

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВОДНОГО ГОСПОДАРСТВА ТА
ПРИРОДОКОРИСТУВАННЯ**

“До захисту допущений”

Зав. Кафедри комп’ютерних наук

та прикладної математики

д.т.н., професор Турбал Ю.В.

«___» _____ 2023 р.

КВАЛІФІКАЦІЙНА РОБОТА

ТЕМА:

«ЗАСОБИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ «SMALL DATA»»

Виконав: Бондарчук Віктор Олегович

студент навчально-наукового інституту автоматки, кібернетики та
обчислювальної техніки

групи ПЗ-41

підпис

Керівник: старший викладач Турбал Маріанна Юріївна

підпис

Рівне 2023

ЗМІСТ

Завдання на кваліфікаційну роботу	2
Реферат	5
Вступ	6
1 Огляд сучасного стану сфери інтелектуального аналізу даних	8
1.1 Огляд наукової літератури	8
1.2 Огляд предметної галузі	11
1.3 Ринок послуг та технологій інтелектуального аналізу даних в Україні	13
1.4 Бізнес-аналітика у сфері роздрібно́ї торгівлі	16
1.5 Можливості застосування інтелектуального аналізу даних у сфері рітейлу	22
1.6 Методи інтелектуального аналізу даних	25
2 Проектування автоматизованої інформаційної системи	27
2.1 Постановка завдань та аналіз існуючих рішень	27
2.2 Огляд існуючих рішень	29
2.3 Постановка завдань	32
2.4 Обґрунтування та вибір способу вирішення зазначеної проблеми	33
2.5 Високорівневий опис алгоритму	34
2.6 Пошук асоціативних правил	38
3 Програмна реалізація інформаційної системи	45
3.1 Робота з базою даних	45
3.2 Інтелектуальний аналіз даних	51
ВИСНОВОК	56
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	57

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВОДНОГО ГОСПОДАРСТВА ТА
ПРИРОДОКОРИСТУВАННЯ**

НН інститут автоматички, кібернетики та обчислювальної техніки

Кафедра комп'ютерних наук та прикладної математики

Освітньо-кваліфікаційний рівень **бакалавр**

Галузь знань 12 "Інформаційні технології"

Спеціальність 121 Інженерія програмного забезпечення (Інтернет речей)

ЗАТВЕРДЖУЮ

Зав. Кафедри

Доктор технічних наук,

Професор Турбал Ю.В

“ _____ ” _____ 2023 року

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Бондарчуку Віктору Олеговичу

1. Тема роботи «Засоби інтелектуального аналізу «Small Data»»

керівник роботи Турбал Маріанна Юріївна, ст.викл.

затверджені наказом вищого навчального закладу від "19" квітня 2023 року
№ 449

2. Термін подання роботи студентом 19 червня 2023 р.
3. Вихідні дані до роботи

В даний час технологія інтелектуального аналізу даних представлена цілою низкою комерційних та вільно розповсюджуваних продуктів. Вивчаючи предметну область, я дійшов висновку, що вони мають одні загальний недолік – вони вимагають наявності спеціальних знань і навичок до роботи з ними. Великі фірми та підприємства вирішують цю проблему наймаючи кваліфікованих спеціалістів чи навчаючи своїх. У той час як представники малого підприємництва, часто, не мають у

своєму розпорядженні необхідних для цього ресурсів. За підсумками всього вищесказаного переді мною було поставлено таку мету: забезпечення інформаційної підтримки представників малого підприємництва шляхом надання їм спрощеного доступу до технології інтелектуального аналізу даних.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

Переді мною була поставлена задача спроектувати та реалізувати інформаційну систему, що вирішує певні завдання та виконує конкретно зазначені функції.

А саме:

- система повинна мати можливість роботи з базою транзакцій фірми. Повинні бути реалізовані можливості створювати базу даних, додавати до неї нові транзакції, редагувати та видаляти існуючі, проводити вибірку з бази даних за заданими умовами;

- система має містити засоби інтелектуального аналізу даних, а саме пошуку асоціативних правил у базі даних. Повинні бути реалізовані пов'язані з цим підготовка та перетворення даних;

- користувач повинен мати можливість отримувати результати роботи програми як у вигляді звіту, так і за допомогою наочної та зрозумілої візуалізації.

Проектована інформаційна система має бути виконана максимально просто і зрозуміло для користувача, реалізуючи досягнення поставлених завдань, не ускладнюючи себе зайвими функціями. Інтерфейс системи має бути інтуїтивно зрозумілим

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Турбал М.Ю ст. викладач	08.03.2023	20.05.2023
2	Турбал М.Ю ст. викладач	21.05.2023	01.06.2023
3	Турбал М.Ю ст. викладач	01.06.2023	20.06.2023

7. Дата видачі завдання 8 березня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Отримання теми кваліфікаційної роботи	8.03.2023	виконано
2	Дослідження літератури по темі роботи	8.03.2023 - 10.04.2023	виконано
3	Дослідження конкурентів на ринку України	10.04.2023 – 20.05.2023	виконано
4	Проектування інформаційної системи	20.05.2023 – 01.06.2023	виконано
5	Створення та тестування системи	01.06.2023 – 20.06.2023	виконано
6	Розробка документації кваліфікаційної роботи	20.06.2023 – 26.06.2023	виконано

Студент

Бондарчук В.О

(підпис) (прізвище та ініціали)

Керівник роботи

Турбал М.Ю

(підпис) (прізвище та ініціали)

РЕФЕРАТ

Кваліфікаційна робота: 58 сторінок, 13 рисунків, 2 таблиці, 14 джерел

Мета роботи: забезпечення інформаційної підтримки представників малого підприємництва шляхом надання їм спрощеного доступу до технології інтелектуального аналізу даних

Засоби розробки: C++, SQLite, Qt.

Актуальність роботи: полягає в тому, щоб надати представникам малого підприємництва спрощеного доступу до інструментарію технології інтелектуального аналізу даних.

ВСТУП

За останні кілька десятиліть ми стали свідками повсюдного поширення новітніх інформаційних технологій та їх впровадження у різні галузі соціальної та економічної сфер. Все це разом з наявністю вільного доступу до Інтернету призвело до величезного зростання обсягу інформації, виробленої людиною.

Тільки за кілька днів людство сьогодні виробляє стільки інформації, скільки було створено за весь час існування Землі аж до нинішнього тисячоліття. За твердженнями досліджень, сьогодні обсяг даних на планеті подвоюється кожні кілька років. І така тенденція збережеться як мінімум до 2025 року.

Через величезну кількість існуючої інформації, лише мала її частина буде побачена людським оком, і, як наслідок, значні обсяги практичної та корисної інформації просто пройдуть повз. Це призводить до розуміння важливості проблем, пов'язаних з аналізом накопичених даних для отримання нових знань.

Наша єдина можливість зрозуміти і знайти щось корисне у цьому величезному океані – це широке застосування методів інтелектуального аналізу даних.

Інтелектуальний аналіз даних широко використовується в багатьох областях з великим обсягом даних. У науці – біології, медицині, астрономії тощо. У бізнесі – насамперед у торгівлі, телекомунікації, промисловому виробництві та інших. В рамках цієї роботи я обмежусь сферою малого підприємництва, або, інакше, малого бізнесу.

Зараз йде багато розмов про підтримку та сприяння розвитку малого бізнесу в Україні, і я зі свого боку зробив спробу внести в цю справу посильний внесок відповідно до своїх можливостей і ситуації, що склалася в галузі.

В даний час технологія інтелектуального аналізу даних представлена цілою низкою комерційних та вільно розповсюджуваних продуктів. Вивчаючи предметну область, я дійшов висновку, що вони мають одні загальний недолік – вони вимагають наявності спеціальних знань і навичок до роботи з ними. Великі фірми та підприємства вирішують цю проблему наймаючи кваліфікованих спеціалістів чи навчаючи своїх. У той час як представники малого підприємництва, часто, не мають у своєму розпорядженні необхідних для цього ресурсів.

За підсумками всього вищесказаного переді мною було поставлено таку мету: забезпечення інформаційної підтримки представників малого підприємництва шляхом надання їм спрощеного доступу до технології інтелектуального аналізу даних.

Для досягнення поставленої мети було висунуто такі завдання:

- створення вільно розповсюджуваного програмного продукту, що не потребує спеціальних навичок для роботи та максимально зрозумілого;
- надання користувачеві інструментарію для роботи з базою даних;
- застосування методів інтелектуального аналізу даних до цієї бази даних, з метою аналізу накопиченої інформації для отримання нових та потенційно корисних знань.

1. Огляд сучасного стану сфери інтелектуального аналізу даних

1.1 Огляд наукової літератури

При написанні даної роботи були використані наукова та навчально-методична література, статті на електронних ресурсах мережі Інтернет.

Основним джерелом послужив науковий посібник «Інформаційні системи міжнародного бізнесу» Ірини Вікторівни Журавльової. У цьому навчальному посібнику автор перерахував основні технології, які використовуються сьогодні у сфері бізнес-аналітики. Привів наочні приклади способів реалізації даних технологій, втілення їх у життя та впровадження з виробництва. Сама сфера бізнес-аналітики знаходиться на перетині таких різних галузей науки як інформаційні технології, економіка та статистика. Їхнє об'єднання дозволило створити потужні інструменти та інформаційно-аналітичні системи для сучасного бізнесу, про які і йдеться у навчальному посібнику. Книга знайомить читача з областю бізнес-аналітики поступово, все далі поринаючи в основні проблеми та пояснюючи шляхи їх вирішення. Книга буде корисною як навчальний посібник як студентам інформаційно-технічних спеціальностей, так і студентам економічних напрямів та програм навчання. Також вона буде корисна вже дійсним бізнес-аналітикам, фахівцям з аналізу даних та професіоналам, пов'язаним із цією областю для вивчення питань впровадження систем бізнес-аналітики. Також книга стане у нагоді всім тим, хто прагне більше дізнатися про проблеми інтелектуального аналізу даних та бізнес-аналітики.

Відмінним джерелом інформації послужив посібник «Основи інформаційно-аналітичної діяльності» від В. І Захарової та Філіпової Л. Я. Аналітична робота висвітлена у цій книзі з усіх можливих сторін. Порушується широке коло проблем, пов'язаних з аналітичною роботою в цілому, її організацією, методологією та інформаційно-технологічним.

оснащення сучасних аналітиків. Частину книги автори знайомлять читача зі способами та методами ефективної аналітичної діяльності (як у повсякденному житті, так і стосовно професійної діяльності), методів розумової роботи в цілому. Іншу частину книги автори присвятили висвітленню технологічної сторони сучасної аналітики, наочно розібрали питання, пов'язані з проектуванням та розробкою автоматизованих інформаційних систем аналітичної роботи, їх функціонуванням та ефективним використанням. Автори розкривають суть поняття інтелектуального аналізу даних. Показують, як в аналітичній діяльності переплелися і взаємопов'язані такі різні галузі науки, як логіка, математика, філософія, сфера інформаційних технологій і навіть психологія та ін. Розібрані відмінності української та зарубіжних аналітичних шкіл, їх історія та формування. Автори розкривають всю значущість аналітики для всіх сфер нинішньої людської діяльності, як-от соціальна сфера, економічна сфера, політична сфера. Показано можливості використання аналітики у цих сферах для підвищення ефективності прийняття управлінських рішень, прогнозування на основі внутрішніх та зовнішніх даних. Книга буде цікава та корисна в основному фахівцям, які працюють у галузі розробки інформаційно-аналітичних систем, систем підтримки прийняття рішень. Може використовуватися як науковий посібник для аспірантів та студентів інформаційно-технічних напрямків та спеціальностей. Показано можливості використання аналітики у цих сферах для підвищення ефективності прийняття управлінських рішень, прогнозування на основі внутрішніх та зовнішніх даних. Книга буде цікава та корисна в основному фахівцям, які працюють у галузі розробки інформаційно-аналітичних систем, систем підтримки прийняття рішень. Може використовуватися як науковий посібник для аспірантів та студентів інформаційно-технічних напрямків та спеціальностей.

Ще одним корисним посібником стала книга Крістофера Дейта «Введення в системи баз даних». У цьому навчальному посібнику докладно викладаються методи інтелектуального аналізу даних за допомогою систем управління

базами даних. Розглядаються суть та значення інтелектуального аналізу даних, наводиться послідовний опис основних алгоритмів та їх застосування на практиці. Розглянуто способи створення та проектування баз даних та подальшого їх аналізу як за допомогою вже існуючих програмних рішень, так і з використанням основних алгоритмів аналітики даних. Докладно висвітлено проблеми вилучення інформації з баз даних та інших джерел. Автор навчального посібника докладно і зрозуміло ілюструє кожне рішення наочними практичними прикладами, які роз'яснюють суть методів і технологій, що застосовуються. Книга написана простою та зрозумілою мовою і навіть у початківців вивчати області баз даних та інтелектуального аналізу не виникне проблем з розумінням викладеного матеріалу. Книга буде корисною для аспірантів та студентів інформаційно-технічних напрямків та спеціальностей.

Якщо говорити про стан літератури з інтелектуального аналізу даних взагалі, я можу сказати, що під час роботи я відчув нестачу сучасної актуальної літератури українською мовою.

