МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему: **«ОПТИМІЗАЦІЯ МЕТОДІВ ЗБОРУ ТА АНАЛІЗУ ДАНИХ ДЛЯ БІЗНЕСУ З ВИКОРИСТАННЯМ СУЧАСНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ»**

Виконав студент 2 курсу, групи 8.1223
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)
освітньої програми Комп'ютерні науки
(назва освітньої програми)
Н.І. Бородін
(ініціали та прізвище)
Керівник доцент кафедри комп'ютерних наук, доцент, к.т.н. Борю С.Ю.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)
Рецензент завідувач кафедри програмної інженерії , доцент, к.ф.-м.н. Лісняк А.О.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Запоріжжя – 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет математичний

Кафедра комп'ютерних наук

Рівень вищої освіти магістр

Спеціальність 122 Комп'ютерні науки
(шифр і назва)

Освітня програма Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри комп'ютерних наук,
д.т.н., професор

_____ Шило Г.М.
(підпис)

" 5 " березня 2024 р.

З А В Д А Н Н Я

НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТОВІ

Бородіну Нікіті Ігоровичу
(прізвище, ім'я та по-батькові)

1. Тема роботи Оптимізація методів збору та аналізу даних для бізнесу з використанням сучасних інформаційних технологій

керівник роботи Борю Сергій Юрійович, к.т.н., доцент
(прізвище, ім'я та по-батькові, науковий ступінь, вчене звання)

затверджені наказом ЗНУ від « 01 » березня 2024 року № 385-с

2. Строк подання студентом роботи 27.11.24

3. Вихідні дані до роботи джерела Інтернет, спеціалізована література та періодичні видання

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
Кваліфікаційна робота магістра складається з основної частини та додаткової. Основна частина включає такі структурні елементи: вступ, три розділи, висновки.
1 Розділ – Аналіз предметної області; 2 Розділ – Розробка методів збору та аналізу даних
3 Розділ - Реалізація програмного забезпечення для оптимізації методів збору та аналізу

5. Перелік графічного матеріалу презентація

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання 04.03.2024

К А Л Е Н Д А Р Н И Й П Л А Н

№	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Розробка плану роботи.	10.03.2024	
2.	Збір вихідних даних.	15.04.2024	
3.	Обробка методичних та теоретичних	20.05.2024	
	джерел.		
4.	Розробка першого та другого розділу.	27.10.2024	
5.	Розробка третього розділу.	20.11.2024	
6.	Оформлення та нормоконтроль	26.11.2024	
	кваліфікаційної роботи магістра.		
7.	Захист кваліфікаційної роботи.	12.12.2024	

Студент _____ Н.І. Бородін
(підпис) (ініціали та прізвище)

Керівник роботи _____ С.Ю. Борю
(підпис) (ініціали та прізвище)

Нормоконтроль пройдено

Нормоконтролер _____ О.Г.Спиця
(підпис) (ініціали та прізвище)

РЕФЕРАТ

Кваліфікаційна робота магістра «Оптимізація методів збору та аналізу даних для бізнесу з використанням сучасних інформаційних технологій»: 68 сторінок, 7 рисунків, 11 таблиць, 31 джерел, 1 додаток.

АВТОМАТИЗАЦІЯ, АНАЛІЗ ДАНИХ, БІЗНЕС-АНАЛІТИКА, ОПТИМІЗАЦІЯ, СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ.

Об’єкт дослідження – методи збору та аналізу даних, що застосовуються у бізнесі для оптимізації управлінських рішень.

Мета роботи – розробка методів і програмного забезпечення для оптимізації процесів збору, обробки та аналізу даних у бізнес-середовищі із використанням сучасних інформаційних технологій.

Методи дослідження – аналітичні методи для огляду предметної області, алгоритми кластеризації та прогнозування, методи автоматизації процесів збору даних, статистичні та математичні методи для аналізу інформації

У роботі проведено аналіз сучасних технологій збору, обробки та аналізу даних, включаючи хмарні сервіси, BI-платформи та алгоритми машинного навчання. Розроблено архітектуру системи, яка складається з модулів збору, обробки, аналізу та візуалізації даних. Реалізоване програмне забезпечення забезпечує інтеграцію з API, автоматизацію та очищення даних.

Результати роботи підтверджують ефективність розробленого підходу, що дозволяє підвищити точність аналізу, автоматизувати рутинні процеси та скоротити витрати часу. Програмне забезпечення рекомендовано для використання в різних галузях бізнесу, таких як фінанси, логістика, маркетинг та клієнтське обслуговування.

Розроблена система є гнучкою та перспективною для подальшого вдосконалення шляхом інтеграції з хмарними платформами, підтримки великих даних та впровадження методів машинного навчання

SUMMARY

Master's Qualification Thesis "Optimization of Data Collection and Analysis Methods for Business Using Modem Information Technologies": 68 pages, 7 figures, 11 tables, 31 sources, 1 annex.

AUTOMATION, DATA ANALYSIS, BUSINESS ANALYTICS, OPTIMIZATION, MODERN INFORMATION TECHNOLOGIES.

Object of research – methods of data collection and analysis used in business to optimize managerial decision-making.

Purpose of the study – to develop methods and software for optimizing the processes of data collection, processing, and analysis in the business environment using modern information technologies.

Research methods – analytical methods for reviewing the subject area, clustering and forecasting algorithms, automation methods for data collection processes, statistical and mathematical techniques for data analysis.

The study includes an analysis of modern technologies for data collection, processing, and analysis, such as cloud services, BI platforms, and machine learning algorithms. The architecture of the system was developed, consisting of modules for data collection, processing, analysis, and visualization. The implemented software ensures integration with APIs, automation, and data cleaning.

The results confirm the efficiency of the developed approach, allowing for improved accuracy in analysis, automation of routine processes, and reduced time costs. The software is recommended for use in various business sectors, such as finance, logistics, marketing, and customer service.

The developed system is flexible and has potential for further enhancement through integration with cloud platforms, support for big data, and the introduction of machine learning methods.

ЗМІСТ

Завдання на кваліфікаційну роботу…… ……2
Реферат 4
Summary 5
Вступ..... 7
1 Аналіз предметної області..... 9
1.1 Огляд методів збору даних для бізнеса 9
1.2 Огляд сучасних інформаційних технологій 14
1.3 Огляд систем для аналізу даних і методів. 19
1.4 Можливі альтернативи для оптимізації збору та аналізу даних 29
2 Розробка методів збору та аналізу даних 37
2.1 Постановка технічного завдання..... 37
2.2 Інструменти та середовище розробки..... 41
2.3 Методи аналізу та обробки даних 43
2.4 Розробка алгоритму оптимізації процесів збору даних 46
3 Реалізація програмного забезпечення для оптимізації методів збору та аналізу даних..... 52
3.1 Проектування архітектури системи 52
3.2 Проектування архітектури клієнтського застосунку 53
3.3 Реалізація клієнтського застосунку..... 55
3.4 Порівняння розробленої системи з існуючими рішеннями..... 59
Висновки 62
Перелік посилань..... 64
Додаток А Діаграми класів модулю..... 67

ВСТУП

У сучасному бізнес-середовищі ефективність прийняття рішень безпосередньо залежить від якості зібраних даних. Інформація стала важливим ресурсом, що визначає конкурентоспроможність компанії, дозволяє адаптуватися до змін ринку, розробляти нові стратегії та задовольняти потреби клієнтів.

Дипломна робота присвячена оптимізації методів збору та аналізу даних для бізнесу із застосуванням сучасних інформаційних технологій. У сучасних умовах бізнес стикається зі значними обсягами даних, які необхідно обробляти, аналізувати та використовувати для прийняття обґрунтованих управлінських рішень. Використання інноваційних технологій дозволяє автоматизувати процеси збору даних, підвищувати точність аналізу та знижувати витрати ресурсів.

Об’єктом дослідження є процеси збору, обробки та аналізу даних у бізнесі.

Предметом дослідження є використання сучасних інформаційних технологій для оптимізації процесів роботи з даними, включаючи автоматизацію, застосування алгоритмів аналізу та візуалізацію результатів.

Метою роботи є розробка підходів і програмного забезпечення, які дозволяють ефективно збирати, обробляти й аналізувати дані для бізнесу, інтегруючи сучасні інформаційні технології, такі як автоматизація, алгоритми кластеризації та інструменти візуалізації.

Завдання роботи:

Провести аналіз сучасних методів збору та аналізу даних, огляд існуючих технологій і інструментів.

Розробити технічне завдання для програмного забезпечення, визначивши методи оптимізації.

Реалізувати алгоритми збору, обробки та аналізу даних із застосуванням автоматизації та інтеграції з сучасними платформами.

Провести порівняльний аналіз розробленого рішення з існуючими альтернативами.

Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків. У першому розділі проведено аналіз предметної області та сучасних технологій збору й аналізу даних. У другому розділі сформульовано технічне завдання, визначено методи реалізації програмного забезпечення. Третій розділ присвячено реалізації системи, аналізу її ефективності та порівнянню з іншими рішеннями.

Дана робота є важливим внеском у розвиток автоматизованих рішень для бізнесу, оскільки вона демонструє можливості сучасних технологій у оптимізації роботи з великими обсягами інформації, сприяючи підвищенню ефективності прийняття управлінських рішень.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Огляд методів збору даних для бізнеса

У сучасному бізнес-середовищі ефективність прийняття рішень безпосередньо залежить від якості зібраних даних. Інформація стала важливим ресурсом, що визначає конкурентоспроможність компанії, дозволяє адаптуватися до змін ринку, розробляти нові стратегії та задовольняти потреби клієнтів.

Методи збору даних, які використовуються у бізнесі, охоплюють широкий спектр інструментів і підходів, що дозволяють отримувати інформацію як про поведінку споживачів, так і про внутрішні бізнес-процеси. Усі методи поділяються на якісні та кількісні, залежно від типу інформації, що збирається, і способу її подальшого аналізу[1].

Якісні дослідження спрямовані на отримання глибокого розуміння поведінки, мотивів і цінностей респондентів. На відміну від кількісних методів, вони базуються на неформалізованому підході, який включає гнучкі польові методи та нестандартизований аналіз отриманої інформації. Як зазначає український дослідник О. І. Мороз у своїй роботі "Якісні методи в соціологічних дослідженнях", якісні методи дозволяють отримати інсайти, які респонденти часто не можуть або не хочуть висловлювати у стандартних опитуваннях. Подібної думки дотримується й американський вчений Д. Лінкольн, який підкреслює, що якісні дослідження є основою для виявлення прихованих цінностей і культурних відмінностей у поведінці споживачів. За своєю структурою якісні методи поділяються на дві основні групи: прямі та непрямі[2].

Прямі (незасекречені) методи збору даних – це дослідницькі підходи, за яких респонденти знають про свою участь у дослідженні. Їх інформують про мету, формат або специфіку процесу збору даних. Ці методи характеризуються

прозорістю і використовуються для отримання відкритих і добровільних відповідей.

До основних прямих методів належать:

Фокус-групи: дискусії в малих групах (7–12 осіб) під керівництвом модератора, які спрямовані на виявлення колективної думки щодо певного питання, продукту або послуги.

Глибинні інтерв'ю: індивідуальні бесіди з акцентом на деталізовані відповіді, які розкривають глибокі мотиви, досвід або ставлення респондентів[3].

Як відзначає Н. Г. Чернишенко, фокус-групи є ефективним інструментом для дослідження групових норм і тенденцій, тоді як глибинні інтерв'ю дозволяють зосередитися на індивідуальних особливостях кожного респондента [4].

Непрямі (засекречені) методи збору даних – це способи отримання інформації, за яких справжня мета дослідження приховується від респондентів. Ці методи застосовуються в ситуаціях, коли учасники не бажають або не можуть надавати відкриту інформацію через соціальний тиск, сором'язливість чи небажання розкривати деталі[5].

До таких методів належать:

Легендовані методи: респондентам пропонується взяти участь у заході чи дослідженні під вигаданим приводом, який не розкриває справжньої мети. Наприклад, фокус-група організовується під виглядом обговорення загальної теми, а дослідник отримує інформацію, яка стосується іншого питання.

Проективні техніки: дозволяють отримати непряму інформацію через інтерпретацію відповідей на абстрактні запитання або завдання. Основні види:

Асоціативні вправи: респонденту пропонується швидко відповісти на певні слова або фрази.

Завершення ситуації: респонденти доповнюють історії або пропонують розв’язання уявних ситуацій.

Конструювання ситуації: учасники створюють сценарій чи уявний розклад подій.

Експресивні методи: учасники висловлюють свою думку через творчість (малюнки, письмові твори).

Ці методи корисні для виявлення глибинних установок та емоцій, які респонденти можуть не усвідомлювати. Вони часто використовуються в маркетингових дослідженнях, психології та соціології.

Підбиваючи підсумки прямі та непрямі методи збору інформації є взаємодоповнюючими інструментами дослідження, кожен із яких має свої переваги та обмеження.

Прямі методи характеризуються прозорістю й відкритістю для респондентів. Вони дозволяють отримати точні та чіткі відповіді, спрямовані на безпосереднє розв’язання поставлених дослідником завдань. Прямі методи, такі як фокус-групи чи глибинні інтерв’ю, підходять для ситуацій, де довіра між дослідником і респондентом є ключовою. Їхнім недоліком є можливість впливу соціальної бажаності на відповіді учасників, а також певні обмеження у виявленні прихованих мотивів.

Непрямі методи дозволяють отримувати інформацію навіть за умов небажання респондентів співпрацювати відкрито. Завдяки засекреченим або проективним підходам, такі методи допомагають виявити приховані мотиви, установки чи упередження, які важко виявити за допомогою прямого запитання. Їхня головна перевага – можливість уникнути впливу соціального контролю чи упередженості у відповідях. Однак ці методи складніші в організації та інтерпретації даних.

Успішність дослідження залежить від правильного вибору методу, який базується на специфіці завдання. У багатьох випадках ефективним є поєднання прямих і непрямих підходів для комплексного аналізу досліджуваного явища.

