

**МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ  
ПВНЗ  
«МІЖНАРОДНИЙ НАУКОВО-ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ АКАДЕМІКА Ю. БУГАЯ»**

*Кафедра інформаційних та комунікаційних технологій*

**Освітня компонента: «Інтелектуальний аналіз даних»**

*Дослідницька РГР – Глосарій*

Виконав:  
студент 4 курсу денної форми навчання  
група І26  
Бондаренко О.В.

Перевірила:  
доц. Гриб'юк О.О.

Київ – 2025

**Тема проекту:** Аналіз поведінкових даних та розробка аналітичних моделей для прийняття рішень.

**Прізвище та ім'я виконавців проекту:** Бондаренко Олег.

**Концепція проекту:** посилання на [github](#).

**Мета проекту:** Формування системи інтелектуального аналізу поведінкових та предметно-орієнтованих даних з метою моделювання сценаріїв розвитку бізнес-рішень, виявлення трендів та оптимізації взаємодії.

**Завдання проекту:**

1. Отримати вибіркові дані поведінкових метрик користувачів веб-проекту з різних джерел (первинна предметна область) та провести їх первинну обробку.
2. Здійснити регресійний та survival-аналіз для виявлення впливу поведінкових ознак на ймовірність покупки або конверсії.
3. Провести розкладання часових рядів поведінкових метрик на складові (тренд, сезонність).
4. Провести апробацію розробленого комплексу методів на новій предметній області – ідентифікація діамантів.
5. Сформувати висновки та рекомендації щодо оптимізації бізнес-рішень на основі виявленіх закономірностей, трендів та результатів моделювання.

## ГЛОСАРІЙ

### 1. Загальні поняття Data Mining

**Data Mining (Інтелектуальний аналіз даних).** Процес виявлення прихованих, нетривіальних та корисних знань у великих наборах даних.

**Business Intelligence (BI).** Процеси та інструменти для перетворення "сирих" даних на корисну інформацію для прийняття бізнес-рішень.

**Класифікація (Classification).** Задача віднесення об'єкта до одного із заздалегідь відомих класів (наприклад, `is_investment_grade`. 1 або 0).

**Регресія (Regression).** Задача прогнозування неперервного числового значення (наприклад, `price`).

**Кластеризація (Clustering).** Задача групування об'єктів без заздалегідь відомих класів (навчання без вчителя).

**Асоціативні правила (Association Rules).** Пошук закономірностей типу "Якщо A, то B" (наприклад, "Якщо `color=D` і `clarity=IF`, то `price > 10000`").

**ETL (Extract, Transform, Load).** Процес витягання, перетворення та завантаження даних з операційних систем у Сховище Даних.

### 2. Регресія та Класифікація (Статистичні методи)

**Кореляція (Correlation).** Статистичний зв'язок між двома змінними.

**Pearson (Пірсон).** Для лінійного зв'язку та нормальних даних.

**Spearman (Спірмен).** Для нелінійного (рангового) зв'язку та ненормальних даних.

**R<sup>2</sup> (Коефіцієнт детермінації).** Метрика якості регресії (від 0 до 1), що показує, яку частку варіації Y пояснює модель.

**Дерево рішень (Decision Tree).** Модель, що буде ієрархію правил "Якщо-Тоді".

**Індекс Джині (Gini Index).** Метрика для вибору найкращого розділення у вузлі дерева класифікації.

**Логістична регресія (Logistic Regression).** Метод класифікації, що моделює ймовірність належності до класу.

**Нейронна мережа (Neural Network / MLP).** Модель, натхнена біологічним мозком; потужний, але складний класифікатор/регресор.

**LDA (Лінійний дискримінантний аналіз).** Метод класифікації, що проєктує дані на нову вісь для максимізації розділення класів.