Більшість літератури з Data Mining українською була написана в нульових роках, і в основному це були оригінальні наукові роботи, а не переклад зарубіжних книг, але, на жаль, вони вже втратили свою актуальність. Якщо говорити про сучасну, не старшу за п'ять років літературу, що знаходиться у вільному доступі і може бути безкоштовно завантажена або придбана в магазинах та онлайн-сервісах – українською мовою вона представлена переважно перекладами зарубіжних та науково-популярних книг, що дають загальний огляд предметної області, і не може бути використана для серйозного навчального посібника.

На сьогоднішній день наукова література з Data Mining представлена в основному англійською мовою. Це з тим, що англійська мова де-факто стала мовою сучасної інформаційної науки, у сфері інформаційних технологій більшість розробок, статей, наукової літератури та технічної документації ведуться англійською мовою.

Щодо можливості навчання методам інтелектуального аналізу даних українською мовою – з розвитком мережі Інтернет з'явилося безліч онлайн-курсів від великих вітчизняних та зарубіжних компаній, які ведуть підготовку фахівців як своїх потенційних працівників.

1.2 Огляд предметної галузі

За останні десятиліття людство накопичило величезні обсяги інформації у цифровій формі. Різні дані органів державного управління, промислових підприємств, великих та малих фірм. Очевидно, що серед цього моря інформації міститься значний прихований потенціал знань, володіючи яким можна підвищити ефективність своєї діяльності, хоч би якою вона була. Сьогодні вся діяльність будь-якого підприємства фіксується, записується та зберігається в архівах та електронних сховищах. Зрозуміло, що такі дані без необхідного аналізу та переробки є просто марним звалищем. Це і призводить до того, що завдання отримання прихованих знань із накопичених даних є на сьогоднішній день гостро необхідною.

До процесу переробки сирих даних на корисні знання пред'являються особливі вимоги:

- оброблювані дані можуть бути як завгодно великого обсягу;
- дані за своєю суттю не однорідні;
- виявлені знання повинні бути зрозумілі та недвозначні;
- засоби аналізу даних повинні бути прості і зрозумілі

Довгий час саме математична статистика була основним інструментом аналітиків. Однак сьогодні, перед новими проблемами, вона справляється вже не так переконливо. Основна причина цього в тому, що в математичній статистиці прийнято концепцію усереднення за вибіркою, що змушує аналітиків оперувати неіснуючими значеннями типу середньої зарплати двірника та олігарха. Це не означає, що математична статистика є слабким

інструментом, але кожен інструмент повинен використовуватися для своєї роботи. Не треба забивати цвяхи мікроскопом. Математична статистика корисна для аналізу вже висловлених гіпотез і розвідувального аналізу, що є основою технології OLAP (аналітична обробка у часі).

Сьогодні для вирішення актуальних проблем з'являються інші методи та технології аналізу даних, що отримали назву Data Mining, або, що є найбільш узвичаєним перекладом українською мовою, інтелектуальний аналіз даних. Ці методи спрямовані на ті завдання, яких сьогодні найбільше потребують – аналіз великих обсягів даних та пошук прихованих знань усередині цих обсягів. Рівні знань та застосовні до них аналітичні інструменти наведено у таблиці 1.

Таблиця 1 – Рівні застосування аналітичних інструментів

Рівень знань, що здобувається з даних	Аналітичні інструменти
Поверхневий	Мова простих запитів
Неглибокий	Оперативна аналітична обробка (OLAP)
Прихований	Інтелектуальний аналіз даних (Data Mining)

Головна особливість інтелектуального аналізу даних – нетривіальність знань, що виявляються. Це означає, що здобуті знання повинні містити у собі неочевидні здавалося б зв'язку, складові звані приховані знання. Зараз більшості людей стало очевидно, що глибоко під поверхнею накопичених даних знаходяться приховані знання, які можуть бути вилучені за належної розкопки. Виходячи з наведеної інформації, при використанні методів та технологій інтелектуального аналізу даних з'являється дійсна можливість відкривати неочевидні закономірності між накопиченими даними та використовувати отримані знання в системах прийняття рішень, допомагаючи у вирішенні управлінських завдань будь-де: на підприємствах, у бізнесі, муніципальних установах .

Сьогодні технології та методи Data Mining використовуються практично скрізь, де зібрана будь-яка інформація. Data Mining може застосовуватись скрізь, де є дані.

Найбільше застосування та поширення інтелектуальний аналіз даних отримав у галузі бізнес-аналітики. Причина в тому, що грамотний аналіз даних фірми дозволяє не тільки підвищити ефективність виробництва, але й значно підвищити його прибутковість, що робить аналітичні системи швидко окупними.

Широке застосування технології інтелектуального аналізу даних отримали у сфері роздрібної торгівлі та маркетингу. Основні прийоми Data Mining, на кшталт кластерного аналізу та пошуку асоціативних правил, дозволяють здорово допомогти при вирішенні основних щоденних проблем, таких як: закупівлі та витрачання товарів, їх розміщення у торговому залі. Вони також дозволяють підвищити лояльність клієнтів та збільшити дохід від них, аналізуючи їхню активність.

1.3 Ринок послуг та технологій інтелектуального аналізу даних в Україні

Порівняно зі світовим ринком послуг та технологій Data Mining український ринок виглядає мізерно малим. Це виявляється переважно в обсягах обороту світового та нашого ринків. За дослідженнями агентства Wikibon, оборот світового ринку інтелектуального аналізу даних за 2021 рік становив \$33,3 млрд. Для порівняння, компанія IDC оцінила український ринок всього в \$340 млн.

Проведемо невеликий огляд, що дає загальне уявлення про ринок обробки даних у Україні. Для зручності розділимо учасників ринку на категорії, хоч і дуже умовні:

- постачальники (SAP, Oracle, IBM, EMC, Microsoft та ін.);
- розробники алгоритмів (датамайнери) (CleverData та ін);
- інтегратори;
- споживачі (телеком, банки, рітейл та ін);
- окремі розробники.

Постачальники займаються продажем професійних СУБД, інформаційно-аналітичних систем та апаратних комплексів – самостійно чи через своїх агентів. Для того щоб виразно розібратися в цих продуктах і використовувати їх з максимальною для себе вигодою, фірми повинні мати власні відділи експертів і аналітиків. Це під силу тільки великим компаніям, менші компанії користуються послугами інтеграторів та ІТ-консультантів, які підбирають програмне та апаратне забезпечення індивідуально під потреби замовника.

Компанія SAS, одна з перших, що розпочала роботу в сфері систем бізнес-аналітики, зараз працює з такими клієнтами як «Теле2», «Українські Залізниці». Крім продажу своїх систем, SAS також надає клієнтам допомогу у навчанні роботі зі своїми продуктами.

Google вийшла на ринок аналізу даних для бізнесу зовсім недавно – лише 2022 року. Вона зайняла відносно вільну нішу роботи із середніми та дрібними підприємствами, на кшталт «Ельдорадо». Інформації про співпрацю Google із українськими державними структурами у сфері аналітики виявлено не було.

Датамайнери (або розробники алгоритмів) самі займаються виявленням знань у даних клієнта. Деякі їх сервіси дозволяють користувачам просто завантажити вихідні дані в хмару і отримати корисні знання на виході. Головною перевагою такого підходу є те, що немає необхідності набувати інфраструктури та наймати дорогий персонал для роботи з нею. Великі компанії переважно користуються своєю інфраструктурою та фахівцями, тоді як послуги датамайнерів користуються попитом переважно серед представників малого підприємництва.

Серед менших гравців можна виділити такі державні фірми, як IBS, AT Consulting, CleverData, EasyData, Double Data, DataMining Labs, MLClass, BaseGroup Labs, Global Innovation Labs. Більшість компаній досить молоді, але всі вони вже міцно закріпилися на українському ринку та поступово пробиваються за кордон.

Інтегратори та ІТ-консультанти допомагають клієнтам впровадити у себе систему аналітики. Вони є сполучною ланкою між бізнесом та технологіями. Їхніми послугами користуються ті, кого не влаштовують готові рішення постачальників та послуги датамайнерів. Інтегратори займаються переважно тим, що комбінують системи та його частини від різних постачальників на єдине ціле, що дозволяє задовольнити потреби замовника щодо потрібного йому функціонала.

На вітчизняному ринку найуспішнішими інтеграторами є «AM-BITS» та «SI BIS». Обидві вони відносно молоді, і розпочали свою діяльність у сфері аналітики великих даних не так давно.

Компанія "AM-BITS" співпрацює з банками, мережами роздрібною торгівлі, телеком-операторами, держсектором. Крім розгортання рішень сторонніх постачальників, вони займаються власними розробками. "AM-BITS" в основному працюють з продуктами компанії Oracle.

"SI BIS" вже встиг попрацювати з продуктами таких гігантів як з EMC, HP, Oracle та Microsoft, Intel. Послугами «SI BIS» користувалися банки, телекомоператори та сектор охорони здоров'я. Потрібно також кілька слів сказати про готові сервіси на основі великих даних. На технологіях Data Mining працюють такі звичні нам послуги як антиспам-фільтр, контекстна реклама, антифрод. Найбільші українські ІТ-компанії аналізують ці величезні обсяги даних самостійно. Це дозволяє їм розвивати свої послуги, підвищувати ефективність реклами та оптимізувати контент.

Google Group застосовували технології Data Mining ще тоді, коли навіть сам термін не узвичаївся. Система веб-аналітики «Рейтинг Google» була одним з перших подібних сервісів. Сьогодні практично всі їхні сервіси в тій чи іншій мірі займаються інтелектуальним аналізом великих об'ємів даних: пошукова машина, пошта, таргетингова реклама та ін. обробляють інформацію про активність користувачів. Вони також мають свої власні розробки, наприклад – NoSQL СУБД Tarantool.

"Gmail" завжди використовував аналітику для покращення своєї пошукової машини, але в останні роки особливо активно став застосовувати інтелектуальний аналіз. "Gmail" використовує методи Data Mining для тих же цілей, що і Google: оптимізація пошуку, реклами, фільтрація спаму. Також компанія нещодавно придбала фірму, яка виробляє продукти для інтелектуальної аналітики текстів (Text Mining), якими користуються великі державні компанії.

1.4 Бізнес-аналітика у сфері роздрібно́ї торгівлі

На жаль, серед представників малого бізнесу у сфері ритейлу системи бізнес-аналітики ще не досить поширені, і їх можливості ще не повною мірою розкриті користувачами. Хоча вони й допомагають збільшити ефективність ведення бізнесу, але чомусь недостатньо охоче впроваджуються на малих підприємствах.

Торік компанією Qlik було проведено опитування, основними завданнями якого було з'ясувати причини використання аналітичних систем у роздрібній торгівлі та дізнатися, наскільки добре їх користувачі знають про можливості та ефективність таких систем.

Опитування проводилося серед представників керуючого ланки роздрібної торгівлі, які працюють з різними товарами. Респонденти були ранжовані за розмірами їхнього бізнесу, виходячи з торгових точок компанії. З усієї кількості опитаних більшість становлять представники малого підприємництва (29% респондентів) з невеликою кількістю торгових точок. Другими за чисельністю серед опитаних виявилися надвеликі мережі, що мають понад 500 торгових точок. На їхню частку припало 18%. На решту категорій припало приблизно по 10% на кожну.

Проведемо невеликий аналіз отриманих даних та з'ясуємо стан сфери інтелектуального аналізу даних у галузі ритейлу.

Серед респондентів значна частина вже застосовує у роботі системи аналізу даних – 57,1% опитаних. Серед тих компаній, хто ще не користується бізнес-аналітикою, значна частина збирається впровадити такі системи у найближчому майбутньому. А повністю відмовляються працювати в ногу з часом лише 2,4% опитаних. Здебільшого це представники малого підприємництва. Ці результати можна побачити малюнку 1.

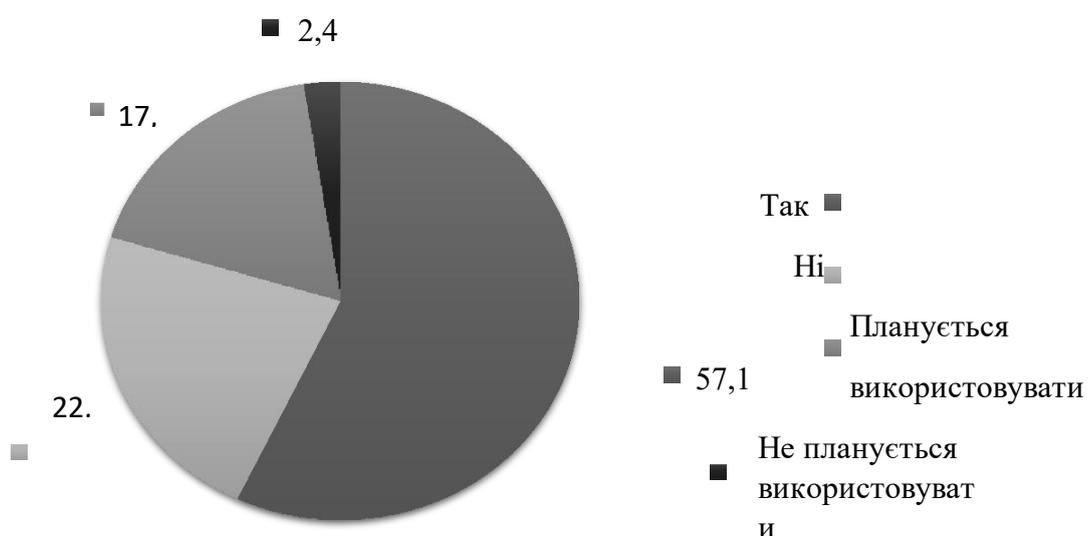


Рисунок 1 – Наявність системи бізнес-аналітики

Що стосується основних приводів, що спонукали компанії використовувати у себе системи аналізу великих даних, тут вони багато в чому збігаються незалежно від бізнесу учасників опитування. Респондентам пропонувалося вибрати кілька головних на їхню думку причин. Наведу три найпопулярніші. Найважливішою причиною, що спонукала до використання бізнес-аналітики, респонденти назвали бажання підвищити швидкість управлінських рішень, що приймаються, і підвищити їх ефективність. Цю необхідність наголосили 88,6% учасників опитування. Трохи менш важливим, але все ж таки значущим, виявилось бажання керівників підвищити продуктивність роботи своїх співробітників, позбавивши їх рутинних дій по роботі зі звітами (72,7%). Серед причин, які вплинули рішення використовувати з виробництва системи бізнес-аналітики також було названо потреба об'єднання інформації з різних джерел – її частку довелося 65,9%. Розподіл топ-3 причин наочно відображено малюнку 2.



