

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

На правах рукопису
УДК 004.942

До захисту допущено
В. о. завідувача кафедри ММСА
О.Л.Тимошук
« ___ » _____ 2019 р.

Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз
на тему: «Математичні методи і моделі для оцінювання кредитних ризиків»

Виконав студент II курсу, групи КА-81 мп
Авраменко Олександр Олексійович _____

Науковий керівник професор кафедри
математичних методів системного
аналізу, д.т.н., проф. Бідюк Петро Іванович _____

Рецензент професор кафедри автоматики та
управління в технічних системах
КІП ім. Ігоря Сікорського,
д.т.н., проф. Теленик Сергій Федорович _____

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

Київ
2019

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

Рівень вищої освіти — другий (магістерський)
Спеціальність — 124 «Системний аналіз»

ЗАТВЕРДЖУЮ
В. о. завідувача кафедри
ММСА

О. Л. Тимощук

«___» _____ 2019 р.

ЗАВДАННЯ

на магістерську дисертацію студента Авраменко Олександр Олександрович

1. Тема дисертації: «Математичні моделі і методи для оцінювання кредитних ризиків», науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор, затверджені наказом по університету від «___» _____ № _____

2. Термін подання студентом дисертації: 13 грудня 2019 р.

3. Об'єкт дослідження: Кредитні ризики, методи аналізу ризиків, статистичні дані стосовно кредитування

4. Предмет дослідження: методи математичного моделювання, ймовірно-статистичні методи аналізу даних

5. Перелік завдань, які потрібно розробити:

1) дослідити сучасний стан та особливості застосування математичного моделювання та оптимізації у вирішенні проблеми кредитних ризиків;

2) розробити математичну модель кредитування фізичних осіб;

3) використати побудовану математичну модель;

4) пошук і накопичення даних для застосування в обчислювальних експериментах;

5) використати евристичний метод побудови байєсівської мережі із застосуванням значення взаємної інформації (ЗВІ) та функції опису мінімальної довжини (ОМД);

6) розробити стартап-проект стосовно виведення на ринок результатів дослідження;

7) розробити концептуальні висновки за результатами наукового дослідження.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:

- 1). Постановка задачі дисертаційного дослідження.
- 2). Методи моделювання кредитного ризику.
- 2). Схема побудованої мережі Байєса (рис.).
- 2). Приклади функціонування створеної моделі.
- 3). Таблиці у розділі стартап-проекту.

7. Дата видачі завдання: 05 вересня 2019 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	05.09.2019—13.09.2019
2.	Перший розділ. Аналіз задачі оцінювання кредитних ризиків	16.09.2019—27.09.2019
3.	Другий розділ. Моделі і критерії для аналізу ризиків	30.09.2019—18.10.2019
4.	Третій розділ. Моделювання кредитних ризиків	21.10.2019—15.11.2019
5.	П'ятий розділ. Стартап-проект	15.11.2019—26.11.2019
6.	Концептуальні висновки. Перспективи розвитку отриманих результатів дослідження.	26.11.2019—04.12.2019

Студент

О.О. Авраменко

Науковий керівник дисертації

П.І. Бідюк

РЕФЕРАТ

Магістерська дисертація : 70 с., 25 рис., 8табл та 16 джерел.

КРЕДИТНИЙ РИЗИК, ЗБІР ДАНИХ, СТАТИСТИЧНА ОБРОБКА ДАНИХ, СКОРИНГОВА КАРТА, БАЙЄСІВСЬКА МЕРЕЖА, ЛОГІСТИЧНА РЕГРЕСІЯ

Об'єкт дослідження – кредитні ризики у фінансовій діяльності

Мета роботи – аналіз кредитних ризиків за допомогою статистичних методів.

У роботі проаналізовано існуючі підходи до оцінки кредитних ризиків, зроблено висновки про те, що всі методи дають доволі різні оцінки, тому при побудові та аналізі моделей краще використовувати різні оцінки для виявлення адекватності моделі, пропонується власний підхід до побудови і виявлення найкращої моделі згідно зібраних даних: зібравши дані, які клієнти банків подують до установи на різних етапах співбесіди, за допомогою критеріальної бази, базуючись на побудованій моделі та байєсівській мережі, зробити висновок щодо надання кредиту тій чи іншій особі, обґрунтувавши усі розрахунки математичними формулами.

У подальшому рекомендується збільшити кількість методів для побудови моделі та застосувати більшу кількість критеріїв для обрання найкращої моделі.

ABSTRACT

Master's Thesis: 70p., 25 fig, 8 table and 16 sources.

CREDIT RISK, DATA COLLECTION, STATISTICAL DATA PROCESSING, SCORING CARD, BAYESIAN NETWORK, LOGISTIC REGRESSION

Object of study - credit risks in financial activities

The purpose of the study is to analyze credit risks using statistical methods to build a scoring card.

The thesis analyzes the existing approaches to credit risk assessment, it concludes that all methods give different estimates of possible losses, so when building and analyzing models, it is better to use different estimates so that to identify the adequacy of the model, propose your own approach to building and identify the best model according to the collected data. By collecting data that bank customers bring to the institution at different stages of the interview, using a benchmark, based on the model built and the scoring card, conclude on granting credit in one person or another, justifying all calculations with mathematical formulas.

In the future it is recommended to increase the number of methods for building the model and apply more criteria to select the best model.

Зміст

Вступ	7
РОЗДІЛ 1 ОЦІНЮВАННЯ КРЕДИТНИХ РИЗИКІВ У ФІНАНСАХ	9
1.1 Актуальність задачі	9
1.2 Існуючий підхід до аналізу кредитних ризиків	11
Висновки.....	21
РОЗДІЛ 2 ІСНУЮЧІ МОДЕЛІ І КРИТЕРІЇ ДЛЯ АНАЛІЗУ КРЕДИТНИХ РИЗИКІВ	23
2.1 Байєсівська мережа.....	24
2.2 Дерева Рішень	33
2.3 Логістична регресія	36
2.4 Скорингове моделювання.....	37
2.5 Критерії для аналізу адекватності моделі	38
2.5.1 Критерій Колмогорова-Смірнова.....	38
2.5.2 Індекс Джині	39
2.5.3 ROC – крива	43
2.5.4 AUC	44
Висновки.....	46
РОЗДІЛ 3 МОДЕЛЮВАННЯ КРЕДИТНИХ РИЗИКІВ	48
3.1 Архітектура і функціональна схема моделі.....	48
3.2 Побудова Байєсівської мережі.....	49
3.3 Аналіз результатів виконання обчислювальних експериментів	51
Висновки.....	55
РОЗДІЛ 4 СТАРТАП АНАЛІЗ ПРОЕКТУ	56
4.1 Інформаційна карта проекту.....	56
4.2. Маркетинговий аналіз стартап – проекту	57
Висновки до розділу	66
ВИСНОВКИ	67
ПЕРЕЛІК ПОСИЛАНЬ	68

Вступ

Фінансові установи відіграють вирішальну роль в економіці, оскільки вони перетворюють кошти з надлишкових одиниць у дефіцитні одиниці. Це відоме, як фінансове посередництво, яке використовувалося всіма фінансовими посередниками. Багато економістів вважають це однією з основних функцій комерційних банків. Популярність фінансових установ виникла завдяки депозитним рахункам, які вони пропонують для надлишкових фінансових одиниць, щоб після цього використати ці кошти для надання кредитів у розмірі та строку погашення, необхідних для отримання прибутку та виплати дивідендів вкладникам, які зробили внески на депозитні рахунки. Вони приймають ризик за наданими кредитами, враховуючи процент маржі та ставку дисконтування, в протиположності процентам, які вони повинні сплатити за депозитами. Також фінансові установи диверсифікують свій позичковий портфель серед декількох позичальників, щоб зменшити ризик за замовчуванням та знизити будь-які ризики за непогашені кредити. Комерційні банки обслуговують як приватний, так і державний сектори, оскільки їхні депозитні та кредитні засоби використовуються фізичними особами, підприємствами та державними установами. Ця функція фінансового посередництва є дуже важливою в економіці, оскільки покращує темпи економічного зростання, надаючи підприємцям капітал, який збільшує інвестиції, зайнятість та виробництво продукції. Потік коштів від депозитарних установ до дефіцитних підрозділів може набувати декількох форм:

- Перший - прямий, через надання позик та кредитних коштів до дефіцитних одиниць.
- Друга тече за рахунок придбання цінних паперів, випущених фінансовими компаніями, які перетворюються на позики фінансових компаній для дефіцитних одиниць.

- Третій набір потоків відображає придбання акцій, випущених фондами, які купують боргові та пайові цінні папери дефіцитних одиниць.

На сьогоднішній день конкуренція серед банків, привела до того, що у населення підвищився попит на банківські послуги, в той же час, банки зацікавлені в максимізації прибутку і мінімізації своїх раптових втрат.

Це змусило фінансові інститути шукати та розробляти більш ефективні шляхи та методи для залучення клієнтів при цьому дотримуватися встановлених жорстких показників по вартості кожного залученого клієнта. Зростання впливу ризиків, які стосуються сфери діяльності:

- Виробничий ризик
- Комерційний ризик
- Фінансовий ризик

на результати економічної діяльності викликана нестабільністю зовнішнього середовища: економічної ситуації в країні, поява і застосування нових інноваційних фінансових інструментів, переорієнтація сфери фінансових відносин, зміна кон'юнктури ринку. Тому оцінка, відстеження, ідентифікація фінансових ризиків є одним із актуальних завдань в діяльності підприємства.

Перший розділ присвячено аналізу існуючих методів і математичних моделей для формального опису кредитного ризику. Розглянуто актуальність цієї задачі у сучасному світі.

У другому розділі наведено математичне формулювання і опис існуючих математичних моделей і методів. Також наведено критерії за допомогою яких можливо оцінити адекватність побудованих моделей.

Третій розділ присвячено побудові байєсівської мережі, скорингової карти. Перевірка результатів на адекватність.

Четвертий розділ присвячено детальному опису стартап проекту на основі тих методів, які ми використовували у цій роботі.

РОЗДІЛ 1 ОЦІНЮВАННЯ КРЕДИТНИХ РИЗИКІВ У ФІНАНСАХ

1.1 Актуальність задачі

Ризик - ймовірність того, що фактична прибутковість інвестиції буде нижчою, ніж очікувана віддача [1]. Виконуючи свої функції, банки піддаються різним видам ризиків. Серед цих ризиків - кредитний ризик, ринковий ризик, ризик ліквідності, операційний ризик, юридичний ризик та репутаційний ризик. Хоча банки та інші фінансові установи стикаються з труднощами з багатьох причин, кредитний ризик є найпоширенішою причиною банкрутства банку, через що всі регулюючі органи встановлюють мінімальні стандарти управління кредитними ризиками [2].

Кредитний ризик – це ризик невиконання боргу, який може виникнути через позичальника, який не здійснив необхідних платежів. У першу чергу, ризик полягає у ризику кредитора і включає втрачені основні суми та відсотки, порушення грошових потоків та збільшення витрат на стягнення. Втрати можуть бути повними або частковими. В умовах ефективного ринку більш високий рівень кредитного ризику буде пов'язаний з більшими витратами на запозичення.[3].

Управління ризиками в банках істотно змінилося за останні десять років. Регламенти, що вийшли під час світової фінансової кризи, та штрафи, що стягувалися після неї, викликали хвилю змін у функціях ризиків. Вони включали більш детальні та жорсткі вимоги до капіталу, ліквідності та фінансування. Управління не фінансовими ризиками набуло більшого значення, оскільки посилилися стандарти дотримання та поведінки. Стресове тестування стало основним інструментом нагляду, паралельно зі зростанням очікувань щодо банківських заяв про ризик та апетит. Банки також інвестували в зміцнення культури ризиків та стали більш активно враховувати ризики під час прийняття рішень. Вони також прагнули детальніше визначити та окреслити свої лінії

оборони. Враховуючи масштаби цих та інших зрушень, більшість ризикових функцій у банках все ще перебувають у розпалі трансформацій, які відповідають цим підвищеним вимогам [4].

У багатьох фінансових установах активне управління кредитним ризиком отримує все більшу увагу регулятора. Регулятори наводять поганий менеджмент кредитного ризику на рівні портфеля, слабкі кредитні стандарти для позичальників та контрагентів та недостатня увага до змін економічних та інших обставин, що впливають на спроможність позичальників та контрагентів як найвищих учасників неадекватного управління кредитними ризиками. Регулятори змінили капітальні витрати, щоб зробити фінансові установи більш чутливими до фактичного кредитного впливу та встановили нові правила щодо того, скільки банків-капіталів потрібно виділити для покриття потенційних збитків.

Кредитний ризик - це ризик збитку, що виникає через нездатність позичальника чи контрагента виконати свої зобов'язання. Більшість кредитних ризиків фінансової установи пов'язані з її кредитною діяльністю - непогашеними позиками та орендами, активами торгового рахунку, похідними активами та зобов'язаннями без фінансування, що включають зобов'язання з позики, акредитиви та фінансові гарантії. Ризик також існує в інших видах діяльності, таких як міжбанківські операції, торговельне фінансування, роздрібні та інвестиційні розрахунки.

Управління ризиками є фундаментальним у банківській справі. Ключовими елементами для ефективного передбачення ризиків є: добре розвинена кредитна політика та пов'язані з нею процедури; ефективне управління портфелем(грошовим, кредитним, нерухомим); ефективний контроль за кредитами і, що найважливіше, добре підготовлений та навчений для роботи в даній системі персонал.