У процесі проведення досліджень збір якісних даних часто виступає як підготовчий етап перед кількісним аналізом. Цей підхід дозволяє визначити ключові аспекти, які потребують подальшого дослідження, а також

формулювати точні питання для збору більш структурованої інформації. Аналіз якісних даних допомагає краще зрозуміти контекст, поведінкові особливості респондентів, їхні установки та пріоритети, що дозволяє адаптувати кількісні методи до реальних потреб дослідження.

Кількісні дослідження спрямовані на збір даних, представлених у числовій формі, які можна точно виміряти та піддати статистичному аналізу. Такі дані дозволяють отримувати узагальнені результати, ідентифікувати закономірності й робити передбачення. Для їхнього збору використовують опитування, анкетування, експерименти або автоматизовані системи моніторингу, що забезпечують репрезентативність та надійність результатів[6].

Комбінація якісного та кількісного підходів є ефективним інструментом для комплексного аналізу. Вона дозволяє глибше зрозуміти досліджуване явище й отримати результати, які не лише відображають загальні тенденції, але й пояснюють, чому вони виникають.

Спостереження, залежно від частоти та тривалості збору даних, поділяються на поточні, періодичні та одночасні, кожен з яких має свої особливості й застосовується залежно від мети дослідження.

Поточне спостережень передбачає систематичний і постійний збір даних у момент виникнення фактів. Такі спостереження дозволяють фіксувати динаміку подій, забезпечуючи актуальність інформації. Приклади включають реєстрацію народжуваності, смертності чи захворюваності в певних групах населення. Поточні спостереження мають перевагу у швидкості отримання результатів, що особливо важливо у випадках моніторингу кризових ситуацій, таких як спалахи епідемій чи економічні зрушення[7].

Регулярний, але не постійний збір даних є основою періодичних спостережень. Дані фіксуються через певні проміжки часу, наприклад, щоквартально або щорічно, що дозволяє аналізувати довгострокові тенденції. Такі методи використовуються для оцінки економічних показників, соціальних змін або освітніх процесів. Їхньою перевагою є зменшення витрат на проведення досліджень, порівняно з постійним спостереженням [7].

Одночасні спостереження фіксують стан явища в певний момент часу, який називається критичним. Вони відображають статичний зріз ситуації, без урахування змін у часі. Приклади включають переписи населення, реєстрацію пацієнтів, які звернулися до лікарні в конкретний день, чи хронометраж роботи працівників. Такі дані є основою для оцінки поточного стану явищ, де зміни відбуваються поступово [7].

Ефективний збір кількісних даних у бізнесі часто передбачає поєднання різних підходів для отримання максимальної точності й репрезентативності. Наприклад, дані про чисельність клієнтів або філій можна збирати за допомогою одночасних методів, тоді як інформацію про щоденний обіг або продажі – шляхом поточного обліку.

Основні методи збору даних:

Безпосередня реєстрація

Застосовується для отримання точних показників, таких як кількість продажів, тривалість обслуговування клієнтів чи обсяг виготовленої продукції. Інформація фіксується безпосередньо під час процесів за допомогою систем обліку або спеціальних програм.

Документальний облік

Використання внутрішніх документів, таких як фінансові звіти, накладні чи інвентаризаційні відомості, дозволяє аналізувати діяльність бізнесу. Цей метод забезпечує достовірність і деталізацію даних.

Викопіювання даних

Використовується для аналізу даних із різних джерел, наприклад, клієнтських баз чи звітів, що потребують подальшого структурування й перевірки.

Опитування

Збір інформації від клієнтів або партнерів може проводитися за допомогою таких підходів:

Експедиційний метод: дослідник самостійно спілкується з респондентами, уточнюючи деталі їхнього досвіду.

Кореспондентський метод: анкети або форми розсилаються клієнтам для заповнення й повернення.

Самореєстрація: клієнти самостійно надають дані через онлайн-форми чи програми.

Анкетування

Цей метод використовується для збору даних про задоволеність клієнтів, споживчі переваги чи загальні враження про послугу/продукт. Онлайн-анкети є популярним способом завдяки їхній зручності й швидкості обробки даних.

Використання технологій

Автоматизація процесів збору даних, наприклад через CRM-системи або інші цифрові платформи, значно підвищує точність і швидкість аналізу, мінімізуючи людські помилки. Поєднання цих методів допомагає бізнесу краще розуміти свою діяльність і приймати обґрунтовані рішення.

Загалом комбінація якісних і кількісних методів дозволяє бізнесу отримати повну картину реальності: від глибокого розуміння до узагальнених статистичних закономірностей. Це забезпечує точність і ефективність у плануванні стратегії.

1.2 Огляд сучасних інформаційних технологій

Інформаційні технології (IT) відіграють ключову роль у сучасному бізнесі та суспільстві. Вони стали основою цифрової трансформації, яка змінює традиційні моделі роботи, підвищує ефективність процесів і створює нові можливості для розвитку підприємств. Завдяки IT бізнес отримує інструменти для оптимізації операцій, покращення комунікації з клієнтами та зростання конкурентоспроможності.

Упродовж останніх років спостерігається швидкий розвиток таких напрямів, як автоматизація рутинних завдань, впровадження інтелектуальних

систем для аналізу даних та використання хмарних платформ, які дозволяють доступ до ресурсів і сервісів у будь-який час та з будь-якої точки світу.

ІТ забезпечують можливість гнучко адаптуватися до змін у ринкових умовах, реагувати на запити споживачів та ефективно управляти великими обсягами даних. Це сприяє побудові інноваційних бізнес-моделей, які стають вирішальним фактором у збереженні лідерських позицій на ринку.

Окрім бізнесу, ІТ трансформують соціальну сферу, освіту, медицину та навіть міжособистісну комунікацію, що робить їх незамінним інструментом у сучасному світі.

Тепер перейдемо та розберемо класифікації сучасних технологій до яких входять: хмарні обчислення(Cloud Computing), штучний інтелект(AI), машинне навчання(ML), великі дані(Big Data), інтернет речей (IoT), блокчейн.

Хмарні обчислення забезпечують доступ до обчислювальних ресурсів через інтернет. Завдяки цій технології підприємства можуть масштабувати інфраструктуру залежно від потреб, не інвестуючи в дорогу апаратну частину [8].

Особливості: зберігання даних, доступність у режимі 24/7, зменшення витрат на підтримку серверів.

Приклади платформ: Amazon Web Services (AWS), Microsoft Azure, Google Cloud.

Ці сервіси дозволяють розгортати бізнес-додатки, створювати резервні копії даних та обробляти великі масиви інформації.

Штучний інтелект і машинне навчання знаходять застосування в автоматизації процесів, аналізі великих даних і створенні інтелектуальних систем[9].

Застосування: рекомендаційні системи (як у Netflix або Amazon), чат-боти для обслуговування клієнтів, прогнозування ринкових тенденцій.

Переваги: економія часу, підвищення точності аналізу, зменшення людських помилок.

Обробка великих обсягів структурованої та неструктурованої інформації є важливою складовою сучасного бізнесу [10].

Основні функції: виявлення закономірностей, прогнозування поведінки клієнтів, персоналізація послуг.

Інструменти: Hadoop, Apache Spark.

Big Data дозволяють компаніям ухвалювати стратегічні рішення на основі глибокого аналізу даних.

Технології IoT забезпечують підключення пристроїв до мережі, що дозволяє автоматизувати процеси і покращувати їх контроль [11].

Приклади: розумні будинки, моніторинг складів у логістиці, автоматизація виробничих процесів.

Переваги: підвищення ефективності, зменшення витрат, створення інноваційних продуктів.

Технологія блокчейн використовується для створення безпечних і прозорих систем зберігання й передачі даних [12].

Сфери застосування: криптовалюти, управління ланцюгами постачань, захист прав інтелектуальної власності.

Особливості: децентралізація, стійкість до підробок, прозорість транзакцій.

Усі ці напрями є важливими елементами сучасної цифрової трансформації, відкриваючи нові можливості для бізнесу та суспільства. Вони не тільки оптимізують існуючі процеси, а й сприяють створенню принципово нових бізнес-моделей.

Сучасний бізнес використовує численні інноваційні платформи для управління, аналізу й автоматизації процесів. Розглянемо три ключові категорії:

CRM-системи (Customer Relationship Management)

Ці системи оптимізують взаємодію з клієнтами, аналізують їхні потреби та автоматизують продажі.

Основні можливості: збір даних про клієнтів, управління комунікаціями, створення прогнозів.

Популярні системи: Salesforce, Zoho CRM.

Переваги: підвищення рівня задоволеності клієнтів, розширення бази даних і зростання конверсії продажів.

ERP-системи (Enterprise Resource Planning)

ERP-системи інтегрують і синхронізують всі бізнес-процеси в межах однієї платформи.

Основні функції: управління фінансами, постачаннями, персоналом, складом.

Популярні системи: SAP, Oracle ERP Cloud.

Переваги: автоматизація рутинних завдань, скорочення витрат, масштабованість і контроль у реальному часі.

BI-платформи (Business Intelligence)

BI-системи аналізують великі масиви даних і перетворюють їх у зрозумілі аналітичні звіти.

Основні можливості: створення інтерактивних дашбордів, прогнозування тенденцій.

Популярні платформи: Power BI, Tableau.

Переваги: візуалізація складної інформації, швидкість ухвалення рішень, адаптація під різні бізнес-моделі.

Тепер більш детально розглянемо, які саме переваги дає впровадження інформаційних технологій для функціонування бізнесу.

IT автоматизують рутинні завдання, дозволяючи співробітникам зосередитися на стратегічних цілях. Наприклад, ERP-системи інтегрують управління фінансами, постачаннями та виробничими процесами, зменшуючи час на виконання задач і мінімізуючи помилки. Інструменти автоматизації (як-от роботизовані процеси, RPA) дозволяють обробляти великі обсяги даних без втручання людини.

Автоматизація процесів скорочує витрати на робочу силу та оптимізує використання ресурсів. Хмарні обчислення (Cloud Computing) знижують витрати на IT-інфраструктуру, дозволяючи компаніям орендувати необхідні

ресурси замість купівлі обладнання. Наприклад, впровадження CRM-систем дозволяє мінімізувати витрати на маркетинг і підвищити рентабельність за рахунок цільових кампаній.

Завдяки інструментам Business Intelligence (BI) та Big Data, компанії можуть аналізувати ринкові тренди та споживчі вподобання. Прогнозуючи попит, бізнес може уникати перевиробництва або нестачі товарів, що позитивно впливає на дохід. AI-рішення допомагають прогнозувати поведінку клієнтів, економічні коливання та конкурентні дії.

CRM-системи дозволяють персоналізувати комунікацію з клієнтами, аналізувати їхні вподобання та створювати персональні пропозиції. Наприклад, автоматичні відповіді через чат-боти забезпечують цілодобову підтримку клієнтів, зменшуючи час очікування. Усе це сприяє підвищенню задоволеності клієнтів і довгостроковій лояльності.

Попри численні переваги інформаційних технологій (IT), їхнє впровадження може супроводжуватися низкою проблем і викликів, які варто враховувати при плануванні цифрової трансформації.

Впровадження сучасних IT вимагає значних фінансових інвестицій. Наприклад, закупівля ліцензійного програмного забезпечення, оренда хмарних ресурсів чи інтеграція ERP-систем (як SAP чи Oracle) може коштувати компаніям мільйони доларів. Окрім початкових витрат, є витрати на оновлення, технічне обслуговування та підтримку інфраструктури. Це особливо критично для малого та середнього бізнесу, який має обмежені фінансові ресурси.

Успішне використання нових технологій вимагає, щоб співробітники мали відповідні знання й навички. Це створює необхідність у регулярному навчанні, яке може бути складним і тривалим процесом. Відсутність підготовки персоналу може призвести до низької ефективності впровадження або до помилок у роботі, що ставить під загрозу результативність бізнесу.

Широке використання інформаційних технологій підвищує ризики кібератак, витоку конфіденційних даних і збоїв у роботі систем. За даними Cisco, більшість компаній стикаються із загрозами фішингових атак, шкідливих

програм та компрометації облікових записів. Забезпечення безпеки вимагає інвестицій у системи захисту (наприклад, антивіруси, міжмережеві екрани, двофакторну автентифікацію) і регулярних перевірок вразливостей.

Виходячи з вище перерахованих переваг та недоліків, можна зробити такий висновок: переваги ІТ значно перевищують їхні недоліки, особливо для компаній, які прагнуть залишатися конкурентоспроможними в умовах цифрової економіки. Успіх впровадження залежить від ретельного планування та стратегії. Також варто зазначити, що сучасні інформаційні технології є не лише інструментом для підвищення ефективності бізнесу, а й основою для створення інноваційних продуктів і послуг. Вони допомагають компаніям бути гнучкими, реагувати на зміни ринку та задовольняти зростаючі очікування споживачів.

Проте для максимального використання переваг необхідно враховувати довгострокову стратегію розвитку, адаптовану до змін технологічного середовища, та забезпечувати безперервне навчання персоналу. Відтак, грамотне впровадження ІТ є ключовим фактором для досягнення сталого зростання та успіху бізнесу в сучасному світі.

1.3 Огляд систем для аналізу даних і методів.

Аналіз даних став ключовим компонентом успішного функціонування сучасного бізнесу, маркетингу, фінансів та багатьох інших сфер. Здатність використовувати дані для виявлення тенденцій, оцінки ризиків і прогнозування змін дозволяє компаніям адаптуватися до динамічного середовища та приймати обґрунтовані рішення.

Системи аналізу даних відіграють вирішальну роль у трансформації даних у корисну інформацію. Вони дають змогу компаніям оцінювати свої показники в реальному часі, оптимізувати ресурси та підвищувати ефективність управління.