Рисунок 2 – Причини впровадження бізнес-аналітики

Такі, здавалося б, очевидні причини, як підвищення прибутку, збільшення ефективності роботи з клієнтами, на подив, виявилися не такі важливі для учасників опитування. Ці причини вказував значно менший відсоток респондентів.

Така тенденція може говорити про те, що серед тих, хто вже використовує у себе системи бізнес-аналітики, багато хто не до кінця розуміє весь спектр їх можливостей, розглядаючи їх в основному як інструмент для спрощення роботи зі звітами та підвищення ефективності взаємодії співробітників, не більше.

Побічно це підтверджують і результати опитування щодо досягнутих результатів роботи таких систем серед користувачів. Більшість опитаних (70,7%) згадали про скорочення часу, який використовується для збору необхідної для прийняття управлінських рішень інформації. Половина опитаних відзначили підвищення ефективності таких рішень. А підвищення прибутковості бізнесу шляхом зниження витрат, оптимізації логістики та оптимального ціноутворення за допомогою систем бізнес-аналітики вдалося досягти не багатьом. Лише близько 10% учасників опитування відзначили такі досягнення. Причини цього вже пояснювалися вище, просто для цього аналітикою користується лише малий відсоток опитаних, але, як бачимо, своїх цілей вони досягають. За сферами застосування в ритейлі систем бізнес-аналітики в першу трійку увійшли аналіз продажів (93,2% компаній, що використовують), фінансовий аналіз і на третьому місці майже порівну маркетинг і закупівлі. Найрідше можливості інтелектуального аналізу даних використовують у галузі безпеки та мерчандайзингу. Дані ранговані за категоріями. Також серед тих, хто ще не використовує у себе системи бізнес-аналітики, але планує їхнє впровадження, запитали, в якій сфері вони планують їхнє використання. Результати багато в чому збіглися з відповідями тих, хто використовує системи аналізу даних у своїй роботі. На малюнку 3 відображено співвідношення респондентів, які вже

використовують систему бізнес-аналітики та лише планують використання у сферах застосування. Також зазначені відсотки від загальної кількості респондентів.

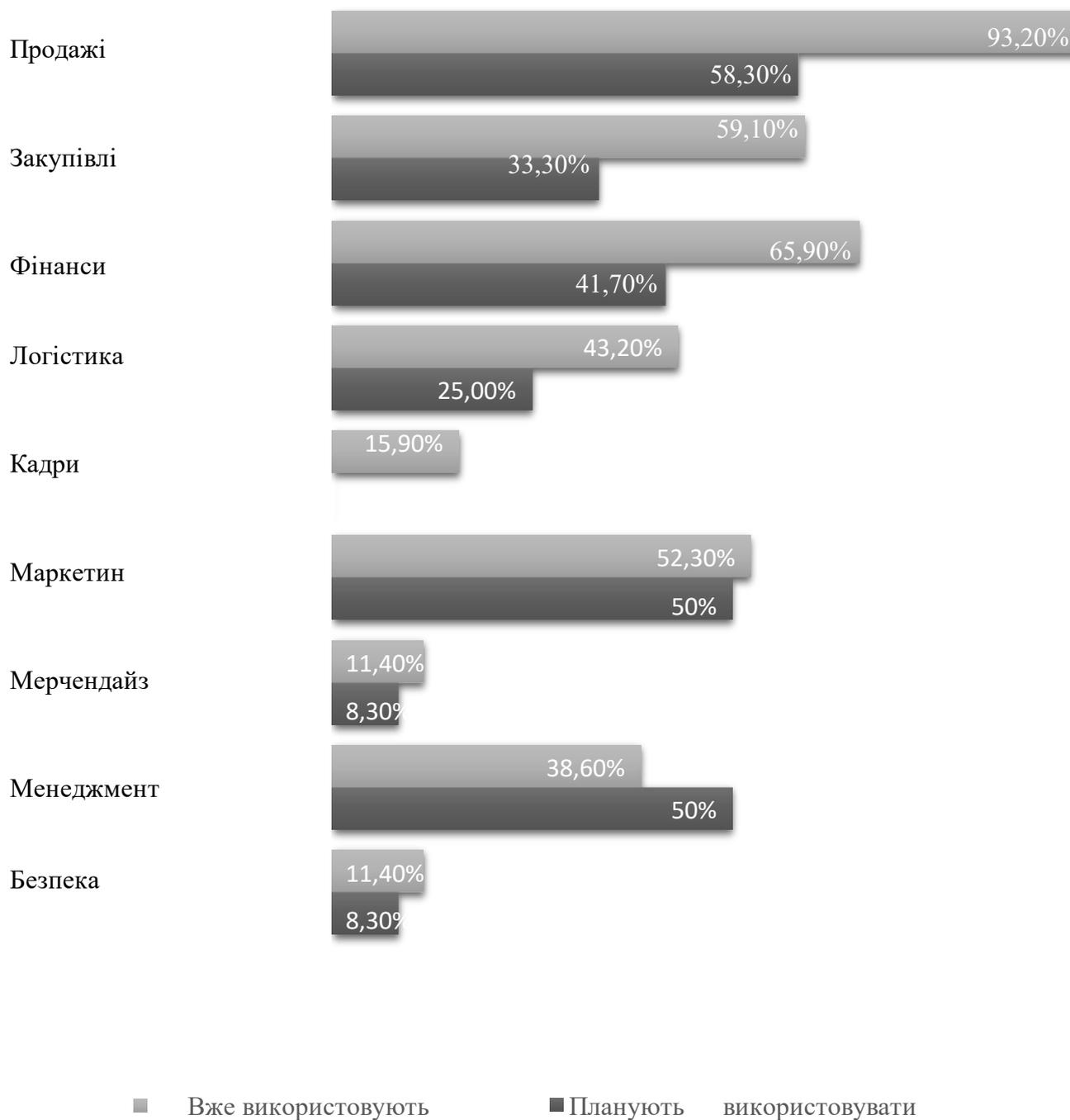


Рисунок 3 – Області використання аналітичних систем

Що стосується основних факторів вибору системи бізнес-аналітики, більшість віддають свою перевагу системам, що дозволяють підвищити швидкість і спростити створення звітів та роботу з ними (65,9%). У необхідності аналізу великих обсягів даних висловилося 59,1% опитаних. Немаловажним фактором також виявилася наявність у системі бізнес-аналітики наочної та простої для сприйняття візуалізації даних.



Рисунок 4 – Основні фактори вибору системи бізнес-аналітики

Також учасників опитування запитали, у яких функціях та можливостях вони відчують нестачу під час використання сучасних систем бізнес-аналітики. Найгостріше респонденти відчули нестачу передиктивного інструментарію, що дозволяє прогнозувати бізнес-процеси на основі накопичених даних. Майже половина опитаних відчувають нестачу у засобах інтелектуального аналізу даних. У топ-5 також увійшли брак засобів візуалізації даних (нагадаю, що наявність у системі засобів якісної візуалізації є одним із головних факторів при виборі системи за результатами

цього ж опитування) та додатків, що працюють на мобільних платформах. Обґрунтованість необхідності мобільних додатків наголошувалась у попередньому абзаці. Результати цього опитування можуть свідчити про те, що багато користувачів систем бізнес-аналітики вже почали осягати весь потенціал можливостей своїх систем.

При цьому в системах бізнес-аналітики залишається величезний потенціал, так як поки що вони використовуються далеко не у всіх напрямках та сферах бізнесу, хоча їхні можливості набагато ширші.

1.5 Можливості застосування інтелектуального аналізу даних у сфері рітейлу

Рітейл є однією з сфер вітчизняної економіки, що найбільш динамічно розвиваються. Постійний розвиток вимагає пошуку нових способів збільшення своєї конкурентоспроможності на загальному ринку і швидкої реакції на умови, що змінюються.

Інтелектуальний аналіз даних є одним із таких способів, дозволяючи фірмам використовувати всю зібрану ними інформацію за своїми транзакціями, клієнтами, звітами та ін. для власного зростання та підвищення продуктивності.

Раніше аналітики ще могли обробляти показники фірми самостійно, але зараз це стало нетривіальним завданням із величезним зростанням накопиченої фірмами інформації та швидкістю її поповнення.

Наслідком цього стало активне впровадження інформаційно-аналітичних систем у галузі роздрібною торгівлі.

Необхідність оперативного отримання нових даних, що дозволяють збільшити ефективність та прискорити процес прийняття управлінських рішень, пояснює зростаючу популярність систем бізнес-аналітики.

Постійна безперервна зміна умов, у яких доводиться працювати бізнесменам, змушує їх шукати способи якнайшвидшого реагування на навколишнє оточення.

Однією з наочних прикладів використання інтелектуального аналізу даних у сфері роздрібно́ї торгівлі є аналіз купівельної кошика, аналізуючий залежності між товарами з погляду попиту них.

Очевидно, що для збільшення прибутку від продажу товари повинні бути розміщені в торговому залі не просто так, а правильно і обдумано. Системи інтелектуального аналізу даних дозволяють вирішити цю проблему шляхом пошуку супутніх товарів - де один об'єкт або набір об'єктів є основним покупцем, а інший - супутній йому, який з великою ймовірністю буде придбаний, якщо придбаний і перший. Такі групи супутніх товарів визначаються за допомогою аналізу купівельного кошика, шляхом аналізу минулих чеків та транзакцій фірми.

Інформація, отримана в результаті аналізу купівельного кошика, може використовуватись:

- маркетологами, збільшення продажів товару шляхом його просування через таргетированную рекламу;
- мерчендайзерами, для ефективного та грамотного розміщення товарів на торговій площі фірми;
- при формуванні цін, знижуючи ціну на основний товар і підвищуючи супутній і навпаки.

На жаль, у нашій країні представники малого бізнесу не помічають цінності інтелектуального аналізу даних на відміну від зарубіжних країн. Це видно хоча б за обсягом коштів, що виділяються на інформаційно-технічну інфраструктуру. Для багатьох бізнес-аналітика залишається чимось чужим та незрозумілим. А серед тих, хто все ж таки впровадив у себе технології інформаційного аналізу, далеко не всі повною мірою розкрили та використовують їх можливості. Найчастіше, аналітикам та ІТ-фахівцям доводиться самим доводити значущість бізнес-аналітики для підвищення виробничої ефективності.

На щастя, в останні роки ситуація почала змінюватися на краще, і все більше управлінців помічають переваги використання систем бізнес-аналітики.

Зараз аналіз даних, накопичених фірмою, умовно можна розділити на дві категорії:

- аналіз внутрішніх даних, накопичених фірмою протягом років своєї роботи. Сюди відносять усі звіти, реквізити, чеки, транзакції, анкети клієнтів, складський облік та інше;

- аналіз зовнішніх даних, сюди відносять інформацію про конкурентів та сферу діяльності в цілому, аналіз цільової аудиторії.

Зрозуміло, це лише основні напрямки, виділені дуже умовно.

Кожне з них можна розглядати більш детально та поглиблено.

Аналіз внутрішніх даних є сьогодні основним джерелом інформації для роздрібної торгівлі, на його частку припадає близько 90% усієї аналітики, яку вирішують засоби інтелектуального аналізу даних у сфері рітейлу.

Аналіз споживчого кошика та активності покупців є на сьогодні головним завданням аналітичних систем роздрібної торгівлі. Вони залишаться такими і в найближчому майбутньому, лише зміцнивши свої позиції, ставши незамінною складовою бізнес-аналітики в роздрібній торгівлі.

1.6 Методи інтелектуального аналізу даних

Головна перевага методів Data Mining полягає в тому, що вони поєднують у собі як математичний інструментарій (починаючи від статистичного аналізу та закінчуючи новітніми евристичними методами), так і останні розуміння в галузі IT. У методах інтелектуального аналізу даних переплелися формалізовані та неформальні способи та техніки аналітики, різні способи аналізу даних.

Методи Data Mining є логічним продовженням та розвитком методів математичної статистики, у зв'язку з чим багато її методів знайшли тут застосування.

Основними методами інтелектуального аналізу даних є насамперед методи, що ґрунтуються на переборі. Звичайний перебір всіх варіантів займає $O(2^N)$ операцій (де N - загальна кількість об'єктів), а це означає, що зі зростанням числа об'єктів перебору обчислювальна складність зростає експоненційно, що при значній кількості об'єктів робить вирішення такого завдання практично нездійсненним.

Щоб зменшити кількість переборів і знизити обчислювальну складність таких алгоритмів, методи інтелектуального аналізу даних використовують різні евристичні підходи.