Є багато підходів для аналізу кредитних ризиків:

- Кредитний (невиконання позичальником взятих зобов'язань);

- Процентний (Різка падіння або ріст ринкових ставок);
- Ринковий (Падіння курсу цінних паперів);
- Валютний (Зміна курсу валют).

Величина банківського ризику описується трьома положеннями: повний рівень ризику, помірний рівень ризику і низький рівень ризику.

Повний ризик передбачає втрати, рівні банківським вкладенням в операцію. Так, сумнівний або втрачений кредит має повний, тобто 100-відсотковий ризик. Банк прибутку не отримує, перебуває в зоні неприпустимого або критичного ризику.

Помірний ризик (до 30%) виникає при неповерненні невеликої частини основного боргу або відсотків за позикою, при втраті лише частини суми по фінансовим або іншим операціям банку. Ризик перебуває в зоні допустимого. Банк отримує прибуток, що дозволяє покрити допущені втрати і мати доходи.

Низький ризик - незначний ризик, що дозволяє банку не тільки покрити втрати, але і отримати високі доходи. Одна і та ж сама операція може одночасно бути пов'язана з різними ступенями та зонами ризику.

Усі банки намагаються мінімізувати свої ризики, це необхідно для мінімізації втрат і подальшого розвитку фінансової установи. Мінімізація ризиків складається з : передбачення ризиків та втрат, які може нанести ризик, розмірів, наслідків, а також заходів щодо запобігання та мінімізації або усунення втрат.

1.2 Існуючий підхід до аналізу кредитних ризиків

Фінансові установи повинні встановити кредитні ліміти для контролю ризику в усіх видах діяльності, пов'язаних з кредитуванням. Ліміти за галузями, географічним регіоном, продуктом, клієнтом та країною повинні бути визначені разом із підходами, що використовуються для обчислення експозицій

щодо цих лімітів, і становлять частину кредитної політики. Слід також враховувати поширення по галузях чи регіонах, оскільки дефолт однієї фірми чи галузі також може вплинути на інші. Більші фінансові установи можуть також розглянути кілька лімітів для кожного позичальника чи групи позичальників, за продуктами, операційною одиницею та членом позичальника, щоб банківська та торгова діяльність тих позичальників чи груп позичальників, що створюють кредитний ризик, була більш адекватною моніторингом. Хоча тенденція полягає в тому, що багато фінансових установ здійснюють моніторинг загальної експозиції за цими категоріями, більшість не встановлює максимальних обмежень для цих ризиків [5].

Кредитний ризик у комерційному портфелі може бути керований на основі профілю ризику позичальника, джерела погашення та характеру базового застави з урахуванням поточних подій та умов. Управління комерційним кредитним ризиком слід починати з оцінки профілю кредитного ризику окремого позичальника чи контрагента на основі поточного аналізу фінансового стану позичальника у поєднанні з поточними галузевими, економічними та макрогеополітичними тенденціями. Як частина загальної оцінки кредитного ризику боржника, кожній комерційній кредитній операції чи транзакції слід присвоювати рейтинг ризику та підлягати затвердженню на основі стандартів затвердження, визначених у кредитній політиці. Після видачі позики рейтинги ризику слід постійно коригувати, як це необхідно, щоб відобразити зміни у фінансовому стані боржника, грошовому потоці чи постійній фінансовій життєздатності. Регулярний моніторинг спроможності позичальника чи контрагента виконувати за своїми зобов'язаннями дозволяє вносити коригування, які впливатимуть на оцінку кредитного ризику.

Для вимірювання та оцінки концентрацій у межах портфелів слід враховувати агрегацію рейтингів ризику. Рейтинги ризику також є фактором, що визначає рівень присвоєного економічного капіталу та резерв для кредитних втрат. Для управління відносним ризиком у комерційному портфелі багато фі-

нансових установ використовують участь чи синдикацію впливу інших фінансових установ чи організацій, продаж кредитів та сек'юритизацію, а також кредитні деривативи для управління величиною кредитного портфеля та відносним пов'язаним кредитним ризиком. Ці види діяльності можуть відігравати важливу роль у зменшенні кредитних ризиків з метою зменшення ризику або там, де було визначено, що концентрація кредитного ризику небажана.

Таблиця 1.1 - Класифікація кредитних ризиків

Класифікація кредитних ризиків	Вид кредитних ризиків
За сферою появи	Ризик позичальника Ризик за страхуванням Ризик кредитного продукту Ризик зміни оточуючого середовища банку
За позицією	Ризик за ринковими позиціями Портфельний ризик
За рівнем ризику	Макроекономічний ризик Ризик усієї банківської установи Ризик особи, яка приймає рішення.
За сферою використання Кредитів	Ризик по споживчим кредитам Ризик за промисловим кредитами Ризик стосовно інвестиційних кредитів.
За рівнем прогнозованості ситуації	Передбачуваний ризик Непрогнозований ризик
За причинами виникнення	Суб'єктивний ризик Об'єктивний ризик Законодавчий ризик
За розмірами можливих втрат	Великий ризик Незначний ризик Критичний ризик
За терміном дії кредитної угоди	Ризик по короткостроковими угодами Ризик сьосово довгострокових угод.
За фінансовою відповідністю	Ризик, результатом якого можуть бути фінансові втрати Ризик, внаслідок якого є втрачена вигода Ризик, внаслідок якого може бути фінансовий дохід

За видом операційного кредитного характеру	<p>Ризик при кредитуванні</p> <p>Ризик при операціях з застосуванням векселей</p> <p>Ризик при лізингових операціях</p> <p>Ризик за акредитивами</p>
Ризик за способом впливу	<p>Ризик, що регулюється банком самостійно</p> <p>Ризик, що страхує втрати перед траховою компанією (гаранту, поручителю)</p> <p>Ризик, який стосується кожного банку в системі.</p>

Управління кредитним ризиком для споживчого кредиту повинно починатися з початкового андеррайтингу та тривати протягом усього кредитного циклу позичальника. Споживчі та інші поширені ознаки для оцінки кредитного ризику. Статистичні методи можуть використовуватися для встановлення цін на продукцію, апетиту до ризику, операційних процесів та показників для відповідного збалансування ризиків та вигод. Можна придбати або створити статистичні моделі, що використовують детальну поведінкову інформацію із зовнішніх джерел, таких як кредитні бюро, а також внутрішній історичний досвід. Ці моделі слід періодично перевіряти, щоб переконатися, що вони продовжують бути статистично дійсними та відображати результати роботи клієнтської бази установи, особливо якщо вони використовуються для оцінки кредитів. При використанні ці моделі стануть основою ефективного процесу управління ризиками споживчого кредиту і можуть бути використані при визначенні схвалення / відхилення кредитних рішень, процедурах управління стягненнями, рішеннях щодо управління портфелем, адекватності резерву на втрати позики та оренди та економічному капіталі розподіл на кредитний ризик.

Забезпечення точних розрахунків експозиції за лімітами має вирішальне значення для управління кредитним ризиком. Методології будуть залежати від типів продукції. Що стосується позичкових продуктів та поточних рахунків, балансовий баланс вважається відповідним заходом, при цьому відповідні нарахування включаються як частина експозиції, оскільки дефолт контрагента за первинною експозицією також може призвести до втрати процентного доходу. Поточну ринкову вартість слід використовувати для впливу емітентів на облигації та акції, при цьому вартість заміщення торгів використовується як міра для будь-яких неурегульованих торгів. Що стосується іноземної валюти та похідних інструментів, експозиція повинна вимірюватися за заміною вартості торгів плюс додаткової вартості на основі номінальна вартість для відображення можливих майбутніх несприятливих змін у вартості заміщення.

Кредитний ризик портфеля слід оцінювати, щоб переконатися, що концентрація кредитного впливу не призводить до небажаного рівня ризику або до порушення нормативних вимог. Необхідно проводити регулярний огляд та вимірювання концентрацій кредитного впливу відповідно до встановлених лімітів у залежності від товару, галузі, географії та відносин із клієнтами. Для спеціалізованих галузей промисловості можуть бути доцільними додаткові категорії вимірювань, такі як географічне розташування та тип власності для позик на комерційну нерухомість

Періодична повторна перевірка встановлених лімітів була б доречною для того, щоб гарантувати, що ліміти продовжують відповідати стратегічному апетиту до ризику, передбачити цільову суміш активів та визнати потенційну експозицію такою, що передбачалася.

Діяльність з питань проведення регуляторних експертиз використовує різноманітні методи оцінки кредитного ризику фінансової установи, включаючи вибірку позик та огляд процесів управління кредитними установами. Розглядається складність продуктів та діяльності фінансової установи та загальної практики управління ризиками. Розробка, впровадження та коригування про-

цесів та практик ефективного управління кредитним ризиком обмежить непередбачувані ризики.

В основу банківського управління ризиками повинні бути покладені наступний принципи:

- Прогнозування ситуацій, які можуть принести збитки;
- Проведення заходів щодо зменшення втрат;
- Чітка політика і механізми управління ризиками.

Ступінь кредитного ризику залежить від таких факторів:

- Ціль кредиту;
- Розмір кредитної суми;
- Кредитна історія позичальника;
- Кредитоспроможність позичальника (DI, DTI)
- Ситуація в галузі, в якій працює;
- Форма та якість застави за кредитом.

Методи для оцінки кредитних ризиків банком бувають різні: Байєсівські мережі, Моделі двійкового вибору, Дерева рішень та інше.

В іноземних банках прийняті моделі, подані у Таблиці 1.2 для оцінювання кредитоспроможності клієнта.

Моделі для комплексного аналізу кредитоспроможності позичальника досить привабливі для використання в українській практиці (Таблиця 1.2). Однак механічно переносити зарубіжний досвід в економічний простір Україна не враховуючи особливостей і специфіки функціонування українських підприємств неефективно. Крім того, слід брати до уваги ряд недоліків, що можуть виникнути при здійсненні комплексного аналізу кредитоспроможності: це недостатнє використання математичного апарату, суб'єктивізм експертів і відсутність методологій для проведення аналізу.

Таблиця 1.2 – Системи для оцінки кредитних ризиків:

Правило шести «СІ»	CAMPARI	PARTS	PARSER
1С – репутація позичальника	С – (character) – репутація позичальника	Р – purpose – цільове призначення кредиту	Р – person – репутація позичальника
2С – фінансові можливості	А – (ability) – здатність повернути кредит	А – amount – обґрунтування суми кредиту	А – amount – обґрунтування суми кредиту
3С – Капітал, власність	М – (means) – (аналіз необхідності звернення за кредитом)	Р – repayment – умови погашення	Р – repayment – умови погашення
4С – забезпечення	Р – (purpose) – цільове призначення кредиту	Т – term – термін кредиту	С – security – забезпечення, застава
5С – загальні економічні умови	А – (amount) – обґрунтування суми кредиту	S – security – забезпечення, застава	Е – expediency – доцільність кредиту
6С – контроль	Р – (repayment) – умови погашення І – (insurance) – спосіб страхування кредитного ризику		Р – remuneration – винагорода банку

В основі наведених систем оцінки покладено базові критерії, які доповнюються.

Основним джерелом інформації для банків стосовно позичальників є внутрішня кредитна історія банку і інформація, яка була отримана в бюро кредитних історій на момент подання анкети. У таблиці 1.3 наведено список методів, які використовуються для аналізу кредитного ризику, та інформація яка необхідна для цього. Для побудови скорингової моделі можливо використовувати такі змінні: кількість рахунків клієнта існуючих на даний момент часу, кількість та наявність кредитних карт, рейтинг бюро кредитних історій, загальна сума всіх кредитних коштів, дата отримання останнього кредиту, наявність у клієнта інших кредитних продуктів або пропозицій, цієї фінансової установи, баланс і інформація стосовно поточного рахунку.

