# Звіт: Задача класифікації

## Мета

Розв'язати задачу класифікації на основі датасету Churn\_Modelling.csv за допомогою нейронних мереж. Метою є передбачення, чи клієнт припинить користування послугами банку. Виконано попередній аналіз даних, крос-валідацію, оцінку метрик якості моделі та зроблено висновки.

## Препроцесинг даних

**Видалення зайвих стовпців:**

Було видалено: RowNumber, CustomerId, Surname.

**Кодування категоріальних змінних:**

Перетворено змінні Geography та Gender за допомогою pd.get\_dummies, використовуючи drop\_first=True.

**Нормалізація числових змінних:**

До стовпців CreditScore, Age, Tenure, Balance, NumOfProducts, EstimatedSalary застосовано MinMaxScaler.

**Розділення даних:**

Ознаки (X) та цільова змінна (y), де y = Exited (чи залишив клієнт послуги).

## Кореляційний аналіз

Обчислено матрицю кореляції для оцінки взаємозв'язків між змінними. Основні висновки:

Найвища позитивна кореляція з вихідною змінною Exited:

* Age (0.2853)
* Geography\_Germany (0.1735)
* Balance (0.1185)

Найбільша негативна кореляція:

* Geography\_Spain (-0.0527)
* NumOfProducts (-0.0478)

Кореляція з іншими змінними є слабкою.

## Крос-валідація

Було виконано 5-кратну стратифіковану крос-валідацію:

**Архітектура моделі:**

* Вхідний шар: кількість ознак X.shape[1].
* Приховані шари: 16 і 8 нейронів (активація ReLU).
* Вихідний шар: 1 нейрон (активація Sigmoid).

**Навчання:**

* Оптимізатор: Adam.
* Функція втрат: Binary Crossentropy.
* Кількість епох: 50.
* Розмір батчу: 32.

## Метрики якості

**Точність (Accuracy):** Середня точність моделі: 85.43%.

**ROC AUC:** Середній ROC AUC: 0.8529.

Модель демонструє високу точність та стабільність на різних фолдах. Значення ROC AUC (0.8529) вказує на гарний баланс між чутливістю та специфічністю. Кореляційний аналіз показав, що ключовими ознаками для передбачення є Age, Geography\_Germany, та Balance.