

Державний торговельно-економічний університет

Кафедра цифрової економіки та системного аналізу

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Аналіз кредитоспроможності клієнтів банку»

Студентки 4 курсу, 14 групи,
першого (бакалаврського)
рівня вищої освіти
спеціальності
124 «Системний аналіз»
освітньої програми
«Інформаційні технології та
бізнес-аналітика (Data Science)»

підпис студента

Курочкіна Дмитра
Олександровича

Науковий керівник
кандидат економічних наук,
старший викладач

підпис керівника

Лазоренко Віталій
Валерійович

Гарант освітньої програми
кандидат економічних наук,
доцент

підпис гаранта

Кулаженко Володимир
Валерійович

Київ 2023

Державний торговельно-економічний університет

Факультет інформаційних технологій

Кафедра цифрової економки та системного аналізу

Освітній ступінь бакалавр

Спеціальність 124 «Системний аналіз»

Освітня програма «Інформаційні технології та бізнес-аналітика (Data Science)»

Затверджую

Зав. кафедри _____ Роскладка А.А.
«15» грудня 2022 р.

Завдання на випускню кваліфікаційну роботу студентці

Куручкін Дмитро Олександрович
(*прізвище, ім'я, по батькові*)

1. Тема випускної кваліфікаційної роботи

«Аналіз кредитоспроможності клієнтів банку»

Затверджена наказом ДТЕУ від «09» грудня 2022 р. № 3333

2. Строк здачі студентом закінченої роботи «09» червня 2023 року

3. Цільова установка та вихідні дані до роботи

Мета роботи полягає у дослідженні оцінювання кредитоспроможності клієнтів банку та інтерактивній візуалізації аналізу

Об'єкт дослідження інформація про клієнтів банку

Предметом дослідження є сукупність аналітичних даних банку для оцінки кредитоспроможності клієнтів.

4. Зміст випускної кваліфікаційної роботи (перелік питань за кожним розділом)

ВСТУП

РОЗДІЛ 1. ОСНОВНІ МОЖЛИВОСТІ ТА ХАРАКТЕРИСТИКА АНАЛІЗУ ДАНИХ

1.1. Основні положення аналізу даних та основні задачі консолідації даних

1.2. Інструменти та сфера застосування Data Mining для аналізу даних

1.3. Основи програмування мовою DAX та технології Data Science

Висновки до розділу 1

МЕТОДОЛОГІЯ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ НА ОСНОВІ ДАНИХ ПРО КЛІЄНТА

2.1. Поняття кредитоспроможності клієнтів банку

2.2. Аналіз існуючих методів для оцінки кредитоспроможності клієнтів банку

Висновки до розділу 2

РОЗДІЛ 3. ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ

3.1. Постановка задачі, опис вхідної та вихідної інформації

3.2. Опис початкових етапів побудови аналітичного звіту

Висновки до розділу 3

ВИСНОВКИ ТА ПРОПОЗИЦІЇ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ



5. Календарний план виконання роботи

№ пор.	Назва етапів випускної кваліфікаційної роботи	Строк виконання етапів роботи	
		за планом	фактично
1	2	3	4
1	<i>Вибір теми випускної кваліфікаційної роботи</i>	01.12.2022	01.12.2022
2	<i>Розробка та затвердження завдання на випускну кваліфікаційну роботу</i>	15.12.2022	15.12.2022
3	<i>Вступ</i>	01.02.2023	
4	<i>Розділ 1. Аналіз діяльності банків у сучасних умовах</i>	13.03.2023	
5	<i>Розділ 2. Методологія оцінки кредитоспроможності клієнтів банку на основі даних про клієнта</i>	24.04.2023	
6	<i>Розділ 3. Побудова аналітичного звіту оцінки кредитоспроможності клієнтів банку</i>	01.05.2023	
7	<i>Висновки та пропозиції</i>	08.05.2023	
8	<i>Здача випускної кваліфікаційної роботи на кафедрі науковому керівнику</i>	22.05.2023	
9	<i>Попередній захист випускної кваліфікаційної роботи</i>	30.05.2023	
10	<i>Виправлення зауважень, зовнішнє рецензування випускної кваліфікаційної роботи</i>	06.06.2023	
11	<i>Представлення готової зшитої випускної кваліфікаційної роботи на кафедрі</i>	09.06.2023	
12	<i>Публічний захист випускної кваліфікаційної роботи</i>	За розкладом роботи ЕК	

6. Дата видачі завдання «15» грудня 2022 р.

7. Науковий керівник випускної кваліфікаційної роботи

_____ (підпис)

Лазоренко В. В.
(прізвище, ініціали)

8. Гарант освітньої програми

_____ (підпис)

Кулаженко В. В.
(прізвище, ініціали)

9. Завдання прийняв до виконання студент

_____ (підпис)

Курочкін Д. О.
(прізвище, ініціали)

Анотація

В даній роботі було розглянуто основні задачі та етапи проведення аналізу даних, основні задачі консолідації даних, методи та алгоритми переробки даних, методи публікації поширення та захисту аналітичних звітів, характеристика аналітичної платформи Microsoft Power BI, досліджено аспекти оцінки кредитоспроможності клієнтів банку на основі даних про клієнтів, розглянуто існуючі методи та моделі для оцінки кредитоспроможності.

У практичній частині роботи побудовано інтерактивний звіт для оцінки кредитоспроможності клієнтів банку на основі даних про клієнта з врахуванням доходу, наявності майна, та попередньої кредитної історії та побудовано профіль клієнта для визначення оцінки кредитоспроможності.

Ключові слова: кредитоспроможність, аналіз даних, візуалізація даних, інтерактивний звіт.

Annotation

In this work, the main tasks and stages of data analysis, the main tasks of data consolidation, methods and algorithms of data processing, methods of publication, distribution and protection of analytical reports, characteristics of the Microsoft Power BI analytical platform, aspects of assessing the creditworthiness of bank clients based on client data were considered. , existing methods and models for assessing creditworthiness are considered.

In the practical part of the work, an interactive report was built to assess the creditworthiness of bank clients based on data about the client, taking into account income, property availability, and previous credit history, and a client profile was built to determine the creditworthiness assessment.

Keywords: creditworthiness, data analysis, data visualization, interactive report.

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ОСНОВНІ МОЖЛИВОСТІ ТА ХАРАКТЕРИСТИКА АНАЛІЗУ ДАНИХ	5
1.1 Основні положення аналізу даних та основні задачі консолідації даних	5
1.2. Інструменти та сфера застосування Data Mining для аналізу даних	9
1.3. Основи програмування мовою DAX та технології Data Science	12
Висновки до розділу 1	16
МЕТОДОЛОГІЯ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ НА ОСНОВІ ДАНИХ ПРО КЛІЄНТА	17
2.1. Поняття кредитоспроможності клієнтів банку	17
2.2. Аналіз існуючих методів для оцінки кредитоспроможності клієнтів банку	19
Висновки до розділу 2	21
РОЗДІЛ 3. ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ	23
3.1. Постановка задачі, опис вхідної та вихідної інформації	23
3.2. Опис початкових етапів побудови аналітичного звіту	25
Висновки до розділу 3	34
ВИСНОВКИ ТА ПРОПОЗИЦІЇ	35
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	36

ВСТУП

В теперішній час люди досягли дуже високого розвитку, але і прогрес на місці не стоїть. Розробляються нові технології, що необхідні для оптимізації та полегшення умов для комфортного життя людей, оновлюються та покращуються програми, що вже існують, стаючи більш довершеними.

Погіршення економічної ситуації в нашій країні останні декілька років збільшила попит населення до різноманітних кредитних програм в банках або спеціалізованих організаціях.

В останній час люди почали широко застосовувати інтернет в повсякденному житті. Цей факт сприяє створенню великої кількості веб застосунків та сайтів, спрямованих на використання в банківській сфері.

На сьогодні отримання кредиту є дуже непростю і витратною в часі процедурою, основним елементом якої є перевірка кредитоспроможності позичальника. Даний процес займає дуже багато часу і сил, як в клієнта банку чи іншої спеціалізованої організації, так і в її працівників. Це стосується не тільки черги в даному закладі, а і довгої розмови, після якої і визначається основний елемент описаний вище. У найближчому майбутньому людство планує перевести у режим онлайн користування велику кількість операцій.

Зараз дуже швидко набуває популярності онлайн банкінг.

Це пов'язане з тим, що даний сервіс зберігає час працівників банку та клієнтів; клієнт має доступ у вільний для нього час заповнити форму, яка зайняла б дуже багато часу та нервів в відділенні банку. Коли клієнт перевірить свою кредитоспроможність він матиме змогу оцінити свої можливості по виплаті даного кредиту та, можливо переглянути інші варіанти. Якщо об'єкта все влаштовує, то він може явитися в відділення банку, на умовах якого він аналізував свою кредитоспроможність, і зустрівшись обличчям до обличчя з банківським працівником все що буде потрібно клієнтові – назвати своє прізвище, ім'я та номер паспорта або індивідуальний податковий номер і консультант банку одразу зможе побачити абсолютно всі дані клієнта, які було введено в комфортних для об'єкта умовах.

Мета роботи полягає у дослідженні оцінювання кредитоспроможності клієнтів банку та інтерактивній візуалізації аналізу.

Об'єкт дослідження інформація про клієнтів банку.
Предметом дослідження є сукупність аналітичних даних банку для оцінки кредитоспроможності клієнтів.