**Аналіз виживаності (Survival Analysis).** Статистичний метод для аналізу часу до настання події.

**Регресія Кокса (Cox Regression).** Модель, що оцінює вплив ознак на "ризик" настання події в часі.

**Цензурування (Censoring).** Ситуація, коли подія не відбулася за час спостереження (`is_sold = 0` для сгенерованого датасету).

**Hazard Ratio (Відношення ризиків).** Показник того, у скільки разів ознака прискорює або сповільнює настання події.

### 3. Аналіз часових рядів

**Часовий ряд (Time Series).** Послідовність даних, впорядкована за часом.

**Тренд (Trend).** Довгостроковий напрямок руху ряду (вгору або вниз).

**Сезонність (Seasonality).** Регулярні, повторювані коливання (наприклад, щокварталу або щомісяця).

**Автокореляція (ACF).** Кореляція ряду із самим собою, але зі зсувом у часі (лаг).

**ARIMA (Авторегресія, Інтегроване ковзне середнє).** Потужна модель для прогнозування часових рядів.

### 4. Методи оптимізації

**Градієнтні методи.** Алгоритми, що використовують похідну (градієнт) для пошуку локального мінімуму/максимуму.

**Квазіньютонівські методи (BFGS, L-BFGS).** Просунуті градієнтні методи, які наближено оцінюють другі похідні (кривину) для швидшої збіжності.

**Еволюційні методи.** Методи глобальної оптимізації, натхненні біологічною еволюцією.

**Генетичний алгоритм (GA).** Використовує популяцію, селекцію, схрещування та мутацію для пошуку рішення.

**Популяція (Population).** Набір потенційних рішень ("особин").

**Селекція (Selection).** Процес вибору найкращих "особин" для розмноження.

### 5. Інтелектуальні та Мультиагентні системи

**АОП (Агентно-орієнтоване програмування).** Парадигма, що моделює систему як сукупність автономних, проактивних агентів, які мають цілі.

**ООП (Об'єктно-орієнтоване програмування).** Парадигма, що моделює систему як набір об'єктів, які пасивно реагують на виклики методів.

**Метод мурашиних колоній (ACO).** Мультиагентний алгоритм оптимізації, де "мурахи" знаходять шляхи, залишаючи феромони. Ідеальний для задач на графах (наприклад, задача комівояжера).

**Нечітка логіка (Fuzzy Logic).** Метод, що дозволяє комп'ютеру працювати з нечіткими поняттями (наприклад, "висока" ціна, "шивидке" заповнення).

**Алгоритм Мамдані (Mamdani).** Алгоритм нечіткого виведення, де і умови, і висновки є нечіткими множинами (використовується для експертних систем).

**Алгоритм Сугено (Sugeno).** Алгоритм нечіткого виведення, де висновки є чіткими функціями (використовується в системах керування).

## **6. Аналітичні системи та Архітектури**

**Сховище Даних (Data Warehouse).** Централізована, інтегрована, історична база даних, оптимізована для аналізу.

**Вітрина Даних (Data Mart).** Тематична підмножина Сховища Даних (сгенерований датасет diamonds\_dataset.csv).

**OLAP (On-Line Analytical Processing).** Технологія для швидкого, інтерактивного, багатовимірного аналізу даних.

**OLAP-Куб (OLAP Cube).** Багатовимірна структура даних (Міри та Виміри), що забезпечує швидкі відповіді на запити.

**ROLAP (Relational OLAP).** Архітектура OLAP, де дані зберігаються у реляційних таблицях (схема "зірка"). Наприклад, аналіз в pandas (groupby) є аналогом ROLAP.

**MOLAP (Multidimensional OLAP).** Архітектура OLAP, де дані зберігаються у спеціалізованому багатовимірному кубі.

**Мережевий аналіз (NetworkX).** Аналіз даних, представлених у вигляді графів (вузли та зв'язки).

**Двочастковий граф (Bipartite Graph).** Граф, вузли якого можна розділити на дві множини, де зв'язки існують тільки між множинами (наприклад, граф Expert <-> Clarity).