Ринок систем аналізу даних демонструє стрімке зростання завдяки їхній доступності, ефективності та широкому спектру застосувань. Сучасні інструменти, такі як BI-платформи, Big Data та AI-рішення, інтегруються у бізнес-процеси, забезпечуючи конкурентні переваги. Зростання обсягу даних у цифрову епоху стимулює попит на такі системи, перетворюючи аналіз даних на обов'язкову складову успіху бізнесу.

Сучасні системи аналізу даних поділяються на кілька типів залежно від їхнього функціоналу, технологічного підходу та цільового використання. У кожній категорії існують популярні інструменти, які відрізняються за можливостями, гнучкістю та специфікою застосування. Для кращого розуміння наведемо опис категорій і порівняльні таблиці для ключових платформ.

BI-системи (Business Intelligence) – програмні рішення для збору, систематизації та наочного відображення даних з різних джерел. Інформація у потрібних розрізах відображається в одному вікні, що дозволяє швидко виявити вузькі шийки та знайти точки зростання для бізнесу.Наприклад Tableau, Power BI, QlikView. [13]

У таблиці 1.1 порівняємо ключові платформи для BI-системи.

Таблиця 1.1 – Порівняння платформи для BI-системи

Категорія	Tableau	Power BI	QlikView
Легкість використання	****	*****	***
Візуалізація даних	*****	****	****
Інтеграція з джерелами	****	*****	***
Розширена аналітика	****	***	***
Мобільна версія	****	*****	***

Висновок: Tableau ідеально підходить для глибокої візуалізації, тоді як Power BI пропонує кращу інтеграцію з іншими інструментами Microsoft.

Системи для аналізу великих даних (Big Data). Ці інструменти використовуються для обробки величезних масивів інформації в режимі реального часу. Приклади: Hadoop, Apache Spark.

У таблиці 1.2 порівняємо платформи для систем аналізу великих даних

Таблиця 1.2 – Порівняння платформ для систем аналізу великих даних

Категорія	Hadoop	Apache Spark
Обробка даних	*****	*****
Швидкість	***	****
Інтеграція	****	****
Підтримка реального часу	**	*****
Масштабованість	*****	****

Висновок: Hadoop підходить для тривалого зберігання великих даних, тоді як Apache Spark краще для обробки даних у реальному часі.

Платформи машинного навчання застосовуються для створення прогнозів і аналізу даних за допомогою алгоритмів.

Приклади: TensorFlow, Scikit-learn.

У таблиці 1.3 порівняємо платформи для машинного навчання

Таблиця 1.3 – Порівняння платформ для машинного навчання

Категорія	TensorFlow	Scikit-learn
Легкість використання	***	****
Швидкість навчання	*****	***
Розширення	*****	****
Інтеграція	****	****
Підтримка бібліотек	*****	*****

Висновок: TensorFlow забезпечує кращу продуктивність для великих моделей, тоді як Scikit-learn орієнтований на простіші та швидші проекти.

Інструменти статистичного аналізу призначені для роботи зі статистикою та аналізу результатів опитувань, досліджень.

Приклади: SPSS, Stata.

У таблиці 1.4 порівняємо інструменти статистичного аналізу

Таблиця 1.4 – Порівняння інструментів статистичного аналізу

Категорія	SPSS	Stata
Легкість використання	****	****
Швидкість аналізу	****	***
Розширені моделі	****	*****
Візуалізація результатів	****	****

Висновок: SPSS підходить для широкого кола користувачів, тоді як Stata забезпечує кращі можливості для побудови економетричних моделей.

Хмарні платформи для аналізу даних дозволяють обробляти дані онлайн без необхідності в локальних потужностях.

Приклади: Google BigQuery, Amazon Redshift.

У таблиці 1.5 порівняємо хмарні платформи для аналізу даних

Таблиця 1.5 – Порівняння хмарних платформ для аналізу даних

Категорія	Google BigQuery	Amazon Redshift
Швидкість обробки	*****	****
Простота інтеграції	*****	****
Масштабованість	*****	*****
Вартість	****	****

Висновок: Google BigQuery краще підходить для аналітики з використанням Google Workspace, а Amazon Redshift є оптимальним для інтеграції з AWS.

Цей аналіз демонструє, як різні платформи відповідають потребам користувачів, дозволяючи їм обирати рішення залежно від цілей та завдань бізнесу.

Далі розглянемо методи аналізу даних, застосовані в системах, а саме: описовий, діагностичний та прогнозний аналіз, машинне навчання і візуалізація даних.

Описовий аналіз використовується для виявлення базових закономірностей, структури та статистичних характеристик даних. Він відповідає на запитання "що сталося?". Інструменти, як-от Power BI або Tableau, активно застосовують описові показники для побудови дашбордів [14].

Діагностичний аналіз спрямований на вивчення причин явищ чи подій. Наприклад, аналізуючи зниження продажів, системи (QlikView) можуть допомогти знайти ключові фактори, що призвели до такого результату [15].

Прогнозний аналіз використовує моделі, які дозволяють передбачати майбутні тренди на основі історичних даних. Наприклад, у TensorFlow чи Scikit-learn прогнозування застосовують для визначення майбутнього попиту, ризиків чи фінансових показників [16].

Машинне навчання використовується для автоматизації складних завдань, таких як кластеризація клієнтів, класифікація запитів чи регресійний аналіз для визначення зв'язків між змінними. TensorFlow та Apache Spark є популярними платформами для цих завдань.

Візуалізація даних забезпечує зрозуміле представлення даних у вигляді графіків, діаграм і дашбордів. Tableau, Power BI і QlikView пропонують інтуїтивні інструменти для створення візуалізацій, що допомагають краще зрозуміти тренди та залежності [17].

Ось детальні приклади застосування цих методів у різних галузях:

В e-commerce: прогнозний аналіз для передбачення покупок.

Задача: Прогнозування, які товари можуть зацікавити покупців, щоб запропонувати їм релевантні рекомендації або створити персоналізовані пропозиції.

Метод: Використання алгоритмів машинного навчання, таких як регресія або дерева рішень, для аналізу історії покупок клієнтів, їх поведінки на сайті (клікання, перегляди) і демографічних даних. Це дозволяє передбачити, які товари, ймовірно, будуть придбані в майбутньому.

Приклад: Платформи як Amazon чи eBay використовують ці методи для персоналізації пропозицій на основі даних про попередні покупки і перегляди.

У фінансах: класифікація кредитоспроможності клієнтів.

Задача: Оцінка кредитоспроможності клієнтів для надання позик або кредитних карток.

Метод: Класифікація на основі методів машинного навчання, таких як логістична регресія або дерева рішень. Алгоритми аналізують історію кредитних операцій, доходи, витрати, наявність боргів і інші фінансові показники для оцінки ймовірності того, що клієнт поверне кредит.

Приклад: Багато банків і фінансових установ використовують ці методи для автоматичного схвалення або відхилення заявок на кредит.

У маркетингу: кластеризація споживачів для персоналізації кампаній.

Задача: Сегментація споживачів для створення персоналізованих маркетингових кампаній, які сприяють збільшенню продажів або залученню нових клієнтів.

Метод: Кластеризація, наприклад, за допомогою алгоритму K-середніх або ієрархічної кластеризації. Цей метод дозволяє групувати споживачів за схожістю у поведінці, інтересах, демографічних характеристиках або покупках.

Приклад: Онлайн-магазини можуть створювати персоналізовані email-розсилки або рекламні кампанії, націлені на конкретні групи споживачів, наприклад, на основі їхніх покупок, щоб збільшити ймовірність покупки.

Ці методи, комбінуючись, створюють потужний інструментарій для оптимізації бізнес-рішень і покращення операційної ефективності.

Тепер роздивимося як саме ці методи використовуються у бізнес сферах

BI платформи надають компаніям інструменти для збору, аналізу та візуалізації великих обсягів даних, що дозволяє приймати обґрунтовані бізнес-

рішення. Вони мають великий вплив на маркетингові кампанії завдяки своїй здатності:

Аналізувати ефективність кампаній: BI-системи дозволяють моніторити такі показники, як конверсії, доходи, витрати, залученість аудиторії, що дає змогу визначити, які маркетингові стратегії працюють найкраще.

Побудова звітності в реальному часі: Візуалізація даних у вигляді дашбордів дозволяє маркетологам оперативно виявляти проблеми або можливості, коригуючи стратегії в реальному часі.

Цільова аудиторія та персоналізація: Аналізуючи історію покупок та демографічні дані клієнтів, BI-платформи допомагають точно налаштувати рекламні кампанії під конкретні групи споживачів.

Big Data – це великий обсяг структурованих і неструктурованих даних, які можуть бути використані для глибокого аналізу та розуміння споживачів. Основні напрямки використання:

Аналіз споживчих уподобань: Використовуючи дані з різних джерел – онлайн-покупок, відгуків, соціальних мереж – компанії можуть вивчати уподобання і звички споживачів для створення персоналізованих пропозицій.

Прогнозування майбутніх потреб: Завдяки аналізу історії покупок та поведінки клієнтів, можна передбачити, які продукти чи послуги будуть популярні в майбутньому, і адаптувати стратегії.

Покращення лояльності клієнтів: Використовуючи дані з Big Data, компанії можуть створювати персоналізовані програми лояльності, що стимулюють повторні покупки і підвищують задоволення клієнтів.

Приклад: Netflix і Spotify використовують великі дані для аналізу вподобань користувачів і надають рекомендації, що можуть спонукати їх до підписки на нові серіали чи плейлисти.

Машинне навчання (ML) є потужним інструментом для аналізу та прогнозування майбутніх тенденцій, зокрема в прогнозуванні продажів та управлінні ризиками:

Прогнозування продажів: Машинне навчання дозволяє побудувати моделі, які прогнозують обсяги продажів на основі історичних даних, сезонних змін, маркетингових активностей та інших факторів. Це дозволяє бізнесам приймати стратегічні рішення щодо запасів, ціноутворення і розподілу ресурсів.

Приклад: Алгоритми можуть прогнозувати, які товари будуть найбільш популярні в найближчі місяці, що дозволяє компаніям оптимізувати свої складські запаси і знизити витрати на зберігання.

Управління ризиками: Машинне навчання може автоматично аналізувати дані, щоб виявляти ризики, наприклад, у фінансових операціях або в діяльності компанії. Це дозволяє швидко реагувати на потенційні проблеми, такі як шахрайство чи неплатоспроможність клієнтів.

Приклад: У банківській сфері алгоритми машинного навчання використовуються для оцінки кредитоспроможності клієнтів і виявлення аномальних транзакцій, що можуть вказувати на можливе шахрайство.

У великих онлайн-роздрібних мережах, таких як Amazon або Walmart, аналіз великих даних, BI-платформи та алгоритми машинного навчання використовуються для:

Створення персоналізованих пропозицій для кожного користувача, аналізуючи його поведінку та історію покупок.

Прогнозування попиту на товари, щоб уникнути дефіциту або надлишку товарів на складах.

Моніторингу і управлінню ризиками, такими як неплатежі або повернення товарів.

У сучасному бізнес-середовищі, яке характеризується швидкими змінами та високою конкуренцією, ефективне використання даних стає ключем до успіху. Системи аналізу даних забезпечують компаніям можливість не лише зібрати та організувати інформацію, але й перетворити її на цінні інсайти, які допомагають ухвалювати стратегічні рішення. Використання таких систем дозволяє суттєво економити час і ресурси, підвищувати точність рішень і

оптимізувати бізнес-процеси. Тому далі розглянемо, як ці технології сприяють покращенню ефективності бізнесу через автоматизацію, знижуючи витрати та підвищуючи продуктивність на всіх рівнях організації.

Економія часу та ресурсів

Використання систем аналізу даних дозволяє компаніям значно зекономити час та ресурси в кількох аспектах:

Автоматизація процесів збору та обробки даних: Завдяки впровадженню систем BI, Big Data або машинного навчання, зменшується потреба в ручному зборі і обробці даних. Це дозволяє уникнути людських помилок, прискорює обробку великих обсягів інформації і дозволяє фахівцям зосередитись на важливіших завданнях.

Швидкість доступу до інформації: Системи аналізу даних забезпечують оперативний доступ до потрібної інформації та наочні звіти у реальному часі. Це скорочує час, який необхідно витратити на пошук і аналіз даних, дозволяючи приймати швидші та ефективніші рішення.

Оптимізація витрат: Аналізуючи дані щодо витрат, системи можуть виявити неефективні процеси та надлишкові витрати, що дає змогу знижувати витрати на операційну діяльність і забезпечувати більш ефективне використання ресурсів.

Підвищення точності ухвалення рішень

Системи аналізу даних допомагають ухвалювати більш точні та обґрунтовані рішення завдяки:

Аналізу великих обсягів даних: В системах аналізу даних можна обробляти величезну кількість різноманітних джерел інформації (фінансові дані, поведінка клієнтів, ринкові умови, тощо). Це дозволяє робити висновки на основі більш повної і точнішої картини.

Прогнозування майбутніх тенденцій: За допомогою алгоритмів машинного навчання та статистичного аналізу можна прогнозувати майбутні події, наприклад, попит на товар, зміни на ринку чи фінансові результати. Це

дає можливість приймати рішення на основі передбачених результатів, а не лише на поточних даних.

Мінімізація людської помилки: Системи автоматично обробляють та аналізують інформацію, знижуючи ймовірність помилок, які можуть виникнути при ручному введенні або інтерпретації даних.

Збільшення ефективності бізнесу через автоматизацію

Системи аналізу даних дозволяють автоматизувати багато бізнес-процесів, що веде до значного підвищення ефективності:

Автоматизація звітності: За допомогою BI-платформ можна автоматично генерувати звіти і дашборди, які відображають важливі показники ефективності (KPI). Це знижує час, витрачений на створення звітів, і дозволяє керівникам швидше реагувати на зміни.

Інтелектуальна автоматизація: Використання машинного навчання дає змогу автоматично оптимізувати робочі процеси, наприклад, розподіл ресурсів, планування виробництва чи управління запасами. Це дозволяє знижувати витрати та підвищувати точність операцій.