РОЗДІЛ 1. ОСНОВНІ МОЖЛИВОСТІ ТА ХАРАКТЕРИСТИКА АНАЛІЗУ ДАНИХ

1.1 Основні положення аналізу даних та основні задачі консолідації даних

Методи аналізу даних широко використовують при дослідженні різноманітних систем і процесів – природних, технічних, екологічних, економічних, соціальних тощо. З огляду на це формування відповідних знань та навичок є необхідною складовою підготовки фахівців у галузі системних наук і кібернетики, інформатики та багатьох інших галузей знань.

За допомогою способів аналізу даних, проводять пошук раніше невідомих закономірностей у великих масивах інформації. Ці закономірності дають можливість прийняття ефективних управлінських рішень і оптимізації бізнес-процесів. В таблиці 1 зазначено задачі аналізу даних та описано їх.

Таблиця 1.1

Задачі аналізу даних

Назва задачі	Опис
1. Класифікація	Визначення класу об'єкта по його характеристикам. Для вирішення завдання класифікації можуть використовуватися такі методи як: найближчого сусіда (Nearest Neighbor); k-найближчого сусіда (k-Nearest Neighbor); байесовські мережі (Bayesian Networks); індукція дерев рішень; нейронні мережі (neural networks).
2. Кластеризація	Логічне продовження ідеї класифікації. Ця задача більш складна, особливість кластеризації полягає в тому, що класи об'єктів спочатку не визначені. Результатом кластеризації є розбиття об'єктів на групи.
3. Прогнозування	Вважається однією з найбільш складних задач Data Mining, оскільки потребує ретельного дослідження вихідного набору даних і методів. Метою прогнозування є передбачення майбутніх подій.
4. Асоціація	Метою пошуку <i>асоціативних правил</i> є знаходження частих залежностей (або асоціацій) між об'єктами або подіями. Знайдені залежності представляються у вигляді правил і можуть бути використані як для кращого розуміння аналізованих даних, так і для передбачення появи подій.
5. Візуалізація	В результаті <i>візуалізації</i> створюється графічний образ аналізованих даних. Для вирішення даної задачі використовуються графічні

	методи, що показують наявність закономірностей в даних. Як приклад методів візуалізації – представлення даних в 2-D і 3-D вимірах.
--	--

1. Джерело: [1]

Виділяють наступні етапи аналізу даних (рис. 1):

1. Аналіз предметної області. Вирішення задачі повинне починатися з вивчення предметної області. Предметна область складається з об'єктів, що розрізняються за властивостями і знаходяться у визначених відносинах між собою або взаємодіючих яким-небудь чином. В процесі аналізу предметної області повинна бути побудована її модель.
2. Постановка задачі. Даний етап включає в себе формулювання та формалізацію задачі, опис статичної та динамічної поведінки дослідженого об'єкта. В результаті етапу визначається яку саме задачу необхідно вирішити.
3. Підготовка даних. При підготовці даних обов'язково потрібно описати фактори, що впливають на процес аналізу. Також варто зауважити, що недостатня кількість даних може відобразитися на коректності моделі. Точність моделі збільшується зі збільшенням кількості актуальних даних.
4. Побудова моделі. Для побудови моделі використовуються методи та алгоритми Data Mining. Для того, щоб вибрати метод або комбінацію методів, необхідно проаналізувати наскільки побудована модель буде точно описувати досліджуваний об'єкт.
5. Перевірка і оцінка моделі. При перевірці вивчають наскільки модель відповідає реальності, тобто модель перевіряють на адекватність та достовірність. Оцінка моделей передбачає перевірку її правильності. Це відбувається шляхом її тестування. При тестуванні визначаються характеристики моделі, а також її працездатності. Кінцевий результат оцінює спеціаліст предметної області. У випадку, коли результат незадовільний, варто повернутися на попередні етапи [1].
6. Вибір моделі. Якщо під час моделювання було побудовано декілька моделей, то вибір відбувається за оцінками таких характеристик як точність і ефективність роботи алгоритму [1].

7. Застосування моделі. На цьому етапі обрана модель використовується відповідно до нових даних, з метою вирішення задач, що були поставлені на початку.

8. Корекція і оновлення даних. Через деякий час з моменту використання моделі, необхідно проаналізувати отримані результати та визначити наскільки ефективно працює модель і чи були проблеми при її використанні. З часом побудована модель втрачає свою точність і для виконання подальших функцій, модель варто покращувати. У випадку, коли з'являються нові дані, необхідне повторне навчання моделі. Даний процес називається оновленням моделі [1].

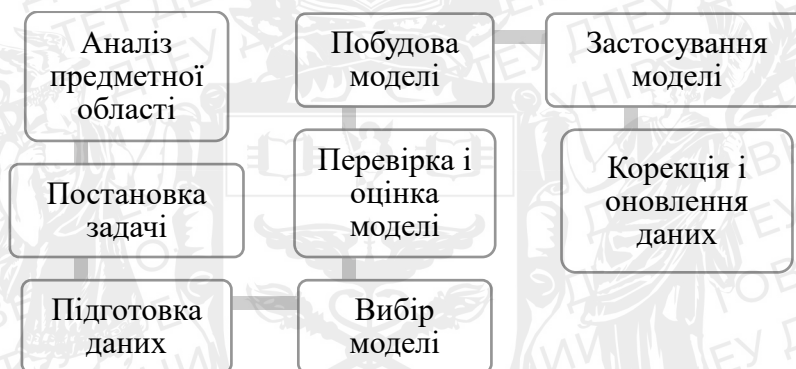


Рис. 1.1 Етапи аналізу даних

Консолідація – це комплекс методів та процедур, що направлені на вилучення даних із різних джерел та перетворення їх в єдиний формат, в якому вони будуть завантажені в базу даних або аналітичну систему. Основу консолідації складає процес збору і організації зберігання даних в тому вигляді, який є оптимальний для їх обробки [2].

Задачі консолідації даних:

1. Вибір джерел даних – джерела, що містять дані, які стосуються вирішення поставленої задачі
2. Розробка стратегії консолідації даних – враховується характер розміщення джерел даних: розміщення на тому ж самому персональному комп'ютері (ПК), що і аналітична платформа. Даний характер називається локальним. Другий тип характеру – віддалений – джерела дані доступні через локальну або глобальну мережу [2].
3. Оцінка якості даних – у випадку, коли дані мають фактори, що заважають якісному аналізу, встановлювати зв'язки між елементами даних і виконувати дії,

які необхідні для отримання аналітичного рішення, необхідно провести очистку даних. До таких факторів відносяться: помилки вводу, пропущені дані, шум, відсутність повноти даних, протиріччя, некоректні формати та представлення даних. Для знищення цих факторів використовують очистку та збагачення. Очистка даних – комплекс методів, які спрямовані на усунення причин, що заважають коректній обробці даних. Збагачення – процес доповнення даних інформацією. Даний процес підвищує інформаційну насиченість і як наслідок значимість для вирішення аналітичної задачі.

2. Перенос даних в сховище даних [2].

З метою підвищення якості даних використовується комплекс методів і алгоритмів, який має назву «очищення даних». Перед застосуванням методів очищення потрібно виконати оцінку якості даних. Місця проведення оцінки якості даних:

Безпосередньо в джерелах даних	В процесі Extract transform load (ETL)	В аналітичній системі
пошук орфографічних помилок, пропущених, аномальних значень, протиріччя і дублікатів на рівні записів і таблиць.	виявлення проблеми в даних і за результати задіяти методи очищення, щоб забезпечити завантаження в сховище вже очищенні дані.	оцінка може проводитися аналітиком візуально за допомогою таблиць, графіків і діаграм або на основі статистичних оцінок і характеристик.

Рис. 1.2 Місця проведення оцінки якості даних

Профайлінг – один з найбільш поширених методів перевірки якості даних в Data Mining. Профайлінг виконується на основі аналізу метаданих, що описують структуру даних. В процесі профайлінгу аналізується наступна інформація:

Таблиця 1.2

Аналіз інформації в процесі профайлінгу

Тип атрибута (поля)	причиною спотворення типу може бути неправильне використання символу – роздільника цілої і дробової частин числа, коли замість крапки використовується кома, і навпаки; також невідповідність
---------------------	---

	роздільників груп розрядів в числах – пробілу, крапки, коми – або їх відсутність
Довжина значень	визначається максимальна допустима кількість символів в значенні атрибутів
Дискретні значення	перевірка частоти їх появи та унікальність
Діапазон допустимих значень	задається мінімальне та максимальне значення, яке може приймати атрибут
Аналіз рядкових шаблонів	обробка рядка у відповідності з деяким завчасно визначеним правилом

Джерело: [12]

Профайлінг розрахований на виявлення «штатних», технічних проблем, які відносяться до найбільш поширених.

Для оцінки якості даних (якщо їх об'єм невеликий) використовуються візуальні методи. Для цього можна використати вбудовані засоби візуалізації аналітичної платформи, а також сторонні програмні засоби, в яких можна переглядати дані в табличній та графічній формах.

Оцінка якості за допомогою таблиць – табличне представлення дозволяє робити висновки про наявність в даних структурних порушень, пропусків, аномальних та фіктивних значень. Оцінка якості за допомогою графіків – оцінка наявності шумів в даних, а також пропущених та аномальних значень [12].

