття остаточного рішення щодо підтвердження запиту на позику, слід враховувати межу (поріг) відсічення.

Слід зазначити, що лінійна модель є простотою в реалізації та для подальших розрахунків, в той час як не завжди прийнятною є якість моделі. Таким чином, при побудові моделі лінійної регресії необхідно знайти оцінки вектора параметрів  $(b_1, b_2, \dots, b_k)$ , використовуючи експериментальні значення  $y$  та  $(x_1, x_2, \dots, x_k)$ ; а потім застосовувати цю модель для оцінювання кредитоспроможності клієнта.

### 1.3.3 Логіт та пробіт моделі

Оскільки лінійні методи прогнозування являються недосконалыми, це підштовхнуло до аналізу та пошуку більш ефективних методів. Зазначимо, що недосконалість полягає в тому, що оцінка ймовірності не обов'язково може знаходитися в інтервалі  $[0; 1]$ . Це питання можна вирішити, якщо вдасться виконати відповідне перетворення, яке гарантує знаходження оцінки ймовірності в цьому інтервалі [21]. У лінійній моделі залежна змінна  $y$  є бінарною, тобто вона може

приймати тільки два значення і, крім того, є функцією параметрів особи-заявника.

Зазначимо, що кумулятивні функції розподілу є не що інше як набір перетворень, які виконують переведення значення ймовірності в інтервал  $[0; 1]$ . Крім того, ці функції мають властивість монотонності, тобто вони монотонно зростають, або монотонно спадають. Можна припустити, що для вираження ймовірності доцільно використати стандартний нормальний розподіл:

$$p = \Phi(b'x) = \int_{-\infty}^{b'x} \varphi(z) dz, \quad (1.5)$$

де  $\varphi(z)$  – функція щільності нормального розподілу.

У випадку, коли ймовірність має логістичну функцію розподілу, то можна отримати логіт-модель:

$$p = \Phi(b'x) = \int_{-\infty}^{b'x} \varphi(z) dz = \frac{1}{1 + e^{-b'x}}, \quad (1.6)$$

або

$$p = \frac{e^{b_1x_1 + \dots + b_kx_k}}{1 + e^{b_1x_1 + \dots + b_kx_k}}. \quad (1.7)$$

У більшості випадків оцінку вагових коефіцієнтів для обох моделей можна отримати з використанням методу максимальної правдоподібності (ММП), враховуючи те, що його реалізація є відносно простою і не вимагає значних обчислювальних витрат. Крім того, цей метод дозволяє виконувати оцінювання параметрів моделі в автоматичному режимі.

#### 1.3.4 Моделі на основі експертного оцінювання

При дослідженні складних систем, якими є фінансові системи, можуть мати місце випадки, коли виникають проблеми, вирішення яких може знаходитися поза межами формальних математичних задач. Тоді необхідне використання інтелекту людей та їх здібностей для знаходження рішень задач, які є слабо формалізованими. Необхідне проведення експертного оцінювання моделей. Методика проведення такого оцінювання включає в себе наступні етапи:

- формулювання цілей оцінювання;
- постановка задачі;
- створення груп експертів для управління процесом оцінювання;
- опис форм отримання необхідних результатів;
- вибір експертів та визначення їх компетентності.

Вибір експертів здійснюється в такий спосіб, щоб фахівці, що входять до груп, розуміли специфіку задач, не були зацікавлені в результатах оцінювання, а також були ознайомлені з об'єктом прийняття рішень. Тому в групі експертів включають, як правило, 2–3-х зовнішніх експертів та 2–3-х фахівців зі своєї організації. Об'єктивно можна визначити ступінь компетентності експертів та їх кваліфікації у певній області знань по їх професійній та науковій діяльності. Суб'єктивність оцінювання компетентності експертів полягає у взаємному оцінюванні кожним з них кваліфікації своїх колег і своєї за відповідною шкалою. Потім встановлюється компетентність експертної групи після обробки результатів опитування, яка може визначити можливу похибку оцінювання.

Використання експертного підходу передбачає, що в області кредитування індивідуальних позичальників фахівці визначають суттєві характеристики клієнта банку, згідно яких можна впливати на повернення ним кредиту, і виставляють певні ваги у відповідність цим характеристикам. За всіма цими характеристиками фахівці для кожного клієнта проставляються бали згідно їх вагових коефіцієнтів, які були встановлені експертами, і підраховується сума всіх балів.

Як правило, завчасно встановлюється для банку відповідне порогове значення, яке визначає правило, згідно якого, якщо сума всіх балів менше цього порогового значення, то клієнту не слід видавати кредит. Якщо сума всіх балів є більшою, ніж це значення, то тоді клієнт може отримати кредит. Крім того, банк може встановити верхню межу обмеження, яке визначає наступне правило: якщо сума всіх балів більше встановленого значення, то клієнту без надмірної обережності можна видавати кредит. За умов, коли верхня і нижня межі щодо видачі кредиту задані, то тоді значення між цими межами буде знаходитися у компетенції менеджера банку, який визначає на власний розсуд – видавати кредит клієнту, чи ні [21, 22].

### 1.3.5 Байєсівські моделі на основі мереж довіри

Байєсівські мережі довіри (БМ) дозволяють відображати причинно-наслідкові зв'язки між різними впливовими чинниками ризику та змінами середовища. Байєсівські мережі, на відміну від регресійних моделей, дозволяють враховувати не тільки безпосередні залежності рівня ризику від факторів ризику, а також і залежності між факторами ризику. Слід зазначити, клас байєсівських моделей надає більше можливостей на основі неповних даних формувати висновки [23]. Байєсівські мережі довіри з математичної точки зору являють собою орієнтований граф, вершини якого означають чинники ризику та зміни середовища, а ребра графа – це виявлені або передбачувані взаємозв'язки. БМ можна описати множиною умовних розподілів випадкових величин, якими можна охарактеризувати чинники ризику та змінні середовища.

Необхідно розв'язувати дві математичні задачі, коли використовуються байєсівські мережі як інструмент аналізу даних:

- 1) побудова структури БМ;
- 2) формування ймовірнісного висновку.

Слід зазначити, що побудова БМ за заданими навчальними даними є NP-складною задачею, тобто задачею нелінійної поліноміальної складності.

### 1.4 Моделі ринкового ризику

Під ринковим ризиком розуміється ризик втрати позицій, які виникають під час руху ринкових цін. Сьогодні немає єдиної класифікації ринкових ризиків, тому що вона може стосуватися різних його аспектів. Разом з тим найбільш часто використовуються наступні ринкові ризики:

- ризик власного капіталу;
- ризик зміни цін на акції;
- процентний ризик, тобто ризик можливих змін процентних ставок або їх нестабільність;

- валютний ризик, або ризик зміни валютних курсів;
- товарний ризик, або ризик, що пов'язаний зі зміною цін на товари;
- маржинальний ризик, який може бути наслідком невизначених майбутніх відтоків грошових коштів, що викликані маржинальними викликами, які повинні покривати несприятливі зміни цінності даної позиції.

Ринковий ризик вважають ризиком втрат в торговельній книзі банку, які мають місце через зміни цін на акції, кредитні спреди, процентні ставки, цін на сировинні товари, курси валют та інших показників, значення яких зазвичай встановлюються на публічному ринку. На сучасному етапі для менеджменту та аналізу ринкового ризику банки застосовують різні досить складні математичні та статистичні методи, за допомогою яких забезпечується створення адекватних математичних моделей та обчислення високоякісних оцінок. Розглянемо їх.

#### 1.4.1 Value-at-Risk

Value-at-Risk (VaR) оцінюється у грошовому вираженні як кількісна міра потенційного збитку згідно позицій фактичної вартості активів внаслідок руху ринку. Ця міра не повинна перевищуватися з певним рівнем достовірності протягом певного періоду часу. Для торгових підприємств важливість цього ризику базується на їх власній внутрішній моделі. Згідно рекомендацій [23] для розрахунку регулятивного капіталу була затверджена внутрішня модель для загальних та специфічних ринкових ризиків, яка постійно удосконалювалася.

VaR характеризується трьома параметрами, які наведені нижче:

1. Часовий горизонт, що залежить від розглядаємої ситуації, яка за Базельським документом складає 10 днів, а за іншими методиками це може бути 1 день. Термін 10 днів використовуються при розрахунках величини капіталу, який покриває збитки. Але найчастіше використовується розрахунок з часовим горизонтом 1 день.

2. Довірчий рівень (confidence level), який є рівнем допустимого ризику. За Базельським документом слід використовувати величину 99%.

### 3. Базова валюта, в якій найкраще можна вимірювати показник ризику.

Методика аналізу VaR за останні роки стала нормативним і галузевим стандартом для вимірювання ринкового ризику. Останнім часом вимоги, які пред'являються до VAR і до інших методів аналізу дуже зросли, завдячуючи новим продуктам, а саме таким як: кореляційна торгівля, реверсивні двовалютні свопи, опціони з багатьма активами, номінальна вартість яких амортизується непередбачувано, а також десятки інших нових інновацій.

Фактори ризику, які в торговельній книзі глобальних установ необхідні для аналізу ціноутворення, зросли до десятків тисяч. Крім того, моделі, що використовуються для оцінювання можливих втрат, стають все більш складними. Тому більшість банків натеper знаходяться в процесі інтеграції з новою аналітикою стрес-тестування, згідно якої передбачається широкий спектр макроекономічних змін.

Використання VaR та інших методик аналізу ризику, незважаючи на існуючі досягнення, можуть призводити до некоректних результатів. Фінансова криза, що відбулася 2008 року, продемонструвала певні обмеження моделювання ризиків [24]. Відзначимо, що напередодні початку європейської суверенної кризи у 2011 році моделі ризиків більшості банків Єврозони розглядали державні облігації як практично безризикові.

Було вказано на залежність VAR від нормального розподілу ринкових процесів, а також на фундаментальне припущення, що їх позиції можуть бути легко ліквідовані. Деякі з цих обмежень органи регуляції намагалися компенсувати, наприклад, через Базель II.5, а також було зроблене комплексне оновлення системи ринкових ризиків, що вступила в грудні 2011 року в силу. Деякі елементи в структурі нової методики, зокрема, вимога до розрахунку стресового значення VAR, дозволили підвищити активи (RWA), зважені з урахуванням ризику, а також підвищити в два-три рази вимоги до об'єму резервного капіталу.

Більш безпечною може зробити фінансову систему введення більш високих вимог до капіталу, але це досить недосконалий інструмент з точки зору моделювання. Корисним доповненням до основної методики VaR можуть бути

тривалі удосконалення в реалізацію стрес-тестування. Разом з тим, більшість банків згодні з тим, що вимагають подальшого удосконалення майже всі моделі оцінювання ризику [25]. Для банків цікавими є варіанти, які пов'язані з адекватним оцінюванням і моделюванням, що дозволяє отримати необхідний баланс між точністю оцінювання і системністю з одного боку, а також прозорістю, простотою і швидкодією процедур з іншого. Важливим є наявність високоякісних ринкових даних, так само як і самих моделей. Разом з тим багато банків не впевнені, де можна провести межу, яка буде між прийнятним і неприйнятним рівнями якості.