Однією з переваг методів інтелектуального аналізу даних є їхня легкість як для сприйняття, так і для програмної реалізації. З мінусів методів Data Mining можна назвати відсутність формалізованого математичного апарату і суворо оформленої теорії виходячи з яких будуються методи, і з цього випливає – проблеми у розвитку.

Перерахуємо основні типи знань, які здобуваються за допомогою методів інтелектуального аналізу:

- асоціативні правила;
- дерева рішень;
- кластери;
- математичні функції.

Методи пошуку цих знань використовують напрацювання різних сфер науки, включаючи теорію ймовірностей, математичний аналіз, статистичний аналіз, теорію множин, нейронні мережі, генетичне та еволюційне програмування.

Багато технологій, що використовуються в інтелектуальному аналізі даних, вже давно відомі і широко використовуються в математиці. Однак у сфері Data Mining вони знаходять нове застосування, що відкриває раніше невідомі можливості, що в першу чергу пов'язане зі зростанням рівня технічних та програмних засобів за останні десятиліття. Багато методів

інтелектуального аналізу даних працюють в рамках теорії штучного інтелекту. Основні завдання, які вирішуються методами інтелектуального аналізу даних, це регресійний аналіз, завдання класифікації, завдання кластеризації та пошук асоціативних правил в базах даних.

Ці завдання прийнято ділити на описові та предиктивні.

Описові завдання спрямовано на вдосконалення осмислення інформації. Дані, що отримуються в результаті роботи над цими завданнями, повинні бути прості і зрозумілі для людини. До цього виду належать пошук асоціативних правил у базах даних та завдання кластеризації.

Завдання класифікації та регресійний аналіз відносять до предиктивних завдань. Для їхнього вирішення використовуються методи, створені задля аналізу поточної інформації з метою прогнозу у майбутньому.

2 Проектування автоматизованої інформаційної системи

2.1 Постановка завдань та аналіз існуючих рішень

Перш ніж приступати до аналізу існуючих рішень і постановки завдань, перерахуємо вимоги до проектованої інформаційної системи.

Раніше ми вже встановили, які вимоги пред'являються до систем аналітики в ритейлі, чого від них очікують користувачі та в яких функціях відчують нестачу.

На підставі цих даних виділимо основні критерії, яким повинна задовольняти проектована інформаційна система.

Серед основних стримувальних факторів для впровадження системи бізнес-аналітики респондентами було названо такі як: висока вартість, тривалий період впровадження, неструктурованість даних компанії. Проектована інформаційна система має бути максимально позбавлена цих недоліків.

По-перше, вона має поширюватися вільно. Це задовольнить майже половину (41,7%) респондентів, які сумніваються у виборі системи бізнес-аналітики через високу вартість більшості аналогів. Так як 20% опитаних були зацікавлені у швидкій окупності системи, вільне поширення

проектованої інформаційної системи так само буде перевагою, крім етапу окупності системи.

По-друге, необхідно максимально спростити етап впровадження, невдоволення тривалістю якого висловили чверть респондентів. Впровадження інформаційної системи, розробленої самостійно, нерідко призводить до зміни процесів роботи, що вже склалися на підприємстві. Доводиться змінювати їх відповідно до стандартів та принципів роботи інформаційної системи. Незважаючи на те, що впровадження інформаційної системи допомагає вирішити багато управлінських завдань, основною проблемою залишається людський фактор. Крім того, при впровадженні ІС на виробництво необхідно обов'язково навчити персонал роботі з цією системою, але, нерідко трапляється, що персонал не дуже хоче переучуватися. Зміна старих звичок та розвиток нових – довгий та важкий процес. На підставі всього вищесказаного, для проекрованої ІС необхідно максимально спростити взаємодію з кінцевим користувачем, зробити його інтуїтивно зрозумілим, а призначення функцій та елементів інтерфейсу програми очевидними та однозначними. Це дозволить значно скоротити етап підготовки та навчання співробітників роботі з ІС, і як наслідок – скоротити час її впровадження.

По-третє, інформаційна система повинна мати власні інструменти роботи з даними компанії, зберігати та обробляти їх у чітко структурованому вигляді. До цього можна пред'явити такі основні вимоги:

- дані мають бути певним чином структуровані, це підвищить ефективність та швидкість пошуку;
- дані мають зберігатися у одному місці;
- інформація, що зберігається, повинна бути корисною в даний момент, або мати перспективи застосування в майбутньому;
- має бути реалізована можливість фільтрації даних;
- нова інформація має записуватися максимально просто та швидко.

Серед тих самих учасників опитування майже половина (44,2%

респондентів)

стурбовані нестачею інструментарію інтелектуального аналізу даних у сучасних системах бізнес-аналітики. Проектована ІС міститиме у собі можливість для інтелектуального аналізу даних у вигляді пошуку асоціативних правил з урахуванням транзакцій підприємства, так як це дозволить у перспективі підвищити рентабельність бізнесу. Дане рішення є особливо актуальним і затребуваним, враховуючи, що 93,2% бізнес-аналітики, які вже впровадили у себе системи, використовують їх саме для аналізу продажів, а серед тих, що тільки планують впровадження, більше половини (58,3%) збираються використовувати їх з тією ж метою.

Отже, позначимо основні функції, якими має володіти проєктована ІС:

- система повинна містити базу даних транзакцій фірми і надавати користувачеві інструментарій для роботи з цією базою;
- система має містити засоби інтелектуального аналізу даних, а саме пошуку асоціативних правил у базі даних.

Для проєктування та розробки ІС необхідно конкретизувати ці функції та грамотно поставити завдання, що й буде зроблено далі.

Проведений раніше аналіз області дозволив сформулювати основні вимоги до ІС на основі актуальних потреб користувачів, їхнього досвіду та побажань. Для подальшої роботи необхідно провести аналіз вже існуючих на ринку продуктів, виявити їх плюси та мінуси, що дозволить об'єднавши переваги наявних аналогів та виправити їх недоліки створити конкурентоспроможну інформаційну систему.

2.2 Огляд існуючих рішень

Ринок програмного забезпечення засобів інтелектуального аналізу даних представлений безліччю варіантів. З кожним роком їхня кількість лише зростає, але, на жаль, не завжди зростає якість.

Сьогодні найбільш популярними на ринку є такі продукти як Oracle Data Mining, STATISTICA Data Miner, Microsoft SQL Server Analysis Services та інші. Основні постачальники готових товарів були перераховані раніше

під час аналізу ринку.

Розглянемо можливості найпопулярніших на сьогоднішній день продуктів.

SAS Enterprise Miner. Програма американської компанії включає широкий інструментарій для інтелектуального аналізу даних, включаючи такі поширені на сьогоднішній день методи інтелектуального аналізу даних, як:

- дерева рішень;
- нейронні мережі;
- регресійний аналіз;
- кластеризацію;
- пошук асоціативних правил;
- секвенціальний аналіз.

STATISTICA Data Miner. Продукт компанії СтатСофт, про яку я згадував під час огляду наукової літератури. За роки свого існування платформа STATISTICA обросла безліччю різноманітних модулів, що реалізують різні методи та технології аналітики та статистики. Дуже важко не загубитися у такому розмаїтті. Програмний продукт включає роботу з нейронними мережами, побудова та аналіз різноманітних моделей, регресійний аналіз, дерева рішень, методи кластеризації, класифікації, асоціації, широкий набір засобів візуалізації. Основні компоненти програми забезпечують роботу з такими завданнями:

- класифікація;
- моделювання;
- передиктивний аналіз;
- нейронні мережі;
- пошук асоціативних правил;
- побудова різноманітних дерев рішень.

KXEN Analytic Framework. KXEN Analytic Framework є своєю платформою для різних модулів, що дозволяють проводити описовий та

предиктивний аналіз. Продукт використовує регресійні алгоритми, дозволяє виявити природні кластери у наборі даних, дозволяє проводити бінарну класифікацію, дозволяє проводити передиктивний аналіз.

Microsoft SQL Server Analysis Services. Продукт компанії Microsoft, як і раніше перелічені продукти, включає основні технології для інтелектуального аналізу баз даних на своїй же платформі.

Програмні продукти SPSS Вони дозволяють будувати різноманітні дерева рішень, проводити кластерний аналіз, знаходити асоціативні правила у базах даних та проводити передиктивний аналіз. Містить більшість технологій, які надаються його прямими конкурентами.

Порівняльний аналіз методів Data Mining, які є у найбільш поширених продуктах отримання знань можна узагальнити у таблиці 2.

Таблиця 2 – Методи Data Mining у програмних продуктах

	SAS Enter- prise Miner	STATI- STICA Data Miner	Oracle Data Mining	KXEN Analytic Frame- work	Micro- soft SQL Server Analysis Services	ПП SPSS	Зустрі ч- чайніс ть
Адаптивна Байєсівська мережа			+				1
Аналіз тимчасових рядів		+		+	+	+	4
Граничні методи		+	+	+			3
Дерева рішень	+	+			+	+	4

Продовження таблиці 2

	SAS Enterprise Miner	STATISTICA Data Miner	Oracle Data Mining	KXEN Analytic Framework	Microsoft SQL Server Analysis Services	ПП SPSS	Зустріч- чайніс- ть
Ієрархічна кластеризація	+	+	+	+			4
Лінійна регресія	+	+	+	+	+	+	6
Логістична регресія	+	+	+	+	+	+	6
Наївний Байєсовський алгоритм			+		+	+	3
Неієрархічна кластеризація	+	+	+	+	+		5
Нейронні мережі	+	+			+		3
Пошук асоціативних правил	+	+	+		+		4

Як видно з таблиці 2, обране нами для реалізації в ІВ рішення реалізувати пошук асоціативних правил є одним із найбільш вдалих. З одного боку, дана технологія зустрічається досить часто, щоб вважатися затребуваною, але не так часто реалізується програмно.

2.3 Постановка завдань

У своїй роботі я поставив за мету спроектувати та реалізувати інформаційну систему, що забезпечує інформаційну підтримку представників малого підприємництва шляхом надання їм

можливостей доступу до методів та технологій інтелектуального аналізу даних.

Провівши аналіз предметної області, вивчивши вимоги кінцевих користувачів стан ринку програмних продуктів такої призначення, можна поставити більш конкретні завдання.

Таким чином переді мною була поставлена задача спроектувати та реалізувати інформаційну систему, що вирішує певні завдання та виконує конкретно зазначені функції. А саме:

- система повинна мати можливість роботи з базою транзакцій фірми. Повинні бути реалізовані можливості створювати БД, додавати до неї нові транзакції, редагувати та видаляти існуючі, проводити вибірку з БД за заданими умовами;

- система має містити засоби інтелектуального аналізу даних, а саме пошуку асоціативних правил у базі даних. Повинні бути реалізовані пов'язані з цим підготовка та перетворення даних;

- користувач повинен мати можливість отримувати результати роботи програми як у вигляді звіту, так і за допомогою наочної та зрозумілої візуалізації.

Проектована ІС має бути виконана максимально просто і зрозуміло для користувача, реалізуючи досягнення поставлених завдань, не ускладнюючи себе зайвими функціями. Інтерфейс має бути інтуїтивно зрозумілим.

2.4 Обґрунтування та вибір способу вирішення зазначеної проблеми

З урахуванням поставлених цілей та завдань, необхідно спроектувати систему, що надає користувачеві можливості ведення бази транзакцій компанії та отримання з неї прихованих корисних даних через використання методів інтелектуального аналізу даних.

При розгляді вимог, що висуваються до системи, мною були перераховані основні вимоги як до самої системи, так і до структури інформації, яку потрібно обробляти проектною системою. Це веде до необхідності використання проектною ІС максимально простої вбудованої БД, що відповідає поставленим вимогам. Застосування вбудовуваної СУБД дає проектованій ІС такі переваги:

- система управління базою даних пов'язана з прикладним додатком та працює на тій же робочій станції, не вимагаючи спеціального адміністрування;

- відсутня програма-сервер. Це значно полегшує роботу з БД, так як до неї все одно не потрібен доступ із кількох робочих станцій;

- висока швидкість роботи та невелике споживання ресурсів системи. Завдяки спеціалізованому АРІ кількість операцій читання-запису є мінімальною.

При постановці завдань було наведено високорівневий опис основних функцій системи, обґрунтовано їх вибір.

Вибір методу інтелектуального аналізу даних, що застосовується в системі, був обґрунтований під час огляду існуючих рішень. Для цієї роботи було обрано спосіб пошуку асоціативних правил у основі транзакцій фірми.

2.5 Високорівневий опис алгоритму

Основні функції проектною ІС: ведення бази даних, що містить відомості про транзакції фірми; пошук асоціативних правил у базі даних.

Робота з базою даних, що містить відомості про транзакції фірми, здійснюватиметься за допомогою SQL запитів програми до своєї вбудованої бази даних, що містить транзакції фірми. При виборі користувачем певної дії, спрямованої на роботу з базою даних (таких як додавання нового запису або будь-яка інша дія) програма формує запит мовою SQL до вбудованої СУБД, спрямований на здійснення

обраної користувачем дії. Застосування СУБД, що вбудовується, було обґрунтовано раніше.

Серед дій, спрямованих на роботу користувача з базою, ІС має містити такі:

- створення нової БД;
- відкриття вже існуючої БД;
- додавання нової транзакції до БД;
- редагування існуючої транзакції у БД;
- видалення із БД існуючої транзакції;
- відображення всього вмісту БД як наочної таблиці;
- фільтрація відображуваного вмісту за номером транзакції;
- фільтрація відображуваного вмісту за найменуванням товарів;
- фільтрація відображеного вмісту за датою транзакції.