Оптимізація маркетингових кампаній: Аналіз даних з різних каналів дозволяє автоматично налаштовувати рекламні кампанії, визначати найбільш ефективні стратегії і навіть персоналізувати пропозиції для клієнтів, що збільшує їх залучення та конверсію.

Покращення логістики: За допомогою аналізу даних можна прогнозувати попит, оптимізувати запаси, автоматизувати маршрути доставки і зменшити витрати на транспортування товарів.

Приклад використання переваг:

Компанія Amazon використовує системи аналізу даних на всіх етапах бізнесу:

Вони автоматично аналізують поведінку користувачів, щоб рекомендувати персоналізовані продукти, що підвищує конверсії та збільшує продажі.

Машинне навчання прогнозує попит на товари і допомагає оптимізувати запаси, що дозволяє знизити витрати на складування та транспортування.

Використання BI-систем для моніторингу фінансових показників і ефективності кампаній дозволяє швидко коригувати стратегії і приймати більш точні бізнес-рішення.

Таким чином, використання систем аналізу даних дозволяє компаніям не тільки значно заощаджувати час та ресурси, але й приймати більш обґрунтовані, точні рішення, що сприяють підвищенню ефективності їхнього бізнесу.

1.4 Можливі альтернативи для оптимізації збору та аналізу даних

Збір і аналіз даних є ключовими елементами сучасного бізнесу. Однак традиційні методи можуть бути недостатньо ефективними через обмеження у швидкості, масштабованості та адаптивності. Тому з’являються нові альтернативи, які дозволяють оптимізувати ці процеси.

Зважаючи на зростання обсягів інформації та важливість швидкого доступу до релевантних даних, автоматизація стає ключовим інструментом для досягнення конкурентних переваг. Використання сучасних технологій, таких як роботизована автоматизація процесів (RPA), спеціалізовані API та інструменти для автоматичного витягнення даних, дозволяє бізнесу не лише зекономити ресурси, а й підвищити ефективність прийняття рішень. У цьому контексті досліджуються основні методи та технології для поліпшення збору й обробки інформації.

Автоматизація процесів – це ефективний спосіб оптимізації збору даних, який дозволяє зменшити витрати часу та ресурсів. Завдяки сучасним технологіям, зокрема RPA (Robotic Process Automation), можна значно підвищити точність і швидкість збору інформації [18].

RPA дозволяє автоматизувати повторювані завдання, наприклад:

витягнення даних із різних систем або веб-ресурсів;

заповнення таблиць або баз даних;

оновлення звітів у реальному часі.

Ці інструменти імітують дії людини, працюючи з інтерфейсами програмного забезпечення без необхідності прямої інтеграції.

Автоматичне витягнення даних

Застосування інструментів автоматизованого витягнення, таких як web scraping (наприклад, BeautifulSoup або Scrapy), дозволяє зібрати релевантну інформацію з веб-ресурсів. Це може бути: збір контактних даних клієнтів, аналіз цін конкурентів та витяг трендових даних у реальному часі.

API (Application Programming Interface) спрощують доступ до даних у зовнішніх системах або платформах.

Приклади:

Google Maps API – для збору геолокаційних даних.

Twitter API – для аналізу соціальних мереж.

CRM API – для автоматичного збору даних клієнтів.

Інтеграція API дозволяє не лише отримувати дані, але й автоматично їх обробляти та передавати у потрібний формат.

Переваги автоматизації збору даних

Економія часу – процеси, які займали години, виконуються за хвилини.

Точність – мінімізація людських помилок.

Масштабованість – можливість обробляти великі обсяги даних.

Реальний час – автоматизація дозволяє оновлювати дані в режимі реального часу, що критично для конкурентного бізнесу.

Автоматизація процесів значно підвищує ефективність збору даних, але для повного розкриття потенціалу цих даних важливо забезпечити їхнє зберігання та доступність у масштабованому середовищі. Саме тут на допомогу приходять хмарні платформи, які стали стандартом для сучасного бізнесу. Завдяки таким рішенням, як AWS, Google Cloud та Microsoft Azure, компанії

отримують можливість безперервного доступу до даних, масштабування ресурсів відповідно до потреб і скорочення витрат на інфраструктуру.

До основних можливостей хмарних платформ входять: збір і зберігання даних, швидкий доступ до даних та масштабованість

Збір і зберігання даних

Хмарні платформи дозволяють зберігати великі обсяги структурованих і неструктурованих даних у безпечному середовищі. Дані з різних джерел інтегруються та систематизуються для подальшого аналізу.

AWS: пропонує Amazon S3 для зберігання даних із високою доступністю.

Google Cloud: використовує BigQuery для аналізу великих масивів даних.

Microsoft Azure: забезпечує інтеграцію з іншими бізнес-сервісами через Azure Storage.

Швидкий доступ до даних

Однією з ключових переваг хмарних платформ є можливість доступу до даних із будь-якого місця та пристрою. Це особливо важливо для компаній із розподіленими командами чи міжнародним охопленням.

Масштабованість

Хмарні сервіси дозволяють автоматично збільшувати чи зменшувати ресурси залежно від навантаження. Наприклад, в пікові періоди (розпродажі, акції) компанії можуть збільшити потужності, а після завершення повернутися до базових налаштувань.

Хмарні платформи відкривають широкі можливості для бізнесу, дозволяючи зберігати, обробляти та аналізувати дані в єдиному середовищі з високим рівнем доступності та масштабованості.

Штучний інтелект (AI) став важливим інструментом для оптимізації процесів збору, очищення та аналізу даних. Завдяки його можливостям компанії можуть обробляти великі обсяги інформації швидко, точно та з мінімальним людським втручанням.

Автоматичне очищення, фільтрація та категоризація даних

Штучний інтелект дозволяє:

Очищення даних: Виявлення й усунення дублікатів, пропущених значень або некоректних записів.

AI-системи, наприклад, на основі TensorFlow, автоматично аналізують структуру даних і забезпечують їхню цілісність.

Фільтрація даних: Відсівання нерелевантної інформації на основі заданих критеріїв. Наприклад, у фінансових операціях можна автоматично ідентифікувати аномалії.

Категоризація: AI використовує алгоритми кластеризації для поділу даних на групи. Це корисно в маркетингу для сегментації клієнтів.

Використання чат-ботів і віртуальних асистентів

AI-технології, такі як чат-боти та віртуальні асистенти, активно застосовуються для збору даних про клієнтів.

Чат-боти: автоматично збирають інформацію через діалог із користувачами. Наприклад, чат-боти в e-commerce можуть зберігати дані про запити, вподобання та потреби клієнтів.

Віртуальні асистенти: такі як Alexa чи Google Assistant, збирають голосові команди для подальшого аналізу. Вони допомагають компаніям зрозуміти, як покращити свої сервіси.

Переваги використання AI

Швидкість: AI обробляє інформацію в десятки разів швидше, ніж людина.

Масштабованість: здатність аналізувати великі обсяги даних одночасно.

Точність: мінімізує людські помилки.

AI у зборі та аналізі даних стає ключовим фактором для підвищення ефективності бізнесу. Завдяки його впровадженню компанії можуть не лише автоматизувати рутинні завдання, а й вивчати глибокі закономірності у своїй діяльності.

Мобільні додатки стають все більш популярним інструментом для збору даних безпосередньо від користувачів. Вони забезпечують зручний спосіб інтерактивної взаємодії, дозволяючи компаніям отримувати актуальну інформацію в реальному часі.

Збір даних від користувачів

Анкети та опитування: Мобільні додатки дозволяють швидко збирати відповіді користувачів через інтерактивні форми. Це зручно для маркетингових досліджень, тестування продуктів або оцінки клієнтської задоволеності.

Відстеження активності: Додатки можуть автоматично фіксувати дії користувачів, наприклад, частоту відвідувань, тривалість сеансів чи геолокацію.

Розширення функціональності через інтеграцію

Підключення до аналітичних платформ: Інтеграція мобільних додатків із сервісами, такими як Google Analytics, Firebase або Mixpanel, забезпечує деталізований аналіз зібраних даних.

Інтеграція з CRM-системами: Дозволяє автоматично передавати дані користувачів до клієнтських баз, оптимізуючи управління взаємовідносинами з клієнтами.

Інтеграція з IoT: Мобільні додатки можуть працювати в парі з пристроями IoT, наприклад, для моніторингу даних із сенсорів чи фітнес-трекерів.

Приклади застосування

Ритейл: Збирання відгуків про покупки або поведінку користувачів у магазинах.

Медицина: Моніторинг симптомів пацієнтів через спеціалізовані додатки.

Транспорт: Відстеження маршрутів та отримання зворотного зв'язку від пасажирів.

Open Source інструменти є доступними та гнучкими рішеннями, які дозволяють компаніям економити бюджет на ліцензуванні та адаптувати програмне забезпечення до власних потреб.

Основні переваги Open Source

Економія бюджету: Відсутність витрат на ліцензії робить ці інструменти ідеальними для малого та середнього бізнесу.

Кастомізація: Код із відкритим доступом дозволяє модифікувати функціонал відповідно до специфічних потреб компанії.

Широка спільнота підтримки: Завдяки великій кількості користувачів та розробників доступні регулярні оновлення й виправлення помилок.

Популярні Open Source інструменти

RapidMiner: Платформа для аналізу даних із можливостями машинного навчання. Застосовується для класифікації, регресії, кластеризації.

KNIME: Інструмент для побудови аналітичних процесів із простим drag-and-drop інтерфейсом. Ідеальний для аналізу великих обсягів даних та інтеграції з іншими платформами.

Apache Superset: Інструмент для візуалізації даних. Підтримує інтеграцію з популярними базами даних та дозволяє створювати інформативні дашборди.

Приклади використання Open Source інструментів

Маркетинг: Аналіз клієнтської поведінки.

Фінанси: Побудова прогнозів для управління ризиками.

Логістика: Оптимізація маршрутів і планування постачань.

У таблиці 1.6 підсумуємо їхні переваги, щоб показати, які методи найбільш ефективні залежно від потреб бізнесу.

Таблиця 1.6 – Найбільш ефективні методи залежно від потреб бізнесу

Категорія	Переваги	Недоліки	Найбільш ефективно для
Хмарні платформи	Масштабованість, зручність, безпека, інтеграція з іншими сервісами.	Залежність від провайдера, потенційні витрати на зберігання великих обсягів даних.	Організації, які працюють із розподіленими командами або мають значні обсяги інформації.
Штучний інтелект	Очищення, категоризація, прогнозування, автоматизація аналізу.	Висока вартість впровадження, потреба в експертних знаннях для налаштування алгоритмів.	Компанії, які працюють із великими даними або потребують персоналізації для клієнтів.

Продовження таблиці 1.6

Розподілені системи IoT	Дані в реальному часі, моніторинг, оптимізація процесів.	Високі витрати на впровадження сенсорів і мережі, залежність від стабільності передачі даних.	Виробництво, логістика, енергетика, де потрібен реальний час для моніторингу процесів.
Мобільні додатки	Зручність для користувачів, інтеграція з аналітикою, швидкість збору.	Потреба в розробці, технічній підтримці, ризики втрати даних через мобільні пристрої.	Ритейл, маркетинг, сфери з великим залученням клієнтів або збиранням зворотного зв’язку.
Open Source інструменти	Економія бюджету, гнучкість, широка спільнота підтримки.	Потреба у кваліфікованих фахівцях для кастомізації, менша підтримка, ніж у комерційних рішень.	Компанії зі строго обмеженим бюджетом або потребами в адаптації програмного забезпечення.

Сучасний бізнес має широкий вибір технологій для оптимізації збору та аналізу даних, кожна з яких має унікальні переваги й обмеження.

Автоматизація процесів (наприклад, RPA) забезпечує швидкість і точність, що особливо корисно для великих компаній із рутинними завданнями. Хмарні платформи, такі як AWS та Google Cloud, відкривають доступ до даних із будь-якого місця, пропонуючи масштабованість і безпеку. Застосування штучного інтелекту дозволяє не лише автоматизувати аналіз, а й прогнозувати поведінку клієнтів, підвищуючи якість управлінських рішень.

Розподілені системи, засновані на IoT, є ідеальними для реального часу моніторингу процесів, таких як логістика чи виробництво. Мобільні додатки сприяють прямій взаємодії з клієнтами, швидко збираючи зворотний зв'язок і доповнюючи його аналітичними платформами.

Open Source інструменти, як-от KNIME або Apache Superset, надають економічно вигідні рішення, які дозволяють компаніям з обмеженим бюджетом залишатися конкурентоспроможними, водночас зберігаючи гнучкість для кастомізації.

Таким чином, вибір оптимального інструменту залежить від специфіки бізнесу, його бюджету та стратегічних завдань. Поєднання декількох технологій дозволяє досягти максимальної ефективності й створити основу для сталого розвитку в умовах цифрової економіки.

Висновок до розділу 1

У розділі проведено всебічний аналіз предметної області, який охоплює методи збору, обробки та аналізу даних, сучасні інформаційні технології та можливості їхнього впровадження в бізнес. Якісні та кількісні методи збору даних показали свою ефективність у різних контекстах, забезпечуючи глибоке розуміння ринку, поведінки клієнтів та ключових показників ефективності.

Сучасні інформаційні технології, такі як BI-системи, хмарні платформи, системи для аналізу великих даних і платформи машинного навчання, забезпечують можливість роботи з великими обсягами інформації, підвищують точність аналізу та спрощують ухвалення стратегічних рішень. Методи аналізу даних, включаючи описовий, діагностичний, прогнозний аналіз і візуалізацію, дозволяють не лише оцінювати поточний стан, а й прогнозувати майбутні тенденції.

Розглянуто альтернативні підходи для оптимізації збору й аналізу даних. Використання автоматизації процесів, хмарних платформ, штучного інтелекту, мобільних додатків і інструментів із відкритим кодом відкриває нові перспективи для бізнесу, сприяючи зменшенню витрат і підвищенню гнучкості. Кожна з альтернатив має свої переваги та обмеження, що дозволяє адаптувати рішення під конкретні завдання компанії.