1.2. Інструменти та сфера застосування Data Mining для аналізу даних

Apache Spark - це платформа паралельної обробки з відкритим кодом для запуску великомасштабних програм аналізу даних. Він може обробляти як пакетну аналітику, так і роботи в режимі реального часу та обробки даних. Дана платформа підтримує такі мови програмування, як R, Python, Java, Scala, а також мову програмування структурованих запитів SQL. Здатність обчислювати дані з багатьох типів сховищ даних вказує на те, що Apache Spark можна застосовувати в багатьох сферах для вирішення різних проблем. Наприклад, фінансові компанії використовують платформу для отримання фінансових даних та створення моделей для керівництва інвестиційної діяльності. Для

компанії споживчих товарів Apache Spark має цінність як програма для узагальнення аналізу споживачів та здійснення прогнозу. Компанії, що займаються цифровою рекламою застосовують платформу для аналізу веб-активності кампаній. Серед компаній, що використовують Apache Spark є Yahoo, Cisco Systems, Alibaba, Intel, IBM. Аби зрозуміти можливості платформи, варто вказати, що в 2014 році було встановлено рекорд при сортуванні 100 терабайт даних [3].

R – провідна безкоштовна мова програмування і програмне середовище для статистичних обчислень, аналізу даних і роботи з графіками. R є досить поширеною мовою програмування за рахунок широкого спектру статистичних (лінійне та нелінійне моделювання, статистичні тести, аналіз часових рядів, класифікація, кластеризація) та графічних методів. R використовують такі компанії як, Google – для ефективності реклами і прогнозування, Facebook – для аналізу поведінки, пов’язаної із оновленням статусу та зображенням профілю, Uber та Airbnb – для статистичного аналізу [6].

Python – мова кодування загального призначення, яка, на відміну, від HTML, CSS, JavaScript, може використовуватися для інших типів програмування та розробки програмного забезпечення. Дана мова широко розповсюджена, оскільки є адаптованою мовою програмування, що дозволяє розробляти та підтримувати проекти різної складності. Python використовує NASA для математичних розрахунків параметрів польоту, а також Python однією з ключових мов Google. Відомо, що Python також використовує Youtube в цілях адміністрування відео, перегляду відео, керування шаблонами для веб-сайту. Python використовується для загальної веб-розробки, розробки веб-додатків; для наукових досліджень та обчислень; для машинного навчання; для фінансової сфери.

Deductor – серед можливостей платформи виділяють інтеграцію даних, візуалізацію, очистку даних, Data Mining, сценарний підхід. В склад комплексу входить Deductor studio – автоматизація рутинних операцій по обробці даних, формалізація правил прийняття рішень, тобто це робоче місце для аналітика; Deductor Viewer – отримання звітів без необхідності занурення в обробку;

Deductor Warehouse – сховище даних, яке дозволяє користувачу, не маючи спеціальних знань, оперувати звичними бізнес-термінами; Deductor Analytic Server – служба, яка забезпечує можливість «прогону» даних через сценарії аналізу, обробки великих масивів інформації; Deductor Integration Server – механізм обміну даними зі сторонніми додатками, реалізації сервіс-орієнтованої архітектури і вбудовування в корпоративні системи, побудовані на базі Enterprise Service Bus.

SAS - програмний пакет, керований командами, який використовується для статистичного аналізу та візуалізації даних. SAS може зчитувати дані із електронних таблиць, баз даних та виводити статистичного аналізу в таблиці, графіки та в документи форматів HTML, PDF, RTF. SAS широко використовується у фармацевтичній промисловості, BI та прогнозуванні погоди. Серед відомих компаній, що використовують SAS це: Google, Facebook, Netflix, Twitter [6].

Технологію Data Mining використовують скрізь, де є дані. Серед найпоширеніших сфер застосування виділяють:

- фінансову – фінансові дані користувачів можуть використовуватися для таких цілей, як прогнозування виплат за кредитом, визначення кредитних рейтингів. В свою чергу, методи класифікації дають змогу полегшити відокремлення факторів, що впливають на банківську рішення клієнтів. Сегментація клієнтів дає змогу розділити клієнтів на категорії і ,таким чином, банки створюють маркетингову політику, яка є більш цілеспрямованою та результативною, оскільки пропонує різні послуги окремим категоріям клієнтів. Аналіз даних може допомогти виявити фінансові злочини;

- роздрібної торгівлі - аналіз даних допомагає мати справу з даними, пов'язаними з різними типами клієнтів, продуктами, регіонами та часовими поясами. Прогнозування з використанням часових рядів дозволяє побудувати модель прогнозу продаж на основі даних за попередній період. За допомогою алгоритму «Пошук асоціативних правил» можна здійснити аналіз корзини клієнта. Кластерний аналіз дозволяє провести аналіз поведінки покупця через розбиття клієнтів на кластери. Даними зазвичай виступають результати

опитування та інформація з дисконтної системи. Технологію Data Mining також використовують для аналізу продажу товарів і формування асортименту, аналізу рекламних кампаній;

- телекомунікаційну – дана галузь швидко розвивається, оскільки відбувається стрімкий розвиток комп'ютерних та телекомунікаційних технологій. В даній галузі Data Mining допомагає аналізувати записи викликів, і таким чином виявити категорії клієнтів зі схожими шаблонами використання їх послуг та створення спеціальних пропозицій. OLAP та засоби візуалізації дозволяють порівняти таку інформацію, як поведінка групи користувачів, трафік даних, перевантаження системи. Аналіз даних сегментує бази абонентів та дозволяє виявити шахрайські дії [12];

- медичну – інструментальний аналіз даних дозволяє розрізнити патологічні дані від нормальних, обрати засоби медикаментозного призначення. Регресійні моделі, дерева рішень, нейронні мережі, карта Кохонена дозволяють виявити ознаки хвороби і змодельовати залежності між симптомами та захворюваннями. Логістична регресія використовується для підбору оптимальних порогів діагностичних показників і, таким чином, для оцінки діагностичних тестів. Асоціативні правила застосовуються для виявлення зв'язку між прийомом ліків та побічних ефектів [12];

- кримінологію – аналіз злочинності включає вивчення та розкриття злочинів та їх зв'язок із злочинцями. Визначення характеристик злочинів є першим кроком для подальшого аналізу. Після обробки даних, процес ідентифікації та вивчення злочинів відбувається шляхом виявлення закономірностей у сховищах даних. Кластеризація може застосовуватися для того, щоб виконати процес збігу злочинів [12];

освіти – навчальні заклади можуть використовувати аналіз даних для прогнозування, які студенти вступлять до певної навчальної програми. Дані можуть бути використані для класифікації та прогнозування поведінки, успішності та відсіву учнів, а також результатів діяльності вчителів [12].

1.3. Основи програмування мовою DAX та технології Data Science

Щорічно кількість інформації у світі зростає на 30%, що призводить до ускладнень процесу управління даними. Якщо враховувати, що підприємства постійно збирають дані про обсяги продажів, кількості споживачів, конкурентів, діляться новинами компанії та здобутками в технологічному та корпоративному плані, то для ефективного управління корпоративними даними потребується проведення бізнес-аналізу.

Business Intelligence (бізнес-аналіз) включає в себе аналіз даних з використанням дієвих методів Data Science, зокрема те, як обробляти дані, збирати статистичні дані та використовувати їх для прийняття обґрунтованих рішень та прогнозів.

Перед тим, як Data Science починає використовувати метод прогнозування в аналітиці даних компанії, бізнес-аналіз фокусується на поведінці різних показників в минулому, аналізує їх, щоб представити шляхи рішення проблем та вдосконалення бізнесу. Бізнес-аналіз надає відповіді на такі питання[9]: Яка динаміка продажів продукції порівняно з іншим звітним періодом? Чи впливає корпоративна культура на становлення бізнесу? В якому географічному регіоні найбільші продажі? Який сегмент товару лідирує на ринку продажів? Як продажі впливають на рух грошових коштів? Які чинники могли вплинути на зміну минулорічних результатів?

Business Intelligence є частиною Data Science, а отже, він може застосовуватись для вимірювання ефективної діяльності бізнесу, використовуючи конкретні методи обробки даних (консолідація, обчислення даних, вибір моделі даних тощо).

Консолідація даних – комплекс методів і процедур, спрямованих на вилучення даних з різних джерел, забезпечення необхідного рівня їх інформативності та якості, перетворення їх в єдиний формат, в якому вони можуть бути завантажені в сховище даних або аналітичну систему[2].

У процесі консолідації даних вирішуються наступні задачі[2]:

- 1) Вибір джерел даних.
- 2) Розробка стратегії консолідації даних.
- 3) Оцінка якості даних.

- 4) Збагачення.
- 5) Очистка даних.
- 6) Перенесення в сховище даних.

Сховище даних - предметно-орієнтований, інтегрований, незмінний, що підтримує хронологію набір даних, організований для цілей підтримки прийняття рішень.

Сховище даних є підсистемою підтримки прийняття рішень, що необхідне для збереження даних, а отже, є однією з головних задач консолідації даних. Завдяки структурі сховища даних зменшується час на пошуки необхідної інформації та отримання додаткових запитів.

Структура сховища даних виокремлює основні її архітектури[1]:

- 1) Реляційні СД використовують класичну реляційну модель, де дані зберігаються в таблицях.
- 2) Багатовимірні СД дозволяють представляти багатовимірні дані.
- 3) Гібридні СД поєднують в середині себе властивості як реляційної, так і багатовимірної моделі даних.
- 4) Віртуальні СД не є сталим елементом сховища даних, тому що дані не консолідуються, а збираються в режимі реального часу.

Після консолідації даних, відповідні дані необхідно кількісно та якісно обчислити, та простежити їх зміни для одержання результатів за розрахунками, з подальшим формулюванням висновків. Тут можна використати мову програмування функцій та формул DAX, що широко застосовується в додатках від компанії Microsoft: Power BI, надбудови Excel (Power Pivot) та таблиці SSAS (SQL Server)[15].

Мова DAX дає змогу створювати запити, брати з вихідних моделей даних необхідні відредаговані дані та проводити з ними обчислення. В DAX можна обчислювати текстові, математичні, логічні, статистичні, інформаційні, табличні функції, також функції фільтрів, логіки операцій з датами, функції дати і часу.