Методи, які використовуються на території СНД для оцінки клієнтів (Рис. 1.1)

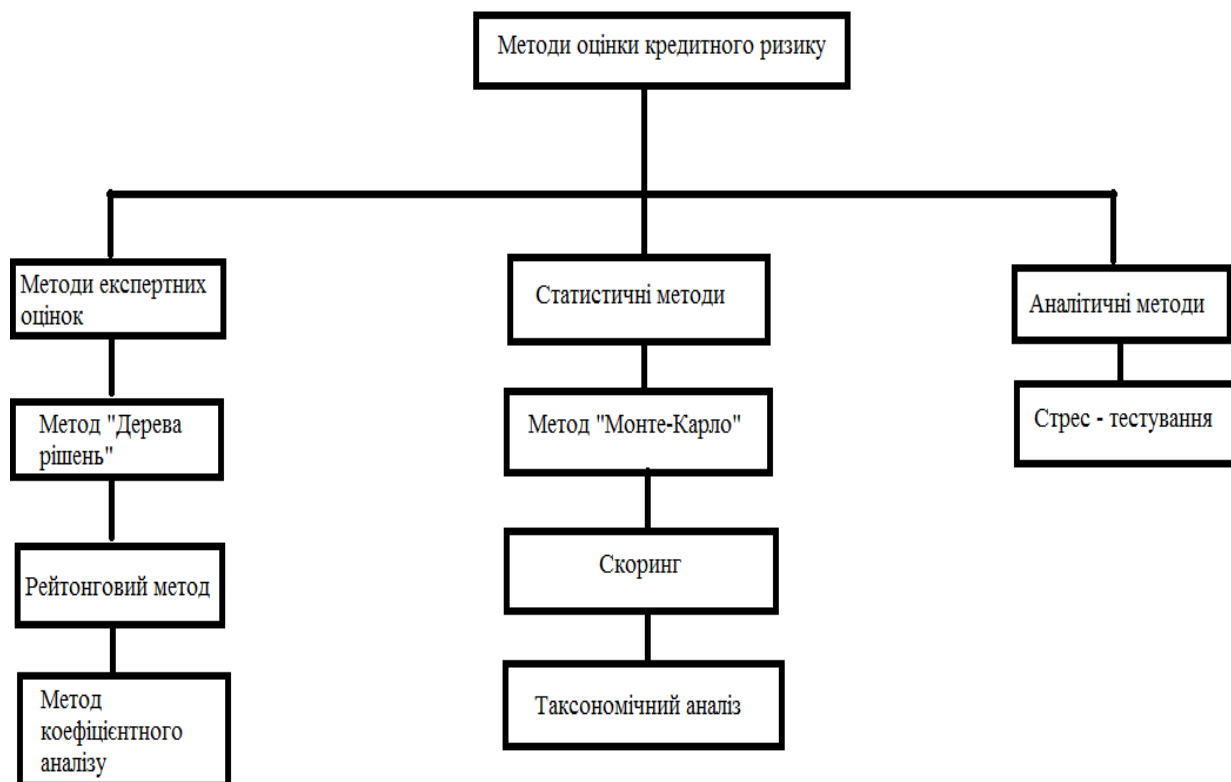


Рисунок 1.1- Методи оцінювання клієнтів у банках на території

СНД

Таблиця 1.3 - Опис методів та інформації яку вони використовують:

Скорингові моделі та методики для аналізу кредитних ризиків	Скоринг являє собою математичну або статистичну модель, за допомогою якої банк намагається визначити ймовірність повернення кредиту конкретним позичальником на основі кредитної історії	Використовуються інтегральні показники стосовно кожного клієнта: вік, професія, дохід, термін перебування клієнтом банку та інше
Метод «Монте - Карло»	Моделювання випадкових процесів за заданими характеристиками.	Аналізуються різні варіанти реалізації проекту. Ймовірні характеристики: ймовірна кредитна процентна ставка
Метод «Дерева рішень»	Графічна побудова варіантів рішень, які можуть бути прийняті.	Показники ризику для кредитного портфеля банку.
Метод рейтингів	Усі банк за певною методикою розраховує рейтинг і робить висновок про кредитоспроможність по-	Показники фінансового стану позичальника.

	зичальника.	
Таксономічний аналіз	Для групування об'єктів, що характеризуються великим числом ознак, застосовуються таксономічні процедури. Показник дає змогу впорядкувати елементи даної сукупності.	Показники ризику діючого кредитного портфеля; Показники прибутковості кредитного портфеля.
Метод коефіцієнтного аналізу	Експертний аналіз динаміки економічних коефіцієнтів, які характеризують кредитоспроможність позичальника через порівняння його показників з середніми показниками по галузі.	Показники кредитоспроможності та фінансового стану позичальника
Стрес-тестування	Дозволяє в динаміці проаналізувати зміни кількісних показників ризику.	Показники ризику кредитного портфеля.

В основу більшості опублікованих методик (Таблиця 1.4) покладено однакові характеристики для аналізу фінансового стану позичальника, а саме за фінансовими характеристиками його діяльності. Основні етапи вибору і побудови моделі складаються з обрання вхідних даних, визначення незалежних змінних, за якими буде будуватися модель. Основним джерелом для отримання даних про клієнта є його анкетні дані на момент подачі кредитної заявки, також інформація яка наявна на клієнта у бюро кредитних історій, наприклад:

- Демографічні показники (Вік, національність, стать, освіта, місце проживання, тривалість проживання в актуальному місці проживання, професія, стаж роботи, наявність власності, сімейний стан, наявність дітей і ін.);
- Дані які стосуються запитуваного кредиту (Мета кредиту, початковий внесок, загальний розмір позики, термін фінансування, відношення розміру позики до коштів якими проходить забезпечення за кредитом та інше.);
- Фінансові показники (при обліку кількісних характеристик в моделі не рекомендується використовувати абсолютні величин і використовувати коефіцієнти , наприклад: сума заборгованості до доходу, місячний вільно ная-

вний бюджет до щомісячного доходу, щомісячні виплати по кредиту до щомісячного доходу, щомісячні виплати по кредиту до щомісячного вільного бюджету та ін.

– Маркетингові показники: мотив що спонукає звернутися до банку, джерело надходження кредитної анкети, програма та ін.

Висновки

Кредитна діяльність є однією з головних операцій на фінансовому ринку. Операції пов'язані з кредитуванням — це, як правило, основна стаття банківського бізнесу, яка приносить найбільший прибуток. Саме за рахунок правильної кредитної діяльності і формується основна частина прибутку, з якої потім і складається резервний фонд, той, з якого виплачуються дивіденди акціонерам банку. Але неповернення кредитів, може призвести банк спочатку до втрати прибутку, а потім і до банкрутства, а через це можуть постраждати ряд підприємств, банків, приватних осіб. Тому правильне поводження з кредитними ризиками — це основна проблема банку, а грамотне управління ними є фундаментальною частиною стратегії виживання банку на ринку, а також розвитку.

Кредитний ризик — це ймовірність отримання збитків фінансовою установою у зв'язку з несвоєчасним поверненням або відсутністю повернення позичальником суми основного боргу і відсотків за основною сумою. Кредитний ризик включає в себе стратегічний ризик, відсотковий ризик, інноваційний ризик, ризик незбалансованої ліквідності і ризик формування ресурсної бази, валютний ризик, операційний і технологічний ризики, ринковий ризик. Вираженням ступеню ризику є відсоткова ставка за операціями, що мають кредитну природу це:

- кредит, факторинг,
- облік векселів,
- надання гарантій.

Ставки за кредитом мають компенсувати банку вартість ризиків, а також покрити втрачені гроші у разі неповернення кредиту. Ризик невиконання позичальником взятих на себе зобов'язань визначається такими факторами, які разом об'єднуються у поняття кредитоспроможності: репутація та фінансове положення, якість запропонованого забезпечення за кредитом тощо. Правиль-

ність і точність оцінки втрат залежить від обґрунтованості обрання методики оцінювання, а також швидкого реагування на зміни у фінансовому положенні клієнта.

У розділі було розглянуто актуальність задачі оцінювання кредитних ризиків, а також оглянуто сучасні методи і моделі для оцінювання кредитного ризику.

РОЗДІЛ 2 ІСНУЮЧІ МОДЕЛІ І КРИТЕРІЇ ДЛЯ АНАЛІЗУ КРЕДИТНИХ РИЗИКІВ

Проблема якісного прогнозу для економічних і фінансових процесів потребує застосування нових сучасних методів, які ґрунтуються на системному підході до розроблення відповідного програмного забезпечення. Частіше це програмне забезпечення реалізується у формі сучасних форм систем підтримки прийняття рішень, які набувають поширення як інструмент і засіб розв'язання великої кількості практичних задач. Математичне моделювання знайшло широке застосування у розв'язанні задач з економіки, біології, екології та інше. Завдяки системному підходу до моделювання та прогнозування, спрямованого на зменшення або усунення невизначеності різних типів, заповнення пропусків, обробку екстремальних значень, врахування можливих типів розподілів статистичних даних, а також використанні статистичних критеріїв для підвищення якості оброблення даних на кожному етапі з виконанням обчислень. Доцільно використовувати різні ідеологічні методи моделювання для однакових даних за для комбінування оцінок прогнозу.

Мірами якості системи є:

- надійність оцінки економічного та фінансового стану замовника;
- правильність оцінки ризику, віднесеної до конкретного класу економічних та фінансових позицій.

Беручи до уваги безпеку банку, рекомендується дотримуватися суворих заходів в стратегії кредитування. Однак дотримуватися їх занадто суворо, інколи буває зайвим, оскільки це може призвести до обмеження доступності до кредитів, погіршити показники банку та скоротити його частку на ринку. Крім зниження ризиків, дуже важливою є і якість обслуговування.

Рівень якості можна виміряти, використовуючи такі критерії:

- кредитна доступність;
- час, необхідний банку для опрацювання єдиної заявки на кредит;
- вартість кредиту.

Інструмент, що покращує якість обслуговування та мінімізує ризик кредитування, може істотно підвищити конкурентоспроможність банку.

Методи для оцінювання кредитних ризиків банком бувають різні, наприклад такі: Байєсівські мережі, Древа рішень, Моделі двійкового вибору та інші.

2.1 Байєсівська мережа

Байєсівські мережі класифікуються як графічні аналітичні моделі прийняття рішень. Їх найвидатнішою особливістю є можливість кодування як кількісних, так і якісних знань, а також можлива дуже висока розмірність моделі. Це означає, що ми можемо спиратися в цій моделі на статистичні дані та досвід експертів. Застосування БМ відбувається у багатьох сферах: медицина, фінанси, економіка, програмне забезпечення, військова справа, космічні польоти та інші. Відомі застосування БМ у системах технічної діагностики – система моніторингу космічного корабля багаторазового використання, діагностика двигунів різних типів та призначення, аналіз стану технологічних процесів і технічних систем. Своє застосування БМ знайшла в системах класифікації даних різної природи, системах автоматичного розпізнавання мовних сигналів, маркетингу і бізнесі, а також у багатьох інших сферах діяльності. Загалом Байєсівська мережа дає можливість встановити причинно-наслідкові зв'язки між подіями та визначити ймовірності настання тієї чи іншої ситуації при отриманні нової інформації стосовно зміни стану будь-якого вузла (змінної) мережі. Ступінь успішності застосування даного методу моделювання та формування ста-

тистичного висновку залежить від вміння коректно сформулювати постановку задачі, вибрати змінні процесу, які в достатній мірі характеризують його динаміку або статику, зібрати дані та використати їх для навчання мережі, а також коректно результат – висновок за допомогою побудованої мережі.

Мережа в цілому втілює структуру проблемної області, в той час як локальні взаємодії між параметрами кількісно визначаються у вигляді умовної таблиці ймовірностей. Байєсівські моделі засновані на надійних статистичних методах.

БМ необхідно розглядати як модель представлення ймовірнісних залежностей між його вершинами. $A \rightarrow B$ називають причинним зв'язком, якщо подія A є причиною виникнення B , тобто якщо існує метод впливу значень змінної A на значення в рамках яких може бути змінна B . Тоді БМ називають причинною, коли всі зв'язки в ній є причинними. Формально, БМ – це трійка $N = (V, G, J)$

- V – множина змінних,
- друге – спрямований ациклічний граф G , вузли якого відповідають першій компоненті – випадковим змінним модельованого процесу;
- J - це розподіл ймовірностей змінних $V = \{X_1, X_2, \dots, X_n\}$.

При цьому стосовно множини змінних виконується марковська умова, тобто кожна змінна в мережі залежить лише від батьківської змінної, попередника цієї змінної. І не залежить від усіх інших змінних.

Спочатку ставиться задача обчислення значень взаємної інформації між усіма вершинами мережі. На наступному кроці знаходимо оптимальну структуру мережі з використанням за критерій якості оцінку опису мережі мінімальної довжини (ОМД), значення якої аналізується і оновлюється на кожній ітерації алгоритму навчання.

Байєсівська мережа – це високо ресурсний метод ймовірнісного моделювання процесів довільної природи з невизначеностями різних типів, який забезпечує можливість достатньо точно описувати їх точне функціонування та оцінювати прогнози і на їх основі будувати системи управління.

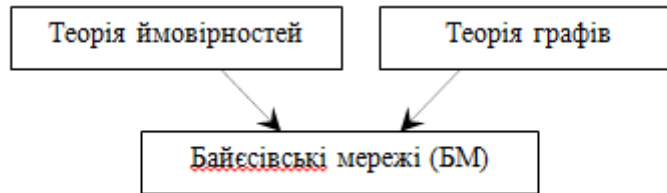


Рис.1.2 Походження байєсівських мереж

Практичне застосування байєсівських мереж впливає з їх аналітичних та діагностичних можливостей. Побудована модель дає нам стан знань кодування модельованого явища.

Перевірка нових розподілів ймовірностей робить можливим висновок щодо змін, що не спостерігаються. БМ має здатність виявляти невідомі та нетривіальні зв'язки між змінними, про які навіть експерти не завжди мають уявлення.

Кількість спостережуваних фактів може бути різною, залежно від швидкості зміни явища / процесу, витрат на збір інформації або кількість змінних, які неможливо спостерігати безпосередньо для різних причини.

Байєсівська модель (БМ) виявляє в одну з найсильніших своїх переваг - можливість робити висновки, коли наявні виміри у шумах або неповні дані.

Типи Байєсівських мереж :

1. Дискретні БМ – це такі мережі, де змінні вузлів представленні дискретними величинами. Дискретні БМ мають властивості :

– кожна вершина це є подія, яка описується за допомоги випадкової величини, яка може мати кілька станів;

- всі вершини, які мають зв'язок з «батьківськими», визначаються за допомогою таблиці умовних ймовірностей (ТУЙ) або використовуючи функцію умовних ймовірностей;

- для вершин у яких не має «батьків», ймовірності їх станів є маргінальними.

2. Дискретні БМ – мережі, у яких змінні вузлів представленні дискретними величинами. Дискретні БМ мають властивості :

- кожна вершина являє з себе подію, що описується завдяки випадковій величині, яка може мати декілька станів;

- всі вершини, пов'язані з «батьківськими», визначаються за допомогою таблиці умовних ймовірностей (ТУЙ) або використовуючи функцію умовних ймовірностей;

- для вершин у яких не має «батьків», ймовірності їх станів є маргінальними.