Існує декілька способів оцінювання VaR:

- параметричний, коли оцінка виконується за припущенням, що відомий розподіл прибутків (найчастіше розподіл передбачається нормальним або логнормальним);
- непараметричний (історичний), коли розподіл прибутку можна взяти з реалізованого часового ряду, тобто передбачається, що в майбутньому розподіл прибутку також буде аналогічний історичному;
- метод Монте-Карло.

Практично використовуються зазвичай два параметричних методи розрахунку Value-at-Risk: дельта-нормальний VaR та дельта-гамма VaR.

Найбільше поширення серед параметричних методів розрахунку VaR отримав дельта-нормальний метод [25, 26], при застосуванні якого використовується припущення про нормальність розподілу всіх ринкових чинників які можуть впливати на вартість портфеля, а також за складовими портфеля про лінійність зв'язку між змінами факторів ризику та фінансовими результатами. Тоді результат по портфелю буде являтися сумою нормально розподілених величин, тобто результат буде також нормально розподіленою величиною.

Згідно з дельта-нормальним методом значення Value-at-Risk можна розрахувати в такий спосіб:

$$VaR = K \sqrt{\sum_{i,j=1}^N \sigma_{ij} D_i D_j}, \quad (1.8)$$

де  $D_i$  – дельта-чутливість портфеля до  $i$ -го фактору ризику (розраховується за

всіма складовими портфеля як сума коефіцієнтів лінійного зв'язку за  $i$ -м фактором результатів;

$K$  – коефіцієнт, що залежить від обраного інтервалу довіри, і показує у скільки разів втрати для заданої ймовірності довіри більше, ніж стандартне відхилення нормального розподілу;

$\sigma_{ij}$  – коваріація  $i$ -го та  $j$ -го факторів ризику;

$N$  – кількість факторів ризику.

Отже, щоб використовувати даний метод, необхідно знати матрицю коваріацій ринкових факторів, при цьому волатильність окремих факторів буде враховуватися в матриці як коваріація факторів самих з собою. Дану матрицю можна отримати як на основі прогнозів, так і на основі історичних даних.

Два значення коефіцієнта  $K$ , які найбільш часто використовуються, такі: 2,33 для ймовірності 0,99 та 1,65 для ймовірності 0,95.

Зазначимо, що дельта-нормальний метод має наступні переваги: швидкість обчислень; відносна простота реалізації; він дозволяє використання різних варіантів значень кореляцій та волатильності.

Можна відзначити наступні недоліки дельта-нормального методу [26]:

- неможливість використання інших розподілів, крім нормального, тому що неможливо враховувати «важкі хвости»;
- при використанні нелінійних інструментів неможливо коректно врахувати ризику;
- у використовуваних моделях є імовірність появи суттєво значущих похибок.

Слід зазначити, що метод Монте-Карло являється найскладнішим методом, який можна використати для розрахунку VaR. Але даний підхід, незважаючи на складність, дозволяє отримати точність вище, ніж іншими методами. Використовуючи метод Монте-Карло, можна зробити значну кількість випробувань, тобто будувати разові моделі для різних ситуацій на ринку, враховуючи розрахунки результатів по портфелю. Отримані розподіли можливих фінансових результатів будуть результатом проведення таких випробувань, а також втрати, на основі яких можна отримати VaR-оцінки, відсікаючи найгірші результати для обраної

ймовірності довіри.

Застосування методу Монте-Карло, як і параметричного VaR, дозволяє побудувати такі моделі:

- модель залежності вартості фінансового результату від змін факторів ризику по портфелю;
- модель кореляцій і волатильності факторів ризику.

Щоб отримати аналітичну оцінку портфеля в цілому, метод Монте-Карло не реалізує узагальнення формул, ось чому для отримання результатів згідно портфелю, а також для кореляцій і волатильності слід використовувати більш складніші моделі. Так як оцінювання VaR з використанням методу Монте-Карло майже завжди проводиться з застосуванням програмних засобів, то такі моделі слід реалізовувати складними підпрограмами. Таким чином, метод Монте-Карло може використовуватися при розрахунку ризиків практично будь-якої складності [26]

Метод Монте-Карло має наступні переваги [26]:

- можна розраховувати ризики для нелінійних інструментів;
- можна використовувати будь-які розподіли ймовірностей;
- можна моделювати складні поведінки ринків: тренди, кластери високої або низької волатильності, нестационарні кореляції між факторами ризику, сценарії типу «що-якщо» та ін.

- можливий подальший, нічим практично не обмежений, розвиток моделей.

Методу Монте-Карло притаманні наступні недоліки [26]:

- висока складність реалізації;
- вимогу до великої потужності обчислювальних ресурсів;
- імовірність наявності у використовуваних моделях значущих помилок.

#### 1.4.2 Stressed Value-at-Risk

Stressed Value-at-Risk використовується для розрахунку вартості під ризиком за умов наявності на ринку значного стресу протягом одного року. Величина такого ризику обчислюється при використанні рівня достовірності 99%. При цьому термін

зберігання результатів становить для внутрішніх цілей один день, а для цілей регулювання – десять днів. В якості вихідних даних для моделювання методом Монте-Карло використовуються історичні ринкові дані, а також спостереження за кореляцією за термін значного фінансового стресу. В цьому випадку фінансові процеси характеризуються високою волатильністю.

У випадку, коли основні фактори оцінювання ризику характеризується високим рівнем волатильності, процес вибору часового вікна ґрунтується на ідентифікації вікна для розрахунку стресового значення ризику. При управлінні ризиками стрес-тестування включає в себе часто дуже складні, генеровані комп'ютерами імітаційні моделі, які можуть використовувати в якості тестової бази різні сценарії. При цьому проводиться аналіз того, як баланс організації відповідає конкретним ситуаціям.

Наприклад, фінансові організації можуть використовувати ці моделі, щоб аналізувати ринковий та портфельний ризики, а також приймати обґрунтовані рішення у період невизначеності на основі отриманих результатів. Після виконання стрес-тестів у менеджменті ризиками наявність високоякісних даних у своєму розпорядженні дозволяє фінансовим організаціям відчувати себе більш ефективними. Це дозволяє також пом'якшувати ризики та виявляти на ранніх стадіях проблеми.

Умови, коли великі банки надмірно сприймали ризики, відсутня прозорість, наявність неспроможності фінансового регулювання та суб-стандартного кредитування призвели до катастрофічної ситуації, наслідком чого стали значні зміни у фінансовому регулюванні. Це призвело до значних зрушень у банківському регулюванні та нагляді, особливо регуляторної звітності та звітності стосовно менеджменту ринкових ризиків.

Було переглянуто систему правил Базель II та посилена рамками Базеля III. Для розширення регуляторної сфери управління ризиками був розроблений Базель III, який фінансовим організаціям гарантував юридичне право проводити комплексне стрес-тестування, що їм дозволило керувати ризиками та визначати здатність їх виживати в реальних фінансових умовах.

Згідно з Базельськими правилами з липня 2018 року, банки повинні проводити стрес-тестування. Крім того, повинна повністю документуватися інформація про те, в який спосіб були отримані результати стрес-тестів. Для управління балансом фінансовим організаціям необхідне єдине, скоординоване рішення щодо стрес-тестування. Це забезпечить більш ефективне управління процесом стрес-тестування, а також пом'якшить ризики та забезпечить постійне дотримання нормативних вимог.

Стрес-тестування має переваги рішення. Існує конкретна потреба в комплексному стрес-тестуванні, щоб вирішувати проблеми управління ризиками, так як нові вимоги до стрес-тестування пов'язані з управлінням і мають тенденцію оновлюватися. Слід зазначити, що для великих фінансових організацій стрес-тестування не є новим. Новими є рівні прозорості та складності, які вимагають тести.

Вимоги до коректного стрес-тестування та чинні положення показують, що для отримання необхідних результатів фінансові організації мають бути послідовними та прозорими. Тобто, надання даних, централізація даних, їх агрегація та консолідація є ключовими аспектами, що необхідно обов'язково враховувати при додержанні нормативних вимог.

У прагненні до прозорості та точності деякі фінансові організації стикаються з проблемами, якими є неефективна інфраструктура, старі системи, а також проблеми щодо впровадження та звітування про моделі ризику. Неефективна інфраструктура може призвести протягом всього життєвого циклу моделі до відключення деяких процесів, що в свою чергу може призвести до невідповідності стрес-тестування результатам. Старі системи можуть призвести до непослідовного використання та визначення даних, що може ускладнити їх консолідацію, а також розробку повторюваного і послідовного процесу їх аналізу.

Нарешті, можливі ситуації, коли фінансові організації не володіють добре структурованим, документованим та прозорим процесом, який повинен відповідати нормативним вимогам. В цьому випадку результати стрес-тестування будуть помилковими та фрагментованими. Банківське управління ризиками може проаналізувати результати стрес-тестування на одній централізованій платформі. Проте, фінансові організації можуть виконувати індивідуальні та конкретні

фінансові сценарії, маючи рішення стосовно стрес-тестування. Ці сценарії повинні узгоджуватися з останніми правилами менеджменту ризиків та використовуватись для оцінювання ризику портфеля, аналізу тверджень «що-якщо», та планування управління капіталом.

Отримані в результаті стрес-тестування рішення можуть використовувати всі учасники, а також вони можуть працювати з повністю інтегрованим, однаковою і узгодженим джерелом даних, яке може включати логічні структури даних та їх належну ієрархію, що забезпечено надійністю, точністю та повторюваністю стрес-тестування. Інформація є легко доступною для пошуку та аналізу в будь-який момент процесу стрес-тестування, забезпечуючи при цьому необхідну прозорість.

## 2 МОДЕЛЮВАННЯ КРЕДИТНИХ РИЗИКІВ

Банк зобов'язаний у випадку надання кредитів дотримуватись основних принципів кредитування, тобто, дотримуватись вимог, встановлених банком, щодо концентрації ризиків, а також проводити аналіз кредитоспроможності позичальників. При видачі кредиту вагомим моментом являється кредитоспроможність та її загальна оцінка. Крім того, важливим є гарантування платоспроможності фізичної особи, яке може бути представлене платою поручителів, високим рівнем заробітної плати, або заставою особистого майна. Але гарантія для юридичних осіб не завжди являється повною мірою, тому що можуть виникати обхідні варіанти, а суми позик зростати.

Слід зазначити, що різні фактори фінансового стану можуть впливати на кредитоспроможність позичальника, якими можуть бути ліквідні активи, власні кошти, або фінансова дисциплінованість. Оцінку кредитоспроможності можна виконати в два етапи: проведення аналізу ділового ризику та проведення аналізу фінансового стану позичальника, згідно якого слід враховувати грошові потоки і фінансові коефіцієнти.

### 2.1 Оцінювання інформативності змінних

Однією з проблем при побудові моделей є проблема відбору вхідних ознак, значущих для моделі. Для зменшення розмірності моделі необхідне зменшення числа незалежних змінних з метою видалення з неї всі незначущих ознак, які не несуть в собі інформації, корисної для аналізу, і таким чином, спростити модель та видалити надлишкові ознаки. Слід зазначити, що дублювання інформації в надлишкових ознаках не тільки не покращує якість моделі, а може її погіршити, наприклад, коли має місце мультиколінеарність.