Усі формули в мові DAX мають свій синтаксис, він містить різні елементи, які входять до складу формули, тобто, як пишеться сама формула.

Синтаксис мови DAX має наступні елементи[17]:

- 1) Оператор знаку рівності (=) вказує на початок формули, а після обчислення цієї формули поверне результат або значення. Усі формули, які обчислюють значення, починаються зі знаку рівності.
- 2) Стовпець, на який посилається посилання, містить значення, які потрібно обчислити. Посилання на стовпець у формулі завжди оточений квадратними дужками. На відміну від формул Excel, які посилаються на клітинку, формула DAX завжди посилається на стовпець.
- 3) Оператор мови DAX (логічний, арифметичний, оператор порівняння, конкатенації тексту).
- 4) Стовпець, на який посилається посилання, містить значення, які потрібно відняти/порівняти/додати від значень у стовпці.

Використання мови DAX для проведення бізнес-аналізу значно полегшує процес створення нових даних для подальшої роботи з ними.

Окрім обробки суто числової інформації, Data Science, а особливо бізнес-аналітика, полягає у візуалізації висновків та створенні легкозасвоєваних зображень[13]. Врешті-решт, усі групи людей, які будуть використовувати ці дані для спостережень, як наприклад, клієнти компаній, або для використання їх у професійних рішеннях (менеджери з розвитку бізнесу, маркетологи, верхні ланки управління), повинні мати змогу побачити і зрозуміти суть даних, щоб доводити до відома співробітників про процес прийняття відповідних рішень. Бізнес-аналітики розробляють аналітичні звіти, що супроводжуються різними способами візуального представлення даних. Прикладами візуалізацій можуть бути діаграми, графіки, таблиці, картки, карти, матриці та інші порівняльні візуалізатори.

Завершальними етапами аналізу даних є публікація, поширення та захист аналітичних звітів, які утворились в ході розроблення візуалізації даних.

Бізнес-аналіз також використовує Data Science для оптимізації руху грошових коштів на продукцію та управління запасами на підприємстві[9]:

Оптимізація руху грошових коштів та наука про дані – застосування науки про дані для інформування методів оптимізації грошей на підприємстві, шляхом збору минулих даних, зіставляючи їх з нинішніми.

Управління запасами та наука про дані. Data Science та бізнес-аналітика важливі для впровадження ефективних методів управління запасами, що відповідають вимогам при мінімальних витратах, використовуючи поглиблений аналіз минулих операцій з продажу, визначаючи сезонні закономірності обсягів продажів на періоди року.

Висновки до розділу 1

Вивчаючи науку про дані, можна навчитись трансформувати дані з різних джерел у представленні технологічного ландшафту з метою проведення безпосередніх досліджень, які допоможуть людям та організаціям розуміти проблеми їх діяльності, передбачити, сформувані майбутні цілі та досягнення.

За допомогою методів аналізу на основі Data Science можна перетворювати дані на знання, які потім застосовуються компаніями для прийняття економічних рішень, прогнозування показників продажів, дослідження минулих досягнень та сьогоденних позицій на ринку, а також для створення нових технологій виробництва чи представлення послуг та продуктів конкретних клієнтам, та споживачам.

Бізнес аналіз нерозривно пов'язаний з наукою про дані, також можна сказати, що він охоплює аналіз бізнес-процесів компанії, починаючи від обробки даних – до пояснення нововиявлених закономірностей з метою прогнозування подальшої їх поведінки.

МЕТОДОЛОГІЯ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ НА ОСНОВІ ДАНИХ ПРО КЛІЄНТА

2.1. Поняття кредитоспроможності клієнтів банку

Зважаючи на поточний стан економіки України, стан фінансів багатьох вітчизняних компаній та значну кількість проблемних кредитів, українські банки постійно вдосконалюють свою стратегію та тактику у сфері кредитування. В контексті сучасних тенденцій у банківському секторі, аналіз та управління кредитним ризиком з метою його зниження є актуальним завданням. Оцінка кредитоспроможності позичальників займає провідне місце серед інструментів для мінімізації ризику в кредитній діяльності банку. Реальні висновки та пропозиції, що випливають з аналізу кредитоспроможності, допомагають уникнути непотрібних ризиків при здійсненні кредитних операцій. Отже, кредитоспроможність позичальника завжди була й залишається одним з основних критеріїв для визначення доцільності встановлення кредитних відносин з клієнтом.

Кредитоспроможність суб'єкта відображає його здатність отримати кредит і своєчасно його повернути згідно з угодою. Кредитоспроможність позичальника характеризується кількома ключовими аспектами:

Репутація позичальника: Цей аспект визначається швидкістю і надійністю погашення попередніх кредитів. Якщо позичальник вчасно виконує свої фінансові зобов'язання, це підвищує його репутацію та довіру банків.

Стійкий фінансовий стан. Це вказує на фінансову стабільність позичальника, його здатність забезпечити сталі доходи та управляти фінансовими ризиками. Банк оцінює фінансові показники, такі як доходи, витрати, активи та зобов'язання, для визначення стабільності фінансового стану позичальника.

Можливість мобілізації грошових коштів: Це вказує на здатність позичальника отримати додаткові грошові ресурси з різних джерел у разі необхідності. Це може включати наявність активів, доступ до ліквідних коштів або здатність залучити фінансування від інших джерел.

У той же час, банк оцінює потенційний прибуток, який він отримає від кредитування конкретних витрат позичальника. Банк оцінює розмір прибутку,

який позичальник може здобути, з урахуванням його спроможності сплатити відсотки за умови нормальної фінансової діяльності. Крім того, банк визначає ступінь ризику.

При роботі і дослідженні кредитування, банк:

- оцінює причину отримання кредитних ресурсів;
- кількість кредитних коштів, яку оцінюють коли приймають кінцеве рішення по роботі з позичальником, визначають спрямовання цих коштів та визначають ліквідність балансу;
- погашення кредиту, яке визначається аналізуючи попередню кредитну історію та наданих гарантій;
- забезпечення кредиту.

Під час аналізу кредитоспроможності клієнта, розглядаються різні показники, причому двома найважливішими є норма прибутку на вкладений капітал і ліквідність. Аналіз коефіцієнтів кредитоспроможності виконується на декількох етапах.

Перший етап полягає у визначенні коефіцієнтів кредитоспроможності та класу позичальника. Система коефіцієнтів кредитоспроможності складається з п'яти таких коефіцієнтів.

Другий етап аналізу кредитоспроможності полягає у визначенні рейтингу позичальника. Під час попередніх етапів вже було визначено попередній клас позичальника. Однак, при аналізі може виникнути ситуація, коли різні коефіцієнти кредитоспроможності вказують на різні класи для одного позичальника. У такому випадку потрібно визначити рейтинг позичальника.

Для оцінювання кредитоспроможності клієнта банк використовує різноманітні джерела інформації. Основним джерелом інформації є баланс позичальника разом з пояснювальною запискою. Аналіз балансу допомагає визначити, які фінансові ресурси має клієнт і до якої суми банк може надати йому кредит. Надаючи кредит, банк бере на себе значну відповідальність, тому важливо чітко та правильно визначити клас позичальника. У випадку, коли одному позичальнику присвоюються кілька класів одночасно, потрібно встановити рейтинг об'єкта.

Джерела інформації про клієнта включають такі ресурси: документи клієнта, архівна історія клієнта, дані, надані партнерами клієнта, звіти і висновки зовнішніх агентств та установ.

При роботі з активами балансу, важливо звернути увагу на наступні фактори:

При формуванні застави з основних коштів, таких як будівлі та обладнання, важливо підтвердити їх вартість шляхом включення до відповідних балансових статей. Це підтверджує втрату права власності заставника на ці активи.

Залишок грошових коштів на розрахунковому рахунку повинен відповідати даним банківської виписки на звітну дату. Це вказує на наявність вірного обліку фінансових ресурсів.

При аналізі дебіторської заборгованості важливо звернути увагу на терміни погашення. Відсоток позичальника від отримання боргів може бути одним із джерел повернення кредиту.

При розгляді пасивної частини балансу, особливу увагу необхідно звернути на наступні аспекти:

- Аналіз кредитних договорів за позиками, на які є заборгованість, що відображена у балансі, і переконатися, що вона не є протермінованою на момент запиту про кредит.
- Наявність протермінованої заборгованості за кредитами в інших банках є негативним фактором, що свідчить про помилки та недоліки в діяльності позичальника, які можуть бути тимчасово компенсовані шляхом отримання нового кредиту.
- Важливо перевірити, чи не була застава на забезпечення кредиту закладена в іншому банку. Оцінка стану кредитної заборгованості повинна включати переконання, що позичальник здатний своєчасно виконати зобов'язання за отриманими кредитами.[5]

2.2. Аналіз існуючих методів для оцінки кредитоспроможності клієнтів банку

Інформаційна система для аналізу кредитоспроможності позичальника надає можливість відхилити або розрахувати суму кредиту для клієнта. Об'єкт

вводить лише необхідні дані для розрахунку його кредитоспроможності та бажаної суми кредиту. Крім того, якщо клієнту підходять умови, він може залишити заявку на обробку своїх персональних даних на веб-сайті.

Коли заявка надходить до консультанта, він має можливість переглянути повну інформацію про клієнта за допомогою спеціальної кнопки. Якщо консультант після огляду заявки підтверджує згоду на кредитування особи, то клієнт може особисто звернутися до найближчого відділення банку, в якому була перевірена його кредитоспроможність. Підійшовши до консультанта, йому потрібно буде навести кілька основних даних для ідентифікації особи. Після цього всі дані, які клієнт вводив у системі, будуть відображені на екрані консультанта. Таким чином, можна уникнути довгих розмов з банківським працівником.