Перерахованих дій достатньо для приведення інформації фірми в структурований вигляд та забезпечення користувача комфортної роботи.

Під час створення нової бази даних користувач вказує її ім'я. Програма посилає SQL запит вбудовуваній СУБД, що створює БД із двома таблицями у ній: перша таблиця зберігає всі транзакції фірми, призначення другий буде сказано пізніше під час розгляду алгоритму пошуку асоціативних правил.

При відкритті БД програма перевіряє наявність у ній необхідних таблиць, у разі відсутності, створює їх.

При додаванні нової транзакції користувач вводить у вікно програми такі дані: список товарів та дату здійснення транзакції.

За промовчанням встановлюється поточна дата, але користувач має право змінити її на власний розсуд.

Після того як користувач підтвердив введення даних.

Програма надсилає запит до БД, що містить введені користувачем дані.

У таблицю транзакцій заноситься рядок, що містить унікальний номер транзакції, список товарів та дату здійснення транзакції.

Унікальний номер надається транзакції автоматично без участі

користувача і не може бути ним змінено.

У разі редагування транзакції користувач спочатку вказує унікальний номер транзакції, яку потрібно відредагувати. Потім у вікні програми користувач вносить зміни до потрібних полів. Після підтвердження програма надсилає запит до БД та запис оновлюється. Номер транзакції у своїй не змінюється.

Видалення транзакції відбувається за унікальним номером транзакції. Користувач вказує номер, після чого програма надсилає до БД запит, видаляючи необхідну транзакцію.

Вміст БД відображається в основному вікні програми. Кожна дія користувача, що призводить до зміни вмісту бази, відразу відображається у вікні програми.

Під час фільтрування вмісту бази в основному вікні програми відображаються лише записи, які відповідають вибраним умовам. При фільтрації за номером користувач вказує інтервали. Транзакції, номери яких потрапляють у вказаний інтервал, відобразатимуться в основному вікні програми. Фільтрування за датою відбувається схожим чином, тільки замість інтервалу номерів користувач вказує інтервал дат. При фільтрації товару користувач пише в полі програми найменування необхідного товару. Усі транзакції, що містять цей товар, будуть відображені в основному вікні програми.

Алгоритм пошуку асоціативних правил буде докладно розглянуто в одному із наступних пунктів. Що стосується опису способу його роботи в проєктованій ІС – в ній реалізується інструментарій, що дозволяє користувачеві виконати пошук асоціативних правил та отримати результат у наочному та зрозумілому вигляді.

Алгоритм через свої особливості повинен працювати з особливим форматом даних, а саме – БД транзакцій має бути перетворена на файл, що містить лише список транзакцій, без зазначення номерів транзакцій та дати

їх вчинення. Найменування товарів мають бути замінені цілими кодами. Використання цілих даних замість рядкових дозволяє колосально збільшити швидкість роботи алгоритму і скоротити кількість необхідних обчислювальних ресурсів. Для перетворення бази транзакцій на цілий формат, кожному товару, що міститься в базі, повинен бути присвоєний унікальний номер. Це відбувається автоматично за допомогою другої таблиці, що містить список товарів та їх унікальні номери. При додаванні нової транзакції до бази програма перевіряє, чи є введені товари в таблиці товарів. Якщо програма знаходить новий товар, який раніше не зустрічався в базі, вона додає його до таблиці товарів і надає йому унікальний номер. Таблиця товарів змінюється лише програмою, і користувач не може отримати доступ до неї.

Користувач повинен вказати програмі шляху до двох файлів: вхідного, що містить список транзакцій у цілісному конвертованому форматі; вихідному, куди буде записано знайдені правила.

Користувач обов'язково має вказати програмі два основні параметри: мінімальну підтримку та мінімальну достовірність. Якщо користувач не вкаже потрібні значення, використовуються значення за промовчанням.

Крім обов'язкових основних, користувач може також вказати необов'язкові додаткові параметри: мінімальний ліфт та максимальну підтримку.

Після вибору всіх потрібних параметрів користувач повинен запустити процес пошуку асоціативних правил шляхом вибору цієї дії в інтерфейсі програми. Процес пошуку асоціативних правил відбувається автоматично і вимагає втручання користувача. Після завершення процедури пошуку користувачеві з'явиться список знайдених правил.

2.6 Пошук асоціативних правил

Для вирішення задачі пошуку асоціативних правил у базах даних, а також для вирішення будь-якого іншого завдання, необхідно обробити вихідні дані та отримати результат. Результати пошуку подаються у формі асоціативних правил. При їх пошуку прийнято виділяти два основні етапи:

- пошук усіх найпоширеніших наборів об'єктів;
- пошук асоціативних правил за знайденими найпоширенішими наборами об'єктів.

Опишемо завдання пошуку всіх найпоширеніших наборів об'єктів у загальному вигляді. Об'єкти, що складають досліджувані набори (itemsets), означимо як множину:

$$I = \{i_1, i_2, \dots, i_j, \dots, i_n\},$$

Де i_j – об'єкти, які входять у набори;

N – загальна кількість об'єктів.

У сфері рітейлу, наприклад, такими об'єктами є товари, що продаються фірмою. Набори об'єктів, що складаються з множини I , що зберігаються в базі даних і згодом аналізовані програмою, називаються транзакціями. Транзакцію можна представити як підмножину множини I :

$$T = \{i_j \mid i_j \in I\}.$$

Транзакції в рітейлі являють собою набори товарів, придбані покупцем одноразово. Транзакції зберігаються у базі даних фірми. Набір транзакцій, що зберігається в базі даних та доступний нам для аналізу ми позначимо множиною:

$$D = \{T_1, T_2, \dots, T_r, \dots, T_m\},$$

Де m – кількість доступних для аналізу транзакцій

Множину транзакцій, до яких входить об'єкт i_j , позначимо наступним чином:

$$D_{ij} = \{T_r \mid i_j \in T_r; j = 1..n; r = 1..m\} \subseteq D.$$

Деякий довільний набір об'єктів (itemset) позначимо наступним чином:

$$F = \{i_j \mid i_j \in I; j = 1..n\}.$$

Набір, що складається з об'єктів k , називається k -елементним набором. Множину транзакцій, до яких входить набір F , позначимо наступним чином:

$$D_F = \{T_r \mid F \subseteq T_r; r = 1..m\} \subseteq D.$$

Відношення кількості транзакцій, до якої входить набір F , до загальної кількості транзакцій називається підтримкою (support) набору F та позначається $\text{Supp}(F)$:

$$\text{Supp}(F) = \frac{|D_F|}{|D|}.$$

Під час пошуку асоціативних правил у базі даних аналітик вказує значення мінімальної підтримки для наборів об'єктів Supp min . Якщо значення підтримки набору більше, ніж мінімальне значення підтримки Supp min вказане аналітиком – то такий набір прийнято називати таким, що часто зустрічається.

Вирішуючи завдання пошуку асоціативних правил потрібно перш за все знайти таку множину найпоширеніших наборів.

Асоціативні привали зазвичай мають такий вигляд «якщо (умова), то (Результат)». Тут «умова» – набір будь-яких об'єктів з множин об'єктів I , з яким безпосередньо пов'язаний «результат», що також складається з об'єктів множини I . Асоціативне правило можна представити як імплікацію над множиною I . Головною перевагою асоціативних правил є їх наочність і

простота як сприйняття людиною, так програмної реалізації та інтерпретації. Однак, не всі виявлені в процесі пошуку правила несуть у собі користь. Прийнято виділяти три типи асоціативних правил:

- Корисні правила. Корисні правила містять у собі інформацію, яка при її застосуванні принесе очевидну користь. Ці правила раніше не були відомі аналітикам і були відкриті в процесі інтелектуального аналізу даних, але ці правила логічно зрозумілі;
- тривіальні правила. Тривіальні правила містять у собі інформацію, яка є істинною, але вже відомою та очевидною аналітикам. Ці правила легко пояснити з логічного погляду, але їх використання не принесе відчутної користі, так як більшість цих правил вже використовуються через їхню популярність. Іноді такі правила можуть застосовувати для перевірки правильності прийнятих рішень;
- Незрозумілі правила. Незрозумілі правила містять інформацію, яка не може бути пояснена аналітиками. Такі правила можуть бути отримані в результаті аналізу глибоко прихованих знань або в результаті аналізу аномальних та нестандартних значень. Такі правила не можна використовувати в чистому вигляді через неможливість пояснити їх з логічного погляду, через це їхнє застосування також призведе до невідомих та непередбачуваних результатів.

Для генерації самих асоціативних правил необхідно попередньо знайти всі найпоширеніші набори. Внаслідок великої кількості таких наборів при великій кількості об'єктів, кількість згенерованих асоціативних правил також може бути занадто великою для розуміння людиною. До того ж, як було зазначено раніше, не всі асоціативні правила корисні. Для скорочення числа правил та відсіювання марних вводяться додаткові величини.

Підтримка (support) – показує, який відсоток транзакцій підтримує це правило. Так як правило будується на підставі набору, то, значить, правило

має підтримку, що дорівнює підтримці набору.

$$\text{Supp}_{X \Rightarrow Y} = \text{Supp}_F = \frac{|D_{F=XUY}|}{|D|},$$

Де X- набір об'єктів, X належить I. Y- набір об'єктів Y належить I.

Очевидно, що правила, побудовані на підставі одного і того ж набору, мають однакову підтримку.

Достовірність (confidence) – показує можливість, що з наявності у транзакції набору X слідує наявність у ній набору Y. Достовірністю правила є відношення числа транзакцій, що містять набори X і Y, до числа транзакцій, що містять набір X:

$$\text{Conf}_{X \Rightarrow Y} = \frac{|D_{F=XUY}|}{|D_X|} = \frac{\text{Supp}_{XUY}}{\text{Supp}_X}$$

Очевидно, що чим більше достовірність, тим правило краще, причому у правил, побудованих на підставі одного і того ж набору, достовірність буде різна.

Достовірність дозволяє відсіяти частину слабких правил, але не показує, наскільки те чи інше правило є корисним. Для визначення корисності правила було запроваджено такий захід, як ліфт правила (оригінальна назва – інтерес, також зустрічається термін «покращення»).

Ліфт правила показує, чи випадкове правило чи воно має під собою логічне обґрунтування. Ліфт правила є відношенням числа транзакцій, що містять набори X та Y, до добутку кількості транзакцій, що містять набір X, та кількості транзакцій, що містять набір Y:

$$\text{Lift}_{X \Rightarrow Y} = \frac{|D_{F=X \cup Y}|}{|D_X| \cdot |D_Y|} = \frac{\text{Supp}_{X \cup Y}}{\text{Supp}_X \cdot \text{Supp}_Y}$$

Якщо значення ліфта правила більше одного, це означає, що правило ймовірніше має під собою логічне обґрунтування, ніж отримано випадково. Якщо ліфт менше одиниці - з великою ймовірністю можна говорити про те, що правило є всього лише випадковим збігом. Ліфт показує наскільки асоційовані між собою об'єкти правила.

Найчастіше кількість одержуваних у результаті роботи будь-якої програми правил заздалегідь обмежується аналітиком для спрощення аналізу та сприйняття. Зазвичай задаються мінімальні та максимальні значення підтримки та достовірності. Якщо встановити занадто великі значення обмеження, то буде знайдено тривіальні та добре відомі правила. Якщо встановити занадто низькі значення - буде згенеровано величезну кількість правил, більшість з яких будуть незрозумілі і необґрунтовані, і знайти в цьому морі корисні правила буде завданням, що важко здійснити. Отже, аналітику необхідно підібрати оптимальні значення обмежень, які призведуть до генерації корисних та необхідних асоціативних правил.

Алгоритм Apriori – найпоширеніший алгоритм пошуку асоціативних правил у базах даних. Алгоритм був вперше описаний у 1994 році і з того часу вийшло безліч його модифікацій.

Перший етап пошуку асоціативних правил – пошук всіх найпоширеніших наборів – нетривіальне завдання, т.к. при порівняно великих обсягах даних вона вимагає чималих обчислювальних ресурсів та часу. Найпростіший спосіб її вирішення - простий перебір усіх можливих варіантів – при великому обсязі даних стає складним завданням, бо зі зростанням кількості об'єктів кількість всіх переборів зростає

експоненційно. Алгоритм Apriori використовує евристичний підхід для вирішення цього завдання. Одна з властивостей підтримки набору полягає в тому, що підтримка будь-якого набору не може бути більшою за підтримку будь-якої з його підмножин. Наприклад, підтримка 3-елементного набору $\{a, b, c\}$ буде завжди меншою або дорівнює підтримці 2-елементних наборів $\{a, b\}$, $\{a, c\}$, $\{b, c\}$. Справа в тому, що будь-яка транзакція, що містить $\{a, b, c\}$, також повинна містити $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, причому протилежне неправильно.

Ця властивість називається властивістю анти-монотонності і значно звужує простір пошуку, роблячи завдання знаходження всіх найпоширеніших наборів відносно легко здійсненним. Якби не властивість анти-монотонності – пошук асоціативних правил був би практично неможливим завданням.

Властивість анти-монотонності можна перефразувати трохи інакше: зі зростанням розмірності набору його підтримка або зменшується, або залишається незмінною. Зростати вона не може. Резюмуючи все вищесказане, можна сміливо стверджувати, що будь-який набір часто зустрічається тоді і тільки тоді, коли всі його підмножини також часто зустрічаються.

