**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп’ютерних наук та кібернетики

Кафедра теорії та технології програмування

Звіт

до лабораторної роботи №1 на тему

« Market Basket Analysis»

Виконав студент 4-го курсу

Групи ТТП-42

Божевський Іван Петрович

**Київ – 2025**

**Опис завдання**

* **Objective:** To uncover associations and correlations between different items purchased in supermarkets or online platforms.
* **Key Concept:** Association Rule Mining, specifically using the Apriori algorithm.

**Tools and Technologies**

* **Libraries**:
  + Python: ***pandas*** for data manipulation, ***mlxtend*** for implementing Apriori.
  + ***matplotlib*** or ***seaborn*** for visualization.

**Dataset**

* A typical dataset for this project is the [Groceries dataset,](https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset) which is a standard dataset used for market basket analysis. It contains a collection of transactions with each transaction listing all items purchased.

**Tasks Breakdown**

1. **Data Preprocessing**

* **Loading Data**: Read the dataset into a suitable format for analysis.
* **Data Cleaning**: Handle missing values, if any.
* **Data Transformation**: Convert the data into an appropriate format for the Apriori algorithm (e.g., one-hot encoding in Python).

2. **Exploratory Data Analysis (EDA)**

* Analyze the most common items and itemsets.
* Visualize the frequency of top items/itemsets.

3. **Implementing Apriori Algorithm**

* **Parameter Setting**: Set appropriate values for support, confidence, and lift.
* **Frequent Itemset Generation**: Use the Apriori algorithm to find frequent itemsets.
* **Rule Generation**: Generate association rules from these itemsets.

4. **Analysis of Results**

* **Interpretation:** Understand and interpret the rules generated. For example, **if {bread, butter} -> {milk}** is a rule, it implies that customers who buy **bread** and **butter** are likely to buy **milk** as well.
* **Filtering Rules:** Filter out the most significant rules based on metrics like confidence and lift.

5. **Visualization**

* Create visual representations of the most important itemsets and rules (e.g., using bar plots, network graphs).

**Теоретичні відомості**

**Market Basket Analysis (MBA)** — це метод аналізу транзакцій, який дозволяє знаходити зв'язки між товарами, що часто купуються разом. Основною метою є виявлення **асоціативних правил** для покращення маркетингових стратегій, рекомендацій товарів, розміщення товарів у магазинах тощо.

**Основні поняття:**

* **Асоціативне правило:** Вираз виду A → B, де A і B — множини товарів.
* **Support (підтримка):** Частота, з якою правило зустрічається в транзакціях.

Зображення, що містить текст, Шрифт, знімок екрана, ряд

Вміст, створений ШІ, може бути неправильним.

* **Confidence (довіра):** Ймовірність того, що товар B буде куплений, якщо вже куплений товар A.

Зображення, що містить Шрифт, текст, ряд, знімок екрана

Вміст, створений ШІ, може бути неправильним.

* **Lift:** Міра, яка показує, наскільки частіше товари купуються разом, ніж якби вони були незалежними.

Зображення, що містить Шрифт, текст, ряд, знімок екрана

Вміст, створений ШІ, може бути неправильним.

**Apriori Algorithm:**

Один з найпопулярніших алгоритмів для пошуку частих наборів товарів та генерації асоціативних правил. Основна ідея: якщо набір елементів є частим, то всі його підмножини також повинні бути частими.

**Опис коду**

1. Імпорт бібліотек

* pandas – для обробки та аналізу даних.
* matplotlib і seaborn – для візуалізації даних.
* mlxtend – для реалізації Apriori алгоритму та створення правил асоціації.
* networkx – для побудови графу зв'язків між продуктами.

2. Попередня обробка даних

* Дані завантажуються з файлу groceries.csv.
* Виконується обробка пропущених значень (видаляються порожні рядки).
* Здійснюється попередній перегляд структури даних через head().
* Дані перетворюються на список транзакцій (список списків) для зручності подальшого аналізу.

Зображення, що містить текст, знімок екрана, Шрифт

Вміст, створений ШІ, може бути неправильним.

3. Перетворення даних

За допомогою класу TransactionEncoder дані конвертуються у формат "one-hot encoding", який потрібен для роботи алгоритму Apriori. Результат міститься у DataFrame, в якому кожен товар має свій стовпець, а кожний рядок представляє транзакцію.

Зображення, що містить текст, Шрифт, програмне забезпечення, знімок екрана

Вміст, створений ШІ, може бути неправильним.

4. Етап дослідження даних (EDA)

* Обчислюється і сортується кількість появ кожного товару у транзакціях.
* Побудовано графік найпоширеніших товарів (топ-10) за допомогою бібліотеки Seaborn.

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Вміст, створений ШІ, може бути неправильним.

Зображення, що містить текст, знімок екрана, Шрифт

Вміст, створений ШІ, може бути неправильним.

Зображення, що містить знімок екрана, текст, схема, дизайн

Вміст, створений ШІ, може бути неправильним.

5. Алгоритм Apriori

* Використовується метод Apriori для пошуку частих наборів товарів із мінімальною підтримкою (support) у 2%.
* На основі отриманих частих наборів генеруються правила асоціації з такими метриками:
  + Підтримка (support).
  + Достовірність (confidence).
  + Зображення, що містить текст, знімок екрана

    Вміст, створений ШІ, може бути неправильним.Зображення, що містить текст, знімок екрана, Шрифт

    Вміст, створений ШІ, може бути неправильним.Зростання (lift).

6. Аналіз результатів

* Виводяться метрики, які оцінюють отримані правила асоціації:
  + Підтримка (Support): Частка транзакцій, у яких зустрічається набір товарів.
  + Достовірність (Confidence): Ймовірність придбання набору товарів-наслідків за умови придбання товарів-передумов.
  + Зростання (Lift): Відображає, наскільки товари в наборі купуються разом частіше, ніж очікувалося б при їх незалежності.
* Виділено значущі правила асоціації: ті, що мають достовірність > 50% і зростання > 1.2.
* Надано приклад інтерпретації знайдених правил.

Зображення, що містить текст, знімок екрана, Шрифт

Вміст, створений ШІ, може бути неправильним.

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Вміст, створений ШІ, може бути неправильним.

7. Візуалізація результатів

* Побудовано графік зв'язку правил, де підтримка винесена на одну вісь, достовірність – на іншу, а зростання відображено через колір та розмір маркерів.
* Створено мережу відносин між товарами у вигляді графа:
  + Вершини графа – це товари.
  + Ребра графа – зв'язки між товарами у знайдених правилах.
  + Зображення, що містить текст, знімок екрана, Шрифт

    Вміст, створений ШІ, може бути неправильним.Товщина ребер відображає значення зростання (lift).
  + Зображення, що містить знімок екрана

    Вміст, створений ШІ, може бути неправильним.

8. Побудова та візуалізація графу

За допомогою бібліотеки networkx реалізовано функціонал:

* Побудова орієнтованого графу, який відображає найсильніші правила асоціації (топ-10 за значенням зростання).
* Візуалізація цього графу з використанням циркулярного макету.

Зображення, що містить текст, знімок екрана, програмне забезпечення

Вміст, створений ШІ, може бути неправильним.

Зображення, що містить знімок екрана, текст, схема, ряд

Вміст, створений ШІ, може бути неправильним.