Проблеми, які висвітлюються в цьому контексті, включають:

1. Велика затрата працездатності на підготовку звітності або збір інформації з декількох джерел, навіть в межах одного філіалу.
2. Складність сприйняття необхідної інформації для аналізу та прийняття рішень в цій сфері.
3. Наявність недостовірних даних у джерелах інформації.
4. Втрата актуальності інформації через довгий процес підготовки звітності або відсутність оновлення на протязі тривалого періоду.
5. Відсутність первинного джерела достовірної інформації.

Переваги системи включають:

1. Розробка єдиного джерела достовірних даних в межах системи, на основі якого формується представлення для підтримки прийняття рішень. Ця система може скоротити час, потрібний для пошуку достовірної інформації з метою досягнення певної мети.
2. Можливість аналізувати історію великих обсягів даних в режимі реального часу. Швидкість запису, витягування та обробки необхідних даних дозволяє прискорити роботу працівників банку і зберегти час не лише для них, але й для їх клієнтів, які вже мають доступ до даної системи.

Так, система аналізу кредитоспроможності позичальника включає в себе

взаємопов'язані методи і способи збору, зберігання, накопичення, пошуку та обробки інформації, які базуються на використанні засобів обчислювальної техніки.

Ці системи розвиваються, щоб забезпечити ефективну обробку і аналіз великого обсягу даних про позичальників, їх фінансовий стан, ризики, розрахункові здібності та інші фактори, які впливають на їх кредитоспроможність.

Засоби обчислювальної техніки, такі як комп'ютери, програмне забезпечення та спеціалізовані алгоритми, дозволяють автоматизувати процеси обробки даних, розрахунку фінансових показників, створення моделей кредитоспроможності та оцінки ризиків. Вони сприяють покращенню швидкості, точності та надійності аналізу, а також забезпечують зручний доступ до необхідної інформації.

Зв'язок в системі аналізу кредитоспроможності використовується для обміну даними між різними джерелами, такими як банки, фінансові установи, кредитні бюро, що дозволяє отримувати повну та актуальну інформацію про позичальників для прийняття кредитних рішень.

Отже, системи аналізу кредитоспроможності позичальників є важливим інструментом для банків та інших фінансових установ при прийнятті рішень про надання кредитів, оцінці ризиків та забезпеченні стійкості фінансової системи. [1].

Висновки до розділу 2

У даному розділі було розглянуто різні підходи до вирішення завдання автоматизації процесу аналізу кредитоспроможності позичальника. Була проаналізована сутність кредиту, його типи і форми.

Було проведено дослідження різних підходів до визначення кредитоспроможності, включаючи визначення самого терміну «кредитоспроможність» та опис процесу аналізу коефіцієнтів кредитоспроможності.

Також був здійснений огляд існуючих технологій, що застосовуються для аналізу кредитоспроможності, і виявлено, що їх кількість є досить обмеженою.

Запровадження такої системи аналізу допоможе автоматизувати деякі процеси при отриманні кредиту, що в свою чергу забезпечить економію часу.



РОЗДІЛ 3. ПОБУДОВА АНАЛІТИЧНОГО ЗВІТУ ОЦІНКИ КРЕДИТОСПРОМОЖНОСТІ КЛІЄНТІВ БАНКУ

3.1. Постановка задачі, опис вхідної та вихідної інформації

Задачею випускної кваліфікаційної роботи є проаналізувати дані клієнтів банків, та побудувати на основі аналізу профілі клієнтів для визначення їх кредитоспроможності. Зробити візуалізацію на аналітичній платформі Power BI. Обов'язковими елементами візуалізації є таблиця (або матриця), стовпчикова діаграма, кругова або кільцева діаграма, графік та карта. Результатом виконання задачі є аналітичний звіт.

Дійсно, система аналізу кредитоспроможності позичальника включає в себе взаємопов'язані методи і способи збору, зберігання, накопичення, пошуку та обробки інформації з використанням аналітичних засобів. Ці методи і інструменти дозволяють банкам та фінансовим установам оцінювати кредитоспроможність позичальника для прийняття рішень щодо надання кредиту та визначення ризиків.

Кредитоспроможність позичальника характеризується декількома рисами, включаючи його репутацію, стійкий фінансовий стан та можливість мобілізувати грошові кошти з різних джерел. Репутація позичальника визначається його своєчасністю у виплаті попередніх кредитів. Стійкий фінансовий стан означає, що позичальник має достатні ресурси для покриття зобов'язань та вчасного повернення кредиту. Можливість мобілізувати грошові кошти з різних джерел є важливою для забезпечення фінансової стійкості позичальника.

Банк також оцінює прибуток, який позичальник отримує від кредитування конкретних витрат. Це оцінюється з точки зору здатності позичальника сплатити проценти при здійсненні нормальної фінансової діяльності. Банк також визначає ступінь ризику, який він готовий взяти на себе в даній ситуації.

При оцінці кредитоспроможності банк також досліджує мету використання кредитних ресурсів, суму кредиту та погашення кредиту:

– погашення кредиту, яке досліджують шляхом аналізу повернення кредиту шляхом реалізації матеріальних цінностей та наданих гарантій;

– забезпечення кредиту.

Дійсно, аналіз балансу позичальника разом з пояснювальною запискою є основним джерелом інформації для оцінювання їх кредитоспроможності. Баланс показує фінансовий стан позичальника, його активи, пасиви, капітал та зобов'язання. Пояснювальна записка доповнює баланс, надаючи детальнішу інформацію про розрахункові показники, фактори, що впливають на фінансовий стан позичальника, та пояснюючи особливості його діяльності.

Однак, окрім балансу та пояснювальної записки, існують й інші джерела інформації про клієнта.

Документи клієнта, такі як фінансові звіти, звіти про доходи, договори та інші релевантні документи.

Архівна історія клієнта, яка містить інформацію про попередні кредити, їх повернення та інші фінансові операції.

Дані, які надають партнери клієнта, наприклад, рекомендації від постачальників або покупців.

Звіти і висновки зовнішніх агентств та закладів, які проводять незалежну оцінку кредитоспроможності позичальника.

Рис.3.1 Джерела інформації суб'єкта

Узагальнюючи, для оцінювання кредитоспроможності позичальника використовуються не тільки дані з балансу та пояснювальної записки, але й інші.

Правильно буде зауважити, що аналіз балансу позичальника дає загальне уявлення про його кредитоспроможність, а для отримання більш детальних та якісних показників необхідно використовувати інші джерела інформації. Дійсно, на більш глибокий аналіз кредитоспроможності позичальника впливають такі фактори:

- Застава основних коштів: Забезпечення кредиту може включати основні активи, такі як будівлі, обладнання та інше майно. Вартість цих активів повинна відображатися на відповідних балансових статтях.
- Розрахунковий рахунок: Залишок коштів на розрахунковому рахунку позичальника повинен відповідати даним банківської виписки на звітну дату.

– Дебіторська заборгованість: При аналізі дебіторської заборгованості слід звернути увагу на терміни погашення, оскільки ці надходження можуть бути джерелом для повернення кредиту.

– Пасивна частина балансу: Особлива увага має бути приділена розділам, які характеризують стан позикових коштів.

Додатково, при оцінці кредитної заборгованості варто враховувати такі фактори:

– Аналіз кредитних договорів: Потрібно перевірити наявність кредитних договорів, за якими позичальник має заборгованість, та переконатися, що вона не є протермінованою.

– Протермінована заборгованість за кредитами інших банків: Наявність протермінованої заборгованості у позичальника за кредитами в інших банках свідчить про можливі помилки та проблеми при поверненні коштів[21].

3.2. Опис початкових етапів побудови аналітичного звіту

Для роботи використовуються чотири Ексел-таблиці з даними, що є джерелами інформації для аналізу кредитоспроможності. Перед початком аналізу важливо виконати попередню підготовку даних, яка включає наступні кроки:

Перевірку правильності та повноту даних: Потрібно переконатися, що всі необхідні дані наявні у таблицях і вони відповідають вимогам аналізу. перевірити, чи немає порожніх полів або помилок у введенні даних.

Об'єднання таблиць: Якщо дані розподілені по кількох таблицях, переконайтеся, що вони можуть бути злиті разом для подальшого аналізу. Використовуйте відповідні ключі або ідентифікатори для злиття таблиць у відповідних колонках.

Форматування даних: Виконайте необхідні операції форматування даних, які можуть включати видалення зайвих символів, зміну типів даних (наприклад, текст у числа) та інші операції для полегшення подальшого аналізу.

Створення нових змінних: В разі потреби створіть нові змінні або показники, які можуть бути корисними для оцінки кредитоспроможності.

Наприклад, розрахуйте показники ліквідності, платоспроможності, прибутковості тощо на основі наявних даних.

Фільтрація та сортування: Використовуйте фільтри та сортування для вибору необхідних даних та групування їх за потрібними критеріями. Це допоможе у визначенні підгруп даних для подальшого аналізу.