На початковому етапі роботи алгоритму Apriori знаходяться всі 1-елементні набори та підраховується їхня підтримка. На цьому етапі відбувається звернення до бази даних; надалі воно не буде потрібно. Набори, підтримка яких менша за мінімально допустиму – відсікаються.

Наступні кроки складаються з двох етапів:

- генерація потенційно найпоширеніших наборів (такі набори називаються кандидатами);
- підрахунок підтримки кандидатів.

Функція генерації потенційних кандидатів виглядає наступним чином. Кожен кандидат формується з допомогою розширення вихідного набору

додаванням до нього елемента іншого набору. Потім, виходячи з вищеприписаної властивості анти-монотонності відбувається видалення надлишкових правил. Набір видаляється зі списку кандидатів, якщо хоча б одна з його підмножин не відповідає вимогам мінімальної підтримки і не часто зустрічається.

Після того, як усі можливі кандидати були згенеровані, підраховується їхня підтримка. Кандидати, підтримка яких виявляється меншою за допустиме значення – відсікаються, інші ж використовуються в подальшій роботі алгоритму.

Коли етап пошуку всіх найпоширеніших наборів завершено, переходять до етапу пошуку асоціативних правил.

Генерація правил не така складна задача, як пошук наборів, що часто зустрічаються. Для обчислення достовірності правила необхідно знати підтримку набору та підтримку умови. Так як всі підмножини - часто зустрічаються, їх підтримка нам вже відома, що позбавляє нас необхідності кожного разу звертатися до бази даних.

Щоб згенерувати правило для множини, потрібно знайти всі його непусті підмножини. Правила генеруються для кожної підмножини, і правила, що задовольняють умову мінімальної достовірності, заданої аналітиком, – зберігаються. Правила, які мають достовірність нижче за мінімально допустиму – відкидаються. Тут слід зазначити одну важливу особливість: будь-яке правило, складене з певного набору, повинно містити всі елементи цього набору. Наприклад, якщо набір складається з елементів $\{a, b, c\}$, правило $a \wedge b$ не повинно розглядати

3 Програмна реалізація інформаційної системи

3.1 Робота з базою даних

У попередніх розділах роботи було поставлено завдання, які має вирішувати інформаційна система; визначено її функції, принципи побудови та роботи, обрано та обґрунтовано методи програмної реалізації. Відповідно до всієї проведеної роботи була програмно реалізована інформаційна система, що відповідає всім раніше поставленим вимогам.

Інформаційна система «Майстер транзакцій» була реалізована мовою програмування C++ з використанням кроссплатформенного інструментарію Qt. В якості бази даних була обрана SQLite – компактна встроєна реляційна база даних.

Програма надає користувачеві можливості роботи з базою даних, а саме:

- створювати нові бази даних;
- додавати, редагувати та видаляти транзакції з бази;
- проводити вибірку за базою з використанням фільтрів.

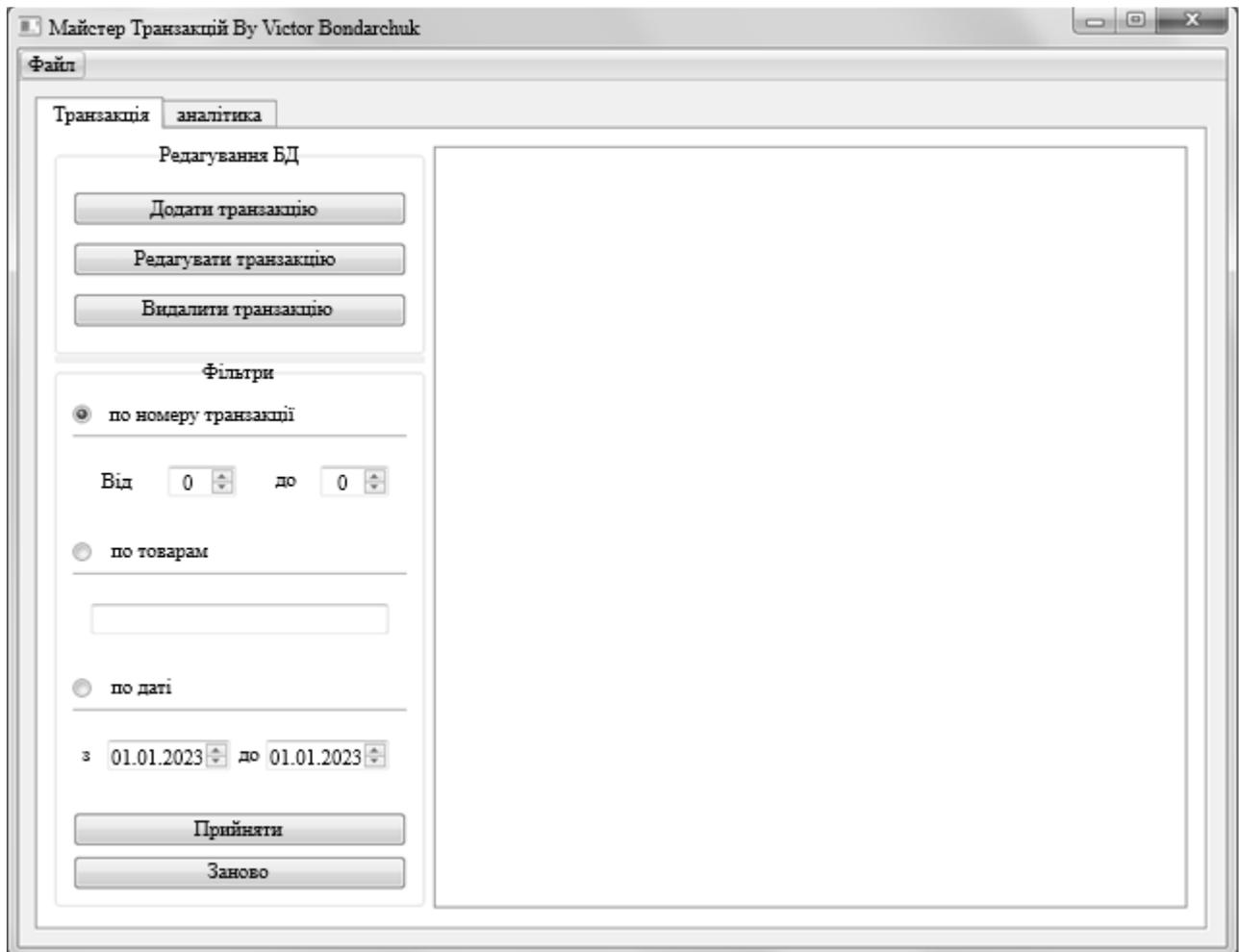
Також програма має аналітичний інструментарій, а саме:

- пошук асоціативних правил за базою даних;
- візуалізація динаміки продажу товарів.

На малюнку 5 зображено вікно програми після запуску. Основну область головного вікна програми займає об'єкт класу QTabWidget, що є набором віджетів із вкладками, по суті є контейнером. Вміст першої вкладки, названої

"Транзакції", дозволяє користувачеві працювати з базою даних. Друга вкладка, «Аналітика», яка впливає з назви, надає користувачеві доступ до аналітичного інструментарію інформаційної системи. Функції, представлені у цих вкладках були перераховані вище, трохи пізніше я докладно опишу

роботу кожної з них.



Малюнок 5 – Вікно програми

Щоб розпочати роботу, необхідно підключити базу даних. Користувач може відкрити існуючу базу даних або створити нову через пункт меню «Файл», що містить два підпункти: «Створити базу транзакцій» і «Відкрити базу транзакцій».

При виборі підпункту «Створити базу транзакцій» користувачу відкривається стандартне діалогове вікно вибору каталогу, де буде створено базу даних. Користувачеві залишається лише ввести ім'я нової бази та натиснути кнопку підтвердження.

При виборі підпункту «Відкрити базу транзакцій» користувачу відкривається стандартне діалогове вікно вибору файлу, де користувач пропонує вибрати базу даних для використання в програмі.

На малюнку 6 зображено процес відкриття вже існуючої бази даних.

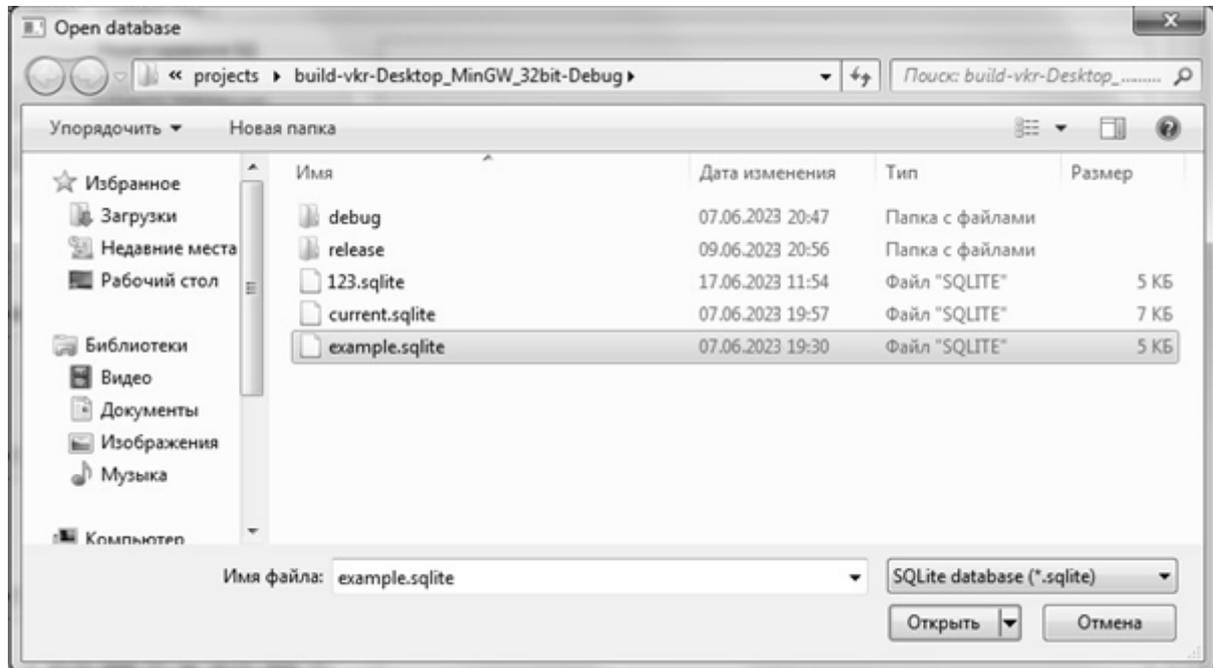


Рисунок 6 – Відкриття бази транзакцій

Програма працює з файлами з розширенням ".sqlite", яке є стандартним розширенням файлів для баз даних SQLite.

З'єднання з базою даних здійснюється за допомогою внутрішніх засобів Qt. При відкритті вже існуючої бази даних потрібно лише відкрити з'єднання з нею. При створенні нової бази ця база спочатку створюється за допомогою виконання SQL запити, потім зі створеною базою встановлюється з'єднання.

Редагування бази даних транзакцій відбувається за допомогою елементів керування, об'єднаних у групу "Редагування БД". Тут розташовуються три елементи управління, названі відповідно до виконуваних ним функцій:

- додати транзакцію;
- редагувати транзакцію;
- видалити транзакцію.

Ці дії виконуються за допомогою діалогових вікон, які викликаються при натисканні користувачем одного з цих елементів керування.

Для того, щоб додати транзакцію до бази даних, користувачеві необхідно натиснути кнопку «Додати транзакцію». Відкриється діалогове вікно, куди користувачеві потрібно буде ввести параметри транзакції, що додається. Це діалогове вікно зображено малюнку 7.

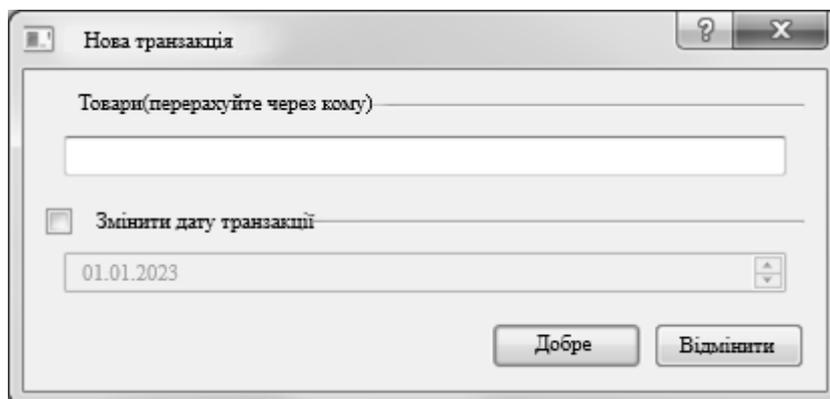


Рисунок 7 – Діалогове вікно додавання нової транзакції

У полі «Товари» користувач вводить список товарів транзакції, перераховуючи їх через кому. Програма сприймає всі символи, укладені між комами, як найменування одного товару. Далі користувач вводить дату здійснення транзакції, активувавши поле редагування дати вибором прапорця «Змінити дату транзакції». При натисненні кнопки «Відмінити», введена користувачем інформація нікуди не зберігається. Транзакція додається до бази даних після натискання кнопки «Добре».