3. Динамічні байєсівські мережі являють з себе такі мережі, у яких значення вузлів може змінюватися з часом, тобто ця байєсівська мережа являє собою мережу, яка описує стани динамічної системи.

Процеси, які змінюються у часі можливо описати за допомогою динамічно байєсівської мережі. Перевага цих мереж є у тому, що вони застосовують табличне представлення розподілу умовних ймовірностей, що значно полегшує, сприйняття і оцінку різних нелінійних явищ. В багатьох випадках параметри моделі можуть не змінюватися з часом, але до структури мережі в разі необхідності є можливість додати додаткові приховані вузли для покращення опису поточного стану процесу.

4. Дискретні БМ – мережі, у яких змінні вузлів представленні дискретними величинами. Дискретні БМ мають властивості :

- кожна вершина уявляє з себе подію, що описується за допомогою випадкової величини яка може мати різні стани;

– всі вершини, які пов’язані з «батьківськими», визначаються за допомогою таблиці умовних ймовірностей (ТУЙ) або використовуючи функцію умовних ймовірностей;

– для вершин у яких не має «батьків» ймовірності їх станів є маргінальними (безумовними).

5. Динамічні БМ – мережі, у яких значення вузлів має змінюватися з часом, тобто це мережа, яка має описувати стани динамічної системи.

Динамічні БМ використовують для моделювання процесів, які змінюються у часі. Їх перевагою є те, що вони використовують табличне представлення умовних ймовірностей, що спрощує, представлення різних нелінійних явищ. В багатьох випадках параметри моделі не змінюються з часом, але до структуру мережі можливо вдосконалити за рахунок додавання прихованих вузлів для уточнення опису поточного стану процесу.

6. *Неперервні БМ* – мережі, де змінні вузлів – є неперервними величинами. Часто події можуть приймати будь-які стани з заданого діапазону. Тобто змінна X – це неперервна величина, яка є випадковою і простором можливих станів цієї змінної є діапазон допустимих значень які вона може приймати $X = \{ x \mid a \leq x \leq b \}$, що містить у собі нескінченну множину точок. Тоді некоректно зауважувати про ймовірності окремо взятого стану, тому що якщо їх нескінченно велика кількість, то вага кожного буде прагнути до нуля. Тому ймовірнісний розподіл для неперервної випадкової величини визначається не так, як для дискретної випадкової; для їх опису потрібно використовувати функції розподілу ймовірностей і щільності розподілу ймовірностей. Неперервні Байєсівські Мережі найчастіше використовують для моделювання стохастичних процесів з неперервним часом.

7. *Гібридні БМ* – мережі, в яких знаходяться вузли з дискретними і неперервними змінними. При використанні байєсівських мереж, що містять неперервні та дискретні змінні, можуть виникати ряд обмежень:

а– дискретні змінні не можуть мати неперервних батьків;

b– неперервні змінні, повинні мати умовний закон розподілу на значеннях батьків і нормальний закон розподілу на своїх

c– розподіл неперервної змінної X з неперервними батьками Z та дискретними батьками Y є нормальним:

$$P(X|Y=y, Z=z) = N(\mu_x(\mu_y, \mu_z), \sigma_x(\sigma_y)),$$

де μ_x, μ_y, μ_z математичні сподівання, σ_x, σ_y – дисперсії, σ_x, σ_y – середньоквадратичне відхилення, μ_x лінійно залежить від неперервних батьків, а σ_x взагалі не залежить від неперервних батьків. Це обмеження дає можливість формування точного висновку.

Більшість існуючих методів оцінювання (побудови) структури БМ можливо умовно розділити на такі категорії: (1) на основі *оціночних функцій* (search & scoring) та (2) основані на *тесту на умовну незалежність* (dependency analysis). Більшість із існуючих методів можуть зустрічатися з такими проблемами:

1. Наявність упорядкованої множини вершин (УМВ). У великій кількості методів, особливо розроблених раніше, вважається, що УМВ задана, але при використанні реальних даних це дуже часто не співпадає з дійсністю.

2. Низька обчислювальна ефективність. Деякі сучасні методи працюють без використання УМВ, а натомість використовують тест на умовну незалежність (ТУН). Однак в такому випадку необхідно зробити експоненціальну кількість таких тестів, що призведе до зменшення ефективності роботи методу через значне зростання об'єму обчислень.

3. Проблема побудови великих байєсівських мереж. Існують методи, з допомогою яких можливо побудувати структуру байєсівської мережі з декількома сотнями вершин, використовуючи для цього навчальну вибірку з мільйонів записів. Це методи Tetrad II [3] та SopLeq [4].

Методика побудови Байєсівської мережі

1. Виконати аналіз проблеми і зробити формалізовану постановку задачі; сформулювати питання, на яке під час аналізу має бути отримана ймовірна відповідь в результаті формування ймовірного
2. Визначити множину даних, що характеризують змінні задачі, отримати їх експертні оцінки або статистичні дані.
3. Поставити у відповідність до отриманих даних взаємовиключні змінні.
4. Побудувати ациклічний граф, що буде відображати існування причино – наслідкових зв'язків та істотні умови незалежності змінних.
5. Визначити апріорні ймовірності та зробити оптимізацію топології мережі на основі наявної інформації.
6. Провести навчання мережі і сформулювати висновок по відношенню до відповідних станів процесу.
7. Проаналізувати отриманні результати і зробити висновки щодо ймовірностей очікуваної події.

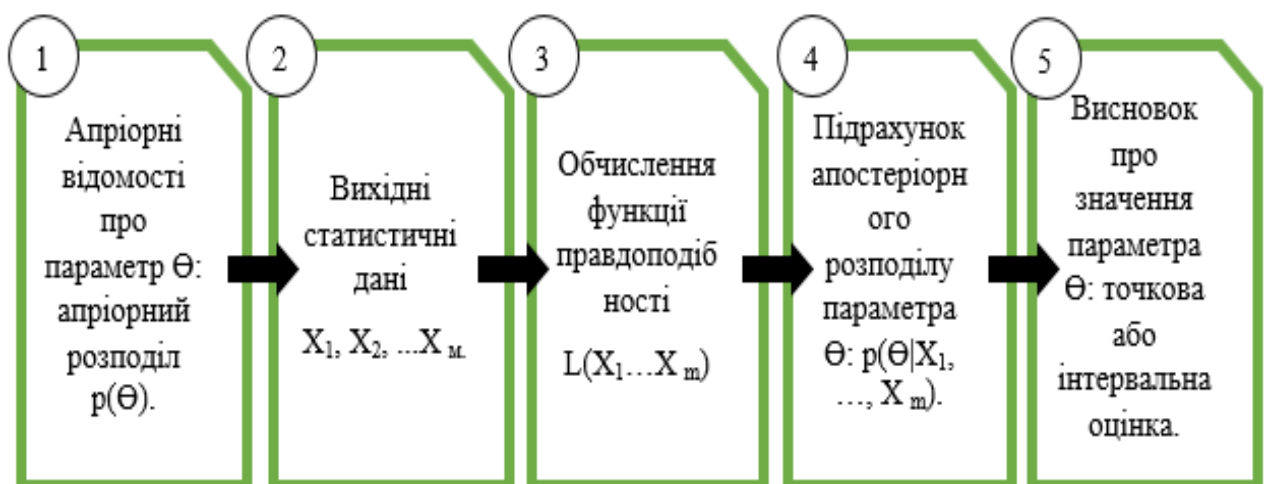


Рисунок 1.3 – Загальна логічна схема байєсівського методу оцінювання

В силу того, що БМ - це повна модель для змінних і їх відносин, вона може бути використана для того, щоб давати відповіді на ймовірнісні запити.

Наприклад, мережу можна використовувати, щоб отримати нове знання про стан підмножини змінних, спостерігаючи за іншими змінними (змінні-свідчення). Цей процес обчислення апостеріорного розподілу змінних за змінними-свідченнями називають імовірнісним висновком. Цей наслідок дає нам універсальну оцінку для додатків, де потрібно вибрати значення підмножини змінних, яке мінімізує функцію втрат, наприклад, ймовірність помилкового рішення. БМ може також вважатися механізмом для автоматичної побудови розширення теореми Байєса для більш складних завдань.

Одним з наслідків байєсівської теореми є те, що оцінка графа підтримується в обох напрямках. Процес ймовірнісного висновку в БМ супроводжується поширенням по мережі новоприбулих свідочств.

Введення в БМ довіри нових даних приводить до виникнення перехідного процесу поширення по БМ довіри свідочства, що тільки що надійшло. Після завершення перехідного процесу кожному висловленню (асоційованого з вершинами графа) приписується апостеріорна ймовірність (2.1):

$$\text{Bel}(V_j^i) = P(V_j^i|D), \quad (2.1)$$

де D – об'єднання усіх даних, які надійшли до мережі; V_j^i – композиційні висловлення, які побудовані з елементарних висловлень, тобто {значення $X_i \in V_j^i$ }; X_i – пропозиційні змінні (тобто ті змінні, значеннями яких є висловлення), які визначають стани вершин БМ довіри.

При цьому процес розповсюдження ймовірностей у БМ довіри ґрунтується на механізмі перерахунку основою для функціонування котрого є така послідовність дій:

- з кожною вершиною наявною в мережі асоційований обчислювальний процес (процесор), якому надходить повідомлення від сусідніх, які пов'язані з ним дугами процесорів;

- цей процесор проводить перерахунок апостеріорних ймовірностей $Bel(V_j^i)$ для всіх можливих значень V_j^i даної змінної X_i та відправляє сусіднім вершинам повідомлення-відповіді;

- діяльність процесора запускається при порушенні умов узгодженості між станами сусідніх процесорів і триває до відновлення цих умов.

У деяких системах, що реалізують байєсовські мережі довіри використовується метод *noisy or gate*, який дозволяє значно спростити обчислювальний процес. Суть його полягає в тому, що серед ряду прикладів вершина "у" може бути умовно незалежна від цілого ряду вершин "x_i", де $i = 1, 2, \dots, n$. Для того, скорочення оцінки 2^n ймовірностей, які необхідні при використанні таблиць умовних ймовірностей, використовується даний метод. Згідно нього, ймовірність "у" у залежності від n вершин "x_i" оцінюється як (2.2):

$$p(y|x_1, x_2, \dots, x_n) = 1 - \prod_{i=1}^n (1 - p(y|x_i)), \quad (2.2)$$

що дає змогу оцінити тільки $p(y|x_1), p(y|x_2), \dots, p(y|x_n)$ та на їх підставі сформулювати оцінку $p(y|x_1, x_2, \dots, x_n)$.

Для проведення ймовірнісного висновку в БМ використовуються точні та наближені алгоритми.

Точні алгоритми:

- висновок методом грубої сили шляхом маргіналізації повного спільного розподілу;
- алгоритми усунення змінних і символічні обчислення;
- кластеризація;
- алгоритми пропагації (передачі) повідомлень між вузлами мережі.

Наближені алгоритми на основі методу Монте-Карло:

- алгоритми формування вибірок з виключенням;
 - метод оцінки вибірок з урахуванням правдоподібності;
- алгоритм МСМС (англ. Markov chain Monte Carlo) та ін.

2.2 Дерева Рішень

Метод дерева рішень - це один з методів автоматичного аналізу який використовується для аналізу величезних масивів даних.

Дерева рішень, як метод data-mining, має багато корисних здібностей, що використовуються в різноманітних прикладних сферах для бізнесу, а не тільки для науки. Ось деякі корисні здібності дерев рішень:

- Надають візуальний максимально зрозумілий результат. Дерева рішень досить просто будувати, розуміти отриманий результат та використовувати. Здатність представити багато пояснюючих факторів процесу в простій формі step-by-step. Корисна здібність дерев рішень – будувати по ітераціям правила високої складності.
- Здібність працювати як з кількісними так і якісними (наприклад, коли цільова зміна приймає тальки два значення - good або bad) даними.

Кількісні дані включають ординарні (наприклад, high, medium, low категорії) та інтервальні (наприклад, температура пацієнта, розмір доходу родини, вага тіла) рівні змінних.

- Легко працюють з даними, що мають ефекти – незбалансованості, гніздові ефекти, взаємного перекриття, пересічення змінних та не лінійності. З цими ефектами прості одно факторні та багатofакторні статистичні підходи не працюють.
- Дерева рішень характеризуються непараметричністю, високим рівнем робастності (наприклад, легко працюють з пропусками в да-

них) та будують дуже близькі структури в незалежності від рівня виміру змінних (наприклад, дерево рішень дасть близький результат для випадку, коли дохід вимірюється в десятках, сотнях, тисячах, та навіть дискретних станах від 1 до 5).

В основі дерев рішень знаходяться алгоритми, що дозволяють розділити багатьма можливими способами набір даних на сегменти, кожен з яких буде містити максимально схожі елементи. Приклад простого дерева рішень наведений на рис. 8.1, як можна побачити відбувається розбиття на дві частини – праву і ліву, відносно критерію розбиття, що за соєю суттю представляє правило вигляду IF-THEN. Набір таких правил призводить до кінцевого листа, що відображає значення ймовірності прийняття того чи іншого рішення – $\text{Target} = X\%$ або $\text{Target} = Y\%$.