Наступним є імпорт даних з Excel-таблиць до системи Power BI та їх подальшу трансформацію за допомогою Power Query. Ось кілька кроків, які можна виконати, використовуючи Power Query:

1. Відкрийте Power BI та створіть новий проект або відкрийте існуючий проект.
2. У меню "Файл" виберіть "Отримати дані" і оберіть тип даних "Excel".
3. Виберіть потрібну Excel-таблицю, з якої ви хочете імпортувати дані, і натисніть "ОК".
4. Power BI запустить Power Query Editor, де ви зможете виконати трансформації даних.
5. Виконайте потрібні зміни в Power Query Editor, щоб трансформувати дані за вашими потребами. Наприклад, змініть типи даних для цілих чисел, змініть формат дати та часу, встановіть грошовий формат та грошову одиницю.
6. Перевірте результати трансформації, щоб переконатися, що дані відображаються правильно. Виконайте необхідні коригування, якщо потрібно.
7. Після завершення трансформації даних натисніть "Застосувати та закрити" для застосування змін до даних і повернення до Power BI.
8. Дані тепер імпортовані до Power BI та готові для подальшого аналізу та візуалізації.

Зверніть увагу, що Power Query дозволяє виконувати багато інших операцій трансформації даних, таких як об'єднання таблиць, фільтрація, сортування, розрахунки нових стовпців тощо. Ви можете скористатися цими функціями, щоб отримати потрібні результати для вашого аналізу кредитоспроможності.

Структурованість даних є важливим аспектом управління базами даних. Для забезпечення цілісності та ефективного пошуку даних в БД, потрібно

враховувати наступні аспекти:

- **Визначення структури даних:** Структура даних визначає, як дані організовані та взаємозв'язані між собою. Вона включає таблиці, поля, ключі, зв'язки та інші складові частини бази даних. Добре спроектована структура даних дозволяє ефективно зберігати, оновлювати та отримувати дані.
- **Типізація даних:** В БД кожен елемент має свій тип даних, наприклад, цілі числа, рядки, дата та інші. Встановлення правильного типу даних для кожного елемента допомагає забезпечити точність та ефективність операцій з даними.
- **Трансформація даних:** При імпорті даних до Power BI за допомогою Power Query ви можете змінити типи даних, формати та інші атрибути для відповідності вашим потребам аналізу. Це дозволяє вам привести дані до необхідного формату та забезпечити їхню правильну інтерпретацію.
- **Забезпечення безпеки даних:** В сучасному світі безпека даних є надзвичайно важливою. Адміністратори баз даних повинні вживати заходів для захисту даних від несанкціонованого доступу, витоків інформації та інших загроз. Це включає реалізацію відповідних прав доступу, шифрування даних та інші заходи безпеки.
- **Оптимізація продуктивності:** Великі обсяги даних та складні запити можуть призвести до зниження продуктивності бази даних. Адміністратори повинні вживати заходів.

Концептуальна модель бази даних – це деяка наочна діаграма, нарисована в прийнятих позначеннях, яка детально показує зв'язок між об'єктами та їх характеристиками.

Основна задача концептуальної моделі – це передати фундаментальні принципи та основні функціональні можливості системи, яку вона представляє. Крім того, модель повинна бути розроблена так, що б забезпечити легко зрозумілу системну інтеграцію для користувачів моделі.

Нижче зображено модель бази даних системи аналізу кредитоспроможності позичальника (див. рис. 3.2).

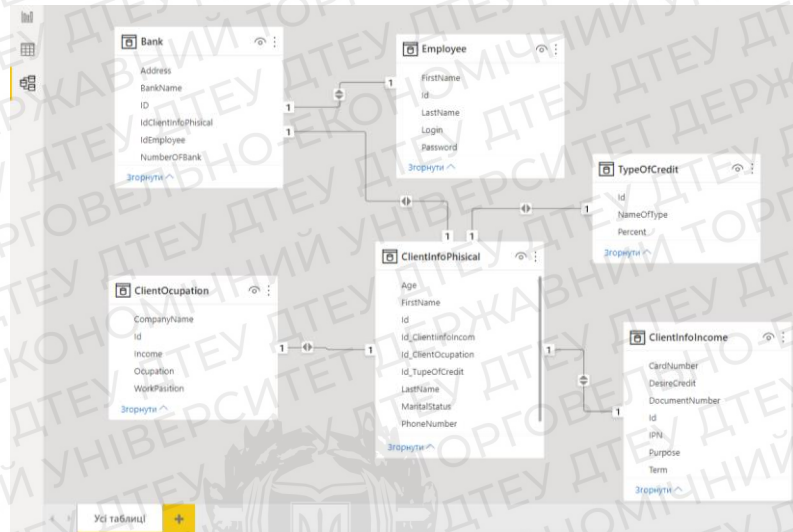


Рис. 3.2. Модель бази даних системи аналізу кредитоспроможності позичальника

Всі поля БД, їх тип та приклад наповнення даними для системи аналізу розписано нижче.

Таблиця 3.1 Фізична модель бази даних системи аналізу кредитоспроможності позичальника

Name of table	Row	Type	Keys	Relationships	Example
Bank	id	INT	Primary key	-	1
	NumberOfBank	TEXT	-	-	47
	BankName	TEXT	-	-	Privat
	Address	TEXT	-	-	langclia 17
	IdEmployee	INT	Foreign key	Employee	3105
	IdClientInfoPhysical	INT	Foreign key	ClientInfoPhysical	0409
Employee	id	INT	Primary key	-	3105
	FirstName	TEXT	-	-	Katya
	LastName	TEXT	-	-	Yanchuk
	Login	TEXT	-	-	Katya Yan
	Password	TEXT	-	-	123456789
ClientInfo Physical	id	INT	Primary key	-	0409
	FirstName	TEXT	-	-	Vladyslav
	LastName	TEXT	-	-	Prysiazhniuk
	Age	INT	-	-	21
	PhoneNumber	TEXT	-	-	0961789539
	Sex	BOOL	-	-	Male
	MaritalStatus	BOOL	-	-	Married
	Id_TypeOfCredit	INT	Foreign key	TypeOfCredit	123
	IdClientInfoIncome	INT	Foreign key	ClientInfoIncome	321
IdClientOccupation	INT	Foreign key	ClientOccupation	135	
TypeOfCredit	id	INT	Primary key	-	123
	NameOfType	TEXT	-	-	Fast Credit
	Percent	INT	-	-	24%

ClientInfo Income	id	INT	Primary key	321
	Purpose	TEXT		Car
	DcsircCrcdit	TEXT		10000
	Term	DATE		05.05.2022
	DocumentNumber	TEXT		00671674772
	IPN	TEXT		37881784741
	CardNumber	TEXT		44411144XX
Client Occupation Client Occupation (продовження)	id	INT	Primary key	135
	Income	INT		12000
	Occupation	TEXT		Trading
	WorkPosition	TEXT		Operator
	CompanyName	TEXT		APG

На основі даних таблиць які представлені на рисунку 3.3 побудовану нову, загальну з інформацією про клієнта, яка буде використовуватись для візуалізації даних та побудові звіту рис. 3.4

Коданкети	Прізвище	ім'я	По-батькові	Сума кредиту, грн #	Кількість років проживання у регіоні	Соціальний статус	Соціальний статус чоловіка(и)	Освіта	Стаж роботи, років
3049	Абаджев	Микола	Васильович	47000	15	Працюю/служу	Працюю/служу	середня	8,5
4000	Широкова	Світлана	Миколаївна	54000	19	Працюю/служу	н/д	початкове	1
3052	Поліва	В'чеслав	Леонідович	38000	8	Працюю/служу	н/д	вища	8
3053	Білець	Юрій	Алефійович	25000	19	Працюю/служу	Працюю/служу	середня	1
3055	Репніков	Ариадій	Іллі	35000	27	Працюю/служу	Працюю/служу	незакінчена вища	2
3056	Калугін	Анатолій	Олександрович	58000	10	Працюю/служу	Працюю/служу	середня	1
3058	Самолічкова	Ніна	Дмитрівна	44000	6	Працюю/служу	Працюю/служу	вища	1
3060	Катков	Андрій	Вікторович	47000	13	Працюю/служу	н/д	середня	4
3061	Абаєв	Олександр	Вікторович	32000	6	Працюю/служу	Працюю/служу	кілька вищих чи вищий ступінь	6
3062	Кудабасєв	Рустам	Альбертович	57000	2	Працюю/служу	н/д	вища	2
3063	Кондратьєва	Ганна	Василівна	47000	8	Працюю/служу	н/д	середня	26
3064	Стрелікова	Міла	Миколаївна	63000	16	Працюю/служу	Працюю/служу	середня	2
4010	Корішова	Олена	Миколаївна	47000	4	Працюю/служу	н/д	незакінчена вища	4
3066	Кардальціш	Микола	Георгійович	64000	17	Працюю/служу	Працюю/служу	вища	2
3067	Шуклін	Віталій	Георгійович	38000	8	Працюю/служу	Працюю/служу	середня	1
3068	Орлова	Марина	Анатоліївна	25000	8	Працюю/служу	Працюю/служу	середня	3
3070	Мендугіна	Світлана	Миколаївна	50000	5	Працюю/служу	н/д	вища	1,5
3071	Корішова	Лариса	Володимирівна	60000	3	Працюю/служу	Не працює	середня	1,5
3072	Миколаєв	Анатолій	Миколайович	54000	34	Працюю/служу	н/д	вища	1,5
3073	Миротайєва	Файруза	Гіланівна	42000	19	Працюю/служу	Не працює	середня	4,5
3074	Іванова	Микола	Данилович	51000	24	Працюю/служу	Працюю/служу	вища	3
3076	Полікова	Тамара	Іванівна	36000	9	Працюю/служу	н/д	вища	2
4000	Ларіна	Олена	Валентинівна	53000	18	Працюю/служу	Працюю/служу	вища	1
4001	Наговицин	Микола	Георгійович	41000	6	Працюю/служу	Не працює	незакінчена вища	1
4002	Сихомова	Ольга	Сергіївна	70000	14	Працюю/служу	н/д	середня	2
4004	Петров	Антон	Сергієвич	86000	16	Працюю/служу	н/д	середня	1,5
4005	Чешкова	Олена	Борисівна	41000	20	Працюю/служу	н/д	вища	4

Рис 3.3 Розрахунок середньої кількості магазинів

В результаті візуалізації даних було створено аналітичний звіт з наступних інформаційних сторінки, що містять в собі:

Кругову діаграму для оцінки профіля клієнта банку, що показує взаємозв'язок між освітою та доходами клієнтів, як одного з факторів прийняття рішення для видачі кредиту, дана діаграма побудована на основі побудови нової міру, рис. 3.4.