Крім переліку товарів та дати здійснення транзакції, всі транзакції, що зберігаються в базі, також мають унікальний номер транзакції. Він надається транзакції автоматично і доступний користувачеві тільки для перегляду, змінити його засобами програми користувач не може.

Видалення транзакції відбувається за унікальним номером. Користувачеві відкривається вікно, де йому пропонується ввести номер транзакції, що підлягає видаленню. Зображення цього вікна наведено малюнку 8.

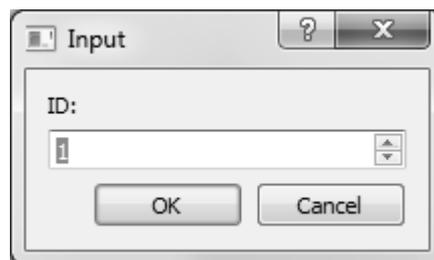


Рисунок 8 – Діалогове вікно видалення транзакції

При натисканні користувачем кнопки ОК транзакція з таким номером видаляється з бази даних.

Для того, щоб внести зміни до вже існуючої транзакції, користувачу необхідно натиснути кнопку «Редагувати транзакцію». Діалог з користувачем комбінує рішення, які застосовуються при видаленні та додаванні транзакції. Редагування відбувається за номером транзакції. Спочатку користувачеві відображається діалогове вікно, зображене на малюнку 8, де йому необхідно ввести номер транзакції, яку потрібно відредагувати. Потім користувачеві відображається діалогове вікно, зображене на малюнку 7, де він вводить нові необхідні дані.

Програмою передбачені випадки, коли транзакції під необхідним номером немає у базі даних, або ж база даних порожня. У цьому випадку користувачеві видається повідомлення про помилку з описом виниклої

проблеми, і програма автоматично повертається до попереднього стабільного стану.

При підключенні бази даних, її вміст негайно відображається у правій частині вікна програми. Для того щоб уточнити вибірку з бази, до неї можна застосувати наявні фільтри. Приклад цього зображено малюнку 9: відображаються лише ті транзакції бази, які серед переліку продуктів містять хліб.

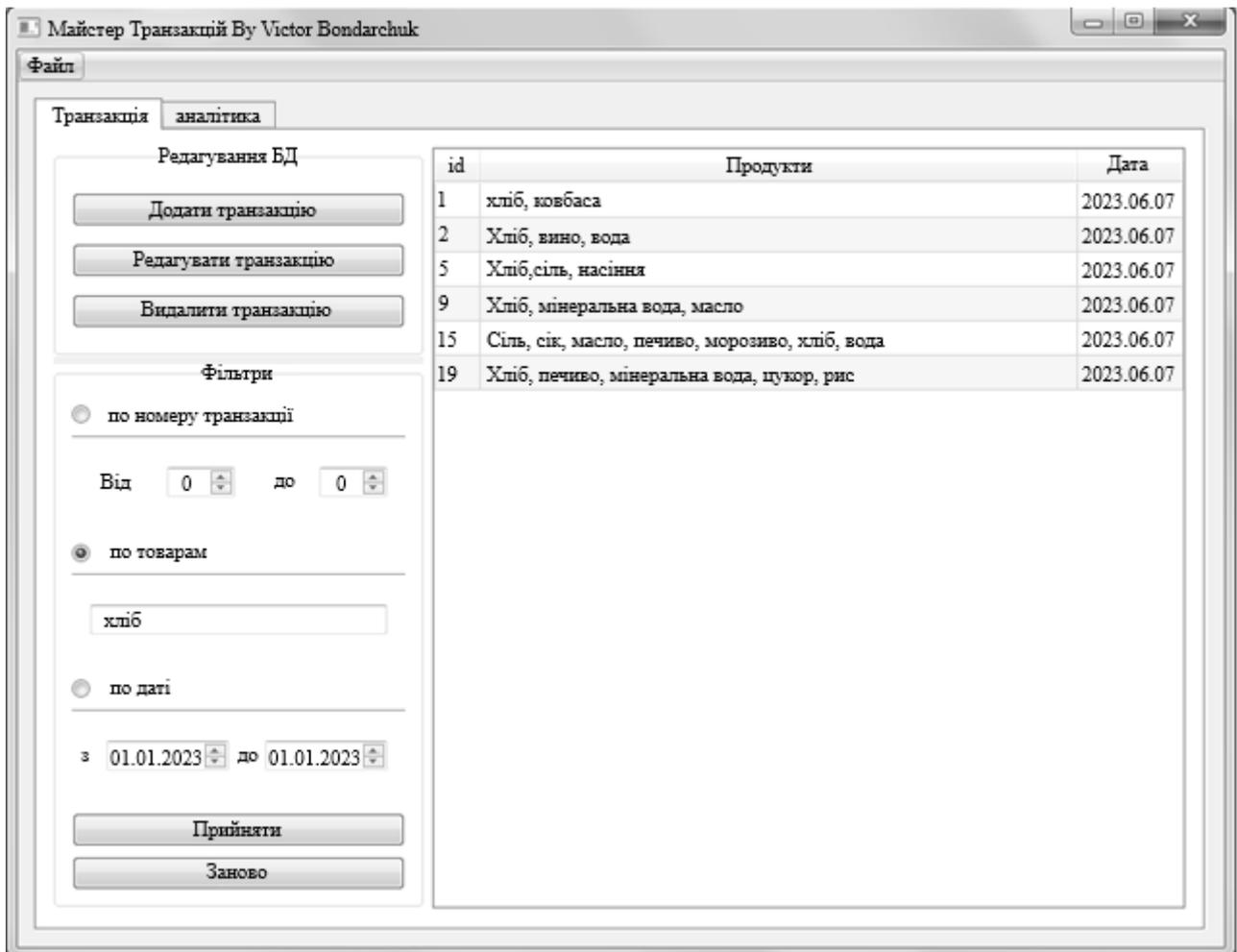


Рисунок 9 – Застосування фільтрів

Програмою передбачена можливість фільтрації транзакцій за трьома параметрами: за номером транзакції, за товарами та за датою.

При фільтрації номера транзакції користувач вказує необхідний інтервал. Усі транзакції, номери яких потрапляють до зазначеного

інтервалу, відображаються у вікні програми. Фільтрація за датою відбувається так само, але замість інтервалу номерів користувач вказує інтервал дат. При фільтрації товару відображаються всі транзакції, що містять вказаний користувачем товар. При натисканні кнопки «Заново» відбувається скидання всіх фільтрів та повернення до відображення повної бази транзакцій.

3.2 Інтелектуальний аналіз даних

Тепер розглянемо вкладку «Аналітика», зображену на малюнку 10:

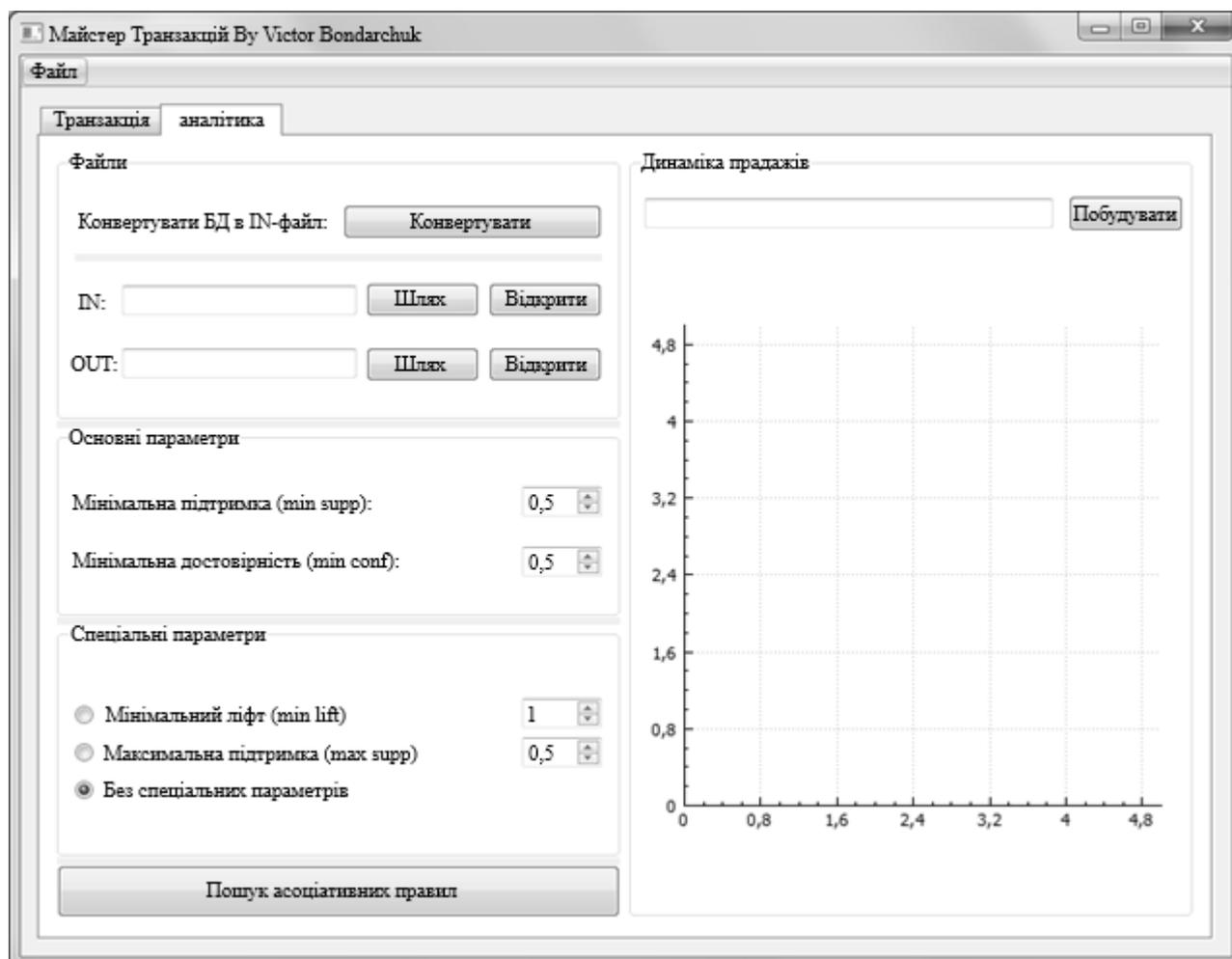


Рисунок 10 – Вкладка «Аналітика»

У лівій частині вікна розташовуються засоби настроювання пошуку асоціативних правил. А у правій – візуалізація динаміки продажу товарів.

Як згадувалося раніше на сторінці, при високорівневому описі алгоритму для своєї роботи алгоритму пошуку асоціативних правил потрібен особливий формат даних. Кнопка «Конвертувати» здійснює перетворення стандартної бази даних у вхідний файл, необхідний алгоритму. У ньому кожен рядок містить лише перелік товарів, без зазначення номера транзакції та дати її здійснення, причому найменування товарів представлені їх цілими кодами. Приклад вхідного файлу, сформованого програмою виходячи з стандартної бази даних, наведено нижче малюнку 11.

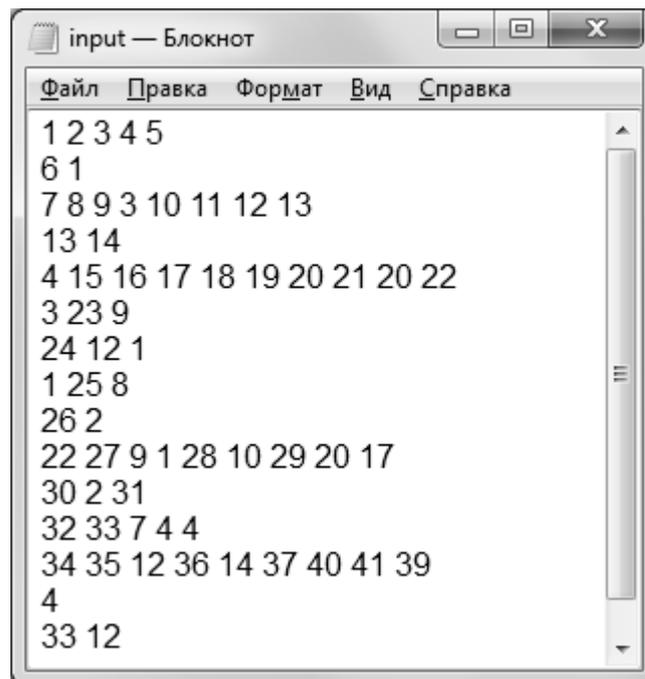


Рисунок 11 – Вхідний файл

У групі керуючих елементів «Файли» користувач вказує вхідний файл, що містить конвертовану базу транзакцій, і вихідний файл, в який будуть записані знайдені в результаті роботи алгоритму асоціативні правила.

При натисканні на кнопку «Шлях» користувачу відкривається стандартне діалогове вікно вибору файлу, де потрібно вибрати потрібний файл.

При натисканні на кнопку «Відкрити» відбувається відкриття файлу засобами ОС для перегляду його вмісту.

Якщо вхідний файл було конвертовано програмою, поле шляху вхідного файлу заповнюється автоматично.

Якщо користувач не вказав вихідний файл, він буде створений автоматично.