Одним з можливих варіантів виходів з такої ситуації можна розглядати побудову моделі для всіх можливих комбінацій набору вхідних ознак. Подальший відбір необхідного варіанту містив би мінімум незалежних змінних і дозволив отримати найкращу описову здатність результуючої ознаки  $y$ . Але таке рішення на

включення у модель може бути тільки за наявності невеликої кількості факторів-претендентів. Коли має місце великий список потенційних ознак, використання подібної методики представляється досить складним, так як кількість моделей, які треба побудувати, буде досить великою. Враховуючи для відбору найбільш важливих факторів це слід використовувати додаткові алгоритми, для реалізації яких необхідно значно менше витрат часу на обчислення.

### 2.1.1 Вибір підмножини ознак

Згідно методу відбору ознак необхідно знайти підмножину вхідних змінних, які являються атрибутами або ознаками. Можна використати три стратегії:

- 1) стратегію фільтра, наприклад, при накопиченні ознак;
- 2) стратегію обгортання, наприклад, при пошуку заданої точності;
- 3) стратегію вкладення, коли при побудові моделей, які засновані на похибках прогнозування, вибираються ознаки, щоб виконати додавання або видалення.

Можна використати наступні підходи для відбору ознак [21, 22]:

- зробити моделі більш простими для інтерпретації користувачами;
- зменшити час на навчання (тренування) моделі;
- зменшити розмірність вхідних даних;
- шляхом скорочення перенавчання покращити узагальнення, тобто зменшити дисперсії.

Застосування різних технік відбору ознак базується на тому, що в даних містяться деякі ознаки, які при формуванні результату являються зайвими або значущими, тому їх можна видалити без великих втрат інформації [24]. Необхідно відрізнити техніки відбору ознак від виділення ознак.

Виділення ознак дозволяє створювати нові ознаки у вигляді функцій від оригінальних ознак, а відбір ознак дозволяє повертати їх підмножину. У випадку, коли є велика кількість атрибутів, а точок даних для вибірки порівняно малі, використовуються техніки відбору ознак.

Для представлення нової підмножини ознак комбінація технік пошуку разом з

обчисленням критеріїв може розглядатися як алгоритм відбору ознак, що відображають відповідні підмножини. Найпростішим алгоритмом являється проведення перевірки кожної можливої підмножини ознак та знаходження серед них такої, яка дозволяє мінімізувати величину похибки. Цей алгоритм являється вичерпним для пошуку по простору, а також обчислювально важким для наборів ознак, які не являються дуже малими. На алгоритм сильно впливає вибір метрики, які являються різними для основних трьох категорій алгоритмів відбору [25].

Для оцінювання наборів ознак можна використати методи обгортання, які застосовують моделі апріорної оцінки результату. Для навчання моделі використовується кожний новий набір та проводиться перевірка на контрольній вибірці, використання якої дозволяє визначити число помилок моделі та отримати оцінку для даної підмножини. З обчислювальної точки зору методи обгортання потребують значних затрат, так як вони тренують модель для кожного набору, але для конкретного типу моделі, як правило, вони дають кращий набір ознак.

Замість показника помилки метод фільтрів для оцінювання набору ознак використовує непрямий показник, який вибирається таким чином, щоб при збереженні показника корисності набору ознак його можна було б легко обчислити. Для цього можна застосовувати наступні показники: коефіцієнт кореляції змішаних моментів Пірсона, поточна взаємна інформація [25], взаємна інформація [24], алгоритм на основі показника Relief [26], відстань між класами або всередині класу; а також критерії значимості для кожної комбінації класів або ознак [25, 27]. Метод фільтрів обчислювально менш інтенсивний, але він дозволяє отримати набори ознак, які не орієнтовані на специфічний тип прогнозуючої моделі [28].

Методи вкладення відносяться до узагальнюючої групи технік, що здійснюють як частину процесу побудову моделі відбору ознак. Метод оцінювання коефіцієнтів лінійної регресійної моделі (Least absolute shrinkage and selection operator, LASSO) являється прикладом такого підходу для побудови лінійної моделі, згідно якого накладається штраф L1 на коефіцієнти регресії, що дозволяє багато з них зменшити до нуля. Алгоритмом LASSO «вибираються» ознаки, для яких мають місце ненульові коефіцієнти регресії [29].

## 2.1.2 Виконання загального статистичного тесту

На тепер можливе опрацювання множини числових показників, які дозволяють аналізувати та вимірювати ступінь взаємозв'язків між двома коефіцієнтами зв'язків як зі змінними. Для оцінювання інформативності моделей таким показником являється критерій  $\chi^2$  та загальний статистичний тест. Особливістю використання цього підходу є знаходження залежностей між екзогенними та цільовою змінними.

Цей критерій було запропоновано Карлом Пірсоном у 1990 році для оцінювання близькості теоретичних розподілів до емпіричних. Критерій  $\chi^2$  може бути використаний у випадку потреби встановлення відповідності між двома порівнюваними рядами розподілу – теоретичного і емпіричного, або між двома емпіричними. Тоді необхідно порівняти частоти названих рядів розподілу, виявити розбіжності між ними та визначити ймовірність цих розбіжностей [27].

Використовуючи цей критерій в розподілах двох емпіричних рядів виявляються відмінності, виконується порівняння вибірок, які можуть мати альтернативні ознаки та виконується оцінювання ймовірностей кореляції між альтернативними ознаками. Критерій  $\chi^2$ , як і інші критерії згоди, що представлені в [27], представляє собою деяку величину, яку можна оцінити з відповідною ймовірністю. Згідно  $\chi^2$ -критерію можна здійснювати статистичну перевірку гіпотез відносно певних розподілів та знаходити відповідність емпіричних даних до деякому розподілу теоретичного закону розподілу.

Отримана оцінка наближення емпіричного розподілу до теоретичного дозволяє отримати суму співвідношень частот. Величину  $\chi^2=0$  також зумовлює збіг теоретичних і емпіричних частот. Наведене являється підтвердженням нульової гіпотези ( $H_0$ ). Величина  $\chi^2$  може також свідчити про неправильність висунутої гіпотези за умов, коли наявною є достовірність різниці в частотах теоретичного і емпіричного ряду [28].

У статистиці сьогодні розподіл  $\chi^2$  являється одним з найбільш широко використовуваним. Зазвичай його можна використовувати при перевірці статистичних гіпотез щодо ймовірного закону невідомого розподілу, а також для

перевірки гіпотез щодо різноманітних розподілів, які розраховується за виразом:

$$X^2 = \sum_{j=1}^n \left( \frac{f_{\epsilon_j} - f^+}{f^+} \right)^2, \quad (2.1)$$

де  $f_{\epsilon_j}$  – емпіричні значення;

$f^+$  – теоретичні значення;

$n$  – число ступенів свободи.

Для перевірки гіпотези слід провести порівняльний аналіз емпіричних (спостережуваних) частот з теоретичними частотами, які вираховані у припущенні, що розподіл, наприклад, нормальний [29].

Якщо має місце повне узгодження частот, які вираховані або очікувані, з емпіричними частотами, то критерій  $\chi^2$  прирівнюється до нуля. Інакше, коли критерій  $\chi^2$  не дорівнює нулю, це вказує на невідповідність вирахованих частот емпіричним частотам ряду. У такому випадку слід оцінювати значимість критерію  $\chi^2$ , який може теоретично змінюватися в межах від нуля до нескінченності. Це можна здійснити порівнянням фактично отриманої величини з її критичним значенням.

Слід зазначити, що критерій  $\chi^2$ , незалежно від того, розподілені частоти нормально чи ні, дозволяє виконувати порівняння розподілів частот. Із частотами виникнення події мають справу у випадках, коли змінні вимірюються в шкалах найменувань або в інших характеристиках, окрім частот, які неможливо або проблематично підібрати, тобто у випадках, коли змінні мають якісні характеристики [34].

### 2.1.3 Рекурсивне виключення змінних

Вибір функції для класифікації з високою розмірністю та невеликими навчальними вибірками відіграє важливу роль, щоб мати можливість уникати проблем із накладанням та отримати покращенні результатів класифікації. Одним найбільш часто використовуваним методом відбору функцій при наявності

невеликих зразків являється метод рекурсивного усунення ознак (RFE). Цей метод вбудований в алгоритми типу SVM (Support Vector Machine) та застосовує можливість узагальнення, тому може використовуватися для задач з невеликими вибірками. Метод RFE, хоча і має хороші показники, намагається відкинути «слабкі» характеристики з метою забезпечення значного підвищення продуктивності у поєднанні з іншими функціями.

Вибір критеріїв здійснюється шляхом вибору підмножини відповідних функцій із набору визначених та оригінальних критеріїв, таких як роздільність класу або ефективність класифікації. Тому цей метод широко використовується у програмах машинного навчання. При наявності невеликих проблем класифікації вибірок вибір особливостей являється дуже важливим. В умовах, коли кількість навчальних зразків, що доступні, порівняно з кількістю ознак є дуже малою. Для багатьох практичних задач поширеною є проблем, коли має місце невелика класифікаційна вибірка. Так, наприклад, отримують дані великих обсягів від мультигіперспектральних датчиків при автоматичному розпізнаванні цілі [30]. Подолання проблем із придатністю отриманих результатів з метою покращення показників прогнозування являється основною перевагою вибору ознак в задачах класифікації на основі невеликих вибірок [29].

Слід зазначити, що рекурсивне усунення ознак (RFE) серед різних методів вибору ознак являється розробленим методом вибору ознак, що застосований для проблем класифікації за невеликими вибірками [30]. Спочатку метод RFE застосовувався для класифікації кредитоспроможності, коли об'єм навчальних даних був меншим 100, а кількість ознак була більше 10, а потім став ефективним підходом до відбору ознак за невеликою вибіркою. Метод RFE спрямовувався на покращення ефективності узагальнення шляхом видалення найменш важливих функцій, що дозволить досягти найменшого впливу на помилки навчання. Крім того, метод RFE тісно пов'язаний з методом SVM, який, як було вказано, добре узагальнює результати за умов класифікації на невеликих вибірках [29].

Зазначимо, що RFE-метод являється важливим і перспективним при розв'язанні задач обробки невеликих вибірок. Але, незважаючи на це, він

намагається видаляти зайві і слабкі функції та зберігати незалежні функції. Імовірно, що функції, які є надлишковими, зможуть забезпечити краще розділення класів. Наприклад, використання двох слабких функцій, які не мають самі по собі великого значення, зможе забезпечити при спільному використанні краще поліпшення результатів. А просте видалення слабких або зайвих елементів може привести до погіршення ефективності класифікації.

Дискримінант функції в задачі лінійного розділення має вигляд:

$$g(x_i) = w \times x_i + b, \quad (2.2)$$

де  $b$  – термін зміщення;

$w$  – ваговий вектор;

$x_i$  – навчальні дані,  $x_i \in R^n, i = 1, \dots, m$ .

Можна ввести зміщені змінні  $\xi$  для лінійно нероздільного випадку, що дозволить вимірювати відхилення точки даних від оптимальної гіперплощини:

$$\Phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i, \quad (2.3)$$

$$y_i(w \times x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad (2.4)$$

де  $y$  відповідає цільовій змінній:

$$y_i = \{\pm 1\}, i = 1, \dots, m. \quad (2.5)$$

За рахунок застосування двоїстого підходу задача оптимізації вирішується в такий спосіб:

$$y_i(w \times x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad (2.6)$$

$$w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y_i y_j \alpha_i \alpha_j (x_i \times x_j), \quad (2.7)$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, m, \quad (2.8)$$

$$\sum_{i=1}^m \alpha_i y_i = 0, \quad (2.9)$$

де  $\alpha_i$  – коефіцієнти Лагранжа.