Для створення правил використовуються змінні (в наборі даних змінна це стовбець). На кожному рівні обирається змінна відносно якої обчислюється точка для розбиття. На рис. 8.2 наведені рівні дерева рішень. Найнижчі вузли (рівень 2) називаються листами дерева або термінальними вершинами. Кожному листу дерева рішень відповідає унікальний шлях до кореня.

Розділення починається з обрання вхідної змінної для розбиття наявних навчальних даних. Якщо шкала виміру обраної вхідної змінної категоріальна, то кожне унікальне значення використовується в якості потенційної точки розділення даних. Якщо вхідна змінна інтервальна, то розглядається середнє значення цільової змінної в кожному категоріальному рівні вхідної змінної. Середні значення мають таку ж саме роль що і унікальні значення категоріальних змінних. Для обраної вхідної змінної та фіксованої точки розділу генеруються дві групи.

Спостереження з вхідними значеннями, меншими точки розділення, мають назву – лівої гілки. Спостереження з вхідними даними, більшими точки розділення, мають назву – правої гілки.

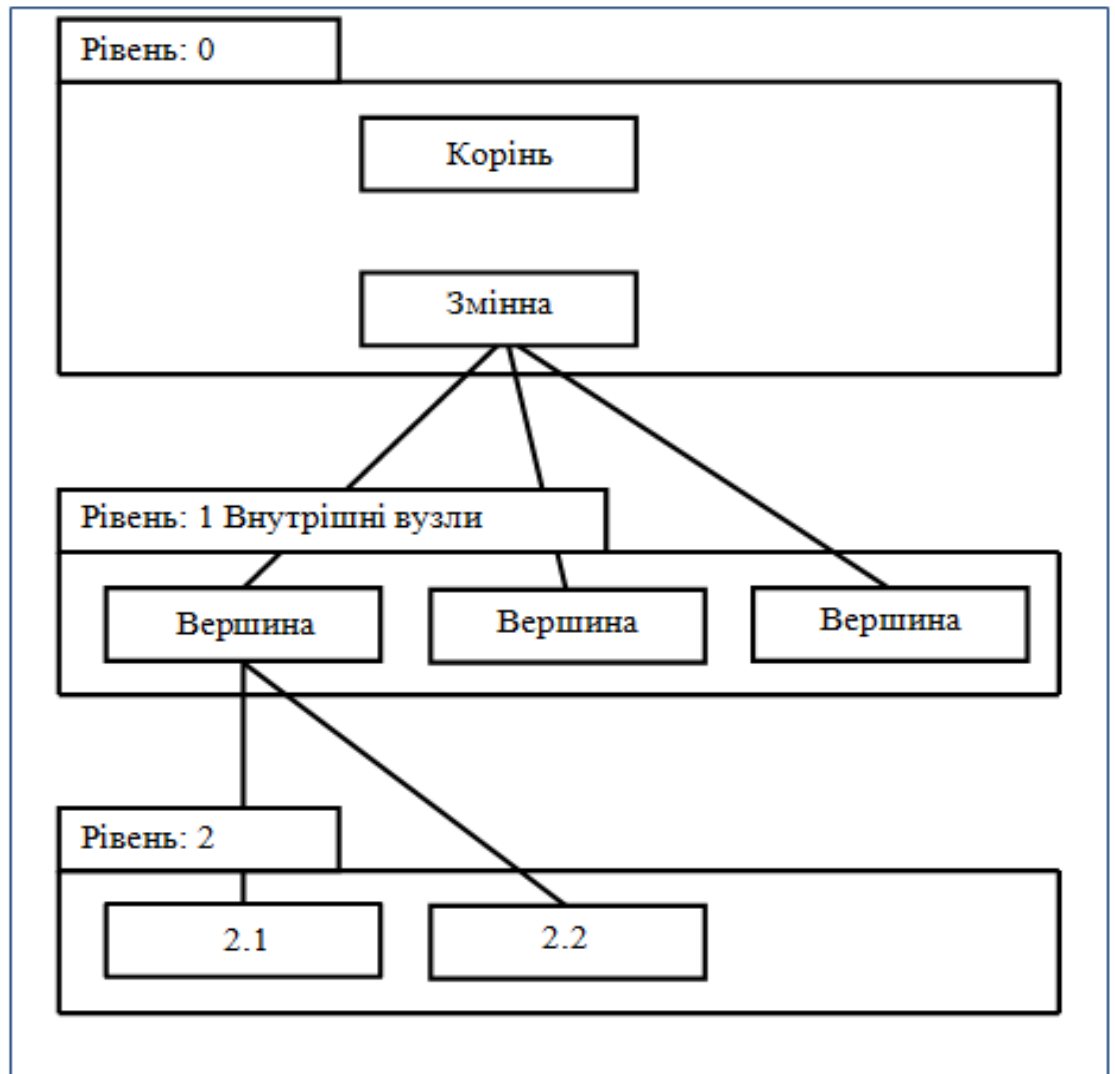


Рис.1.4 Рівні дерева рішень

Велика різниця в пропорціях значень цільової змінної вказує на якісне розділення.

В загальному випадку набір даних для аналізу поділяють на два набори

– тренінгів (навчальний) та перевірочний (валідаційний). Після того як побудоване максимальне дерево рішень на основі тренінгового набору даних починається етап скорочення дерева рішення. Існують різні типи статистики, що відповідають кожному з трьох існуючих в загальному випадку типів прогнозів – рішень, рейтингів та оцінок.

2.3 Логістична регресія

Логістична регресія – різновид множинної регресії, загальне призначення якої є в аналізі зв'язку між кількома незалежними змінними (регресорами або предикторами) і залежною змінною. Якщо залежна змінна є бінарною, тоді застосовується бінарна логістична регресія. З допомогою логістичної регресії з'являється можливість оцінити вірогідність того, що подія може статися для конкретного експеримент (хворий/здоровий та інше). Всі регресійні моделі маємо змогу записати у вигляді формули:

$$y = F(x_1, x_2, \dots, x_n).$$

Існує кілька способів знаходження коефіцієнтів логістичної регресії. На практиці найчастіше використовують метод максимальної правдоподібності. Він застосовується для того щоб отримати оцінки параметрів сукупності за даними, які наявні у вибірці. Основною у методі є функція правдоподібності, яка показує щільність ймовірності появи спільних результатів вибірки Y_1, Y_2, \dots, Y_k :

$$L(Y_1, Y_2, \dots, Y_k; \theta) = p(Y_1; \theta) * \dots * p(Y_k; \theta).$$

На основі методу максимальної правдоподібності, у якості оцінки невідомого параметра беремо таке значення $\theta = \theta(Y_1, Y_2, \dots, Y_k)$, яке максимізує функцію L . Знаходження оцінки стає легше, якщо максимізувати не саму функцію L , а логарифм $\ln(L)$, адже максимум обох функцій буде досягнуто при одному і тому ж значенні θ :

$$L^*(Y; \theta) = \ln(L(Y; \theta)) \rightarrow \max.$$

2.4 Скорингове моделювання

Скоринг уявляє з себе математичну модель у вигляді зваженої суми певних характеристик на основі якої фінансова установа хоче з'ясувати яка вірогідність того, що позичальник не зможе поверне кредит вчасно. Ідея скорингу полягає в тому, що кожному параметру надається оцінка в балах в залежності від значущості цього параметра. Перевага для фінансової установи полягає в тому, що така система оцінки позичальника є знеособленою.

В основу скорингу покладено кредитну історію позичальника, з метою класифікації та визначення характерних ознак надійних та ненадійних клієнтів щодо погашення кредитної заборгованості. Скоринг допомагає оцінити ступінь надійності потенційного клієнта. Результатом реалізації скорингової моделі є інтегральний показник, вище якого позичальник отримує позику, тому що він перевищує рівень беззбитковості, в інакшому випадку йому відмовляють у позиці. Складність побудови скорингової моделі полягає у виборі характеристик, які варто включати в модель.

Успішність скорингових моделей пояснюється такими факторами:

- неупередженість оцінки
- стандартизація кредитних оцінок
- автоматизація процесу

Скорингову модель будують на основі різних статистичних моделей:

- множинна лінійна регресія;
- нейронні мережі;
- множинна логарифмічна регресія;
- дерева рішень.

2.5 Критерії для аналізу адекватності моделі

2.5.1 Критерій Колмогорова-Смірнова

Критерій Колмогорова-Смірнова як відстані між емпіричним та теоретичним законом бере значення, яке обчислене за формулою:

$$D = \sup_x |F_n(x) - F(x)|$$

В наведеній формулі \sup - являє собою точна верхня грань множини $|F_n(x) - F(x)|$. Тобто, йде пошук найбільшої вертикальної різниці між емпіричними $F_n(x)$ та теоретичним $F(x)$ розподілами. Найчастіше використовують статистику виду:

$$S_k = \frac{6nD_n+1}{6\sqrt{n}},$$

$$D_n = \max(D_n^-, D_n^+);$$

$$D_n^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - F(x_i, \theta) \right),$$

$$D_n^- = \max_{1 \leq i \leq n} \left(F(x_i, \theta) - \frac{i-1}{n} \right),$$

де n - об'єм вибірки, x_1, x_2, \dots ,

x_n - впорядковані за зростанням вибіркві значення;

$F(x_i, \theta)$ - функція закону розподілу, узгодженість з яким перевіряється.

Розподіл величини S_k при простій гіпотезі відповідає закону Колмогорова $K(S)$:

$$\forall s > 0: \lim P(\sqrt{n} * D_n \leq S) = K(S) = \sum_{j=-\infty}^{+\infty} (-1)^j * \exp(-2 * j^2 * S^2)$$

Якщо статистика $\sqrt{n} * D_n$ збільшує квантиль розподілу Колмогорова $K(S)$ заданого рівня значимості S , тоді нуль-гіпотеза H_0 (про відповідність закону $F(x)$) відкидається). Інакше, гіпотеза приймається на рівні S .

Чим менше значення отримує статистика Колмогорова-Смірнова, тим краще.

2.5.2 Індекс Джині

Один з найвідоміших і найпопулярніших критеріїв розбиття отримав названий на честь італійського економіста і статиста. Критерій базується на дослідженні різноманітності сукупностей. Вона визначає ймовірність того, що декілька об'єктів обраних з однієї сукупності випадковим чином, відносяться до одного класу. Очевидно, що для зовсім чистої вибірки - ця ймовірність має дорівнювати 1.

При використанні для обчислення регресійного прогнозу та рівнянь інших типів ми отримуємо точкову оцінку. Але, така оцінка є не завжди може бути значущою. При прогнозуванні також необхідно визначити інтервал, всередині якого з достатньою долею впевненості ми отримаємо фактичне значення показника. В регресійному аналізі границі інтервалу задаються використовуючи довірчий інтервал.

Довірчий інтервал – це інтервал, в якому перебуває сам прогноз і в якому з великою ймовірністю можна очікувати появу фактичного значення прогнозованої змінної.

Так, значення прогнозу $\hat{y}(k+1) = 1500$ з довірчим інтервалом ± 150 і ступеню впевненості 95% означає, що з ймовірністю $\approx 0,95$ очікується, що на-

ступне значення прогнозованої змінної буде знаходитись в межах $1350 \div 1650$. Якби довірчий інтервал складав ± 500 , то з ймовірністю $\approx 0,95$ очікується, що майбутнє значення буде знаходитись в інтервалі $1000 \div 2000$.

В регресійному аналізі мінімальна ширина довірчого інтервалу відповідає точці (\bar{y}, \bar{k}) – “середині спостережень”. З обох боків від цієї середини довжина інтервалу збільшується. Для визначення довірчого інтервалу, необхідно знайти стандартну похибку рівняння регресії S_r , яка обчислюється всіма пакетами статистичної обробки даних. Величину S_r називають середньоквадратичним відхиленням і обчислюють за формулою:

$$S_r = \sqrt{\frac{1}{N-p} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}$$

де P – число параметрів моделі, які оцінюються в процесі її побудови.

Якщо є прогнози які не зміщені (це повинен забезпечувати метод прогнозування), то комбінований прогноз теж буде незміщеним. Похибка комбінованого прогнозу:

$$e_c(k) = y(k) - \hat{y}_c(k) = y(k) - \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2} = \frac{e_1(k) + e_2(k)}{2},$$

де $y(k)$ – фактичне значення прогнозованої змінної.

Дисперсія похибки комбінованого прогнозу:

$$\text{var} \left[\frac{e_1(k) + e_2(k)}{2} \right] = E \left[\frac{e_1(k) + e_2(k)}{2} \right]^2 = \frac{1}{4} E [e_1^2(k) + 2e_1(k)e_2(k) + e_2^2(k)] =$$

$$\begin{aligned}
&= \frac{1}{4} \left\{ E[e_1^2(k)] + 2E[e_1(k)e_2(k)] + E[e_2^2(k)] \right\} = \\
&= \frac{1}{4} \left[\sigma_1^2 + 2 \frac{E[e_1(k)e_2(k)]}{\sigma_1 \sigma_2} \sigma_1 \sigma_2 + \sigma_2^2 \right] = \frac{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}{4}.
\end{aligned}$$

Таким чином, дисперсія комбінованого прогнозу обчислюється за виразом:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4},$$

де ρ – коефіцієнт кореляції між похибками прогнозу.