Розділ підкреслює важливість інтеграції новітніх технологій та підходів для покращення бізнес-процесів і підвищення конкурентоспроможності. Використання таких інструментів забезпечує компаніям можливість швидко адаптуватися до змін ринку, знижувати операційні витрати та підвищувати ефективність прийняття рішень.

2 РОЗРОБКА МЕТОДІВ ЗБОРУ ТА АНАЛІЗУ ДАНИХ

2.1 Постановка технічного завдання

Мета програми – створити систему, яка автоматизує процеси збору, обробки, аналізу та візуалізації даних для бізнес-задач. Програма повинна бути орієнтована на підприємства, які потребують ефективної обробки великих обсягів інформації для прийняття обґрунтованих рішень. Метою є не тільки демонстрація можливостей технологій, але й забезпечення реальної вигоди для бізнесу завдяки скороченню часу на ручну обробку даних, підвищенню точності прогнозів і покращенню ефективності роботи.

Цілі програми:

автоматизація збору даних із різних джерел, включаючи веб-ресурси, API, файли;

покращення якості даних через очищення та фільтрацію;

інтуїтивна візуалізація результатів, що допомагає бізнесу швидко приймати рішення;

генерація звітів для аналізу ключових показників.

Функціональні вимоги:

Збір даних з різних джерел (API, веб-ресурси):

Опис вимоги: Програма повинна бути здатною збирати дані з різних джерел, щоб забезпечити всеохоплюючий аналіз для бізнесу. Дані можуть надходити як з відкритих API (наприклад, фінансові дані, метеорологічні дані, інформація про курси валют), так і з веб-сайтів компаній (наприклад, ціни конкурентів, відгуки клієнтів).

Особливості реалізації:

API інтеграція: Використання стандартних бібліотек для роботи з API (наприклад, requests для Python,). Програма повинна підтримувати різні формати даних, як JSON, XML тощо;

веб-скрейпінг: Використання бібліотек для збору даних із веб-сторінок (BeautifulSoup, Scrapy, Selenium), що дозволить автоматично збирати дані з різноманітних веб-ресурсів. Це важливо для збору даних, які не надаються через API;

частота збору даних: Залежно від бізнес-задач, можна налаштувати збір даних за певним графіком (наприклад, щогодини або щодня).

Фільтрація та очищення даних:

Опис вимоги: Після збору даних необхідно провести їх очищення і фільтрацію, щоб забезпечити їх якість для подальшого аналізу. Це включає в себе видалення помилкових, дубльованих або неповних записів.

Особливості реалізації:

створення алгоритмів, які відсіюють нерелевантні дані (наприклад, видалення записів за певними критеріями: дата, значення, регіон і т.д.);

реалізація функцій для обробки пропущених значень, виправлення неточностей, нормалізація форматів даних (наприклад, приведення всіх дат до єдиного формату);

автоматична перевірка на дублікати та їх усунення (наприклад, за допомогою алгоритмів порівняння рядків або хешування);

створення журналу очищення даних для моніторингу процесу та виявлення проблем на етапі обробки.

Візуалізація результатів аналізу:

Опис вимоги: Візуалізація даних має бути зрозумілою і легкою для користувачів, щоб допомогти приймати швидкі і правильні бізнес-рішення на основі аналізу. Візуалізація включає графіки, діаграми, таблиці та інтерактивні елементи.

Особливості реалізації:

використання бібліотек, таких як Plotly, D3.js для створення інтерактивних графіків, що дозволяють користувачу взаємодіяти з даними (наприклад, фільтрувати за часом, категоріями);

виведення результатів у вигляді таблиць або інтерактивних дашбордів. Це дозволяє користувачам отримувати більше деталей про дані;

створення інтерактивних дашбордів (наприклад, у Power BI або Tableau), де користувачі можуть переглядати ключові показники та аналізувати їх у реальному часі.

Генерація звітів для прийняття рішень

Опис вимоги: Система повинна надавати можливість генерувати звіти, що базуються на проаналізованих даних, для покращення прийняття управлінських рішень.

Особливості реалізації:

автоматизація звітів: Можливість автоматично генерувати звіти в різних форматах (PDF, Excel, HTML), які містять основні метрики та візуалізації;

налаштування звітів: Користувач повинен мати можливість налаштовувати, які дані та метрики будуть включені в звіт;

програма має підтримувати налаштування для автоматичної генерації звітів на певну дату або через певний інтервал часу (наприклад, щомісячно).

Нефункціональні вимоги

Легка інтеграція з CRM-системами:

Опис вимоги: Програма повинна підтримувати інтеграцію з популярними CRM-системами (наприклад, Salesforce, HubSpot, Zoho), щоб зібрані дані про клієнтів, транзакції та інші бізнес-метрики можна було використовувати в аналізі.

Особливості реалізації:

розробка модулів для підключення до CRM-систем за допомогою API;

автоматичне оновлення даних із CRM-системи без потреби ручного втручання;

можливість додавання до програми аналітичних функцій, які використовують дані з CRM для сегментації клієнтів, прогнозування попиту тощо.

Висока продуктивність і масштабованість

Опис вимоги: Програма повинна обробляти великі обсяги даних за короткий час і бути здатною масштабуватися з ростом обсягів інформації.

Особливості реалізації:

використання ефективних алгоритмів для обробки даних і зменшення часу виконання (наприклад, багатопоточність або асинхронні операції для збору і обробки даних);

вибір технологій, які дозволяють масштабувати систему з часом, наприклад, за допомогою хмарних обчислень (AWS, Google Cloud) або розподілених систем для зберігання та обробки даних.

Безпека даних

Опис вимоги: Оскільки програма працюватиме з чутливими бізнес-даними, важливо забезпечити їх безпеку та конфіденційність.

Особливості реалізації:

використання сучасних методів шифрування для захисту даних як під час зберігання, так і під час передачі (наприклад, TLS/SSL для API);

розробка системи контролю доступу для визначення, хто може переглядати, редагувати або видаляти дані;

регулярне створення резервних копій важливих даних для уникнення їх втрати в разі непередбачених ситуацій.

Розробка програми для автоматизації збору, обробки та аналізу даних для бізнесу повинна включати низку важливих функціональних вимог, що забезпечать ефективне управління інформацією. Основними завданнями є інтеграція з різними джерелами даних (API, веб-ресурси), очищення та фільтрація отриманих даних для підвищення їх точності, а також створення інтуїтивно зрозумілих візуалізацій результатів аналізу. Програма повинна автоматизувати генерацію звітів, що дозволяють бізнесу швидко приймати обґрунтовані рішення. Ключовим аспектом є забезпечення продуктивності системи, здатної обробляти великі обсяги даних, а також її здатність масштабуватися залежно від зростання потреб бізнесу.

2.2 Інструменти та середовище розробки

Мова програмування: Python

Для розробки програми була обрана Python, і це рішення обґрунтоване кількома ключовими перевагами цієї мови:

Простота та читабельність: Python славиться своєю простотою і зрозумілістю коду, що робить його ідеальним вибором для швидкої розробки програмного забезпечення. Це дозволяє зосередитись на розв'язанні задач, а не на технічних аспектах мови програмування.

Багатий набір бібліотек: Python має великий набір бібліотек для аналізу даних, статистики, машинного навчання та візуалізації. Зокрема, для цього проекту будуть використовуватися:

pandas – для обробки та маніпулювання даними;

NumPy – для виконання числових операцій;

scikit-learn – для алгоритмів машинного навчання;

matplotlib та seaborn – для візуалізації результатів аналізу;

requests – для роботи з API.

Інтеграція з іншими технологіями: Python має велику підтримку інтеграції з іншими мовами та технологіями, такими як SQL, веб-сервіси (API) і різні хмарні платформи, що дозволяє створювати масштабовані та надійні програми.

Масштабованість та продуктивність: Завдяки бібліотекам, таким як Dask та pandas, Python дозволяє ефективно працювати з великими обсягами даних, а також здійснювати паралельні обчислення та розподілену обробку.

Підтримка в бізнесі та наукових дослідженнях: Python активно використовують у галузях, де потрібно обробляти великі масиви даних, створювати моделі машинного навчання та автоматизувати різноманітні бізнес-процеси.

Інтегроване середовище розробки (IDE):

Для розробки на Python буде використовуватись PyCharm – одне з найпопулярніших середовищ розробки для Python. Обрання цього IDE обумовлене такими перевагами:

Потужний набір інструментів: PyCharm пропонує розширену підтримку для розробки Python-додатків, включаючи вбудовані інструменти для налагодження коду, тестування, інтеграції з системами контролю версій (Git) і вбудовану документацію.

Інтелектуальне автозавершення та рефакторинг: PyCharm забезпечує функції автозавершення коду та інтелектуального пошуку, що значно прискорює розробку, підвищує продуктивність і знижує кількість помилок.

Інтеграція з бібліотеками та фреймворками: PyCharm підтримує безперешкодну інтеграцію з популярними бібліотеками Python, такими як pandas, NumPy, scikit-learn та іншими, що робить його ідеальним для задач, пов'язаних з аналізом даних.

Тестування та деплоймент: PyCharm має вбудовану підтримку для написання юніт-тестів, інтеграції з Jenkins і Docker, що дозволяє автоматизувати процеси тестування та розгортання програми.

Підтримка веб-розробки: Якщо в рамках проекту виникне необхідність розробки веб-інтерфейсу для візуалізації результатів аналізу, PyCharm підтримує розробку веб-додатків на Django або Flask, що полегшує інтеграцію з іншими веб-технологіями.

Обрання Python як основної мови програмування для цього проекту зумовлене його зручністю, потужними інструментами для аналізу даних та великою підтримкою в спільноті. Використання PyCharm як IDE забезпечить зручне середовище для розробки, тестування і підтримки коду, а також інтеграцію з іншими необхідними бібліотеками та інструментами. Ці інструменти дозволяють ефективно реалізувати проект для автоматизації збору та аналізу даних для бізнесу.

2.3 Методи аналізу та обробки даних

У рамках цієї програми будуть застосовуватись різноманітні методи обробки та аналізу даних, що дозволяють оптимізувати бізнес-процеси, прийняття рішень та підвищення ефективності діяльності компанії. Для кожного етапу буде використовуватися відповідний набір технік, інструментів та алгоритмів, що відповідають сучасним вимогам до обробки даних.

Очищення даних

Очищення даних є важливою частиною підготовки даних для подальшого аналізу, оскільки від якості вхідних даних залежить точність і коректність результатів. Ось основні етапи очищення:

Видалення дублікатів: Дані можуть містити повторювані записи, що виникають через помилки збору чи обробки. Для їх виявлення та усунення використовуються методи, як-от перевірка унікальних значень або порівняння записів по всіх полях.

Заповнення пропущених значень:

імпутація: Це метод заміни пропущених значень за допомогою середнього або медіаною для числових змінних, або моди для категоріальних. Існують й більш складні методи, як-от застосування моделей для заповнення пропусків (наприклад, використання алгоритмів KNN для заповнення пропусків) [19];

лінійна інтерполяція: Для часового ряду можна застосувати методи інтерполяції для заповнення пропусків на основі тренду даних [20].

Фільтрація аномалій: Наявність екстремальних значень (аномалій) може суттєво вплинути на результати аналізу. Для виявлення аномалій використовуються методи:

метод міжквартильного діапазону (IQR). Виявлення значень, що знаходяться за межами 1,5 разів IQR від нижнього чи верхнього квартилю[21];

Z-оцінка. Для нормальних розподілів можна використовувати стандартне відхилення для виявлення аномальних значень[22];

Трансформація та нормалізація. Після очищення даних може виникнути необхідність застосування перетворень до значень для приведення їх до певного масштабу:

мін-Макс нормалізація: Приведення значень до діапазону від 0 до 1[23];

Z-скоринг: Стандартизація, коли значення зменшуються на середнє значення та діляться на стандартне відхилення[24].

Обробка даних

Після очищення даних наступний етап – це їх трансформація. Трансформація дозволяє привести дані до формату, зручного для подальшого аналізу.

Агрегування даних: Це процес підсумовування або групування даних за певними ознаками для зменшення їх обсягу та спрощення подальшого аналізу.

агрегування по категоріях. Наприклад, підсумування продажів по різних регіонах, або обчислення середнього значення за квартал;

агрегування за часом. Наприклад, перетворення денних даних в місячні або річні показники для аналізу трендів.

Конкатенація та об'єднання даних: Для створення єдиного набору даних часто необхідно об'єднувати дані з кількох джерел (наприклад, API, файли CSV, бази даних). Використовуються методи merge, join, або concat для об'єднання таблиць по спільних колонках.

Трансформація типів даних: Іноді необхідно змінити тип даних (наприклад, змінити текстову категоріальну змінну на числову) для зручності аналізу. Python пропонує можливість перетворення даних за допомогою функцій, таких як astype() в pandas.

Форматування текстових даних: Визначення важливих аспектів для категоріальних змінних, наприклад, приведення всіх текстових значень до одного формату (перетворення в нижній регістр, видалення зайвих пробілів).

Аналіз даних

На цьому етапі здійснюється власне вивчення даних для виявлення корисних патернів, трендів або взаємозв'язків, які можуть допомогти в ухваленні бізнес-рішень.

Описова статистика: Для початкового розуміння структури та характеру даних використовуються основні статистичні показники:

середнє, медіана, мода: Для аналізу центральної тенденції даних;

стандартне відхилення, варіація: Для оцінки дисперсії даних;

квартилі та IQR: Для визначення розподілу та виявлення аномалій.

Кореляційний аналіз: Для виявлення взаємозв'язків між змінними можна використовувати[26]:

кореляційний коефіцієнт Пірсона для кількісних змінних;

коефіцієнт Крамера для категоріальних змінних;

кореляційні матриці для візуалізації взаємозв'язків між різними параметрами.