RFE-метод зберігає найменш неважливі функції, які являються найменш

важливою ознакою, що пов'язана з найменшим значенням ваги в  $w$ . Ваговий вектор для лінійних випадків обчислюється так:

$$w = \sum_{k \in SV} y_k \alpha_k x_k . \quad (2.10)$$

Обчислення для нелінійних випадків виконуються з використанням рівняння:

$$w_i = \frac{1}{2} \alpha^T K \alpha - \frac{1}{2} \alpha^T K(-i) \alpha , \quad (2.11)$$

де  $K(-i)$  вказує шляхом видалення  $i$ -ї функції на кожному вході  $x$  на обчислену матрицю ядра.

Метод RFE на кожному кроці видаляє рекурсивно функції та повторно використовує решту елементів, перенавчаючи на основі інших функцій ці функції. У разі слабкої функції RFE її видалить. Разом з тим слабка функція при спільному використанні з іншими функціями може виявитися важливим елементом. Таким чином, просте видалення слабких або зайвих ознак може призвести до погіршення ефективності класифікації. З метою оцінення важливості потенційно слабкої ознаки слід оцінити ефективність класифікації після того, коли ця ознака з точки зору відповідного значення вперше буде видалена. Якщо ефективність класифікації після видалення цієї функції буде знижуватися, ця функція буде збережена, хоч вона і має найменше значення. Цей процес необхідно повторювати до тих пір, поки не встановиться результат, який вкаже, що без цієї функції ефективність класифікації не буде погіршуватися [30].

#### 2.1.4 Оцінка значущості результатів

Для розв'язання задачі кредитоспроможності платників найбільш перспективним підходом являється використання дерева рішень, яке дозволяє представити правила у вигляді ієрархічної, послідовної структури, в якій кожен об'єкт відповідає єдиному користувачеві, який приймає рішення. Під правилом мається на увазі логічна конструкція, яка представлена у вигляді «якщо ... то ...».

На тепер існує велика кількість алгоритмів, які застосовуються активно для

розв'язання різних задач класифікації: CART, NewId, C4.5, CHAID, CN2, ITrule та ін. [31].

До так званих «жадібних алгоритмів» відноситься велика кількість відомих алгоритмів. Якщо було обрано один раз атрибут, згідно якого було виконано розбиття на підмножини, то алгоритм «не може» вибрати інший атрибут і повернутися назад, який би не був більш значущим атрибут для отримання результату. Разом з тим неможливо сказати на етапі побудови дерева, чи можливе отримання оптимального розбиття.

Для побудови дерева рішень розглянемо алгоритм C4.5, для якого не обмеженою є кількість потомків для вузла. Використання цього алгоритму дозволяє вирішувати тільки завдання класифікації, так як він «не може» працювати з постійним цільовим полем. Для розв'язання поставленої задачі слід, по-перше, внести зміни до процедури розбиття даних згідно знань постійного типу; а по-друге, ввести поняття "значущості" для вхідних атрибутів, що саме головне, та визначити для її розщеплення формулу.

Алгоритм розбиття по значенню неперервного типу наступний:

1) упорядкувати по зростанню всі знання і дані.

2) вихідну множину  $T$  розбити на дві множини:  $T_1$  і  $T_2$ . На першій ітерації в множину  $T_1$  підпаде один лише первинний елемент, а всі інші елементи – до  $T_2$ . На наступній ітерації з  $T_2$  первинний елемент попадає в  $T_1$  і т.д.

3) для кожної множини з  $T$  обчислити індекс  $Gini_{split}$ . Обрати ту множину з  $T$ , для якої індекс  $Gini_{split}$  буде мінімальним. При цьому можуть використовуватися наступні варіанти обчислень:

$$Gini(T) = 1 - \sum_{i=1}^n p_i^2, \quad (2.12)$$

$$Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \frac{N_2}{N} Gini(T_2), \quad (2.13)$$

де  $p_i$  – імовірність знаходження прикладу класу  $i$  у множині  $T$ ;

$N$  – кількість даних у множині  $T$ ;

$N_1$  та  $N_2$  – кількість даних (значень) у множинах  $T_1$  та  $T_2$  відповідно.

4) далі припиняється розбиття вузла при виконанні однієї з умов:

- достатня кількість значень уже міститься у вузлі (параметр налаштування);
- вузол містить приклади одного класу;
- кількість прикладів, що нерозпізнані, менше мінімальної кількості

прикладів у вузлі (параметр налаштування).

Введемо тепер поняття «значущості» вхідного атрибута. Будемо розуміти під значимістю атрибута показник, який характеризує те, наскільки вихідний результат сильно залежить від вхідних показників.

Розрахунок значущості описується формулою:

$$Z_m = \frac{\sum_{j=1}^{k_m} (E_{m,j} - \sum_{i=1}^{n_{m,j}} E_{m,j,i} \times \frac{N_{l,j,i}}{N_{l,j}})}{\sum_{l=1}^g \sum_{j=1}^{k_l} (E_{l,j} - \sum_{i=1}^{n_{l,j}} E_{l,j,i} \times \frac{N_{l,j,i}}{N_{l,j}})} \times 100\% , \quad (2.14)$$

де  $g$  – кількість вхідних атрибутів;

$k_l$  – кількість вузлів, що були розбиті за атрибутами  $l$ ;

$E_{l,j}$  – ентропія батьківського вузла, що розбитий за атрибутом  $l$ ;

$E_{l,j,i}$  – ентропія дочірнього вузла для  $j$ -го елемента, що розбитий за атрибутом  $l$ ;

$N_{l,j,i}$ ,  $N_{l,j}$  – кількість прикладів для відповідних вузлів;

$n_{l,j}$  – кількість дочірніх вузлів для  $j$ -го батьківського елемента.

В цілому, для атрибутів обчислення показників значущості можливе лише після побудови дерева класифікаційних правил.

### 2.1.5 Остаточне включення змінних до моделі

Кожен із розглянутих вище методів дозволяє оцінити вплив параметрів на цільову змінну, тому необхідно враховувати оцінку за кожним з методів. Для цього необхідно за результатами оцінювання провести скоринг за кожним з методів, де надається бінарний індикатор для екзогенної змінної кожного з методів, тобто, чи

варто за версією конкретного методу вводити цю змінну в модель:

$$\vec{\mu}_{i,j} = f_j(\vec{x}_i), \quad (2.15)$$

де  $(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$  – вектори екзогенних змінних;

$n$  – кількість екзогенних змінних;

$(f_1, f_2, \dots, f_m)$  – функція методу, що класифікує значущість змінної;

$n$  – кількість екзогенних змінних.

На рисунку 2.1 показана структурна схема процесу автоматичного вибору змінних до моделі.

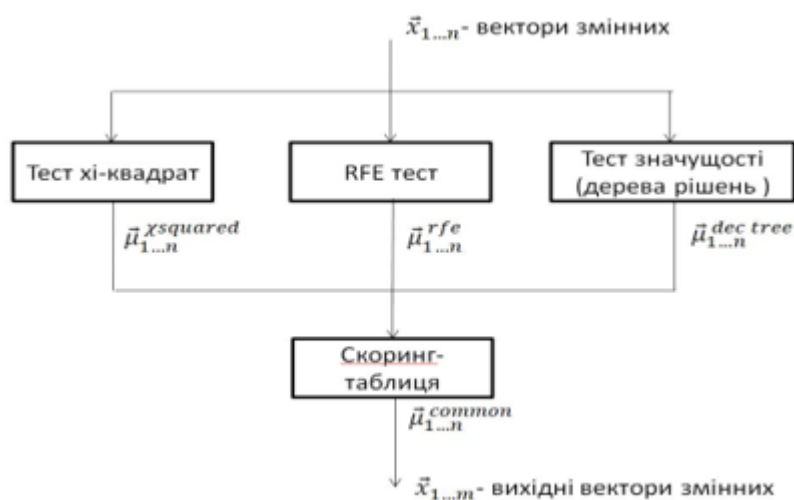


Рисунок 2.1 –Схема процесу остаточного вибору змінних

Таким чином, для кожного використаного методу та для кожної змінної будемо мати кілька бінарних індикаторів. Далі для кожної змінної обчислюється загальна оцінка, згідно якої визначається кількість методів, що обрали відповідну змінну як значущу для моделі. При цьому загальна оцінка буде знаходитися в межах  $[0, n]$ . Якщо значення загальної оцінки буде більше або буде дорівнювати  $n/2$ , то змінна визначається як значуща і автоматично додається в модель.

## 2.2 Оцінювання кредитного ризику з використанням адаптивної мережі Байєса

При застосування мереж Байєса (БМ) як інструменту, що використовується для аналізу даних, необхідно вирішувати відразу дві математичних задачі:

- 1) будується структура БМ;
- 2) формується імовірнісний висновок.

Побудова БМ за навчальними даними являється NP-складною задачею, яка має нелінійну поліноміальну складність. Кількість нециклічних моделей, які можливі та які необхідно проаналізувати, може бути обчислена за рекурентною формулою Робінсона [31]:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \cdot C_n^i \cdot 2^{i(n-i)} \cdot f(n-i), \quad (2.16)$$

де  $n$  – кількість вершин, а  $f(0) = 1$ .

Реально, шляхом практичного виконання повного перебору моделей це можливо зробити лише для мереж не більш ніж з 7 вершинами (вузлами), так як інакше для розв'язання цієї задачі необхідні дуже великі обчислювальні ресурси.

Завданням остаточного формування результату застосування БМ являється імовірнісний висновок, який являється складним і важливим і відноситься до задач прийняття рішень. Щоб реалізувати цей етап, слід структуру БМ привести до так званого об'єднаного дерева (junction tree), а потім слід на об'єднаному дереві використати алгоритм імовірнісного висновку, який заснований на проходженні двох типів повідомлень  $\lambda$  і  $\pi$  по дереву.

Методи, які використовуються при побудові структури БМ, оцінюванні її параметрів та формуванні висновків, вимагають для їх реалізації дуже витратних та трудомістких обчислень. Тому актуальною задачею є створення методів моделювання процесів різної природи із застосуванням ймовірнісних мереж Байєса (або мереж довіри) [32, 33], які дозволяють зменшити обчислювальну складність відповідних процедур.

Задача побудови БМ, тобто оцінювання її структури і параметрів, складається з наступних кроків:

- 1) побудова БМ евристичним методом складається із наступних етапів:
  - на першому етапі виконуються обчислення значень взаємної інформації між усіма вершинами;
  - на другому етапі для оптимальної структури виконується цілеспрямований

пошук, який в якості оціночної функції використовує оцінку мінімальної довжини (ОМД), що аналізується і застосовується на кожній ітерації алгоритму навчання.

2) далі вирішується завдання розробки методу формування (обчислення) імовірнісного висновку в БМ, яке реалізується такими двома кроками:

- на першому кроці обчислюються матриці емпіричних значень для спільного розподілу ймовірностей для всієї мережі;

- на другому кроці обчислюються значення ймовірностей для всіх можливих станів вершин, які є неінстанційованими. Неінстанційованими слід вважати вершини, які, крім апріорно відомої, не отримують додаткової інформації.

