

**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](#).

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 вибірккові 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

Кореляція – статистичний зв'язок між двома змінними.

Коефіцієнт Спірмена – непараметричний показник рангової кореляції.

Коефіцієнт Пірсона – показник лінійної кореляції між змінними.

Коефіцієнт Кендалла – альтернативний показник для рангових даних.

p-value – ймовірність того, що зв'язок є випадковим.

Нормальність розподілу – перевірка відповідності даних нормальному закону.

Scatterplot – графік розсіювання для візуалізації зв'язку.

Гістограма – графік частоти значень змінної.

Heatmap – теплова карта кореляцій між змінними.

Мультиколінеарність – надмірна кореляція між незалежними змінними.

2. Лінійний регресійний аналіз + Інтелектуальна задача

Лінійна регресія – модель, що описує залежність між змінними у вигляді прямої.

R^2 – коефіцієнт детермінації, що показує якість моделі.

MSE – середньоквадратична похибка.

LDA – лінійний дискримінантний аналіз для класифікації.

QDA – квадратичний дискримінантний аналіз.

Класифікація – визначення класу об'єкта на основі ознак.

Сценарне моделювання – імітація дій для оцінки їх впливу.

Конверсія – виконання цільової дії користувачем.

Аналітичний агент – умовна система, що приймає рішення на основі поведінкових даних.

Feature Importance – вплив окремих ознак на результат.

Симуляція – створення штучного середовища для тестування моделей.

3. Нелінійний регресійний аналіз

Нелінійна регресія – модель, де залежність між змінними не є прямою.

Decision Tree – дерево рішень для класифікації або регресії.

Entropy – міра невизначеності, що використовується для побудови дерева.

Gini Index – показник чистоти вузлів дерева.

Overfitting – перенавчання моделі, надмірна адаптація до навчальних даних.

MLP – багат шаровий перцептрон, тип нейромережі.

Activation Function – функція, що визначає вихід нейрона (наприклад, ReLU, sigmoid).

Decision Boundary – межа, що розділяє класи у просторі ознак.

Epoch – один повний прохід по навчальному набору.

Backpropagation – метод оновлення ваг у нейромережі.

4. Логістична регресія

Логістична регресія – модель для оцінки ймовірності події.

Sigmoid-функція – функція, що перетворює значення у діапазон $[0,1]$.

Odds Ratio – відношення шансів події до її непадії.

Confusion Matrix – матриця правильних і неправильних класифікацій.

Precision – частка правильних позитивних передбачень.

Recall – частка знайдених позитивних випадків серед усіх позитивних.

F1-score – гармонічне середнє між precision і recall.

ROC-крива – графік якості класифікації при різних порогах.

AUC – площа під ROC-кривою, загальна якість моделі.

Threshold – поріг, при якому модель класифікує об'єкт як позитивний.

5. Регресія Кокса

Регресія Кокса – модель для аналізу часових подій.

Hazard Ratio – відношення ризиків між групами.

Survival Function – ймовірність того, що подія не відбудеться до певного моменту.

Censoring – часткове спостереження, коли подія не відбулася до кінця дослідження.

Time-to-event – час до настання події.

Log(HR) – логарифм hazard ratio для інтерпретації впливу ознак.

Kaplan-Meier Curve – графік виживання для різних груп.

Proportional Hazards – припущення, що hazard ratio є сталим у часі.

Schoenfeld Residuals – перевірка припущення пропорційних ризиків.

Concordance Index – міра точності прогнозу часових подій.