Моделювання взаємозв'язків:

лінійна регресія для оцінки лінійних залежностей між змінними;

логістична регресія для класифікації та прогнозування категоріальних змінних;

часові ряди для прогнозування майбутніх значень (наприклад, продажів, попиту, фінансових показників) з використанням методів ARIMA, SARIMA.

Візуалізація є критично важливим інструментом для представлення результатів аналізу в зрозумілій та доступній формі.

Графіки:

лінійні графіки для показу змін за часом або відносної динаміки;

стовпчикові діаграми для порівняння категоріальних даних;

гістограми для аналізу розподілу даних;

теплові карти для візуалізації кореляцій між змінними.

Інтерактивні дашборди:

Використання бібліотек Plotly, Dash, Streamlit для створення інтерактивних дашбордів, що дозволяють користувачам фільтрувати дані, взаємодіяти з графіками та отримувати швидкі результати на основі конкретних запитів.

Візуалізація трендів і прогнозів:

Прогнозовані графіки за допомогою методів прогнозування, таких як ARIMA або Facebook Prophet, для відображення майбутніх трендів на основі історичних даних.

Автоматизація ETL-процесів (Extract, Transform, Load)

ETL-процеси – це ключові елементи для забезпечення безперебійного збору та обробки даних з різних джерел [27].

extract: Збір даних з різних джерел: веб-ресурсів, API (наприклад, Google Analytics API або фінансових даних через Yahoo Finance API), файлів CSV або SQL баз даних;

transform: Обробка зібраних даних: очищення, трансформація типів, фільтрація та обробка відсутніх значень;

load: Завантаження оброблених даних у системи для подальшого аналізу або до бізнес-інтелектуальних платформ (наприклад, бази даних або Google BigQuery).

2.4 Розробка алгоритму оптимізації процесів збору даних

У розробленій системі для автоматизації збору та аналізу даних були успішно реалізовані алгоритми оптимізації процесу збору даних, що дозволяє значно підвищити ефективність і точність аналізу для бізнес-задач.

Основні етапи алгоритму оптимізації збору даних:

Алгоритм збору даних передбачає інтеграцію з різними джерелами інформації, такими як API зовнішніх систем (наприклад, фінансові сервіси, соціальні мережі), веб-ресурси (для збору даних через веб-скрейпінг), а також

бази даних і локальні файли (CSV, JSON, Excel). Це дозволяє здійснювати збір даних з найбільш релевантних джерел для конкретних бізнес-потреб.

Програма автоматизує процес збору даних за допомогою API-запитів, які регулярно отримують актуальну інформацію з визначених джерел. Веб-скрейпінг використовується для збору даних з веб-сторінок, де API не доступне або дані не надаються у структурованому вигляді. Алгоритм налаштований на регулярні оновлення даних через cron jobs або за допомогою webhooks, що дозволяє збирати інформацію в реальному часі.

Після збору даних, алгоритм виконує етап очищення, що включає видалення дублікатів, заповнення пропусків і фільтрацію аномальних або некоректних значень. Використання бібліотеки pandas забезпечує швидку та ефективну обробку даних, що дозволяє зібрати тільки коректну та релевантну інформацію для подальшого аналізу.

Алгоритм оптимізовано для збору даних з урахуванням специфіки бізнесу. Для даних, що змінюються не так часто (наприклад, щоденні або щотижневі фінансові звіти), налаштовано періодичне оновлення. Періодичне оновлення даних є ключовим елементом у забезпеченні їхньої актуальності та своєчасності. Цей процес реалізовано через автоматизовані механізми, які забезпечують регулярний збір та оновлення інформації з джерел відповідно до заданих інтервалів.

Алгоритм побудований так, що з його допомогою можна ефективно збирати дані з великої кількості джерел одночасно. Для прискорення збору даних використовується паралельна обробка, яка дозволяє алгоритму виконувати кілька запитів до різних джерел одночасно. Наприклад, у Python це реалізується через бібліотеки на кшталт concurrent.futures або asyncio, які дозволяють асинхронно отримувати дані. Дані зберігаються та обробляються у хмарних сховищах (Google Cloud Storage) що забезпечує масштабованість у залежності від потреб бізнесу. Хмарні обчислювальні сервіси дозволяють динамічно розширювати або скорочувати ресурси, що використовуються, залежно від обсягу оброблюваних даних. Такий підхід дозволяє забезпечити не

лише поточні бізнес-потреби, але й враховувати зростання обсягів даних у майбутньому, що є ключовим фактором для конкурентоспроможності бізнесу.

Програма має вбудовані механізми моніторингу, які відслідковують ефективність збору даних та їх коректність. У разі виявлення аномалій або помилок у зібраних даних, система генерує сповіщення, що дозволяє своєчасно усунути проблеми і забезпечити точність аналізу.

Щоб краще зрозуміти основні принципи роботи такого алгоритму, розглянемо приклад використання статистичного підходу для виявлення аномалій за допомогою меж квартилів (IQR). Цей метод дозволяє виявити значення, що значно відрізняються від основного масиву даних, та усунути їх перед подальшим аналізом.

По шаговий опис алгоритму фільтрації

Алгоритм складається з наступних етапів:

Завантаження даних:

отримання даних із зовнішніх джерел (API, бази даних або файлів CSV);

завантаження даних у вигляді таблиці (наприклад, DataFrame з бібліотеки pandas):

$$D = \{d_1, d_2, \ldots, d_n\}, \tag{2.1}$$

де D – набір сирих даних;

d_i – окремий запис.

Обробка пропущених значень:

видалення записів, у яких відсутня критично важлива інформація. Завантаження даних у вигляді таблиці (наприклад, DataFrame з бібліотеки pandas);

заповнення пропущених значень середнім значенням, медіаною або модою.

Формула для заповнення середнім значенням[28]:

$$d_j = \frac{\sum_{i=1}^{n} d_{ij}}{n}, \quad (2.2)$$

де d_{ij} – наявне значення атрибуту j для запису i.

Виявлення аномалій:

використання статистичних методів (межі квартилів) для визначення викидів;

видалення значень, що виходять за межі діапазону.

Формула для меж квартилів[29]:

$$IQR = Q_3 - Q_1, \quad (2.3)$$

де Q_1 – медіана нижньої половини;

Q_3 – медіана верхньої половини.

Формула діапазонна значень[30]:

$$[Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR]. \quad (2.4)$$

Перетворення даних:

нормалізація числових значень (наприклад, масштабування в діапазон [0, 1]);

кодування категоріальних змінних. Цей процес означає переведення текстових або номінальних даних (наприклад, "Чоловік"/"Жінка", "Місто"/"Село") у числовий формат [31].

Формула для мінмакс-нормалізації [31]:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)}, \quad (2.5)$$

де x – набір значень

Перш ніж дані можуть бути використані для аналізу або моделювання, необхідно виконати етап їхньої підготовки. Сирові дані часто містять помилки, пропущені значення, дублікатні записи або аномалії, які можуть спотворити результати аналізу.

Далі проілюструємо псевдокод алгоритму (рис.2.1).

```
load raw data into memory
handle missing values:
  drop rows with critical missing data
  fill other missing values with the mean
remove duplicate entries
detect and remove outliers using interquartile range (IQR)
filter data based on business logic conditions
normalize numerical values and encode categorical variables
return the cleaned dataset
```

Рисунок 2.1 – Псевдокод алгоритму

Розроблений алгоритм забезпечує оптимізацію процесів збору та обробки даних, що дозволяє зменшити затрати часу та ресурсів. Завдяки використанню статистичних методів і автоматизації, алгоритм покращує якість даних, що є основою для ефективного аналізу та ухвалення бізнес-рішень.

Висновок до розділу 2

У розділі була розглянута розробка та оптимізація методів збору та аналізу даних, що є важливим етапом у підвищенні ефективності бізнес-процесів. Реалізовані алгоритми автоматизації збору даних дозволяють значно знизити час і ресурси, необхідні для отримання актуальної інформації, а також підвищити точність і релевантність даних для подальшого аналізу.

Особливу увагу було приділено вибору інструментів та технологій для збору даних, таких як API, веб-скрейпінг.

Використання Python як основної мови програмування, з бібліотеками для обробки даних (pandas, NumPy) та інтеграції з іншими системами, дозволило створити потужну й ефективну платформу для збору, очищення та аналізу даних.

Розроблений алгоритм збору даних є гнучким і налаштовується під конкретні потреби бізнесу, що дає можливість оптимізувати процеси в реальному часі, забезпечуючи своєчасне отримання необхідної інформації для прийняття стратегічних рішень. Алгоритм також враховує важливі аспекти безпеки даних, що є ключовим для збереження конфіденційності та захисту бізнес-інформації.

Використання розроблених методів дозволяє не тільки автоматизувати процес збору даних, але й здійснити їх детальну обробку та аналіз, що значно покращує точність і ефективність прийняття рішень. В результаті, реалізація цих методів дозволяє підвищити продуктивність і конкурентоспроможність бізнесу за рахунок оптимізації процесів збору та аналізу даних, що є невід’ємною частиною сучасної бізнес-стратегії.

3 РЕАЛІЗАЦІЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ОПТИМІЗАЦІЇ МЕТОДІВ ЗБОРУ ТА АНАЛІЗУ ДАНИХ

3.1 Проектування архітектури системи

Архітектура системи збору та аналізу даних є фундаментом для створення програмного забезпечення, яке забезпечує ефективність та зручність роботи з великими обсягами інформації. Ця архітектура включає чітко визначені компоненти та взаємозв’язки між ними, що дозволяє оптимізувати процеси збору, обробки та аналізу даних. У даному розділі буде описано структуру системи, її ключові елементи та механізми їхньої взаємодії, що створює базу для подальшої реалізації програмного продукту. Компоненти архітектури складаються з трьох ключових елементів: клієнтського застосунку, API і фінансових даних (рис.3.1).

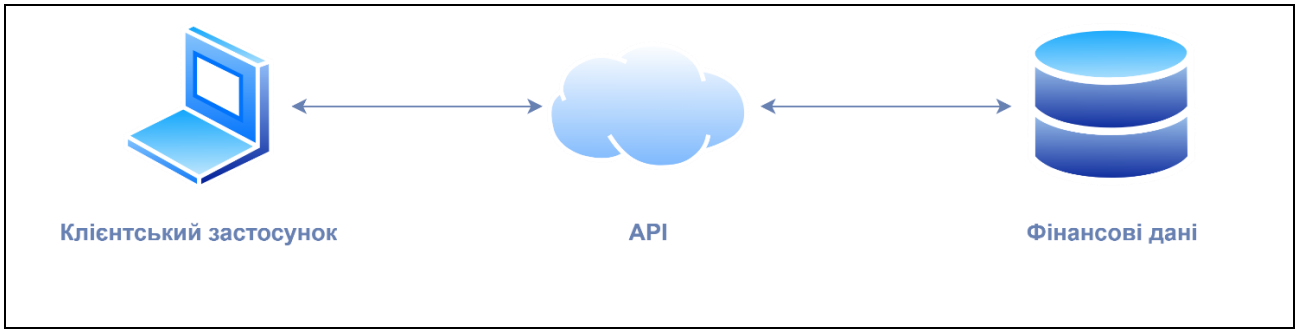


Рисунок 3.1 – Архітектура системи збору та аналізу даних

Клієнтський застосунок є ключовим учасником системи, через який користувач взаємодіє із нею через інтерфейс. Його функції включають надсилання запитів для збору, обробки чи аналізу даних, а також перегляд результатів у вигляді звітів, графіків або дашбордів.

API виступає центральною ланкою системи, яка забезпечує зв'язок між користувачем та фінансовими даними. API приймає запити від користувача, та

повертає результати у зрозумілому вигляді. До його завдань входить отримання фінансових даних із зовнішніх джерел і передача їх користувачеві.

Взаємозв’язки між компонентами побудовані так: клієнтський застосунок надсилає запит до API, яке звертається до фінансових даних для отримання інформації, після чого результати передаються назад через API до застосунку, де вони візуалізуються.

3.2 Проектування архітектури клієнтського застосунку

Система збору та аналізу даних є платформою для автоматизованого збору, обробки, аналізу та візуалізації даних з різних джерел, що допомагає бізнесу приймати ефективні рішення в реальному часі. Основними етапами роботи є: збір даних із API; обробка даних, яка включає очищення, фільтрацію, категоризацію та підготовку до аналізу; аналіз даних за допомогою алгоритмів для отримання інсайтів та прогнозування тенденцій; візуалізація результатів у вигляді графіків, діаграм та інтерактивних дашбордів для зручного сприйняття.

Архітектура побудована на принципі модульності, що дозволяє кожному з компонентів працювати незалежно, зберігаючи при цьому тісну інтеграцію між ними. Це дає можливість гнучко налаштовувати систему під потреби бізнесу та масштабувати її в залежності від кількості джерел даних та обсягу роботи (рис.3.2).

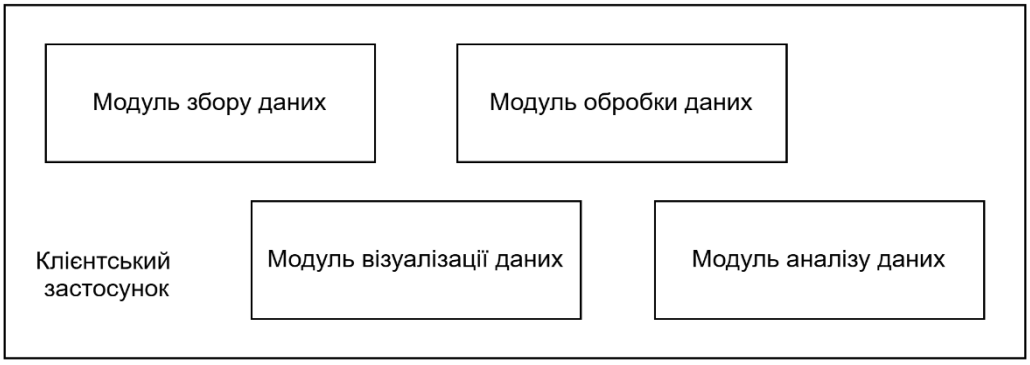


Рисунок 3.2 – Архітектура клієнтського застосунку

Модуль збору фінансових даних відповідає за отримання, обробку та передачу інформації з API сторонніх сервісів до клієнтського застосунку. Його основна задача – безпечне виконання запитів до API, отримання актуальних фінансових даних і забезпечення їх коректного представлення для інших компонентів програми.