БМ являє собою пару  $\langle G, B \rangle$ , в якій компонента  $G$  є спрямований ациклічний граф, в якому випадковим змінним відповідають вершини і в якому формально він записується як множина умов незалежності, тобто кожна змінна в  $G$  незалежна від її батьків.

Друга компонента пари  $B$  являється множиною параметрів, які визначають мережу, тобто, ця множина для кожного можливого значення  $x^{(i)} \in X^{(i)}$  і  $pa(X^{(i)}) \in Pa(X^{(i)})$  містить параметри  $\Theta_{X^{(i)}|pa(X^{(i)})=P(X^{(i)}|pa(X^{(i)}))$ , де  $Pa(X^{(i)})$  та є множиною батьків змінної  $X^{(i)} \in G$ . Кожна змінна  $X^{(i)} \in G$  може бути подана у вигляді вершини. Якщо розглядають більше одного графа для формального представлення задачі, то для визначення на графі батьків деякої змінної  $X^{(i)}$  застосовують позначення  $Pa^G(X^{(i)})$ . Для змінних БМ повна спільна ймовірність або спільний розподіл обчислюється згідно виразу [33]:

$$P_B(X^{(1)}, \dots, X^{(N)}) = \prod_{i=1}^N P_B(X^{(i)} | P_a(X^{(i)})). \quad (2.17)$$

Гібридна БМ  $B=(X,D,P)$  визначається як спрямований ациклічний граф  $G=(X,E)$ , а також його функції  $P_i = \{P(x_i | pa_i)\}$ , в яких  $pa_i$  є множиною батьківських вузлів  $x_i$ .  $X$  є множиною змінних, що розділені на дискретні  $\Delta$  і неперервні  $\Gamma$  змінні, тобто  $X = \Gamma \cup \Delta$ . Структура графа  $G$  являється обмеженою,

тому неперервні змінні не можуть бути дискретно представленими, також як і їх вузли-нащадки. Тоді умовний розподіл неперервних змінних може бути заданий у вигляді лінійної гаусівської моделі:

$$P(x_i / I = i, Z = z) = N(\alpha(i) + \beta(i) \times z, \gamma(i)), \quad x_i \in \Gamma, \quad (2.18)$$

де  $Z$  та  $I$  – відповідно множини неперервних і дискретних батьків  $X_i$ ,

$N(\mu, \sigma)$  – мультиноміальний нормальний розподіл.

Мережа Байеса представляє собою спільний розподіл для усіх його змінних, які задані добутком усіх таблиць умовних ймовірностей, що містять параметри моделі. Таким чином параметрами БМ являються умовні ймовірності у відповідних таблицях.

Можна розглядати формування ймовірнісного висновку за мережею у вигляді процедури перетворення багатовимірних гаусіанів у мережу Байеса. При цьому змінні  $X_1, \dots, X_n$  упорядковуються у певному порядку. Потім цей висновок застосовується для знаходження умовного розподілу:

$$P(X_i / X_1, \dots, X_{i-1}) = N(X_i; \beta_{i,0} + \sum_{j=1}^{i-1} \beta_{i,j} X_j, \sigma_i^2). \quad (2.19)$$

Ребро від  $X_j$  до  $X_i$  ( $1 \leq j < i$ ) створюється тоді і тільки тоді, якщо  $\beta_{i,j} \neq 0$ .

Умовний ймовірнісний розподіл  $X_i$  називають ще спільним розподілом, який після скорочення усіх нульових значень має вигляд певного виразу. Для кореневих вузлів лінійний умовний ймовірнісний розподіл є просто одновимірною гаусіаною.

Мережу Байеса називають багатовимірною лінійною гаусіаною, якщо в ній всі умовні ймовірнісні розподіли є лінійними. У цьому випадку, таким чином, кожна багатовимірна гаусіана може бути подана лінійною гаусіаною. Також справедливим є зворотне твердження. Таким чином, кожна БМ з лінійними умовними ймовірнісними розподілами представляє собою спільний нормальний розподіл.

Умовна лінійна гаусіана являє собою БМ, в якій містяться як дискретні змінні ( $\Delta$ ), так і неперервні змінні ( $\Gamma$ ) з такими обмеженнями [33]:

- дискретний вузол не може мати неперервних батьків, тому для дискретних

вузлів умовні ймовірнісні розподіли можуть бути подані в такий же спосіб, як і в дискретних мережах Байєса;

- умовний ймовірнісний розподіл будь-якої неперервної змінної є лінійним умовним ймовірнісним розподілом, який заданий будь-якою комбінацією дискретних батьків.

Формально, коли вузол  $Y$  має батьків  $\{X_1, \dots, X_k\} \subseteq \Gamma$  і  $D = \{D_1, \dots, D_l\} \subseteq \Delta$ , він може визначатися як умовний ймовірнісний розподіл з використанням наступних параметрів: для кожного  $d \in Dom(D)$ ,  $\beta_{d,0}, \dots, \beta_{d,k}$  і  $\sigma_d^2$ :

$$P(Y / x, d) = N(Y; \beta_{d,0} + \sum_{i=1}^k \beta_{d,i} x_i, \sigma_d^2). \quad (2.20)$$

Цей вид гібридних мережевих моделей являється найбільш популярний. Дані моделі дозволяють відтворити лише лінійні відношення між неперервними змінними а також не передбачують те, що дискретні змінні можуть мати неперервних батьків. Завдячуючи математичній зручності ці моделі отримали широке застосування.

В умовній лінійній гаусіані можна у вигляді багатовимірної гаусіани задати значення дискретних змінних та їх розподіл. Спільний розподіл можна представити у вигляді композиції гаусіан, які дозволяють працювати з ними, застосовуючи аналітичні інструменти. За умов, що всі дискретні змінні будуть задані, тоді умовні ймовірнісні розподіли неперервних змінних будуть являти собою лінійні умовні неперервні розподіли. За умов надання будь-яких значень дискретним змінним лінійна гаусіана та умовна лінійна гаусіана будуть представляють собою нормальний розподіл. З цього слідує, що спільний розподіл, який представлений умовною лінійною гаусіаною, буде являтися композицією цих гаусіан, кожна з яких буде відповідати реалізації дискретних змінних.

Як слідує з наведеного, використання БМ дозволяє проводити аналіз причинно-наслідкових зв'язків між окремими змінними, подіями та даними, а також на цій основі формулювати обґрунтований ймовірнісний висновок або прогноз. Використання гібридних БМ дозволяє коректно проводити у моделі аналіз неперервних змінних (вік чи суму кредиту) і досягати значно вищої точності

остаточного результату [33].

Розглянемо за допомогою БМ реалізацію алгоритма прогнозування кредитоспроможності індивідуального позичальника. Допустимо, що за рік банк видав 1600 кредитів. Кожний клієнт описується набором із 18 характеристик (вік, сімейний стан, сума кредиту, наявність заборгованості за іншими кредитами, рівень заробітної плати та ін.).

За отриманими даними будується БМ, яка вказує на зв'язок між характеристиками клієнта і вершиною, що є подією повернення кредиту. Визначається по навчальним даним ймовірність повернення новим клієнтом кредиту, який прийшов вже до банку. Необхідно визначити ймовірність дефолту  $PD_i$  потенційного позичальника. Так як числові характеристики (вік, сума кредиту, дохід та ін.) можуть мати багато значень, тобто є неперервними, то для розв'язання даної задачі слід застосовувати гібридні мережі Байеса. Згідно описаної вище методики слід здійснити детальний аналіз проблеми та зробити формалізовану постановку задачі. Зупинимося на детальному розгляді схеми застосування цієї методики.

*Крок 1.* Щоб розв'язати задачу оцінювання кредитоспроможності позичальника, слід зібрати за виданими кредитами необхідні статистичні дані, одна частина з яких вже була повернута, а друга частина виявилась дефолтами. Таким чином були отримати негативні та позитивні приклади навчальної бази даних. Використовуючи вибрані параметри позичальника та кредитів, необхідно створити формальну модель та здійснити оцінку ймовірності дефолту позичальника  $PD_i$ :

$$PD_i = F(w^j, x_i^j), \quad (2.21)$$

де  $w^j$  – ваги параметрів  $x_i^j$ ,

$i$  – кількість позичальників,

$j$  – кількість параметрів кредиту.

Для оцінювання кредитоспроможності модель на базі БМ описується наступним чином:

$$PD_i = F(v_i^k, G, J) = 1 - PR_i, \quad (2.22)$$

де  $v_i^k$  – змінні, які описують характеристики клієнта і кредиту;

$J$  – імовірнісний розподіл змінних  $v_i^k$ ;

$G$  – ациклічний спрямований граф, де вузли відповідають випадковим змінним  $v_i^k$  модельованого процесу;

$PR_i$  – ймовірність повернення кредиту.

*Крок 2.* Дані, які необхідні, щоб розв’язати задачу, являються початковими статистичними даними за  $N$  виданих кредитів, з яких приблизно 20% випадків – це дефолти, а 80% – це повернуті кредити.

*Крок 3.* Показники та характеристики, які описують клієнта та його стан, являються для цієї задачі взаємовиключними змінними.

*Кроки 4-5.* На цих кроках для неперервних змінних застосовується дискретизація, тобто область значень змінної поділяється на рівні проміжки. Ця процедура виконується тому, що в задачі використовуються неперервні змінні і будується гібридна мережа Байєса. Кількість проміжків визначається користувачем особисто. У випадку, коли для визначених діапазонів не було вказано число проміжків для розбиття, ці діапазони поділяються автоматично на два інтервали.

Слід зазначити, що доречно виконувати дискретизацію даних за однаковою шириною класів, або, залежно від інтерпретації, за однаковою кількістю точок в кластерах. В такому випадку ширина та кількість таких інтервалів буде регламентуватися банком, спираючись на соціологічні або демографічні дослідження для кожної групи клієнтів [32]. При побудові структури МБ необхідно пам’ятати, що алгоритм, який обрано, суттєво впливає як на швидкодію виконання програми, так і на вигляд побудованої структури.

### 3 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

#### 3.1 Критеріальна база для аналізу якості результатів

##### 3.1.1 Вихідні дані для статистичного аналізу та прогнозування

Статистичні методи аналізу та прогнозування зазвичай використовують глибоку обробку статистичних даних, які характеризують досліджуваний процес. При цьому мають місце наступні особливості виконання такого дослідження.

1. Основні джерела вихідних статистичних даних поділяють на первинні та вторинні. Первинними джерелами являються опитування, вибіркові обстеження, переписи, які орієнтовані на отримання різних даних у такій формі, які потрібні для управлінських рішень або для запланованих прогнозних розрахунків. Для отримання з первинних джерел вихідних статистичних даних необхідне виконання спеціально спланованої роботи і, відповідно, виділення спеціальних засобів для цього [17]. Вторинними джерелами являються в тому чи іншому вигляді опубліковані вихідні дані, які були кимось зібрані, але в тій чи іншій мірі надають інформацію, яка є корисною для розв'язання конкретної задачі.

2. Вимоги, що пред'являються щодо вихідних статистичних даних. Формуючи з первинних та вторинних джерел масив вихідних статистичних даних, слід до якості цих даних дотримуватись основні вимоги.