Якщо похибки прогнозування за двома моделями незалежні, тобто, $\rho = 0$, то формула спрощується:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2}{4}.$$

Таким чином, якщо дисперсії близькі за значеннями і незалежні, то дисперсія комбінованої похибки буде значно меншою будь-якої з двох дисперсій. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$:

$$\sigma_c^2 = \frac{100+100}{4} = 50.$$

Але навіть при існуванні досить високої кореляції між похибками прогнозування дисперсія похибки комбінованого прогнозу буде меншою ніж дисперсія кожного методу окремо. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$ і $\rho = 0.8$:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100+100+2\cdot 0.8\cdot 10\cdot 10}{4} = 90.$$

Навіть в цій ситуації спостерігається зменшення дисперсії похибки прогнозу після усереднення оцінок, отриманих за двома методами.

Однак, ситуація змінюється у випадку, коли дисперсії індивідуальних похибок сильно відрізняються. Наприклад, нехай $\sigma_1^2 = 100$, $\sigma_2^2 = 16$; $\rho = 0.8$:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100+16+2\cdot 0.8\cdot 10\cdot 4}{4} = 45.$$

Таким чином, якщо дисперсії похибок сильно відрізняються, то просте усереднення результатів не завжди може допомогти. Просте усереднення потрібно застосовувати лише у тих випадках, коли дисперсії індивідуальних похибок прогнозування приблизно дорівнюють одна одній або не дуже сильно відрізняються між собою.

Якщо інформація стосовно характеристик індивідуальних прогнозів немає, то необхідно присвоїти різні вагові коефіцієнти окремим прогнозам спираючись на суб'єктивних або експертних суджень:

$$\hat{y}_c(k) = w_1 \hat{y}_1(k) + w_2 \hat{y}_2(k),$$

де w_1, w_2 – вагові коефіцієнти. Зрозуміло, що більші значення вагових коефіцієнтів треба присвоювати тим індивідуальним прогнозам, у яких менша дисперсію похибок. Також для коректності обчислень треба, щоб виконувалась

умова: $w_1 + w_2 = 1$.

2.5.3 ROC – крива

Площа під ROC – кривою є однією з найпопулярніших функціоналів якості в задачах бінарної класифікації. ROC крива – графік, який надає змогу оцінити якість бінарної класифікації. Графік відображує співвідношення між об'єктами які були вірно класифіковані від загальної їх кількості. Результат роботи на фіксованій тестовій вибірці візуалізують за допомогою ROC – кривої, а якість самої моделі оцінюють за допомогою площі під цією кривою - AUC (area under the curve). AUC – надає кількісну інтерпретацію ROC – кривої, чим вище показник AUC, тим якісніше класифікатор. Значення яке дорівнює 0,5 вважається хибним оскільки така модель є нездатною для класифікації і класифікатор діє з точністю навпаки.

Класифікація :

- P (True Positives) – приклади, які вірно класифіковані позитивні (так звані істинно позитивні випадки);
- TN (True Negatives) – негативні приклади, які вірно класифіковані(істинно негативні випадки);
- FN (False Negatives) - позитивні приклади, які були класифіковані як негативні (помилка I роду). Це "помилковий пропуск" - коли цікавить нас помилково не виявляється (помилково негативні приклади);
- FP (False Positives) - негативні приклади, які були класифіковані як позитивні (помилка II роду); Це являє з себе помилкове виявлення, адже при відсутності події помилково виноситься рішення про його присутності (помилково позитивні випадки).

Частка істинно позитивних прикладів (True Positives Rate):

$$TPR = \frac{TP}{TP + FN}$$

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

Частка ложно позитивних прикладів (False Positives Rate).

Чутливість (Sensitivity) - це і є частка істинно позитивних випадків:

$$Se = TPR = \frac{TP}{TP + FN} \cdot 100\%$$

Специфічність (Specificity) – це є частка істинно негативних випадків, які модель правильно ідентифікувала:

$$Sp = \frac{TN}{TN + FP} \cdot 100\%$$

ROC – крива – будується наступним чином :

- Для кожного значення з порогу відсікання, яке змінюється від 0 до 1 з кроком dx (наприклад, 0.01) рахуємо значення чутливості Se і специфічності Sp . Як альтернатива, порогом відсікання може бути кожне наступне значення прикладу в вибірці.
- На другому етапі будується графік залежності: за осю Y відкладається чутливість Se , по осі X - $100\% - Sp$, або, що те ж саме, FPR - частка хибно позитивних випадків

2.5.4 AUC

Показник AUC – назначений для порівняльного аналізу декількох моделей, тому на його основі не можливо робити точних висновків стосовно адекватності моделі.

AUC не містить в собі ніякої інформації, щодо чутливість і специфічність моделі:

$$AUC = \int f(x) dx = \sum_i [X_{i+1} + X_i] \cdot (Y_{i+1} - Y_i).$$

Висновки

В даному розділі розглянуто математичні методи оцінювання кредитних ризиків та можливість їх застосування для українських банків та інших фінансових установ.

При побудові системи управління ризиками, треба зважати на описані вище відмінності моделей. Відмінності полягають як в придатності різних моделей для вирішення поставлених задач, що виникають під час управління ризиками. Так і в рівні вимог, які виставляються до обсягу і якості вхідних і вихідних даних.

Перевагою моделей, які засновані на аналізі різного типу ризиків, є можливість глибокої адаптації моделі під особливості бізнесу конкретної банківської системи, а також, якщо необхідно зробити зміни бізнес-моделі - як глобальні, так і локальні (наприклад, покращення системи внутрішніх контролів).

Перевагою моделей, які аналізують наслідки, є їх відносна об'єктивність - в основі розрахунків знаходяться фактичні дані, це дає змогу уникнути помилок, або маніпуляцій з боку фахівців, які відповідають за оцінку ризику.

Разом з тим, моделі, засновані на аналізі наслідків, можуть мати недоліки, характерні для великої кількості статистичних методів. Зокрема, до розгляду береться лише конкретна вибірка, за якою проходить оцінка властивостей генеральної сукупності. При цьому, дана вибірка може бути не достатньо репрезентативною – як внаслідок малого обсягу інформації, так і внаслідок ймовірного відхилення вибірки від середніх значень.

При виборі моделі необхідно зробити оцінку, наскільки наявні дані дозволяють відкалібрувати модель та провести необхідні розрахунки.

Безпосередньо можливо використати такі методи, як регресійні моделі, сценарний аналіз та методи нечіткої логіки. Однак, якщо необхідно застосування сценарного аналізу і методів нечіткої логіки тоді потрібна значна кількість експертних оцінок. Такі оцінки не завжди є цілком достовірними, і

тому всі експерти можуть проводити оцінювання лише в рамках тільки своєї області, і такі оцінки не можуть враховувати існуючих взаємозв'язів, які наявні в процесі в цілому. Для застосування регресійних моделей потрібна мати значну кількість даних для того, щоб відкалібрувати модель.

Також для проведення такого рода досліджень можливо використовувати нейронні мережі або мережі Байєсса.

Для того, щоб оцінити кредитні ризики в рамках даного дослідження було обрано регресійні моделі, які дають змогу враховувати не лише тривіальні залежності а також дають змогу враховувати і залежності між факторами ризику за допомогою створеної карти ризиків.

РОЗДІЛ 3 МОДЕЛЮВАННЯ КРЕДИТНИХ РИЗИКІВ

У цьому розділі представлений опис моделей, які були розроблені в рамках магістерської дисертації за допомогою програмного забезпечення SAS та GENIE. Використанні програмні продукти призначені для спрощення процесу побудови й аналізу моделей, які можна застосувати у різних сферах економічної діяльності.

У Sas Enterprise Miner та Genie процес аналізу даних відбувається за схемою процесу, яку ми створюємо, завдяки інструментам аналізу, які наявні в робочій області програми. Інтерфейс програми є інтуїтивно зрозумілим, як для досвідчених користувачів, так і для людей, які вперше починають користуватися програмою, але мають при цьому базові знання в статистиці.

SAS Enterprise Miner дає змогу автоматизувати процес підрахунку коефіцієнтів, які потрібні аналітику, для прийняття своєчасного рішення і аналізу ситуації. Код який було створено під час моделювання може бути розгорнуто в різноманітних середовищах.

Sas Enterprise Miner та Genie є настільним програмним продуктом, оскільки він не розрахований на роботу в мережі.

3.1 Архітектура і функціональна схема моделі

Моделі, які нами було описано у попередньому розділі, передбачають проведення аналізу величин різних економічних показників. Нам необхідно побудувати модель, що дасть змогу зменшити втрати різним фінансовим установам.

Рисунок 3.1 показує структуру моделі, яка включає у себе великий комплекс з обробки й аналізу даних.

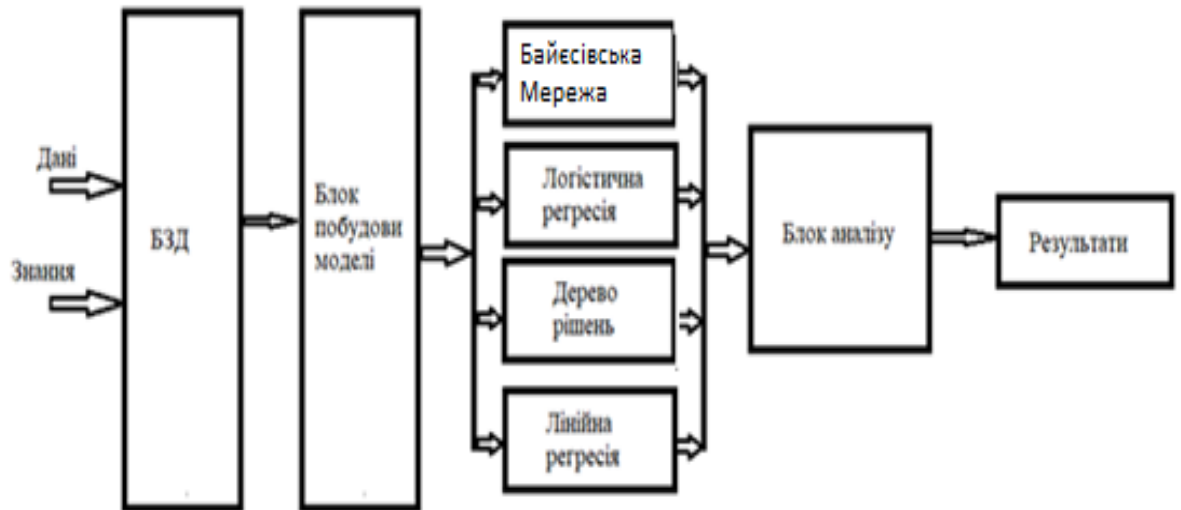


Рисунок 3.1 – Структура моделі

Дані включають у себе:

- Дані які наявні на клієнта;
- втрати внаслідок непогашення кредиту;
- втрати внаслідок дефолту, тощо.

Знання можуть включати в себе:

- обчислювальні процедури;
- константи, які необхідні для обчислень;
- критерії за допомогою яких буде обрано найкращу модель.

У блоці аналізу знаходяться критерії за якими ми проводимо порівняння моделей на адекватність.

А у блоці результатів знаходиться моделі, які ми отримали.

3.2 Побудова Байєсівської мережі

1. В якості проблематики для якої буде побудовано Байєсівську мережу обрано проблему аналізу показників, які клієнти фінансових установ

надають для отримання кредиту, а також виявлення взаємозв'язків, як це впливає на своєчасне погашення кредиту. І яка вірогідність того, що клієнт, який отримав кредит і підходить по усім параметрам скорингової карти поверне кредит до фінансової установи.

2. В якості цільових змінних були обрані такі параметри:
 - Вік
 - Освіта
 - Професія
 - Сімейний стан
 - Історія роботи
 - Історія платежів і кредитна історія у інших фінансових установах
 - Активи
 - Дебітні рахунки
 - DTI – показник, який показує, яке кредитне навантаження припадає на даний момент на позичальника.
 - Сума кредиту
 - Відсоткова ставка.

Для побудови моделі були отриманні експертні оцінки, а також набір даних на яких буде проводитися аналіз та навчання Байєсівської мережі.

3. На цьому етапі для усіх цільових змінних було визначено множину їх значень за рамки яких вони не можуть виходити. Інформація, яка не потрапляє до множини значень будемо вважати збуреннями і при підготовці даних у вибірці будемо за змогою виправляти не форматні данні, в ситуації коли це неможливо, будемо очищувати вибірку від такої інформації.
4. Було знайдено взаємозв'язки між даними і на їх основі було побудовано ациклічний граф, який показує причино-наслідкові зв'язки у нашій вибірці даних.
5. На основі наявних оцінок експертів були визначенні апріорні ймовірності та оптимізовано топологію Байєсівської мережі.
6. На основі наявної вибірки даних було проведено навчання Байєсівської мережі, а також розраховано нові апріорні ймовірності. Модель спочатку навчали на тренувальній вибірці у якій було почищено значення від збурень та пустих значень, та яка є меншою за розміром. Далі було проаналізовано значущість наявних змінних.

7. На останньому етапі Байєсівську мережу було навчено на тестовій вибірці і отриманні результати будемо порівнювати з іншими побудованими моделями.

3.3 Аналіз результатів виконання обчислювальних експериментів

Набір даних містить 4 375 439 записів і 15 показників по клієнтам.

Початковий набір даних ми поділили на тренувальну та тестову вибірки розміром 80% та 20% відповідно.