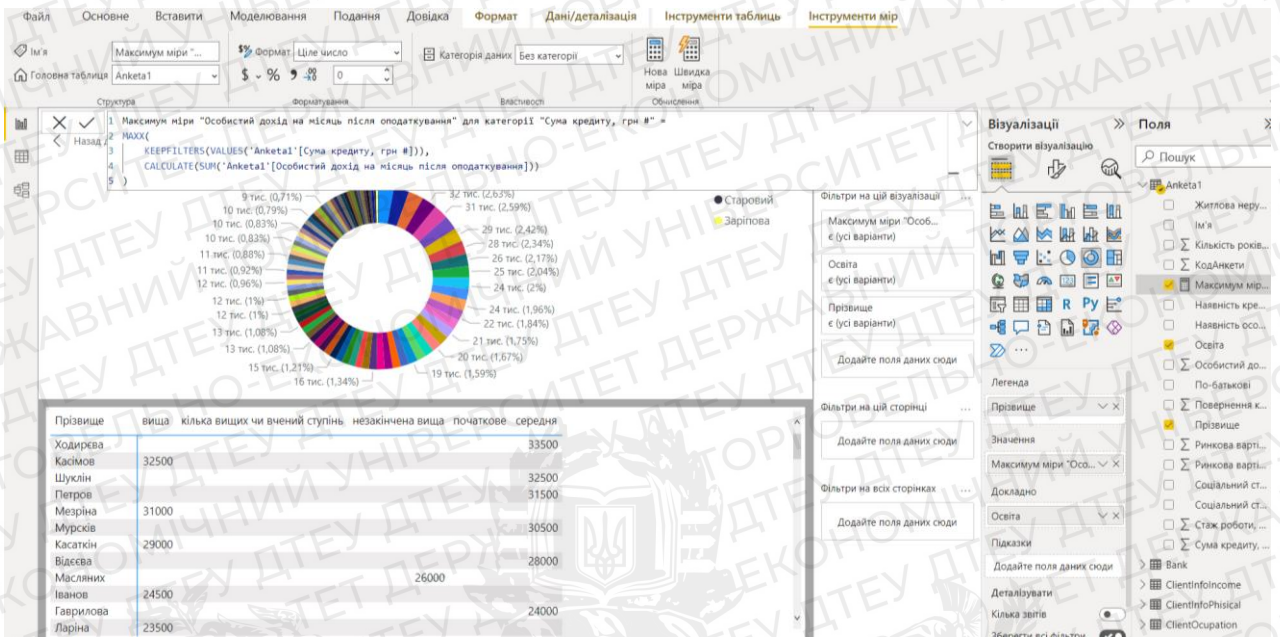


Рис 3.4 Взаємозв'язок доходами та освітою клієнтів банку

Перша сторінка звіту призначена для аналізу профілів клієнтів, для відображення інформації про суму кредиту в залежності від наявності в клієнтів особистого автомобіля, їх освіти та інформації про наявність неповернутих кредитів рис. 3.5.

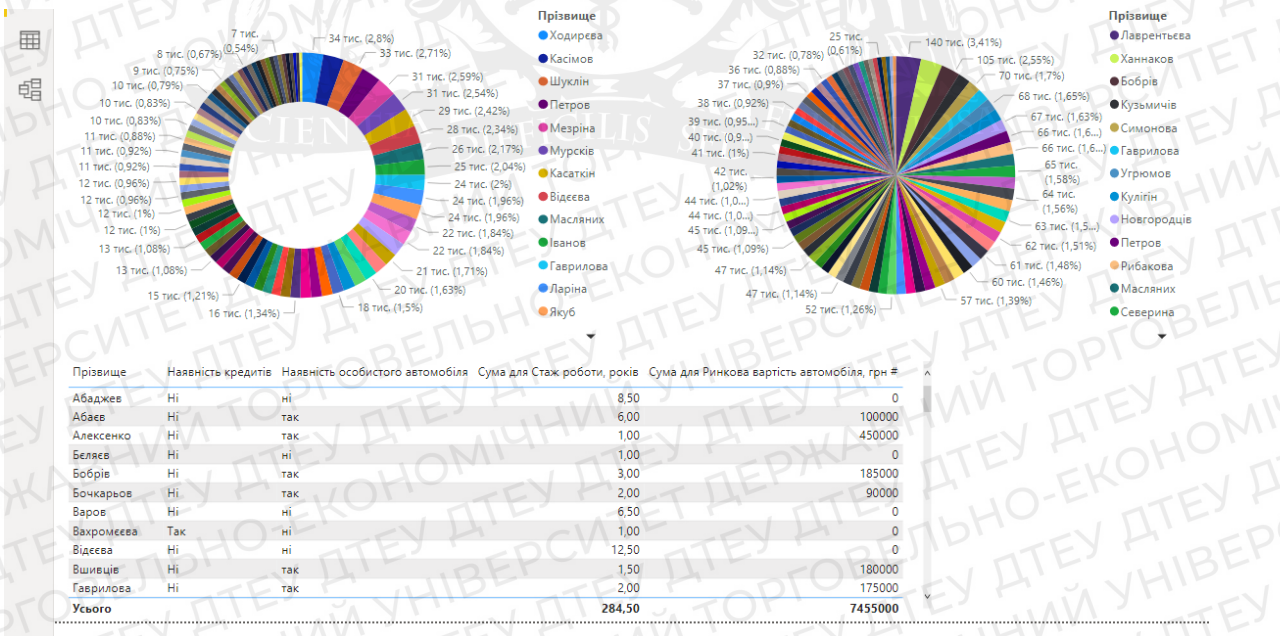


Рис. 3.5 Перша сторінка звіту

Друга сторінка аналітичного звіту демонструє відображення аналітичної інформації про соціальний статус клієнта, його прибуток, суму кредиту та відсоткове співвідношення для відображення по категорії соціального статусу для відображення інформації у відповідності до загального показника категорії рис. 3.6.

кредитоспроможність, і додається ще одне обмеження яке показує вищу кредитоспроможність при умові, що «Дохід» більший за 5000 і сума кредиту не перевищує 30000. Другий показник, від якого залежить вища або ж нижча кредитоспроможність показує наявність у клієнта власного автомобіля, третій, який показує наявність діючих, непогашених кредитів рис. 3.8.

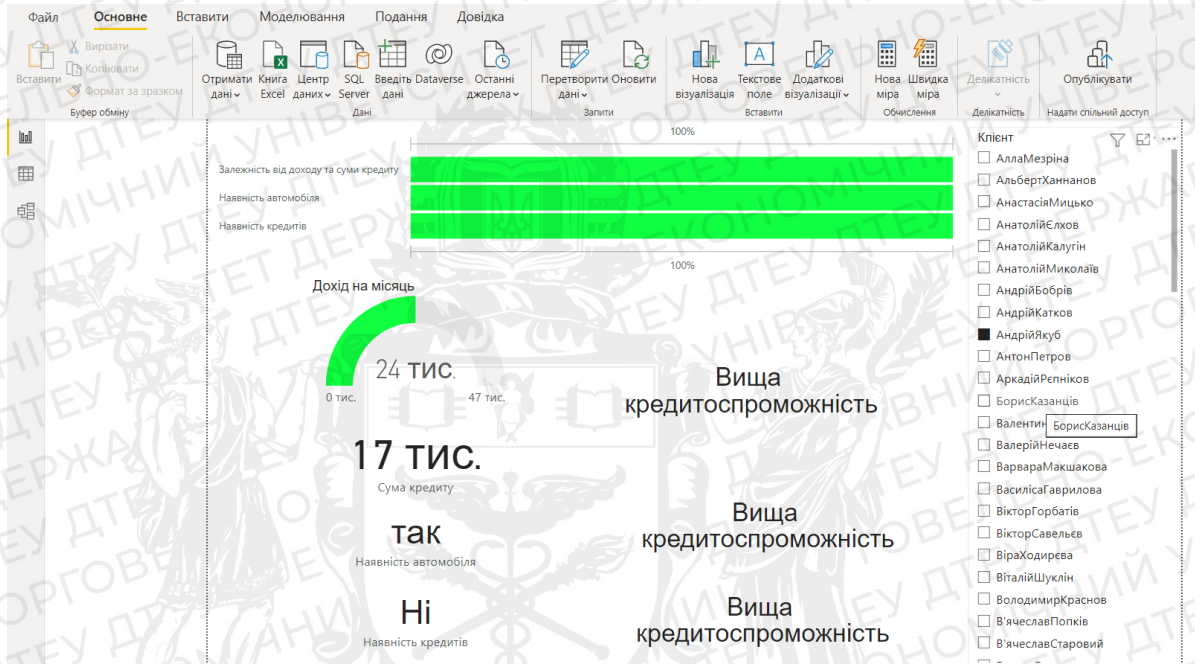


Рис.3.8 Профіль клієнта для оцінки кредитоспроможності

На основі трьох рекомендацій є візуалізатор, який показує зеленим кольором вищу або нижчу кредитоспроможність, чим більше ліній тим вищий статус клієнта для отримання кредиту.