*Релевантність.* Згідно цього повинні дані, що використовуються, відображати аналіз діяльності, а також мають бути «прив'язані» у відповідні моменти часу до потрібних об'єктів.

*Надійність і точність.* Це може досягатися шляхом застосування різних (прямих і непрямих) методів перевірки надійності джерел, що використовуються, дотриманням необхідної методології вимірювань, встановленням можливих помилок і збоїв в їх записах, достовірністю відповідей респондентів та ін.

*Порівнянність.* Дані необхідно супроводжувати відповідними поясненнями та

коментарями, які тлумачать зміст аналізованих показників та методологію їх вимірювання, що дає можливість їх зіставляти у просторі і часі та «приводити до спільного знаменника» в ситуаціях, які можуть характеризуватися змінами в методології їх вимірювання та корекцією складу аналізованих змінних [17].

*Репрезентативність.* Це може досягатися наступним способом організації вибірки, згідно якого вона в повній мірі і адекватно може передавати відповідні властивості всієї сукупності, яка аналізується і яка для цього відбиралася.

Обчислення прогнозу, побудова та експериментальна перевірка (верифікація) ймовірно-статистичної моделі базуються зазвичай на використанні одночасно інформації двох типів:

- апіорна інформація про змістовну сутність і природу аналізованого явища, яка, як правило, представлена у вигляді відповідних теоретичних закономірностей, гіпотез і обмежень;
- вихідні статистичні дані, які характеризують як процес, так і результати функціонування явища або системи, що аналізуються.

Можна виділити наступні основні етапи прогнозування.

Етап 1 – постановка завдання, яка в себе включає

- визначення кінцевих прикладних цілей прогнозування;
- набір показників (змінних) та чинників, а також опис взаємозв'язків між атрибутами, які необхідні;
- ролі цих показників і чинників, які можна вважати вхідними в рамках поставленої конкретної задачі, тобто частково або повністю регульованими і можуть піддаватися реєстрації і прогнозуванню, а які можна вважати вихідними [18].

Етап 2 являється апіорним перед аналізом моделі і ґрунтується на аналізі сутності процесу або явища, що досліджується. На цьому етапі формалізуються та формуються у вигляді гіпотез та припущень апіорні знання про ці явища.

Етап 3 являється інформаційно-статистичним і складається зі збору необхідної статистичної інформації, тобто реєструються значення, які для різних часових і просторових тактів функціонування модельованої системи приймають участь в аналізі показників і чинників.

Етап 4 – специфікація моделі, яка базується на прийнятих на 2-му етапі гіпотезах та початкових припущень. Вона включає в себе формування модельних співвідношень у загального вигляду, які вказують на зв'язок між собою вхідних і вихідних змінних. На даному етапі визначається лише структура моделі та її аналітичний опис, в якій поряд з числовими значеннями, що відомі, будуть величини, наявний змістовний сенс яких не визначено і ще немає числових значень.

Етап 5 – ідентифікація моделі, згідно якої проводиться статистичний аналіз даних для «налаштування» значень невідомих параметрів до тих вихідних статистичних даних, які є в наявності. Згідно реалізації цього етапу слід відповісти на питання, чи є в принципі можливість знайти значення невідомих параметрів моделі по наявним вихідним статистичним даним. Якщо відповідь позитивна, то слід вирішувати проблему параметричної ідентифікації моделі. Якщо проблему ідентифікації неможливо вирішити, то слід повернутися до 4-го етапу і вносити відповідні корективи в розв'язання задачі специфікації моделі [18].

Етап 6 – верифікація моделі, яка використовує різні процедури порівняння висновків, що отримані по моделі, оцінок та наслідків. Цей етап відповідає статистичному аналізу точності та адекватності моделі. Якщо результати виконання цього етапу негативні, то слід повернутися до 4-го етапу, а можливо і до 1-го етапу. Якщо результати виконання цього етапу позитивні, то отриману модель відповідно до описаної загальної схеми можна використовувати для оцінювання прогнозу [17, 18].

Згідно 1-го етапу необхідно визначити кінцеві цілі прогнозування, тобто визначити необхідний тип прогнозу, який визначається двома факторами:

- горизонт прогнозування;
- ієрархічний рівень прогнозованого показника.

Згідно горизонту прогнозування прогнози можна поділити на короткострокові (вперед на 1-2 такту часу), середньострокові (вперед на 3-5 тактів) і довгострокові (вперед більше ніж на 5 тактів часу). Так, як правило, можна визначити періодом дискретизації вимірів.

З вищенаведеного слідує, що для прогнозування створюється математична модель, яка являється математичною конструкцією, що є абстракцією реального

світу як відносини між реальними елементами, що замінені на відповідні відносини між елементами математичної конструкції. Як правило, ці відносини представлені рівняннями або нерівностями між змінними, які моделюють функціонування реальної системи. Складність побудови математичної моделі полягає в необхідності поєднання якомога більшої лаконічності в математичному описі з точністю відтворення відповідних сторін аналізованої реальності, що цікавить дослідника.

### 3.1.2 Вимоги до математичної моделі

Можна використовувати різні математичні способи опису об'єктів: теорію графів, теорію множин, математичну логіку, теорію ймовірностей, диференціальні або інтегральні рівняння, математичне програмування та ін. Загальними вимогами до моделі є наступні:

1. Модель повинна адекватно описувати процес. Згідно адекватності модель повинна:

- відображати взаємодію та найбільш характерні зв'язки між змінними об'єкту;
- враховувати можливі керуючі сигнали;
- враховувати шуми вимірювань і вплив зовнішніх збурень;
- враховувати початкові значення змінних, а також на них обмеження.

Адекватність визначають формально з використанням відповідних статистичних величин [20]. Найбільш часто використовується середньо-квадратична похибка моделі (СКП), що може обчислюватися за формулою:

$$\text{СКП}(x_s, x_m) = \sqrt{\frac{1}{n} \sum_{k=1}^N [x_s(k) - x_m(k)]^2} \quad (3.1)$$

де  $x_s(k)$  – вимір вихідного сигналу об'єкту в момент  $k$ ;

$x_m(k)$  – оцінка виміру за моделлю.

Разом з тим, для визначення ступеня адекватності моделі застосування лише

одного параметра являється некоректним підходом, так як оцінки параметрів являються випадковими величинами, тому підвищенню ймовірності вибору адекватної моделі буде сприяти збільшення кількості критеріїв адекватності.

2. Кожне рівняння математичної моделі мусить мати або хоч один аналітичний розв'язок, або чисельний розв'язок у випадку, коли це неможливо. Для аналізу поведінки процесу, тобто для аналізу збіжності, необхідно мати розв'язок та обчислення оцінок прогнозів. При побудові моделі необхідно дотримуватись принципів, згідно яких крім необхідного в моделі не має бути нічого зайвого. Цього складно дотриматись і на практиці можливо, що модель має дійсно складну структуру, що визначається необхідністю досягнення адекватності процесу її високого ступеня.

3. Модель має бути достатньо універсальною, щоб її можна було б використовувати для опису однотипових процесів або для опису функціонування процесів за різних умов. Для моделюванні технічних систем широко застосовуються рівняння диференціальні або різницеві рівняння першого і другого порядку, з використанням яких можна побудувати моделі будь-якої складності.

4. Вимога робастності, яка означає, що модель повинна забезпечувати необхідний прогноз вихідної змінної не тільки для того відрізка часового ряду, на основі якого вона побудована, але і для будь-якого іншого відрізка, який відповідає обраному режиму роботи. Зазначимо, що робастність можна розглядати також як характеристику стійкості моделі до похибок, збурень та пропусків вимірів. Вимога робастності являється особливо критичною для аналізу систем, які працюють в реальному часі, так як нестійка модель може призвести до створення аварійної ситуації [4, 5].

5. Вимога адаптивності, яка передбачає, що хоча б частина параметрів моделі може уточнюватись при надходженні нових вимірів від об'єкта. Ця вимога важлива при побудові моделей нестационарних об'єктів, в яких параметри є функціями часу. З точки зору збіжності оцінок параметрів та похибок керування такі системи є досить складними. Тому при їх проектуванні слід приділяти особливу увагу питанням вибору методу оцінювання параметрів, збіжності оцінок параметрів, оцінюванню отриманих прогнозів та значень керуючих сигналів [20].

### 3.1.3 Аналіз якості моделі

Можна встановити адекватність регресійних моделей, використовуючи аналіз послідовностей залишків і похибок моделей. Підстановкою в модель фактичних значень всіх включених до неї факторів можна знайти розрахункові значення. Послідовність, що залишилася, перевіряється на виконання статистичних властивостей випадкової компоненти відповідного часового ряду: випадковий характер відхилень, нормальність закону розподілу, близькість нулю математичного очікування, відсутність автокореляції та ін.

Виконати аналіз адекватності моделі можна з використанням відповідних параметрів адекватності, основними з яких являються наступні:

1) сума квадратів залишків, яка є сумою квадратів величин розбіжності між фактичними значеннями змінної та змодельованими в період їх ідентифікації, що розраховується за виразом:

$$R^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3.2)$$

2) коефіцієнт детермінації, який є статистичним параметром, що застосовується в статистичних моделях в якості міри інформативності моделі щодо даних, і показує відповідність моделі наявним спостереженням:

$$R^2 = 1 - \frac{V(y|x)}{V(y)} = 1 - \frac{\sigma^2}{\sigma_y^2}, \quad (3.3)$$

де  $V(y|x) = \sigma^2$  – дисперсія основної змінної, що оцінюється моделлю.

3) критерій Дарбіна-Уотсона, який є статистичним критерієм, який застосовується при тестуванні автокореляції першого порядку елементів послідовності, що досліджується. Його найбільш часто використовують для аналізу часових рядів та залишків регресійних моделей:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \approx 2(1 - \rho_1), \quad (3.4)$$

де  $\rho_1$  – коефіцієнт автокореляції першого порядку. Якщо автокореляція відсутня, то  $d = 2$ , інакше  $d$  прямує до нуля, а при негативній дорівнює 4:

$$\begin{cases} \rho_1 = 0 \rightarrow d = 2 \\ \rho_1 = 1 \rightarrow d = 0 \\ \rho_1 = -1 \rightarrow d = 4 \end{cases} . \quad (3.5)$$

Інформаційний критерій Акаїке, який є мірою відносної якості статистичних моделей при заданому наборі даних. Маючи в наявності сукупність моделей для даних, цей критерій дозволяє оцінити якість кожної моделі відносно до кожної з інших моделей. Цей критерій ґрунтується на теорії інформації і дозволяє характеризувати відносні оцінки втраченої інформації при застосуванні для представлення процесу даної моделі. Тобто він дозволяє отримати компроміс між складністю моделі та ступенем її узгодженості.

Припустімо, що в наявності є статистична модель даних. Нехай  $L$  для цієї моделі являється максимальним значенням функції правдоподібності, і нехай  $k$  у цій моделі є число параметрів, що оцінюються. Тоді для цієї моделі значення критерія обчислюється в такий спосіб:

$$AIC = 2k - 2 \ln(L). \quad (3.6)$$

Зазначимо, що для оцінки адекватності моделі може також використовуватися критерій Байеса-Шварца.

