

**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 проєкту:** Аналіз поведінкових даних веб-проєктів.

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

**Концепція проєкту:** посилання на [github](https://github.com).

**Мета проєкту:** формування системи інтелектуального аналізу поведінкових даних веб-проєктів з метою моделювання сценаріїв розвитку бізнес-рішень, виявлення трендів та оптимізації цифрової взаємодії з користувачами.

**Завдання проєкту:**

1. Отримати вибіркові дані поведінкових метрик користувачів веб-проєкту з різних джерел (Google Analytics, Binotel, CRM) та провести їх первинну обробку.
2. Здійснити регресійний та survival-аналіз для виявлення впливу поведінкових ознак (наприклад, page.view, ad\_click) на ймовірність покупки або конверсії.
3. Провести розкладання часових рядів поведінкових метрик на складові (тренд, сезонність, випадкова компонента) та оцінити наявність тенденцій за допомогою критерію серій.
4. Сформуванати висновки та рекомендації щодо оптимізації бізнес-рішень на основі виявлених закономірностей, трендів та результатів моделювання.

## ГЛОСАРІЙ

### 1. Метод кореляційного аналізу даних

Correlation (Кореляція) – статистичний зв'язок між двома змінними

Heatmap (Теплова карта) – візуалізація кореляцій між змінними

Histogram (Гістограма) – графік частоти значень змінної

Kendall coefficient (Коефіцієнт Кендалла) – альтернативний показник для рангових даних

Multicollinearity (Мультиколінеарність) – надмірна кореляція між незалежними змінними

Normality check (Нормальність розподілу) – перевірка відповідності даних нормальному закону

p-value (р-значення) – ймовірність того, що зв'язок є випадковим

Pearson coefficient (Коефіцієнт Пірсона) – показник лінійної кореляції між змінними

Scatterplot (Діаграма розсіювання) – графік для візуалізації зв'язку між змінними

Spearman coefficient (Коефіцієнт Спірмена) – непараметричний показник рангової кореляції

### 2. Лінійний регресійний аналіз + Інтелектуальна задача

Analytical Agent (Аналітичний агент) – система, що приймає рішення на основі даних

Classification (Класифікація) – визначення класу об'єкта

Conversion (Конверсія) – виконання цільової дії користувачем

Feature Importance (Важливість ознак) – вплив окремих ознак на результат

LDA (Linear Discriminant Analysis) (Лінійний дискримінантний аналіз) – метод класифікації

Linear Regression (Лінійна регресія) – модель залежності між змінними у вигляді прямої

MSE (Mean Squared Error) (Середньоквадратична похибка) – метрика точності

QDA (Quadratic Discriminant Analysis) (Квадратичний дискримінантний аналіз) – метод класифікації

R<sup>2</sup> (Coefficient of Determination) (Коефіцієнт детермінації) – показник якості моделі

Scenario Modeling (Сценарне моделювання) – імітація дій для оцінки впливу

Simulation (Симуляція) – створення штучного середовища для тестування

### 3. Нелінійний регресійний аналіз

Activation Function (Функція активації) – функція, що визначає вихід нейрона  
Backpropagation (Зворотне поширення помилки) – метод оновлення ваг  
Decision Boundary (Межа класифікації) – лінія, що розділяє класи  
Decision Tree (Дерево рішень) – модель класифікації або регресії  
Entropy (Ентропія) – міра невизначеності вузла  
Epoch (Епоха) – один прохід по навчальному набору  
Gini Index (Індекс Джині) – показник чистоти вузла  
MLP (Multilayer Perceptron) (Багатошаровий перцептрон) – тип нейромережі  
Nonlinear Regression (Нелінійна регресія) – модель з нелінійною залежністю  
Overfitting (Перенавчання) – надмірна адаптація до навчальних даних

#### 3.1.Завдання 2: Дерева рішень

Leaf Node (Лист дерева) – кінцевий вузол з прогнозом  
Node Purity (Чистота вузла) – однорідність за класами  
Pruning (Обрізання дерева) – зменшення складності  
Rule Extraction (Витяг правил) – формалізація логіки дерева  
Sample Split (Розділення вибірки) – умовне розгалуження  
Tree Depth (Глибина дерева) – кількість рівнів

#### 3.2.Завдання 3: Нейронні мережі в Python (архітектура 3–1)

Activation = 'logistic' (Логістична функція активації) – сигмоїдна функція  
Bias Toward Positive Class (Зміщення до позитивного класу) – перевага одного класу  
Classification Report (Звіт класифікації) – метрики якості  
Decision Boundary Plot (Графік межі класифікації) – візуалізація результату  
Improvement Strategies (Стратегії покращення) – способи оптимізації  
MLPClassifier (Клас нейромережі з sklearn) – реалізація перцептрону  
Misclassified Samples (Помилки класифікації) – об'єкти з неправильним прогнозом  
Model Evaluation (Оцінка моделі) – аналіз результатів  
Single Hidden Layer (Один прихований шар) – базова архітектура  
Training Set (Навчальна вибірка) – дані для навчання  
hidden\_layer\_sizes=(3,) (Структура шару) – 3 нейрони в прихованому шарі

### 4. Логістична регресія

AUC (Area Under Curve) (Площа під кривою) – загальна якість моделі  
Confusion Matrix (Матриця помилок) – розподіл класифікацій  
F1-score (F1-метрика) – баланс між точністю і повнотою  
Logistic Regression (Логістична регресія) – модель для оцінки ймовірності  
Odds Ratio (Відношення шансів) – співвідношення ймовірностей  
Precision (Точність) – частка правильних позитивних передбачень  
ROC Curve (ROC-крива) – графік якості класифікації  
Recall (Повнота) – частка знайдених позитивних випадків  
Sigmoid Function (Сігмоїдна функція) – перетворення значень у  $[0,1]$   
Threshold (Попіг) – межа для класифікації

## 5. Регресія Кокса

Censoring (Цензурування) – ситуація, коли подія не зафіксована до завершення спостереження

Concordance Index (Індекс узгодженості) – міра точності прогнозу часових подій

Cox Regression (Регресія Кокса) – модель для аналізу часових подій

Hazard Ratio (Відношення ризиків) – порівняння ймовірності настання події між групами

Kaplan-Meier Curve (Крива Каплана-Майєра) – графік виживання для різних груп

Log(HR) (Логарифм hazard ratio) – логарифмічна інтерпретація впливу ознак

Proportional Hazards (Пропорційні ризики) – припущення, що hazard ratio є сталим у часі

Schoenfeld Residuals (Залишки Шенфельда) – перевірка припущення пропорційності ризиків

Survival Function (Функція виживання) – ймовірність того, що подія не відбудеться до певного моменту

Time-to-event (Час до події) – тривалість до настання події