Кожна візуалізація представлена в аналітичному звіті є інтерактивною, при натисканні конкретної ланки, ряду таблиці, частки кругової діаграми та інших елементів буде конкретизовано інформацію у всіх вище перерахованих візуальних об'єктах та показано детальну інформацію по необхідному клієнту, для визначення його кредитоспроможності та прийняття рішення по наданню кредиту.

При завершенні формуванні аналітичного звіту, він був опублікований на порталі Power BI Service.

Отож, для початку публікації звіту необхідно перейти у вкладку «Основне», вибрати категорію «Надати спільний доступ» та натиснути «Опублікувати» рис. 3.9.

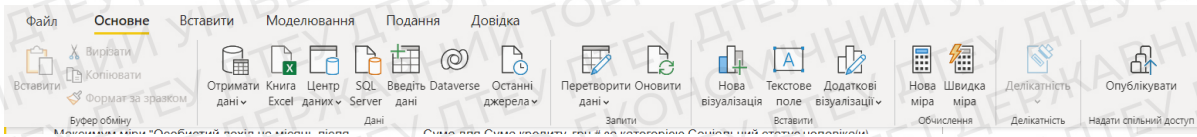


Рис. 3.9 Процедура публікації звіту

Останнім етапом, при успішній публікації, з'явиться готовий звіт на порталі Power BI Service.

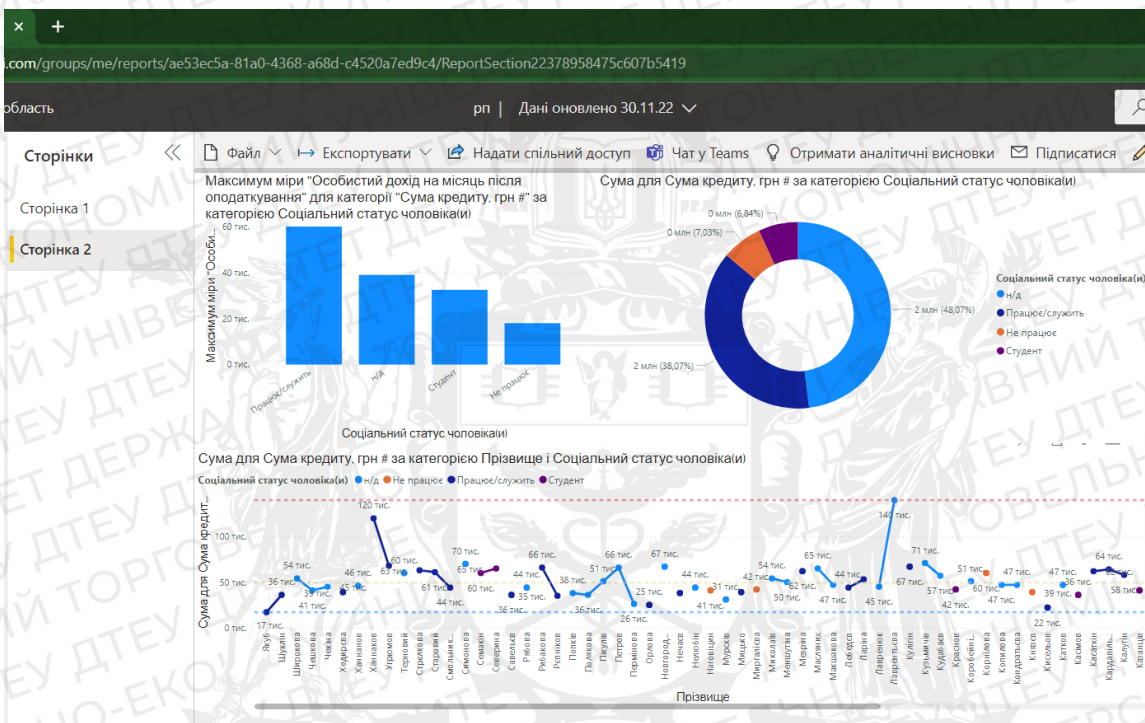


Рис. 3.10 Відображення звіту за допомогою Google Chrome

Функції доступу для опублікованого звіту:

- редагування сторінок аналітичного звіту;
- надання спільного доступу;
- збереження копії;
- друк сторінки;
- завантаження RВІХ-файлу;
- експортування файлу в Power Point, Excel, PDF.

Щоб оновити дані у звітах Power BI на сервері звітів Power BI, потрібно створити розклад оновлення. Щоб створити даний план, треба перейти до налаштувань робочої області Power BI, далі вибрати частоту оновлення, часовий пояс і час.

Також є можливість вказати адресу електронної пошти, якщо щось піде не так під час оновлення.

Висновки до розділу 3

Сформовані статистичні дані які дозволили підготувати їх до аналізу, відредагувати некоректні або пусті значення, розробити нові запити між таблицями, для подальшого їх використання. Використані формули з використанням мови програмування запитів DAX.

Імпортувавши дані, здійснивши необхідні розрахунки та розробивши аналітичний звіт в аналітичній платформі Power BI, з подальшою публікацією на сервері звітів Power BI Service, був проведений аналіз даних по позичальниках в банку, та побудовано профілі клієнтів для визначення кредитоспроможності.



ВИСНОВКИ ТА ПРОПОЗИЦІЇ

Оцінка кредитоспроможності клієнтів банку спрощується можливістю клієнта пройти анкетування банку, дані відразу передаються до бази даних, де з ними може ознайомитись працівник банку і проглянувши їх, визначається чи клієнт кредитоспроможний і пропонує йому умови кредитування, якщо позичальника все влаштовує, то він може явитися в відділення банку, на умовах якого він аналізував свою кредитоспроможність, і зустрівшись обличчям до обличчя з банківським працівником все що буде потрібно клієнтові – назвати своє прізвище, ім'я і консультант банку одразу зможе побачити абсолютно всі дані клієнта, які було введено в комфортних для об'єкта умовах.

У теоретичній частині випускної кваліфікаційної роботи викладено основні завдання та етапи аналізу даних, задачі консолідації даних, їхні типи, методи і структури сховищ (репозиторіїв) даних та алгоритми передобробки даних.

Відповідно до поставлених завдань, розглянуто програмні засоби аналізування даних, побудови аналітичного звіту, в особистості виділено 3 найбільш поширених платформ для аналізу даних та опис їхніх можливостей у цілому.

В процесі побудови аналітичного звіту, було опрацьовано основні можливості використання даних з аналітичною платформи Microsoft Power BI, а саме: трансформація та візуалізація; побудова моделі даних; оновлення даних. Між елементами візуалізації було використано для зображення показників інструменти візуалізації такі елементи як: кругові та стовпчасті діаграми, картки, матричні таблиці. Проаналізовано їх користь, простоту і комфортність у використанні.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Чубукова І. А. Data Mining: навчальний посібник. Київ : КНЕУ, 2016. 328 с.
2. Консолідація даних [Електронний ресурс]. Режим доступу: <https://ppt-online.org/202010>
3. Apache Spark [Електронний ресурс]. Режим доступу: <https://searchdatamanagement.techtarget.com/definition/Apache-Spark>
4. Магічний квадрант The Gartner [Електронний ресурс]. Режим доступу: <https://www.gartner.com/doc/reprints?id=1-24ZXJ0MU&ct=210107&st=sb>
5. Larson B. Data Analysis with Microsoft Power BI
6. Інструменти Data Mining [Електронний ресурс]. Режим доступу: <https://hackr.io/blog/top-data-analytics-tools>
7. Характеристика Microsoft Power BI [Електронний ресурс]. Режим доступу: <https://www.nigelfrank.com/blog/everything-you-ever-wanted-to-know-about-microsoft-power-bi/>
8. Застосування різних візуалізаторів [Електронний ресурс]. Режим доступу: <https://toplead.com.ua/ru/blog/id/vizualizacija-dannyh-kak-pravilno-vybrat-diagrammu-ili-grafik-dlja-godovogo-otcheta-162/>
9. Сфери застосування Data Mining [Електронний ресурс]. Режим доступу: <https://www.researchgate.net/post/14-areas-where-data-mining-is-widely-used>
10. Характеристика аналітичних платформ [Електронний ресурс]. Режим доступу: <https://habr.com/ru/company/newprolab/blog/349186/>
11. O'Connor E. Microsoft Power BI Dashboards Step by Step: 394 p
12. Профайлінг даних [Електронний ресурс]. Режим доступу: <https://panoply.io/analytics-stack-guide/data-profiling-best-practices/>
13. Типи візуальних елементів [Електронний ресурс]. Режим доступу: <https://docs.microsoft.com/ru-ru/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a>

14. Ринок бізнес-аналітики [Електронний ресурс]. Режим доступу: <https://www.mordorintelligence.com/industry-reports/global-business-intelligence-bi-vendors-market-industry>
15. Огляд аналітичних платформ [Електронний ресурс]. Режим доступу: <https://www.wallstreetmojo.com/power-bi-vs-tableau-vs-qlikview/>
16. Складові магічного квадранту [Електронний ресурс]. Режим доступу: <https://uk.icyscience.com/gartner-magic-quadrant>
17. Олійник А. О., Суботін С. О., Олійник О. О. Інтелектуальний аналіз даних: навчальний посібник. Запоріжжя: ЗНТУ, 2012. 278 с.
18. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підручник. Київ: 2010, 841 с.
19. Steven P. Blais Business Analysis: 2011. 432 с.
20. Аналіз кредитоспроможності підприємства. [Електронний ресурс]. Режим доступу: https://pidru4niki.com/10310208/ekonomika/analiz_kreditospromozhnosti_pidpriyemstva