### 3.1.4 Аналіз якості прогнозу

Точність і обґрунтованість прогнозів є обов'язковим етапом прогнозування, на якому може використовуватися певна сукупність підходів, критеріїв та процедур, які дозволяють отримати оцінку якості прогнозу.

Показниками якості прогнозу можуть бути такі:

1. Середній квадрат похибок моделі MSE (Mean Squared Error). Для обчислення MSE зводяться в квадрат всі окремі залишки регресії, потім підсумовуються, а результат суми ділиться на загальне число похибок:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}. \quad (3.7)$$

Корінь квадратний від цієї величини позначають як RMSE (Root MSE):

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}}. \quad (3.8)$$

2. Абсолютна помилка прогнозу визначається як різниця між фактичним значенням ( $y$ ) і прогнозом ( $y^*$ ):

$$\Delta_{pr} = y_t - y^* \quad (3.9)$$

3. Середнє абсолютне значення помилки становить:

$$\bar{\Delta}_{pr} = \frac{\sum_{t=1}^n |y_t - y_t^*|}{n}. \quad (3.10)$$

4. Коефіцієнт нерівності Тейла  $U$ , який являється важливим індикатором якості моделі прогнозу; і за означенням  $0 \leq U \leq 1$ . Якщо  $U = 1$ , то модель практично має нульові, тобто неприйнятні властивості прогнозу, що слідує з виразу для обчислення  $U$ :

$$U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k) + \frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}}. \quad (3.11)$$

При  $U = 0$  прогнозовані значення співпадають з фактичними значеннями ряду, тобто модель ідеальна. Таким чином  $U$  дозволяє в принципі встановити придатність моделі для оцінювання прогнозу.

### 3.2 Результати моделювання кредитних ризиків

Задача оцінювання кредитоспроможності відноситься до класу слабо структурованих задач, що обумовлює складність моделювання кредитних ризиків, оскільки основними їх рисами являються об'єктивна наявність у їх складі як

кількісних, так і якісних показників. Крім того, більшість банківської інформації майже не піддається її точному і формалізованому опису. Разом з тим, вона може бути подана у вигляді нечіткої інформації.

Розглянемо процес аналізу і моделювання кредитних ризиків банку з використанням БМ для управління кредитним ризиком. Ефективність використання цього підходу особливо виявляється за умов, коли процеси, що аналізуються, являються досить складними для аналізу при застосуванні класичних кількісних методів, або за умов, коли джерела інформації можуть інтерпретуватися якісно, невизначено або неточно, інакше кажучи, коли процеси, що розглядаються, є слабо структурованими.

Процес аналізу являє собою виконання декілька етапів, які зображені на рисунку 3.1.



Рисунок 3.1 – Процес аналізу статистичних даних

*Етап 1.* База знань даних, опис даних для експерименту.

Для виконання аналізу та побудови математичних моделей при оцінюванні кредитоспроможності була обрана вибірка із 3350 клієнтів, які характеризувалися такими атрибутами:  $x_1$  – стать позичальника,  $x_2$  – вік позичальника,  $x_3$  – сімейний статус,  $x_4$  – кількість дітей,  $x_5$  – наявність особистого житла,  $x_6$  – працює / не працює,  $x_7$  – освіта,  $x_8$  – вид роботи,  $x_9$  – наявність гарантованої особи,  $x_{10}$  – сума кредиту. Результат аналізу є вихідною змінною  $y$ , яка являє собою оцінку рівня кредитного ризику.

*Етап 2.* Заповнення пропусків даних.

Інформацію, що збирають з вибіркового даних, необхідно обробити для узагальнення на генеральну сукупність. Маючи консолідований масив всіх даних, що отримані в ході обстеження, можна визначити значення всіх узагальнюючих показників.

Найпростіше рішення при обробці неповних даних є виключення некомплектних спостережень, які можуть містити пропуски, та аналіз отриманих таким способом даних. Використання такого підходу може призвести до сильних відмінностей у висновках, зроблених за наявності пропусків в даних. Тому перед аналізом масиву даних будемо рухатися шляхом заповнення пропусків. У початкові дані аналізу кредитоспроможності були штучні додані пропуски у співвідношенні 5%, 10%, 15% та 20%. Заповнення пропусків було виконано з використанням моделі регресії. Значення похибки MAPE при заповненні пропусків даних кредитоспроможності прогнозованими значеннями становить 4,92%.

*Етап 3.* Оцінювання інформативності даних та структурування вибірки.

Банк повинен дотримуватись основних принципів кредитування, коли надає кредити, зокрема, повинен виконувати аналіз кредитоспроможності позичальників, а також дотримуватись вимог щодо концентрації ризиків, які встановлені банком. На кредитоспроможність позичальника можуть впливати різні фактори: фінансове становище, тобто рівень забезпеченості власними коштами, фінансова дисциплінованість, склад родини та ін. Можна для оцінювання значущості використати ряд статистик, що дозволяють виконувати аналіз і знаходити

взаємозв'язки між атрибутами даних.

Відбір значущих змінних для прогнозування відбувається за 3 кроки.

На першому кроці виконувався вибір змінних згідно загального статистичного тесту  $\chi^2$ . При встановленні порогу в 20 змінних було обрано 6 змінних:  $x_4$ - кількість дітей,  $x_5$  - наявність особистого житла,  $x_6$  - працює / не працює,  $x_7$ -освіта,  $x_8$  - вид роботи,  $x_{10}$  - сума кредиту. Кожна із вище зазначених змінних має по одному балу для загального розрахунку значущості.

На другому кроці відбувався вибір змінних з використанням методу рекурсивного виключення. Було обрано дефолтне значення кількості змінних – 6, тобто були обрані наступні змінні:  $x_1$ – стать позичальника,  $x_2$ –вік позичальника,  $x_5$  - наявність особистого житла,  $x_6$  - працює / не працює,  $x_8$  - вид роботи,  $x_{10}$  - сума кредиту.

На основі отриманих статистичними методами розрахунках було розраховане фінальне значення для кожної зі змінних, а потім було отримано результат: чи варто до фінальної моделі обирати дану змінну чи ні. За композицією трьох методів відбору був розраховуємо загальний бал для кожної змінної. Обрано дефолтне значення фінальної функції активації, згідно якого змінні з загальним балом в 2 і більше слід вважати значущими і їх необхідно ввести в модель:  $x_2$ – вік позичальника,  $x_4$ -кількість дітей,  $x_5$  - наявність особистого житла,  $x_6$  - працює / не працює,  $x_7$ – освіта,  $x_8$  - вид роботи,  $x_{10}$  - сума кредиту. Всі інші змінні були відкинуті.

*Етап 4.* Базуючись на отриманих характеристиках інформативності оцінюваних змінних необхідно побудувати модель для прогнозування кредитоспроможності індивідуального позичальника. В якості моделей було обрано байєсівську мережу. За наявними даними була побудована мережа БМ, яка дозволяє встановити зв'язок між характеристиками клієнта і вершиною, яка являється подією повернення кредиту. Також визначалася ймовірність повернення кредиту новими клієнтами, які прийшли в банк,.

*Етап 5.* Так як скорочення кількості атрибутів (змінних) дозволила отримати високу точність моделі, прогноз та аналіз результатів буде виконуватися з використанням 7 вхідних атрибутів. Для прогнозування використовувались

навчальна та перевірна вибірки у співвідношенні (%): 70/30, 80/20, 90/10.

Оскільки такі числові характеристики як сума кредиту, вік, дохід та ін. є неперервними та приймають велику кількість значень, будемо використовувати гібридну мережу Байєса для розв'язання даної задачі. Розглянемо більш детально схему застосування розглянутої в другому розділі методики. Згідно розглянутих основних кроків за початковими даними була побудована структура мережі Байєса, що наведена на рисунку 3.2 та демонструє наочно зв'язки між даними.

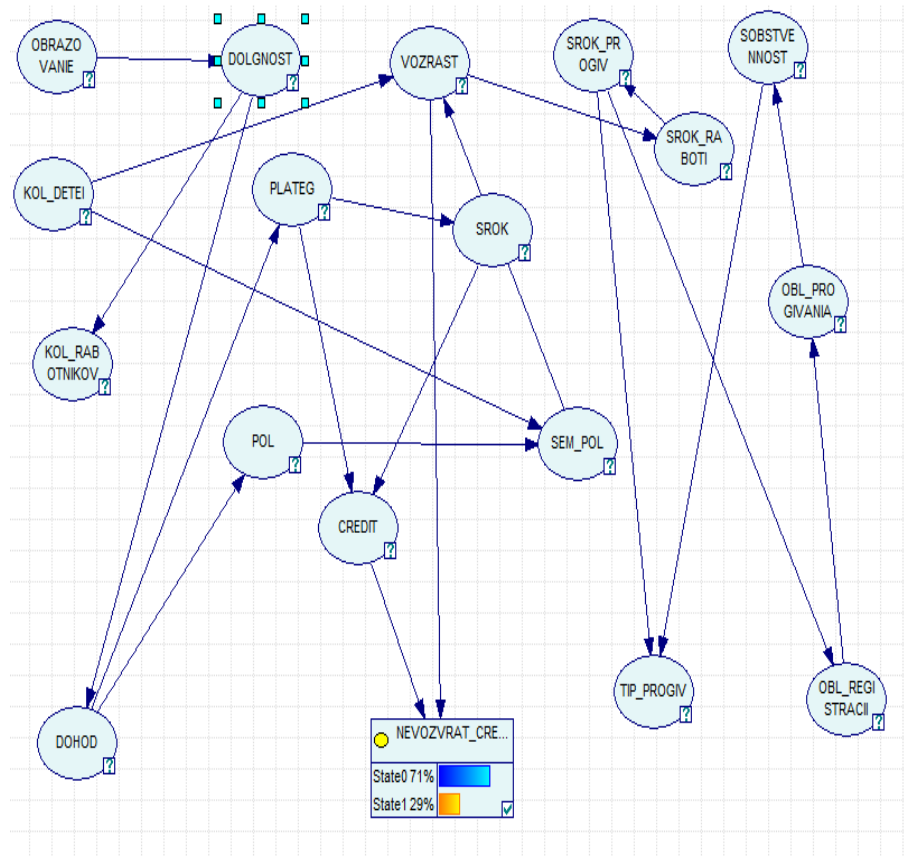


Рисунок 3.2 – Структура побудованої Байєсівської мережі

Значення вершин були використані для знаходження порогу відсікання, згідно компромісу між специфічністю моделі та її чутливістю. В якості критеріїв вибору порогу відсікання можуть бути використані такі:

- 1) вимога мінімальної величини специфічності, тобто чутливості моделі;
- 2) вимога максимальної сумарної специфічності і чутливості моделі, тобто:

$$cut - off = \max_k (Se_k + Sp_k) \quad (3.12)$$

3) вимога балансу між специфічністю і чутливістю, тобто:

$$cut - off = \min_k / Se_k - Sp_k / \quad (3.13)$$

Розглядалися різні рівні відсікання, що були отримані для мережі Байєса. Точність моделей оцінювалася за помилками I-го і II-го роду. Результати моделювання наведені у таблиці 3.1.