Для першої частини аналізу було побудовано 8 моделей, з яких – 3 дерев рішень, 5 регресії (4 логістичні і 1 лінійна). (Таблиця 3.1)

Друга частина складається з побудови Байєсівської мережі. При побудові Байєсівської мережі необхідно визначити взаємозв'язки між змінними моделі. Для однієї частини параметрів такі зв'язки є очевидними для іншої частини параметрів зв'язки можуть бути відсутніми. У роботі використовувалися такі змінні (Таблиця):

Таблиця 7.1 – Вершини байєсівської мережі

Назва вершини	Опис	Множина значень
Історія платежів (PaymentHistory)	Внутрішній показник, характеризує чи були у клієнта заборгованості .	{“acceptable”, “excellent”}
Історія роботи (WorkHistory)	Внутрішній показник діяльності, який характеризує, як часто клієнт змінює роботу	{“justified”, “Stable”, “uniustifi”}
Дебіт (Debit)	Внутрішній показник діяльності, який показує чи є заборгованості у клієнта	{“0-11100”, “11100-”, }
Сума кредиту (Income)	Сума, яку позичальник хоче взяти	{“ 0-1200”, “12000-45000”, “45000-”}
Активи (Active)	Кількість активів, які є у його розпорядженні	{“average”, “poor”, “wealthy”}

Надійність (DTI)	Внутрішній показник, який показує до якого числа клієнтів відноситься позичальник, за своєю історією у цьому банку.	{“reliable”, “unreliable”}
Відсотки (worth)	Відсоткова ставка, яка запропонована за кредитом	{“high”, “low”, “medium”}
Професія (profession)	Показник рівня заробітної плати	{“high”, “low”, “medium”}
Вік (Age)	Показник того, до якої вікової групи відноситься позичальник	{“16-21”, “21-65”, “66-”}
Безпека кредиту (CreditWorthiness)	агальна оцінка ризиків стосовно видачі кредиту	{“rarely”, “neutral”, “often”}

БМ будується на основі карти ризиків – аналізу причинно-наслідкових зв’язків, історії фінансової установи, аналізу основних процесів компанії, оцінки факторів ризику. Навчаючою вибіркою в даному випадку слугує набір експертних оцінок деякого банку. Загальний вигляд БМ представлено на рисунку. Для побудови Байєсівської мережі були обрані найбільш значущі змінні і на їх основі було побудовано модель.

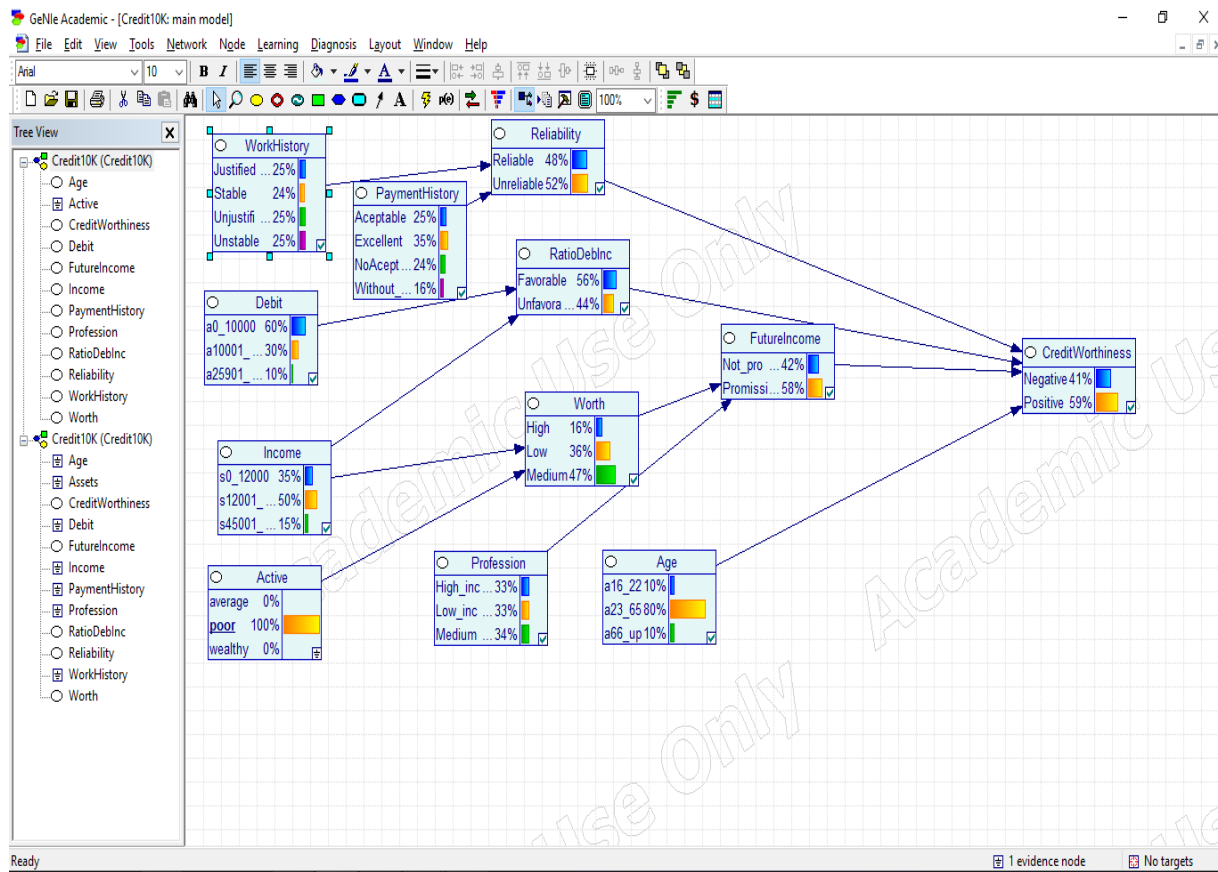


Рисунок 3.2 – Байєсівська мережа

Таблиця 3.2 – Характеристика для порівняння моделей

Тип моделі	Навчальна вибірка (Trainе)			Тестова вибірка (Test)		
	Загальна точність (CA)	Індекс Gini	AUC	Загальна точність (CA)	Індекс Gini	AUC
Дерево рішень	0	0.589	0	0	0.593	0
Логістична регресія	0,485703	0.719	0.844	0,853	0,745	0,853
Лінійна регресія	0,488703	0,696	0,163	0,488469	0,634	0,160
Байєсівська мережа	0,468703	0,654	0	0,489469	0,655	0

Згідно з таблицею 3.1 для подальшого будування моделей були вибрані логістична регресія та Байєсівська мережа.

Для обраних моделей проводимо тестування на іншому розподілі даних, а саме: 85% навчальна та 15% тестова. Ї

На основі проаналізованих даних було побудовано скорингову карту(Рисунок 3.1):

Скорингова карта має допомогти в подальшому більш швидкому аналізу даних, отриманих від осіб, які подають документи для отримання кредиту.

Scorecard		Scorecard Points
Age	Age< 27	59
	27<= Age< 31	27
	31<= Age< 38	73
	38<= Age< 50, _MISSING_	37
	50<= Age	18
CompanyType	ESB, WB	95
	AG, CA, CI, HI, RE	95
	ALC, ASTL, MPP, PSL, RS	62
	0, AP, EOG, FMCG, FMG, FP, HRC, IIT, LI, MFC, OT, _MISSING_, _UNKNOWN_	53
	EC, FS, MFS, MIL, SCI, TSPW	18
Downpayment	Downpayment< 326, _MISSING_	51
	326<= Downpayment< 505	53
	505<= Downpayment< 1098	53
	1098<= Downpayment< 1600	29
	1600<= Downpayment	18
FullEmploymentTerm_Part1	24, 6	39
	12, 60	-496
	120, 18, 24_60, 3, _MISSING_, _UNKNOWN_	18

Рисунок 3.3 – Скорингова карта

goods_type	PK	38
	MOB	93
	BUMAT, HAB	39
	ACS, EL, FURN, HAWB, TV 40 SAMSUNG LE 40 S71BX, _MIS SING_, _UNKNOWN_	20
	HAWB	18
loan_amount	loan_amount< 1439	-126
	1439<= loan_amount< 2718, _MIS SING_	-74
	2718<= loan_amount< 3000	-48
	3000<= loan_amount< 3874	-26
	3874<= loan_amount	18
price	_MIS SING_	18
	price< 1853.55	-15
	1853.55<= price< 3215.83	-9
	3215.83<= price< 5279	-7
	5279<= price	18

Рисунок 3.4 – Скорингова карта

Висновки

У третьому розділі побудовано байєсівську мережу і скорингову карту на основі наявної вибірки, а також використовуючи логістичну і лінійну регресію.

Моделі будували за допомогою Sas Enterprise Miner та Genie. Оюидві програми спрощують процес обробки даних. Допомогають швидко побудувати необхідні моделі. Обидва програмних забезпечення можливо використовувати для аналізу даних різної природи незалежно від сфери. Програми допомагають у вирішенні різних проблем: аналіз ризиків, аналіз фінансового портфеля установи, сегментація даних, обробка даних і заповнення пропусків в них, маркетинг баз даних та інше.

Отримані моделі оцінили на адекватність за допомогою індекса Gini, AUC, загальної точності моделі.

Для побужови байєсівської мережі було обрано 10 значущих змінних на основі яких і було побудовано мережу.

РОЗДІЛ 4 СТАРТАП АНАЛІЗ ПРОЕКТУ

4.1 Інформаційна карта проекту

Ідея реалізації стартап-проекту проілюстровано в таблицях (4.1 – 4.20)

Таблиця 4.1 – Інформаційна карта

1. Назва проекту	Комплексна система оцінювання кредитоспроможності клієнтів фінансової установи
2. Автори проекту	Авраменко Олександр
3. Коротка анотація	Дана система, призначена для забезпечення роботи фінансової установи. На основі наявних даних, система прогнозує з якою вірогідністю клієнт, який просить кредит зможе його повернути, ґрунтуючись на інформацію, яку про себе надав клієнт. Для точності прогнозування система використовує декілька різних моделей, для оцінки клієнта : Байєсівська мережа, Логістична регресія, Дерево рішень. Ці методи доповнюють один одного і роблять висновок системи незалежним від людського фактору.
4. Термін реалізації проекту	<i>12 місяців</i>
5. Необхідні ресурси	Обладнання – комп'ютери, принтер, сканер, ноутбуки, планшети. Програмне забезпечення, операційна система, антивірусне обладнання. Електрика, газ, водопостачання, Інтернет. Фінансові ресурси – заробітна плата працівникам на 6 місяців роботи, гроші на оплату комунальних послуг, аренди, реклами тощо. Приміщення з усіма необхідними комунікаціями.
6. Опис проблеми, яку вирішує проект	Дана програма дозволяє розв'язати проблему отримання точного прогнозу повернення грошей клієнтом.
7. Головні цілі та завдання проекту	Основна мета проекту – отримання диплому автором цього проекту. Додаткові завдання – новий досвід, розробка комплексної системи, робота із реальними даними та створення комерційно успішного продукту.
8. Очікувані результати	Система здатна пришвидшити процес відповіді на запит клієнта про кредит. Зменшити кількість співробітників, які потрібні для розгляду кредитної заявки.

4.2. Маркетинговий аналіз стартап – проекту

Таблиця 4.2- Ідея стартапу

<i>Зміст ідеї</i>	<i>Напрямки застосування</i>	<i>Вигоди для користувача</i>
Дана комплексна система дозволяє розв'язати проблему отримання швидкої відповіді на запит клієнта стосовно редиту до фінансово установи. На основі прогнозу відбувається розрахунок відповіді на запит.	1.Робота із власною клієнтською базою.	Система дозволяє замовникам зменшити витрати часу та людських ресурсів на створення прогнозу, Покращити якість обслуговування клієнтів.
	2. Залучення нових клієнтів.	Система дозволяє замовникам зменшити видатки на працівників відділень.

Таблиця 4.3 характеристики стартапу

№ п/ п	Техніко- економічні характери- стики ідеї	(потенційні) товари/концепції конкурентів			W (слабка сторо- на)	N (нейтральна сторона)	S (сильна сторо- на)
		Мій проект	МБКІ (Міжна- родне бю- ро кре- дит-них історій)	УБКІ (Українське бюро кредит- них історій)			
1.	Точність прогнозу- вання	Засто- сування кількох моделей в залежності від зав- дання	Скоринг кредитної історії МБКІ (Creditinf o Predictor Scoring)	Скоринг за допомогою методу граді- єнт-ний бус- тинг			+
2.	Ризики неві- рної оцінки	Існують, через вели- ку кіль- кість фак- торів, може бути зни- жений за рахунок даних за- мовника	Існують через ви- корис- тання да- них різ- них 20 банків.	Існують за рахунок ви- користання цільової змін- ної «Просро- чка більше 100 грн. бі- льше 90 днів протягом ро- ку».		+	
3.	Доступність, зручність	Графічний інтерфейс	Власний інтерфейс	Власний інте- рфейс			

Таблиця 4.4 - Технологічна здійсненність ідеї проекту

<i>n/n</i>	<i>Ідея проекту</i>	<i>Технології її реалізації</i>	<i>Наявність технологій</i>	<i>Доступність технологій</i>
1	Система, яка на основі наявних даних кредитної історії, здійснює оцінку кредитоспроможності	Використання програмного середовища SAS	Наявна	Доступна
2	фізичних осіб з ціллю мінімізації ризиків	Використання мови програмування R	Наявна	Доступна
3	ймовірних фінансових втрат кредитної організації.	Використання мови програмування Python	Наявна	Доступна
Обрана технологія реалізації ідеї проекту: програмне середовище SAS				

Таблиця 4.5 – Характеристика ринку

<i>№ n/n</i>	<i>Показники стану ринку (найменування)</i>	<i>Характеристика</i>
1	Кількість головних гравців, од	75
2	Загальний обсяг продаж, грн/ум. од	1000
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає
6	Середня норма рентабельності в галузі (або по ринку), %	30%

Таблиця 4.6 – Характеристика клієнтів

<i>n/n</i>	<i>Потреба, що формує ринок</i>	<i>Цільова аудиторія (цільові сегменти ринку)</i>	<i>Відмінності у поведінці різних потенційних цільових груп клієнтів</i>	<i>Вимоги споживачів до товару</i>
	Прогноз кредитоспроможності клієнта	Аналітики, аналітичні відділи	Низька ціна, велика кількість статистичних даних.	Простота використання.
	Створення якісного прогнозу та складання штатного розкладу.	Малі та аутсорсингові фінансові установи	Цікавить простота у використанні, низька ціна швидкодія	Низька ціна, репутація
	Створення точного довгострокового прогнозу, швидка відповідь для уникнення додаткових витрат	Великі Банки	Цікавить передусім точність довгострокових прогнозів, швидкодія	Висока якість, бренд, ім'я на ринку, успішний досвід

Таблиця 4.7 – Фактори загроз

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст загрози</i>	<i>Можлива реакція компанії</i>
1	Конкуренція	Цього року очікується вихід на ринок крупної іноземної компанії-конкурента	Пришвидшити вихід програмного продукту
	Збут	Ускладнення збуту через цінову політику конкурентів	Розміщення додаткових рекламних банерів в Інтернеті

Таблиця 4.8 – Фактори можливостей

<i>n/n</i>	<i>Фактор</i>	<i>Зміст можливості</i>	<i>Можлива реакція компанії</i>
	Гнучкі ціни	Зменшення ціни товару задля збільшення попиту	Введення власних гнучких цін
	Диференціація витрат	Зменшення витрат за рахунок їх перерозподілу	Зменшення витрат на додаткові, непрофільні задачі.