Особливістю модуля є використання механізмів автентифікації (токени, ключі API) для захищеного доступу до даних, а також підтримка обробки великої кількості запитів. Для оптимізації роботи модуль може реалізовувати кешування даних і обробку помилок, щоб зменшити навантаження на API та забезпечити стійкість до збоїв у мережі чи на стороні сервісу.

Модуль обробки даних відповідає за аналіз, трансформацію та підготовку даних для використання в інших компонентах програми. Він приймає "сирі" дані, отримані з API або інших джерел, і перетворює їх у структурований та зручний формат для роботи (наприклад, у вигляді об’єктів, таблиць чи звітів).

Особливістю модуля є підтримка різних форматів вхідних даних (JSON, XML, CSV тощо) та забезпечення коректності й консистентності обробки. Він може включати валідацію даних, агрегацію, фільтрацію за критеріями, а також виконувати складніші операції, такі як математичні розрахунки або застосування алгоритмів машинного навчання. Модуль має бути швидким і надійним, щоб обробляти великі обсяги даних з мінімальними затримками.

Модуль аналізу даних відповідає за виявлення закономірностей, трендів і аномалій у зібраній та обробленій інформації. Він використовує статистичні методи, алгоритми машинного навчання або спеціалізовані підходи для інтерпретації даних та отримання цінних інсайтів.

Особливістю модуля є його здатність працювати з великими обсягами даних і адаптуватися до змін у структурі чи властивостях вхідних даних. Він може виконувати задачі прогнозування, кластеризації, побудови трендів або оцінки ризиків. Результати аналізу зазвичай передаються у вигляді візуалізацій, аналітичних звітів або метрик, що легко інтегруються з іншими компонентами застосунку.

Модуль візуалізації даних відповідає за представлення результатів аналізу та обробки у зрозумілому й наочному форматі. Він перетворює числові або текстові дані на графіки, діаграми, таблиці чи інтерактивні панелі, щоб користувачі могли легко зрозуміти їхній зміст.

Особливістю модуля є гнучкість у виборі типів візуалізації (лінійні графіки, гістограми, теплові карти тощо) і можливість інтерактивної роботи з даними, наприклад, фільтрування, масштабування чи деталізації. Він інтегрується з іншими компонентами програми, забезпечуючи реальний час оновлення та динамічне відображення змін, що робить його ключовим для прийняття рішень.

3.3 Реалізація клієнтського застосунку

Модуль збору даних відповідає за взаємодію з кількома джерелами інформації про компанії та їх фінансові звіти. Основний компонент – клас DataProvider, який отримує інформацію про компанію, перевіряючи кожне джерело в списку доступних джерел, поки не знайде необхідні дані.

Клас джерела даних – DataSource, який має метод get, що повертає дані, там метод exists, що перевіряє, чи існують дані про певну компанію у конкретного джерела. Клас DataSource, може мати безліч нащадків, які реалізують доступ до різних API.

Клас DataProvider має метод get, що приймає клас DataSettings з налаштуваннями отримання даних (дата, тощо) який перевіряє кожне джерело через метод exists (щоб дізнатися, чи є інформація про компанію). Якщо компанія є в джерелі, метод get цього джерела повертає її фінансову звітність у вигляді об’єкту класа Data.

Клас DataProvider отримує список джерел даних (наприклад, ApiSource1, ApiSource2) і за допомогою методу get перевіряє, чи є у цих джерелах інформація про компанію;

Метод get перевіряє кожне джерело через exists. Якщо компанія є в джерелі, викликається метод get, який повертає фінансові дані;

Якщо жодне джерело не містить інформації про компанію, генерується виключення, яке повідомляє, що компанія не знайдена в жодному з джерел.

У таблиці 3.1 розглянемо структуру модулю збору даних

Таблиця 3.1 – Структура модулю збору даних

Клас	Метод	Опис
DataSource	exists	Перевіряє, чи є дані про компанію в конкретному джерелі.
	get	Отримує фінансову звітність для заданої компанії з джерела.
ApiSource1	exists	Реалізує логіку перевірки наявності компанії в першому API.
	get	Реалізує логіку отримання фінансової звітності з першого API.
ApiSource2	exists	Реалізує логіку перевірки наявності компанії в другому API.
	get	Реалізує логіку отримання фінансової звітності з другого API.
DataProvider	get	Перевіряє всі джерела та повертає фінансову звітність компанії.
DataSettings		Зберігає налаштування за якими виконується збір даних.
Data		Зберігає дані про фінансову звітність

Давайте представимо структуру модулю збору даних на рисунку А.1 (додаток А).

Модуль обробки даних призначений для підготовки та очищення отриманої інформації. Він містить один клас DataProcessor, який має метод

proceed, що приймає на вхід отриману інформацію та виконує над нею різні операції для приведення її до належного стану.

Клас DataProcessor приймає на вхід набір сирих даних (наприклад, у вигляді списку словників, де кожен словник містить фінансові показники).

Метод Proceed виконує наступні операції:

Сортування: Дані можна відсортувати за певним полем (наприклад, за датою).

Видалення дублікованих записів: Якщо один і той самий запис з'являється кілька разів, його можна видалити.

Фільтрація: Наприклад, можна відфільтрувати всі записи за певним діапазоном дат або значеннями.

Нормалізація: Форматування числових значень до певної точності або приведення дати до стандартного формату.

Далі розглянемо структуру модулю обробки даних(таблиця 3.2).

Таблиця 3.2 – Структура модулю обробки даних

Клас	Метод	Опис
DataProcessor	proceed	Приймає сирі дані та обробляє їх: сортує, видаляє дублікати, фільтрує, нормалізує.

Давайте представимо структуру модулю обробки даних на рисунку А.2 (додаток А).

Модуль аналізу даних відповідає за виконання різних аналітичних операцій над обробленими даними. Він містить клас DataAnalyzer, який має метод analyze, що проводить аналіз даних та повертає результат у вигляді об'єкта класу AnalysisResult.

Клас DataAnalyzer приймає на вхід оброблені дані та проводить їх аналіз, застосовуючи різні методи або моделі для отримання аналітичних результатів.

Метод analyze аналізує дані, застосовуючи алгоритми статистики або машинного навчання, і повертає об'єкт AnalysisResult, в якому зберігаються результати аналізу.

Клас AnalysisResult містить необхідні дані, що можуть бути використані для подальших операцій або візуалізації.

У таблиці 3.3 розглянемо структуру модулю аналізу даних.

Таблиця 3.3 – Структура модулю аналізу даних

Клас	Метод	Опис
DataAnalyzer	analyze	Проводить аналіз даних і повертає об'єкт класу AnalysisResult.
AnalysisResult		Зберігає результат аналізу, який може містити різні метрики чи показники.

Давайте представимо структуру модулю аналізу даних на рисунку А.3 (додаток А).

Модуль візуалізації даних відповідає за відображення аналітичних результатів у вигляді графіків, діаграм чи інших візуальних елементів. Він містить клас DataVisualizer, який має метод visualize, що приймає налаштування візуалізації та відкриває вікна з відповідними графічними елементами.

Клас DataVisualizer отримує від користувача налаштування для візуалізації через об'єкт класу VisualizationSettings.

Метод visualize приймає ці налаштування і використовує відповідні бібліотеки для побудови графіків або діаграм.

Після цього метод відкриває вікна, в яких відображаються візуалізації відповідно до заданих параметрів.

Налаштування візуалізації (через VisualizationSettings)

Тип графіка (стовпчаста діаграма, лінійний графік, кругова діаграма)

Колір, шрифт, стиль ліній, розмір елементів на графіку.

Інші візуальні параметри, такі як підписи осей, легенди, заголовки та інші деталі для точності відображення даних.

Далі розглянемо структуру модулю візуалізації даних(таблиця 3.4).

Таблиця 3.4– Структура модулю візуалізації даних

Клас	Метод	Опис
DataVisualizer	visualize	Приймає налаштування візуалізації та відображає вікна з графіками чи діаграмами.
VisualizationSettings		Містить налаштування для візуалізації, такі як тип графіка, колір, розміри та інше.

Давайте представимо структуру модулю візуалізації даних на рисунку А.4 (додаток А).

Реалізація модулів для роботи з фінансовими даними охоплює повний цикл: збір, обробку, аналіз та візуалізацію. Модуль збору забезпечує отримання актуальних даних з різних джерел, модуль обробки готує дані шляхом очищення та стандартизації. Модуль аналізу виявляє ключові показники та закономірності, а модуль візуалізації перетворює результати на графіки та діаграми, спрощуючи їх розуміння.

3.4 Порівняння розробленої системи з існуючими рішеннями

У цьому розділі проведено аналіз ефективності запропонованої системи роботи з фінансовими даними шляхом порівняння її з існуючими популярними рішеннями. Мета порівняння – продемонструвати переваги розробленого підходу в контексті автоматизації, гнучкості, швидкості обробки та можливостей візуалізації даних. Для оцінки було розглянуто такі категорії рішень, як готові API для збору даних, традиційні інструменти обробки, такі як електронні таблиці, та платформи для візуалізації.

Опис порівнюваних підходів:

Розроблена система – модулі збору, обробки, аналізу та візуалізації, побудовані з використанням Python, з високим ступенем автоматизації та можливістю налаштування під потреби користувача.

Популярні API для збору фінансових даних (Yahoo Finance, Alpha Vantage) – готові рішення для отримання даних із фіксованих джерел із мінімальними можливостями кастомізації.

Традиційна обробка даних (табличні інструменти, SQL) – обробка вручну або за допомогою базових інструментів для аналізу даних.

Готові платформи візуалізації (Tableau, Power BI) – спеціалізовані рішення для візуалізації даних із мінімальними можливостями інтеграції аналітичних процесів.

Далі порівняємо розроблену систему з існуючими рішеннями (таблиця 3.5)

Таблиця 3.5 – Порівняння розробленої системи з існуючими рішеннями

Критерій	Розроблена система	API(Yahoo Finance тощо)	Табличні інструменти (Exel)	Tableau/Power BI
Автоматизація	Висока, модульний підхід	Середня	Низька	Висока у межах візуалізації
Масштабованість	Легка інтеграція нових модулів	Залежить від API	Обмежена обробкою вручну	Обмежена типом джерел
Гнучкість	Повна кастомізація	Фіксовані формати	Обмежена функціоналом	Обмежена готовими шаблонами
Візуалізація	Гнучка, налаштовувана	Не передбачено	Низька	Висока, але шаблонна
Точність	Залежить від джерела	Залежить від API	Середня (ручні помилки)	Н/д
Швидкість збору	Висока, залежить від API	Висока	Низька (ручний ввод)	Н/д

Швидкість і автоматизація. Розроблена система забезпечує швидкий збір та обробку даних із високим рівнем автоматизації, перевершуючи ручні процеси в таблицях і готові API завдяки гнучкій інтеграції.

Масштабованість і гнучкість. Відкрита архітектура дозволяє легко адаптувати систему до нових джерел і змін у структурі даних, що є недосяжним для фіксованих рішень, таких як готові API чи візуалізаційні платформи.

Точність і інтеграція. Завдяки модульній структурі система забезпечує високу точність даних за умови надійності джерела та дозволяє гнучко інтегрувати дані з різних API.

Візуалізація. Модуль візуалізації підтримує кастомізовані графіки та діаграми, що перевершує більшість статичних платформ у гнучкості, хоча може поступатися їхньому готовому інтерфейсу.

Таким чином, розроблена система демонструє переваги в автоматизації, гнучкості та кастомізації, роблячи її ефективним інструментом для роботи з фінансовими даними у порівнянні з існуючими рішеннями.

Висновки до 3 розділу

У третьому розділі було реалізовано програмне забезпечення для оптимізації методів збору, обробки, аналізу та візуалізації даних. Основна архітектура системи побудована на модульному принципі, що забезпечує її гнучкість, масштабованість та легкість інтеграції з іншими сервісами.

Результати роботи системи були порівняні з популярними інструментами, такими як Excel, Power BI та Tableau. Розроблена система продемонструвала переваги за критеріями швидкості виконання, рівня автоматизації процесів і зручності використання.

Реалізація програмного забезпечення показала, що оптимізація методів збору та аналізу даних дозволяє не лише автоматизувати рутинні процеси, а й значно покращити якість інформації для прийняття рішень. Система є перспективною для подальшого розвитку, включаючи інтеграцію з хмарними сервісами та використання методів машинного навчання, що відкриває нові можливості для бізнесу.

ВИСНОВКИ

У результаті виконання дипломної роботи розроблено методи збору та аналізу даних для бізнесу із використанням сучасних інформаційних технологій. Проведені дослідження та реалізоване програмне забезпечення показали, що інтеграція автоматизованих процесів та сучасних технологій здатна значно підвищити ефективність управління даними в бізнесі, забезпечуючи точність.

У першому розділі роботи було виконано аналіз предметної області, вивчено існуючі методи збору та аналізу даних, оглянуто сучасні інформаційні технології, такі як хмарні платформи, штучний інтелект, системи BI. Було виявлено переваги автоматизації процесів збору та аналізу даних, що дозволяє скоротити час і витрати на обробку інформації.

У другому розділі сформульовано технічне завдання, обґрунтовано вибір інструментів розробки (Python, pandas, scikit-learn, Tableau), розроблено алгоритми обробки даних та представлено інформаційну модель. Особлива увага приділялась автоматизації очищення, аналізу та категоризації даних.