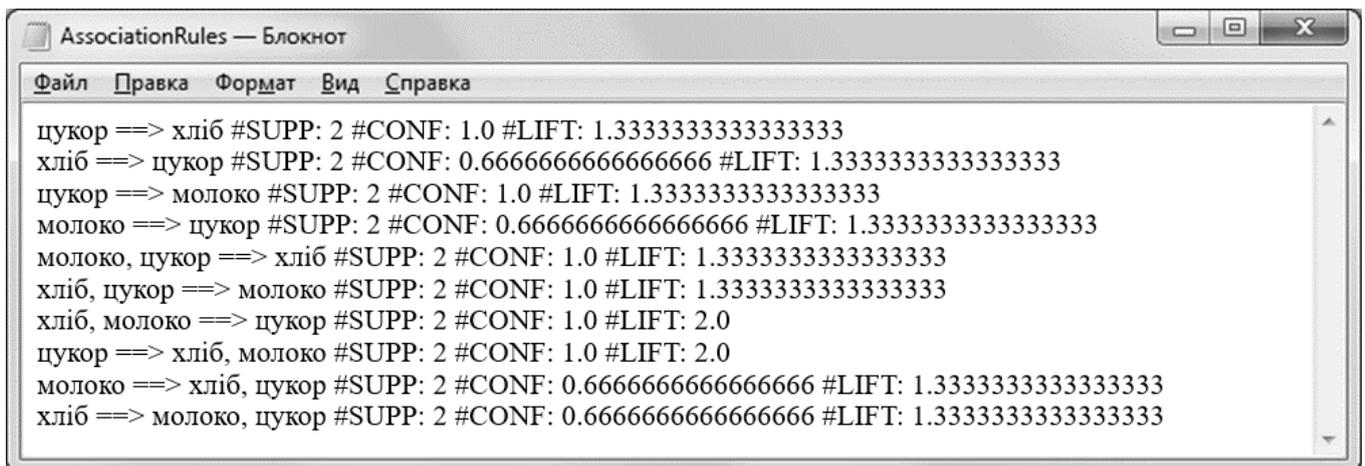
Всі файли створюються програмою в тому ж каталозі, де знаходиться файл, що виконується.

Група елементів «Основні параметри» містить елементи управління, що дозволяють користувачеві встановити основні обов'язкові параметри роботи алгоритму – мінімальну підтримку та мінімальну достовірність. Про сутність та призначення цих параметрів говорилося раніше у цій роботі. За замовчуванням значення встановлені на 0,5. Інтервал значень, які можна вказати – від 0,001 до 1.

Також програма дозволяє зазначити кілька додаткових параметрів роботи алгоритму. Це дозволяє знизити кількість асоціативних правил, відкинувши тривіальні та малокорисні правила. У цій інформаційній системі є можливість вибору мінімального ліфта чи максимальної підтримки. Про сутність та призначення цих правил також говорилося раніше. Значення ліфта за замовчуванням є 1, значення максимальної підтримки – 0,5. Вибір необхідного додаткового параметра роботи алгоритму здійснюється шляхом вибору відповідного перемикача групи елементів «Спеціальні параметри».

Програма має можливість будувати правила не лише за підключеною базою даних. Якщо потрібно аналіз будь-якої іншої бази даних, то необхідний лише її вхідний файл. Проте, у такому разі користувач навряд чи отримає корисну інформацію, так як знайдені асоціативні правила не можуть бути перетворені назад до текстового вигляду відповідно до кодів товарів, бо без самої бази зробити це не вийде, і знайдені асоціативні правила також будуть у вигляді цілих кодів.

Після того, як вибрано вхідний та вихідний файл та вказано параметри роботи алгоритму – можна приступати безпосередньо до пошуку асоціативних правил. Для цього користувачеві необхідно натиснути кнопку "Пошук асоціативних правил". Процедура відбувається автоматично та не вимагає участі користувача. Після закінчення процедури пошуку користувачеві буде представлений текстовий файл, який містить результати роботи програми. Вигляд цього файлу зображено малюнку 12.



```

AssociationRules — Блокнот
Файл  Правка  Формат  Вид  Справка
цукор ==> хліб #SUPP: 2 #CONF: 1.0 #LIFT: 1.3333333333333333
хліб ==> цукор #SUPP: 2 #CONF: 0.6666666666666666 #LIFT: 1.3333333333333333
цукор ==> молоко #SUPP: 2 #CONF: 1.0 #LIFT: 1.3333333333333333
молоко ==> цукор #SUPP: 2 #CONF: 0.6666666666666666 #LIFT: 1.3333333333333333
молоко, цукор ==> хліб #SUPP: 2 #CONF: 1.0 #LIFT: 1.3333333333333333
хліб, цукор ==> молоко #SUPP: 2 #CONF: 1.0 #LIFT: 1.3333333333333333
хліб, молоко ==> цукор #SUPP: 2 #CONF: 1.0 #LIFT: 2.0
цукор ==> хліб, молоко #SUPP: 2 #CONF: 1.0 #LIFT: 2.0
молоко ==> хліб, цукор #SUPP: 2 #CONF: 0.6666666666666666 #LIFT: 1.3333333333333333
хліб ==> молоко, цукор #SUPP: 2 #CONF: 0.6666666666666666 #LIFT: 1.3333333333333333

```

Рисунок 12 – Знайдені асоціативні правила

На малюнку добре видно самі правила у простій та зрозумілій формі, а також їх властивості (такі як підтримка, достовірність та ліфт). Як зазначалося раніше, такі правила мають вигляд «якщо (умова) – то (наслідок)». Наприклад, з результату роботи програми та знайдених правил випливає, що клієнти, які купили хліб і молоко з великою ймовірністю також придбають і цукор. Це правило має дуже добрий показник ліфта. Такі правила після аналізу можуть використовуватися підвищення рентабельності бізнесу. Способи практичного застосування асоціативних правил раніше неодноразово згадувалися та роз'яснювалися протягом усієї роботи.

Тепер розглянемо візуалізацію динаміки продажу. Для того, щоб подивитися, яку динаміку продажів має той чи інший товар, користувач вказує товар, що його цікавить, у полі групи керуючих елементів

"Динаміка продажів" і натискає кнопку "Побудувати". Наприклад, динаміка продажів хліба зображено малюнку 13.

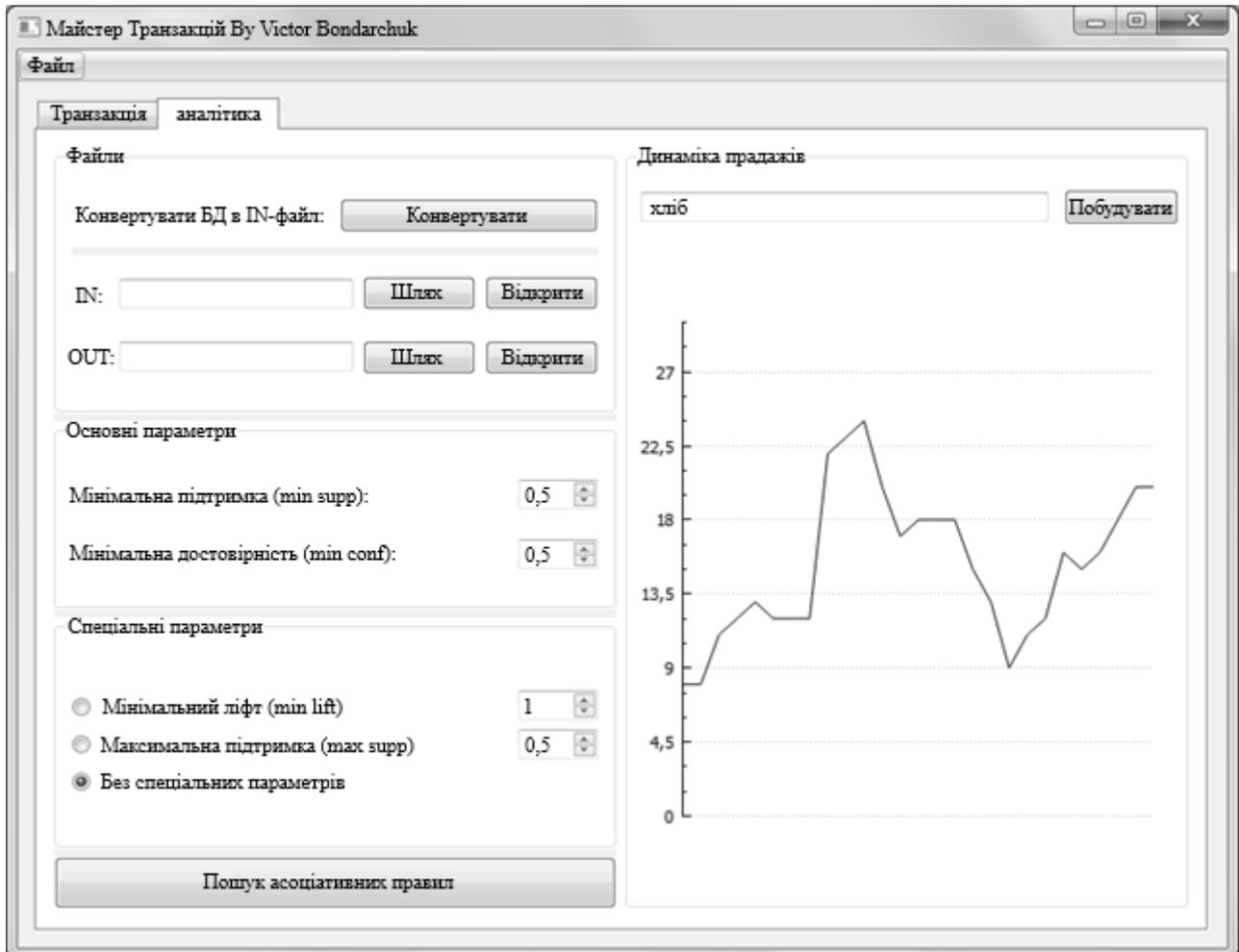


Рисунок 13 – Візуалізація динаміки продажів

Візуалізація динаміки продажу дозволяє користувачу оцінити кількість проданого товару у хронологічному порядку.

Як видно з усього вищесказаного, програма дуже проста у використанні, інтуїтивно зрозуміла і не вимагає наявності будь-яких спеціальних навичок або часу для навчання роботі з нею. Але при цьому вона повністю справляється з покладеними на неї завданнями.

ВИСНОВОК

У ході виконання роботи було вирішено всі поставлені завдання та досягнуто поставленої мети, а саме:

- створено програмний продукт, що не вимагає спеціальних навичок для роботи та максимально зрозумілий для користувача;
- користувачеві надано інструментарій для роботи з базою даних;
- до бази даних застосовані методи інтелектуального аналізу даних з метою аналізу накопиченої інформації для отримання нових та потенційно корисних знань.

Також було зроблено кілька основних висновків щодо предметної галузі:

- повсюдне використання комп'ютерів призвело до розуміння важливості завдань, пов'язаних із аналізом накопиченої інформації для отримання нових знань;
- сфера підприємництва зараз рішучіше, ніж будь-коли, налаштована на підвищення конкурентоспроможності своїх активів за допомогою сучасних технологій, тому ритейл стає однією з передових галузей у сфері використання технології інтелектуального аналізу даних;
- Data Mining представляють велику цінність для керівників та аналітиків у їхній повсякденній діяльності. Ділові люди усвідомили, що за допомогою методів Data Mining можуть отримати відчутні переваги в конкурентній боротьбі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. О.І. Черняк, П.В. Захарченко / Інтелектуальний аналіз даних. Підручник, Київ-2010
2. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
3. О.М. Верес, В.Л. Мельник, Л.Б. Чирун Застосування ms sql server 2005 для побудови інтелектуальної складової інформаційної систем. – Львів – 2008.
4. Fayyad, U. Advances in Knowledge Discovery and DataMining / U. Fayyad, G. PiatetskyShapiro, P. Smyth, R. Uthurusamy.– AAAI/MIT Press, 1996.
5. Харкянен О.В. Інтелектуальний аналіз даних: навчальний посібник / О.В. Харкянен О.В., О.М. М'якшило, С.В.Грибков, – К.: НУХТ, 2019 – 170 с
6. Журавльова І. В. Інформаційні системи міжнародного бізнесу :навч. посібн. / І. В. Журавльова, Р. М. Чен. – Х. : Вид. ХНЕУ, 2006. – 450 с.
7. Інформаційні системи і технології в економіці : підручник / за ред. В. С. Пономаренка. – К. : Академія, 2002. – 544 с.
8. Пономаренко В. С. Проектування автоматизованих економічних інформаційних систем : навч. посібн. / В. С. Пономаренко, О. І. Пушкар,Ю. І. Коваленко. – К. : ІЗМН, 1996. – 312 с.
9. Гужва В. М. Інформаційні системи і технології на підприємствах: навч. посібн. / В. М. Гужва. – К. : КНЕУ, 2001. – 400 с.
10. Олійник А.О. Інтелектуальний аналіз даних: навчальний посібник / А.О. Олійник, С.О. Субботін, О.О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.
11. Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. - 3rd Edition. - Morgan Kaufmann, 2011. - P. 664
12. Буреш, О. В. Інтелектуальні інформаційні системи управління соціально-економічними об'єктами: монографія/О. В. Буреш, М. А. Жук: Красанд, 2022. - 192 с.
13. К. Дж. Дейт. Введення в системи баз даних = Introduction to Database Systems. — 8-ме вид.. — М.: Вільямс, 2006. — С. 1328

14. Основи інформаційно-аналітичної діяльності: нав.посіб. /Захарова В.І, Філіпова Л.Я –К: «Центр учбової літератури», 2013. -336 с.