Таблиця 4.9 – Аналіз конкуренції

<i>Особливості конкурентного середовища</i>	<i>В чому проявляється дана характеристика</i>	<i>Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)</i>
1. Вказати тип конкуренції - Досконала конкуренція	Багато систем/команд аналітиків	Розробити впізнаваний продукт, якість, що вирізнятиме нас від конкурентів
2. За рівнем конкурентної боротьби: міжнародний	На ринку присутні системи, розроблені за кордоном.	Розширення аудиторії, розширення списку мов, які підтримуються системою
3. За галузевою ознакою - міжгалузєва	Робота із бюро даних в різних галузях.	Розширення списку використовуваних бюро, з яких система може брати необхідні данні.
4. Конкуренція за видами товарів: товарно-родова	Конкуренція між прогнозами інформаційних систем та команд аналітиків.	Збільшення функціоналу системи
5. За характером конкурентних переваг: Нецінова	Різні способи прогнозування дають різну точність	Розробка кращих(точніших) алгоритмів
6. За інтенсивністю: марочна	Впізнаваний бренд надає великих переваг	Велику увагу приділити розвитку бренду

Таблиця 4.10- Аналіз конкуренції

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
		Інші комплексні системи	Гнучкі ціни, розмір капіталовкладень	Змінні витрати постачальників, диференціація витрат	Контроль якості, система інформації
Висновки:	Інтенсивна конкуренція	Є як можливість входу на ринок, так і нові потенційні конкуренти	Постачальники не диктують умови роботи на ринку	Клієнти не диктують умови роботи на ринку	Обмеження для роботи на ринку через товари заміники

Таблиця 4.11 – Фактори конкурентноспроможності

<i>n/n</i>	<i>Фактор конкурентноспроможності</i>	<i>Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)</i>
	Багатофункціональність	Жоден конкурент не є настільки багатофункціональним, не здатен на прогноз за декількома моделями одночасно.
	Якість	Висока якість прогнозу, велика кількість допоміжних статистичних даних
	Обслуговування	Робота з клієнтами – передусім фізичні особи.

Таблиця 4.12 – SWOT - аналіз

<p>Сильні сторони:</p> <p>Висока якість прогнозу</p> <p>Якість</p> <p>Багатофункціональність</p>	<p>Слабкі сторони:</p> <p>Інтерфейс користувача</p> <p>Велика конкуренція</p>
<p>Можливості:</p> <p>Попит</p> <p>Зміна рівня доходів населення</p> <p>Вдосконалення системи</p>	<p>Загрози:</p> <p>Конкуренція</p> <p>Збут</p>

Таблиця 4.13 – Альтернативи ринкового впровадження

<i>№ п/п</i>	<i>Альтернатива (орієнтовний комплекс заходів) ринкової поведінки</i>	<i>Ймовірність отримання ресурсів</i>	<i>Строки реалізації</i>
1	Швидкий вихід на ринок із «сирим» продуктом, можливі проблеми із кількістю інформації на якій ґрунтується рішення.	30%	3 місяці
2	Поступовий вихід з готовим, відлагодженим продуктом. Висока якість та конкурентоспроможна ціна.	70%	6 місяців

Таблиця 4.14 – Групи потенційних споживачів

<i>п/п</i>	<i>Опис профілю цільової групи потенційних клієнтів</i>	<i>Готовність споживачів сприйняти продукт</i>	<i>Орієнтовний попит в межах цільової групи (сегменту)</i>	<i>Інтенсивність конкуренції в сегменті</i>	<i>Простота входу у сегмент</i>
	Окремі аналітики та аналітичні відділи невеликих компаній	Низька готовність	25%	Висока	Середня
	Малі та середні фінансові установи.	Висока	40%	Середня	Висока
	Великі топ банки.	Висока	20%	Низька	Середня
Які цільові групи обрано: 2,3					

Таблиця 4.15 – Стратегія розвитку

<i>n/n</i>	<i>Обрана альтернатива розвитку проекту</i>	<i>Стратегія охоплення ринку</i>	<i>Ключові конкурентоспроможні позиції відповідно до обраної альтернативи</i>	<i>Базова стратегія розвитку*</i>
	2 та 3	Стратегія диференційованого маркетингу	Висока універсальність, точність прогнозу, кількість моделей.	Мінімізація витратної частини.

Таблиця 4.16 – Стратегія конкурентної поведінки

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

Таблиця 4.17 – Стратегія позиціонування

<i>№ n/n</i>	<i>Вимоги до товару цільової аудиторії</i>	<i>Базова стратегія розвитку</i>	<i>Ключові конкурентоспроможні позиції власного стартап-проекту</i>	<i>Вибір асоціацій, які мають сформулювати комплексну позицію власного проекту (три ключових)</i>
1	Якість, точність, простота у використанні	Мінімізація витратної частини	Якість прогнозу, універсальність, велика кількість статистичної інформації	По іміджу Позиціонування на низькій ціні позиціонування за сферою застосування

Таблиця 4.18 – Ключові переваги

<i>№ n/n</i>	<i>Потреба</i>	<i>Вигода, яку пропонує товар</i>	<i>Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)</i>
1	Статистичні параметри клієнта	Розгорнута статистика стосовно з/п, кредитної історії, біографія.	Дана статистика має бути максимально розгорнутою та зрозумілою.
2	Створення якісного та точного прогнозу	Точний прогноз – основа функціонування будь-якої фінансової системи.	Розробка коротко та довгострокових прогнозів, використання різних методів для покращення точності прогнозу

Таблиця 4.19 – Межі цін

<i>n/n</i>	<i>Рівень цін на товари-замінники</i>	<i>Рівень цін на товари-аналоги</i>	<i>Рівень доходів цільової споживачів групи</i>	<i>Верхня та нижня межі встановлення ціни на товар/послугу</i>
	1000	1500	300000	800-900

Таблиця 4.20 – Система збуту

<i>n/n</i>	<i>Специфіка закупівельної поведінки цільових клієнтів</i>	<i>Функції збуту, які має виконувати постачальник товару</i>	<i>Глибина каналу збуту</i>	<i>Оптимальна система збуту</i>
	Канал нульового рівня	Доставка товару	0	+

Висновки до розділу

В даному розділі розглянуто стартап-ідею проекту.

З'ясовано, що є можливість ринкової комерціалізації проекту (наявні попит, динаміка ринку, рентабельність роботи на ринку);

З'ясовано, що є перспективи впровадження з огляду на потенційні групи клієнтів, бар'єри входження, стан конкуренції. Конкурентоспроможність проекту є високою;

Подальша імплементація проекту є можливою за рахунок низької конкурентності в даному сегменті. Також є варіанти розвитку проекту за рахунок додавання нових методів для аналізу і тим самим покращувати прогнозуючу здатність програми.

ВИСНОВКИ

Ця магістерська дисертація присвячена процесу аналізу, побудови та застосування моделей, які мають за мету оцінювання кредитних ризиків у банківській сфері. На першому етапі було опрацьовано теоретичний матеріал по визначенню кредитоспроможності та методи його аналізу.

Методом для розв'язання було обрано логістичну регресійну модель, лінійну логістичну регресію, байєсівську мережу, також було побудовано дерево рішень, ці моделі найбільш повно описують наявні дані, та мають найбільш адекватні результати.

На першому етапі були опрацьовано наявні дані. Були усуненні втрати інформації, які були наявні. Також були обрані найбільш значущі змінні, а також усуненні дані, які не мають інформативного впливу, та не допоможуть зробити модель більш адекватною. Для подальшого опрацювання даних було обрано програмний продукт SAS Enterprise Miner. У програмному продукті Genie було побудовано байєсівську мережу.

На основі статистичних показників було оцінено побудовані моделі на адекватність і обрано найкращу. Обрані найкращі моделі на другому етапі тестували на іншому розподілі даних. А саме - тренувальна вибірка 85% та тестова – 15%. Згідно з результатами порівняння моделей, найкращою було обрано логістичну регресійну модель з коефіцієнтом GINI рівним 0,72, що є досить непоганим результатом для задачі прогнозування.

На останньому етапі було побудовано скорингову карту, яка повинна у подальшому зробити аналіз кредитоспроможності більш швидким.

ПЕРЕЛІК ПОСИЛАНЬ

1. Бідюк П.І. Методичні вказівки до вивчення курсу: “Аналіз часових рядів” навч. посібник Київ: ННК ІПСА НТУУ КПІ, 2004. – 148 с.
 2. Терентьев А. Н. Домрачев В. Н. Костецкий Р. И. SAS BASE: Основы программирования Київ: Эдельвейс. 2014. 304 с.
 3. Згуровский М. З., Зайченко Ю. П. Комплексный анализ риска банкротства корпораций в условиях неопределенности. *Системні дослідження та інформаційні технології*, 2012.. С. 113–128.
 4. Романенко В.Д., Бідюк П.І., Подладчиков В.М. та ін. Розробка теоретичних основ аналізу, прогнозування та оптимального прийняття рішень в умовах інформаційної невизначеності / НТУУ "КПІ", звіт по НДР № 0198U000819, 1996. – 186 с.
 5. A Bayesian Network Framework for Project Cost, Benefit and Risk Analysis with an Agricultural Development Case Study, 2016 http://constantinou.info/downloads/papers/Project_ROI_Preprint.pdf.
- (дата звернення : 26.11.2019).
6. Сиддики Н. Скоринговые карты для оценки кредитных рисков. Разработка и внедрение интеллектуальных методов кредитного скоринга. М.: Манн, Иванов и Фербер, 2014. 268 с.

7. Бідюк, П. І. Прикладна статистика / П. І. Бідюк, О. М. Терентьєв, Т. І. Просянкіна-Жарова. – Вінниця : ПП "ТД"Едельвейс і К", 2013. – 304 с.
8. Згуровський М.З., Бідюк П.І., Терентьєв О.М., Просянкіна-Жарова Т. І. Байєсівські мережі в системах підтримки прийняття рішень— Київ : ТОВ «Видавниче Підприємство «Едельвейс», 2015. — 300 с.
9. Friedman N. The Bayesian structural EM algorithm / Friedman N. // Fourteenth conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, Wisconsin, (USA), 24 – 26 July, 1998. – SF. : Morgan Kaufmann, 1998. – P. 129-138.
10. Олексюк О.С. Системи підтримки прийняття фінансових рішень. – Київ: Наукова думка, 1998. – 508 с.
11. Бидюк П.И., Баклан И.В. Системный подход к построению регрессионной модели по временным рядам. *Системні дослідження та інформаційні технології*, 2002, № 3, с. 114-131.
12. Heckerman D. A tutorial on learning with Bayesian networks // *Technical report MSR-TR-95-06*, Microsoft Research, Redmond, Washington, 1995. – <http://www.research.microsoft.com> (дата звернення : 26.11.2019).
13. Chen B., Varshney P.K. A Bayesian Sampling Approach to Decision Fusion Using Hierarchical Models. *IEEE Trans. on Signal processing*, 2002, v.50, No.8, pp. 1809–1818.
14. Ковальов М.С. Методика построения банковских скоринговых моделей для оценки кредитоспособности физических лиц. Интернет-журнал

«Науковедение» Выпуск 2, 2014. <http://www.bsu.by/Cache/pdf/49623.pdf> (дата звернення 25.11.2019)

15. Ишина И. В. Сазонова М. Н. Ишина И. В. Скоринг - модель оценки кредитного риска. Аудит и финансовый анализ. 2007. № 4. <http://www.auditfin.com/fin/2007/4/Ishina /Ishina%20.pdf>. (дата звернення 26.11.2019)

16. Кузнецова Н. В. Бідюк П. І. Порівняльний аналіз характеристик моделей оцінювання ризиків кредитування. *Наукові вісті НТУУ «КПІ»*. 2010. №1. С. 42-53.