У третьому розділі реалізовано програмне забезпечення, яке складається з модулів збору даних, обробки, аналізу та візуалізації. Автоматизація збору інформації дозволила скоротити рутинні операції та забезпечила інтеграцію з API та базами даних. Очищення даних та їх підготовка значно підвищили якість вхідної інформації, що дозволило ефективно використовувати алгоритми кластеризації та прогнозування. Модуль візуалізації даних забезпечив зручний інтерфейс для користувачів, дозволяючи отримувати зрозумілі та інтерактивні звіти. Порівняльний аналіз із такими рішеннями, як Excel, Power BI та Tableau, показав, що розроблена система є конкурентоспроможною завдяки своїй адаптивності, швидкості обробки та низьким витратам на інтеграцію.

Розроблена система показала такі ключові переваги: автоматизація збору даних з різних джерел, включаючи API; висока швидкість обробки завдяки оптимізації алгоритмів очищення та аналізу.

У подальшому систему можна розвивати, інтегруючи технології машинного навчання, додаткові інструменти для роботи з великими даними та хмарні сервіси. Це дозволить розширити можливості аналізу, автоматизації та інтеграції, що зробить систему ще більш універсальною та корисною для вирішення сучасних бізнес-завдань. Дипломна робота підтверджує, що сучасні інформаційні технології відіграють ключову роль у підвищенні ефективності бізнесу та забезпеченні конкурентних переваг у динамічному середовищі.

ПЕРЕЛІК ПОСИЛАНЬ

1. MiO в сфері громадського здоров’я. Profihealth. URL : https://profihealth.org.ua/uk/lessons/140 (дата звернення : 01.08.2024).

2. Методи соціологічних досліджень. Освіта.UA. URL : https://osvita.ua/vnz/reports/sociology/12400/ (дата звернення : 01.08.2024).

3. Кількісні методи досліджень. Stud.com.ua. URL : https://stud.com.ua/19458/marketing/kilkisni_metodi_doslidzhen (дата звернення : 01.08.2024).

4. Чернишенко Н. Г. Основи методології та методів соціологічних досліджень. – Київ: Наукова думка, 2020. – 280 с.

5. Непрямі методики збору даних. Studfile. URL : https://studfile.net/preview/7414707/page:4/ (дата звернення : 01.08.2024).

6. Кількісні дослідження. Pro consulting. URL : http://surl.li/rcfyvf (дата звернення : 02.08.2024).

7. Види спостережень за ступенем охоплення первинної сукупності та часом реєстрації фактів. StudFiles. URL : https://studfile.net/preview/2404026/page:6/ (дата звернення : 04.08.2024).

8. Cloud computing. IT-Enterprise. URL : https://www.it.ua/knowledge-base/technology-innovation/cloud-solutions (дата звернення : 14.08.2024).

9. Штучний інтелект, машинне навчання – різниця, теорія і практика. Evergreen – web розробка і діджиталізація бізнесу. URL : https://evergreens.com.ua/ua/articles/machine-learning-overview.html (дата звернення : 14.08.2024).

10. Що таке Big Data? – Характеристики, приклади, методи обробки великих даних. Sigma Software University. URL : https://university.sigma.software/what-is-big-data/ (дата звернення : 14.08.2024).

11. Що таке IoT (інтернет речей) простими словами?. Atiko. URL : https://www.atiko.com.ua/articles-ua/chto-takoe-iot-prostymi-slovami/ (дата звернення : 16.08.2024).

12. Що таке технологія блокчейн? – Пояснення технології блокчейн. Amazon Web Services (AWS). URL : http://surl.li/jpgbbd (дата звернення : 17.08.2024).

13. Давиденко А. Что такое BI-системы: как работают и зачем они нужны, ТОП-5 систем. Kokoc. URL : https://kokoc.com/blog/bi-sistemy/#:~:text=Системы%20BI (дата звернення : 18.08.2024).

14. Основне про аналіз даних: посібник для початківців. Microsoft. URL : https://www.microsoft.com/uk-ua/microsoft-365/excel/data-analysis#:~:text=Описовий%20аналіз%20у,от%20прибуток%20на%20інвестований%20капітал (дата звернення : 02.09.2024).

15. Вареник В.М. Діагностичний аналіз грошових потоків національної економіки. Економічний нобелівський вісник. 2014. № 1 (7). URL : https://econforum.duan.edu.ua/images/PDF/2014/13.pdf (дата звернення : 03.09.2024).

16. Прогнозування і економічний аналіз. BukLib.net. URL : https://buklib.net/books/30262/#:~:text=аналізу,%20прогнозний%20аналіз%20орієнтований%20на,,%20явища (дата звернення : 02.09.2024).

17. Визначення й приклади візуалізації даних. Microsoft Visio. URL : https://www.microsoft.com/uk-ua/microsoft-365/visio/data-visualization (дата звернення : 03.09.2024).

18. What is Business Process Automation: Definition, examples, and software. Snov.io. URL : https://snov.io/glossary/business-process-automation/ (дата звернення : 05.09.2024).

19. Захаренко Е. Метод K-ближайших соседей (KNN). Принцип работы, разновидности и реализация с нуля на Python. Хабр. URL : https://habr.com/ru/articles/801885/ (дата звернення : 10.09.2024).

20. Інтерполяція. StudFiles. URL : https://studfile.net/preview/9721094/ (дата звернення : 15.09.2024).

21. Міжквартильний розмах. Вікіпедія. URL : http://surl.li/zbijvi (дата звернення : 15.09.2024).

22. Z-критерій. Вікіпедія. URL : http://surl.li/nnpdky (дата звернення : 17.09.2024).

23. Max-Min Normalization. GitHub. URL : https://github.com/UzunDemir/Max-Min-Normalization-Min-Max-Scaling/blob/main/Нормализация.ipynb (дата звернення : 17.09.2024).

24. Модель Альтмана. Вікіпедія. URL : https://ru.wikipedia.org/wiki/Модель_Альтмана (дата звернення : 19.09.2024).

25. Агрегування. Вікіпедія. URL : https://uk.wikipedia.org/wiki/Агрегування (дата звернення : 19.09.2024).

26. Кореляційний аналіз. Wiki ТНТУ. URL : https://wiki.tntu.edu.ua/Кореляційний_аналіз (дата звернення : 27.09.2024).

27. Kaplun G. ETL это? ArtofBA. URL : http://surl.li/kquifk (дата звернення : 27.09.2024).

28. Обчислення середнього значення групи чисел. Microsoft. URL : https://support.microsoft.com/uk-ua/office/обчислення-середнього-значення-групи-чисел-e158ef61-421c-4839-8290-34d7b1e68283 (дата звернення : 17.10.2024).

29. Interquartile range. Wikipedia. URL : https://en.wikipedia.org/wiki/Interquartile_range (дата звернення: 18.10.2024).

30. Intermediate value theorem review. Khan Academy. URL : https://www.khanacademy.org/math/ap-calculus-ab/ab-limits-new/ab-1-16/a/intermediate-value-theorem-review (дата звернення : 19.10.2024).

31. Gupta S. How to apply Min-Max Normalization to your data. Medium. URL : https://sourabharsh.medium.com/how-to-apply-min-max-normalization-to-your-data-f976d1633d2b (дата звернення : 20.10.2024).

ДОДАТОК А

Діаграми класів модулю

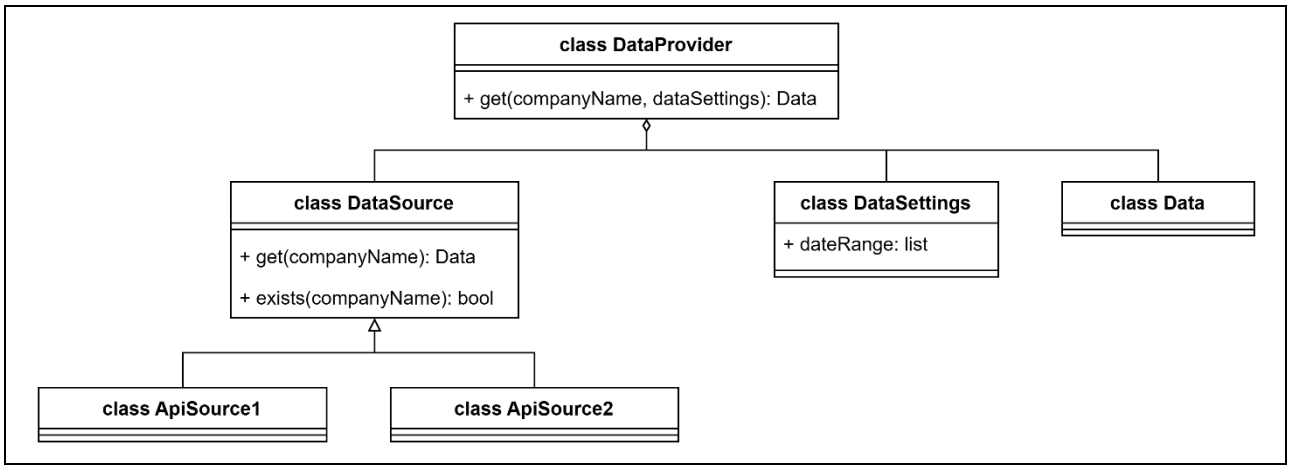


Рисунок А.1 – Діаграма класів модулю збору даних

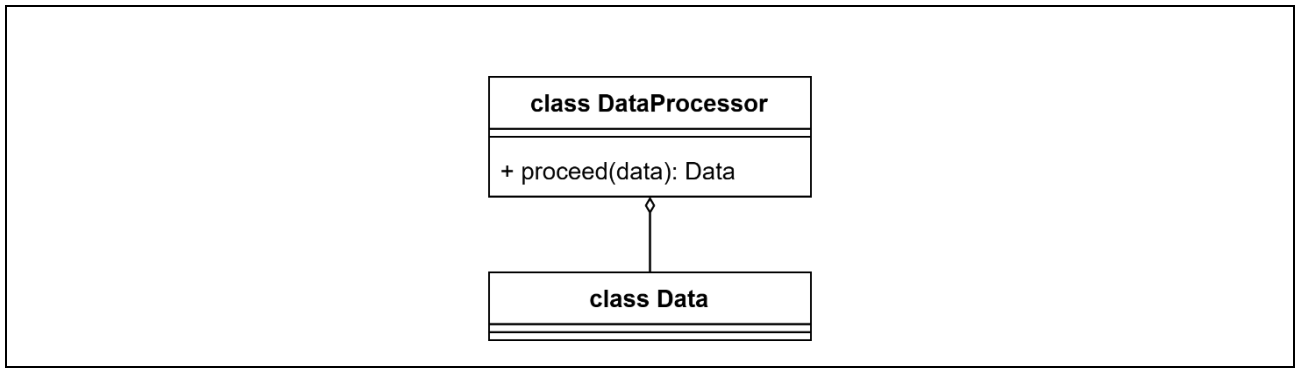


Рисунок А.2 – Діаграма класів модулю обробки даних

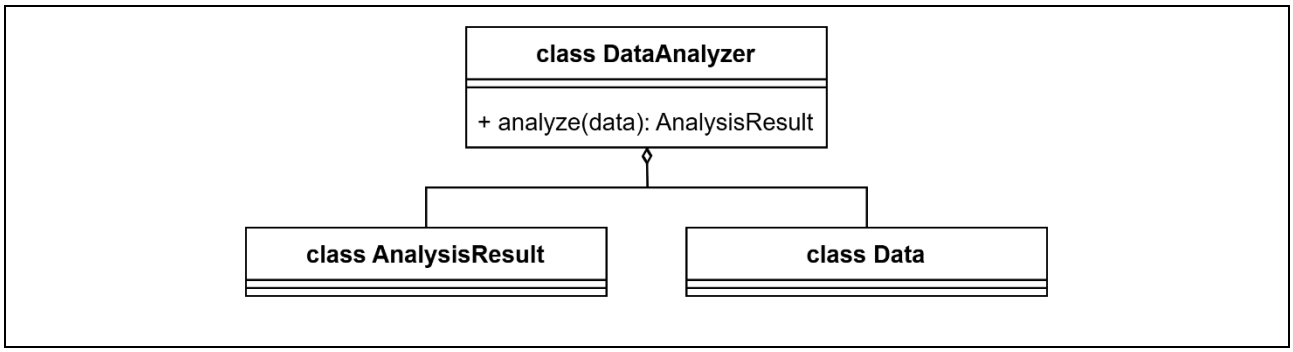


Рисунок А.3 – Діаграма класів модулю аналізу даних

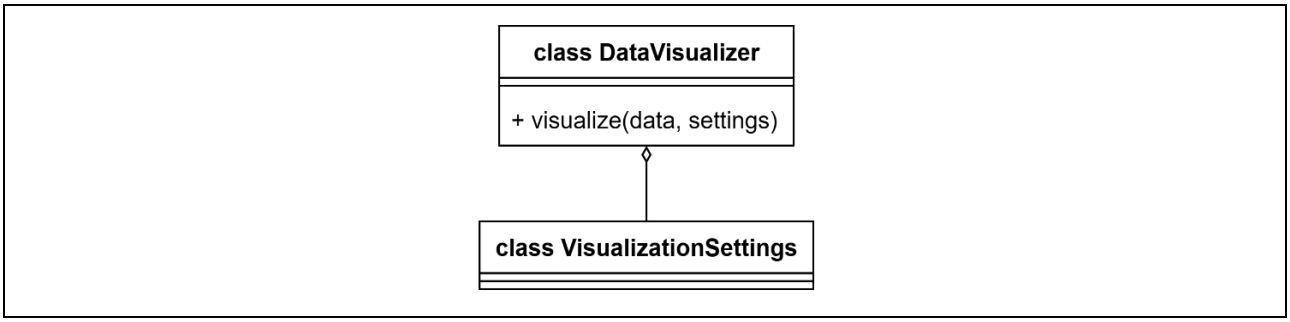


Рисунок А.4 – Діаграма класів модулю візуалізації даних