Таблиця 3.1 – Загальна точність моделі за помилками I-го і II-го роду

	Повернення кредиту (0)	Дефолт (1)
Cut-off=0,1		
Повернення кредиту (0)	3	237
Дефолт (1)	0	60
Загальна точність моделі = 50,6 %		
Cut-off=0,15		
Повернення кредиту (0)	3	237
Дефолт (1)	0	60
Загальна точність моделі = 50,6 %		
Cut-off=0,2		
Повернення кредиту (0)	22	218
Дефолт (1)	0	60
Загальна точність моделі = 54,6 %		
Cut-off=0,25		
Повернення кредиту (0)	23	217
Дефолт (1)	0	60
Загальна точність моделі = 54,8 %		
Cut-off=0,3		
Повернення кредиту (0)	146	94
Дефолт (1)	3	57
Загальна точність моделі = 77,9 %		

З таблиці 3.1 слідує, що найбільша точність моделі була досягнута при рівні 77 % та встановленні порогу 0,3; при цьому були пропущені 3 значення дефолта та відкинуті 39 % добросовісних позичальників.

Отримані результати обчислювальних експериментів свідчать, що використання моделі у формі мережі Байєса дозволяє отримати високі значення точності моделі, що дорівнюють 0,779 або 77,9%. Наведені результати вказують на доцільність використання мереж Байєса при оцінюванні кредитоспроможності позичальників кредитів.

## ВИСНОВКИ

У кваліфікаційній роботі розглядається розв'язок задачі побудови математичних моделей для прогнозування фінансових процесів та оцінювання кредитних ризиків можливих втрат на основі застосування методів інтелектуального аналізу даних, методів аналізу часових рядів і прикладної статистики.

Згідно результатів, отриманих в кваліфікаційній роботі, можна зробити наступні висновки:

1. Ефективність фінансової діяльності банків залежить від якісного управління основними фінансовими ризиками, якими являються кредитний та ринковий ризики. Основним завданням управління фінансовими ризиками являється забезпечення стабільності у всіх сферах. Мінімізація фінансових ризиків може здійснюватися як на рівні кредитного портфеля в цілому, так і на рівні кожної окремої позики.

2. На основі проведеного аналізу встановлено, що при аналізі та моделюванні фінансових ризиків з метою підвищення ефективності прийняття рішень слід використовувати мережі Байєса, які дозволяють враховувати невизначеності різних типів і характеризуються швидкими алгоритмами навчання.

3. Розробка методів, моделей та критеріїв аналізу процесів кредитування являється основою процесу аналізу кредитоспроможності позичальників, при цьому оцінка кредитного ризику виконується з використанням оцінки кредитоспроможності позичальника.

4. Побудова математичних моделей нелінійних нестационарних процесів, якими є фінансові процеси, виконується в декілька етапів. Використання на кожному із етапів статистичних параметрів якості дозволяє отримати високу якість проміжних та кінцевих результатів обчислювальних експериментів та уникнути неадекватності моделей.

5. На кредитоспроможність позичальника можуть впливати різні фактори його фінансового стану, до яких можна віднести такі: власні кошти, фінансова дисциплінованість та ліквідні активи. Оцінка кредитоспроможності виконується в два

етапи, протягом яких необхідно аналізувати діловий ризик, враховувати фінансовий стан позичальника та аналізувати грошові потоки.

6. Оцінювання кредитоспроможності позичальників здійснено з використанням адаптивної Байєсівської мережі, що дозволило підвищити адекватність ймовірнісної моделі та забезпечити зменшення розмірів кредитного ризику.

7. Проведені експериментальні дослідження щодо оцінювання кредитоспроможності, які показали, що оцінка кредитоспроможності з використанням байєсівського підходу дозволила отримати кращі результати, покращити загальну точність результатів даних на 2.5%, створити портрет позичальника та побудувати процес прийняття рішення.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Бланк І. А. Управління фінансовими ризиками. К.: Ніка-Центр, 2014. – 600 с.
2. Клапків М.С. Страхування фінансових ризиків: Монографія. – Тернопіль: Економічна думка, Карт-бланш, 2002. – 570 с.
3. Управління ризиками банків: монографія у 2 томах. Т. 2: Управління ринковими ризиками та ризиками системних характеристик / [А.О. Єпіфанов, Т.А. Васильєва, С.М. Козьменко та ін.] /За ред. Д-ра екон. наук, проф. А.О. Єпіфанова і д-ра екон. наук, проф. Т.А. Васильєвої. — Суми: ДВНЗ "УАБС НБУ", 2012. — 299 с.
4. Allen S. Financial risk management: A practitioner's guide to managing market and credit risk / Allen S. – Hoboken, N.J.: John Wiley & Sons, Inc., 2003. – 567р.
5. Basel Committee on Banking Supervision. Basel III: International framework for liquidity risk measurement, standards and monitoring. – Bank for International Settlements, December 2010. [Електронний ресурс]. – Режим доступу : <http://www.bis.org/publ/bcbs188.pdf>
6. Бобиль В. В. Фінансові ризики банків: теорія та практика управління в умовах кризи: монографія. Дніпропетровськ: Дніпропетр. нац. ун-т залізн. трансп. ім. акад. В Лазаряна, 2016. – 298 с.
7. Верхуша Н. П. Сутність кредитного ризику банку та його фактори. URL:[http://lib.uabs.edu.ua/library/Visnik/Numbers/2\\_31\\_2011/31\\_03\\_04.pdf](http://lib.uabs.edu.ua/library/Visnik/Numbers/2_31_2011/31_03_04.pdf).
8. Бучко І. Є. Скоринг як метод зниження кредитного ризику банку. Вісник Університету банківської справи Національного банку України. 2013. Вип. 2 (17). – С. 178–182.
9. Белевець М. С., Бідюк П. І., Кузнецова Н. В. Розробка системного підходу до менеджменту ризиків. Системні науки та кібернетика. 2015. №1. С. 30 – 48. URL: [http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc\\_1\\_2015.pdf](http://mmsa.kpi.ua/sites/default/files/ssc/issues/ssc_1_2015.pdf).
10. Бідюк П. І. Аналіз кредитоспроможності за допомогою методів інтелектуального аналізу даних / П. І. Бідюк, В. Г. Гуськова // Електронне

моделювання. - 2019. - Т. 41, № 2. – С. 111-120.

11. Гуськова В. Г. Оцінювання кредитоспроможності позичальників кредитів методами інтелектуального аналізу даних / В. Г. Гуськова, П. І. Бідюк. // Міжнародній науково-технічний журнал «Системні дослідження та інформаційні технології». – 2019. – №2. – С. 31–48.

12. Tymoshchuk O. L. A combined approach to modeling nonstationary heteroscedastic processes / O. L. Tymoshchuk, V. H. Huskova, P. I. Bidyuk. // Radio Electronics, Computer Science, Control. – 2019. – №2. – С. 80–89.

13. Гуськова В. Г. Розробка сценарного підходу на основі моделей інтелектуального аналізу даних / В. Г. Гуськова, П. І. Бідюк // Наукові праці Донецького національного технічного університету. Серія : Інформатика, кібернетика та обчислювальна техніка. - 2016. –№2. – С. 158-164.

14. Гуськова В. Г. Аналіз кредитоспроможності позичальників кредитів за допомогою логістичної регресії. / В. Г. Гуськова, П. І. Бідюк // Наукові праці Донецького національного технічного університету. Серія : Інформатика, кібернетика та обчислювальна техніка. – 2017. –№2. – С. 54-60.

15. Jarrow R.A., Turnbull S.M. / Journal of Banking & Finance: North-Holland, 24 (2000) – pp. 271-299.

16. Elena Medova, Robert Smith. A framework to measure integrated risk, Quantitative Finance, Taylor & Francis Journals, vol. 5(1), 2005. – pp. 105-121.

17. Theodore M. Barnhill and William F. Maxwell Journal of Banking & Finance, vol. 26, issue 2-3, 2002. – pp. 347-374.

18. Xeni Kristine Dimakos and Kjersti Aas, Integrated risk modelling in Statistical Modelling 4(4):– December 2004. – pp. 265-277

19. Тен В. В. Проблемы анализа кредитоспособности заемщика / В. В. Тен //Банковское дело. – 2006. – № 3.

20. Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів: навчальний посібник Київ: НТУУ «КПІ», 2013. – 600 с.

21. Галасюк В. В. Методика оцінки кредитоспроможності позичальників // Вісник НБУ. – 2009. – № 2. – С. 39-45.

22. Бідюк П. І. Моделі оцінки ризиків кредитування фізичних осіб / П. І. Бідюк, Є. О. Матрос // Кібернетика та обчислювальна техніка. – 2007. – №153. – С. 87–95.
23. Bermingham, M.L. et al. Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci. Rep.* 5, 10312; doi: 10.1038/srep10312 (2015).
24. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. *J Biomed Inform.* 2018; 85:189-203. doi:10.1016/j.jbi.2018.07.014.
25. Gengjian Xue, Li Song, Jun Sun. Foreground estimation based on linear regression model with fused sparsity on outliers, *IEEE transactions on circuits and systems for video technology.* 2013. – pp. 1346-1357.
26. Орлов А.І. Прикладна статистика/А.І.Орлов. М.: Іспит, 2004. – 284 с.
27. Xiong H. and Chen X., “Kernel-based distance metric learning for microarray data classification,” *BMC Bioinformatics*, 7:299, 2006.
28. Weston J., Elisseeff A., Scholkopf B., and Tipping M., “Use of the zero-norm with linear models and kernel methods,” *Journal of Machine Learning Research*, special Issue on variable and Feature Selection 3, pp.1439-1461, 2003.
29. Guyon I., Weston J., Barnhill S., and Vapnik V., “Gene selection for cancer classification using support vector machines,” *Machining Learning*, vol. 46, no. 1-3, pp. 389-422, 2002.
30. Терентьев А. Н. Сравнение методов интеллектуального анализа данных при оценивании кредитоспособности физических лиц / [ А.Н. Терентьев, П.И. Бідюк, Миронова А.В., Медин Н.Ю.] // Проблемы управления и информатики. – К.: ИКИ НАНУ-НКАУ, 2009. – № 5. – С. 141–149.
31. Бідюк П.І. Основні етапи побудови і приклади застосування мереж Байєса/ Бідюк П.І., Кузнєцова Н.В.// Системні дослідження та інформаційні технології. – 2007. – № 4. – С.26–39.
32. Кузнєцова Н.В. Системний підхід до аналізу кредитних ризиків з використанням мереж Байєса / Н.В. Кузнєцова, П.І. Бідюк // Наукові вісті НТУУ «КПІ». – 2008. – № 3. – С. 11 – 24.

33. Зайченко Ю.П., Оценка кредитных банковских рисков с использованием нечеткой логики// Системні дослідження та інформаційні технології. – 2010. –№2. – С. 37-54.

34. Шовгун Н.В. (Шаповал Н. В.) Аналіз кредитоспроможності позичальника за допомогою методів з нечіткою логікою/ Н.В. Шовгун // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб.: наук. пр. – К.: Век+,- 2012. - №55. – С.169-173

35. Kim, E.H.; Morse, A.; Zingales, L. (2006). «What Has Mattered to Economics since 1970». *Journal of Economic Perspectives* 20 (4): 189—202. doi:10.1257/jep.20.4.189

36. Бідюк П.І., Коновалюк М.М. Прогнозування волатильності валютного ринку за нелінійними моделями // Вісник Національного університету «Львівська політехніка», № 719. –2011. – С. 154 – 163.