**І ПРО ПРОЕКТ**

Метою проекту є розробка прогностичної моделі для ідентифікації ймовірності припинення клієнтами користування телекомунікаційними послугами на основі історичних даних про клієнтів. Проєкт передбачає використання аналізу даних, перед обробки даних, машинного навчання, оцінки моделей, а також пакування моделі у контейнер для забезпечення відтворюваності.

**ІІ ЯК ЗАПУСТИТИ ПРОЕКТ**

Щоб запустити в перший раз або оновити програму:

1) встановіть та запустіть Docker,

2)відкрийте термінал та пропишіть там наступні команди

docker pull maryanashcherbak103/churn:latest

docker run -p 8501:8501 maryanashcherbak103/churn:latest

3)Та перейдіть по посиланню http://localhost:8501

Щоб запустити програму знову

1)запустіть Docker,

2)відкрийте термінал та пропишіть там команду

docker pull maryanashcherbak103/churn:latest

3)Та перейдіть по посиланню <http://localhost:8501>

**ІІІ ПІДТВЕРДЖЕННЯ, ЩО МОДЕЛЬ СПРАВЛЯЄТЬСЯ**

Аналізуючи результати та графіки, можемо зробити висновок, що модель XGBoost демонструє гарні показники якості класифікації:

Метрики якості моделі:

- 1 -Accuracy: 0.823 - вказує на те, що модель правильно класифікує близько 82.3% випадків, це хороший показник для більшості задач класифікації

- 2 -Recall: 0.798 - показує, що модель знаходить близько 79.8% всіх позитивних випадків

- 3 - Precision: 0.861 - означає, що 86.1% випадків, які модель класифікувала як позитивні, дійсно є позитивними

- 4 - F1 Score: 0.828 - є гармонічним середнім Precision і Recall, і значення 0.828 вказує на хороший баланс між ними

- 5 - ROC-AUC: 0.824 - показує, наскільки добре модель розрізняє позитивні та негативні класи. Значення 0.824 є досить високим і вказує на хорошу здатність моделі до класифікації, чим ближче до 1, тим краща модель.

Результати крос-валідації:

- середнє значення крос-валідації (10 фолдів) = 0.821 вказує на стабільну продуктивність моделі на різних наборах даних, що зменшує ймовірність перенавчання

- невеликий розкид між мінімальним (0.810) та максимальним (0.831) значеннями також підтверджує стабільність моделі

Аналіз графіка ROC Curve: Крива ROC показує, як добре модель розрізняє позитивні та негативні класи на різних порогах, площа під кривою (AUC) = 0.824, що є дуже хорошим результатом

Аналіз графіка Confusion Matrix (Матриця неточностей): TP = 6036 і TN = 5556 значно більші за FP = 976 і FN = 1524, це означає, що модель добре справляється з прогнозуванням обох класів (відтік клієнтів і залишення клієнтів) (True Positives (TP) і True Negatives (TN) представляють правильні передбачення моделі; False Positives (FP) і False Negatives (FN) представляють помилки моделі)

Аналіз графіка Feature Importance (Важливість ознак): Діаграми важливості ознак показують, які ознаки є найважливішими для моделі

Аналіз графіка Precision-Recall Curve (Крива точності-відкликання): крива підтверджує високий Precision та Recall, що було показано вище, і демонструє, що модель має високу продуктивність навіть при зміні порогових значень

Підтвердження якості моделі:

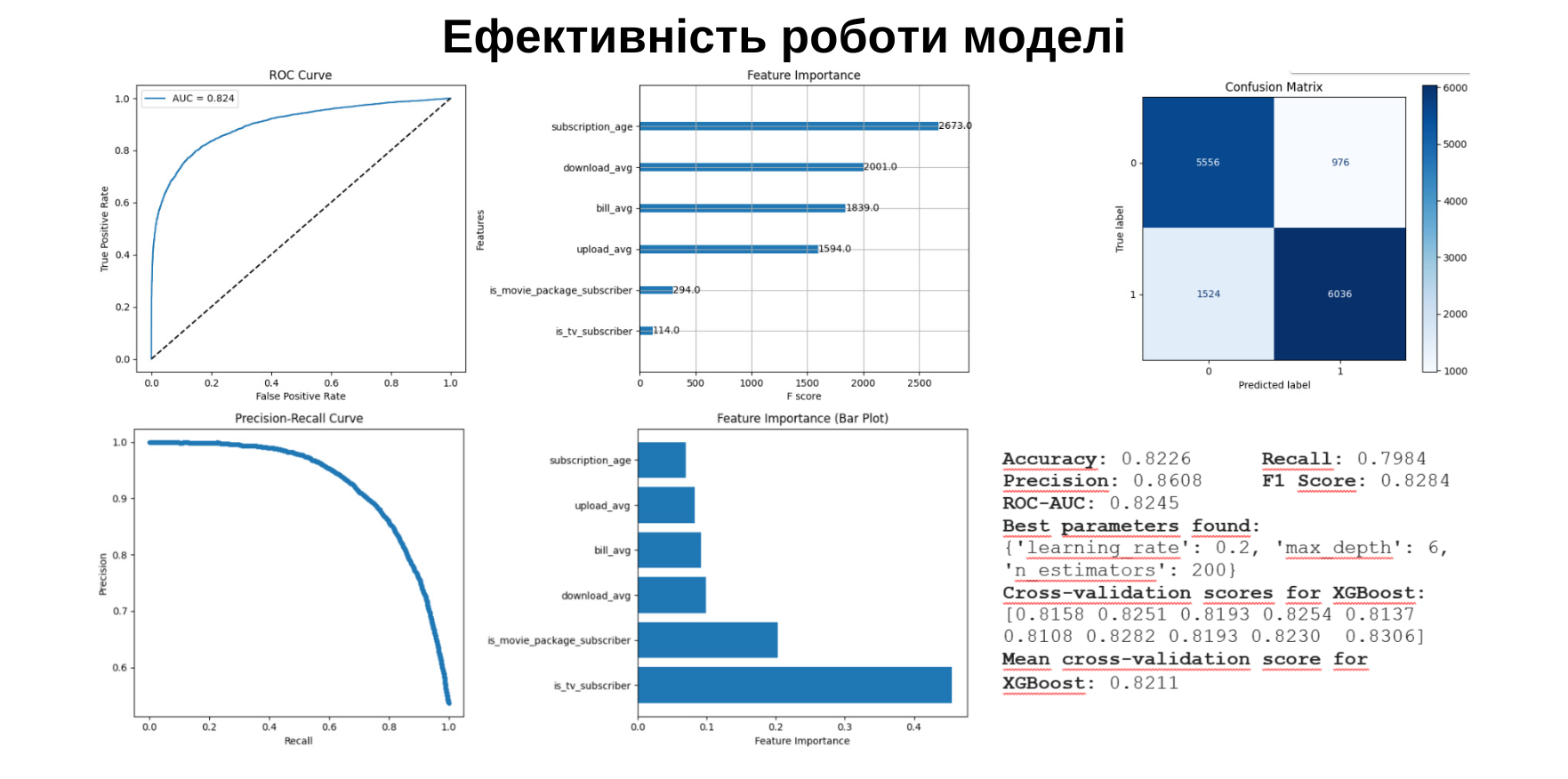
- 1 - Високі значення метрик (особливо AUC і F1 Score) підтверджують, що модель працює добре і справляється зі своєю задачею

- 2- Відносно висока середня оцінка крос-валідації (0.821) говорить про стабільність моделі на різних наборах даних

- 3 - Висока площа під ROC кривою і низька кількість помилок у матриці неточностей свідчать про те, що модель добре розрізняє класи

- 4 - Перенавчання моделі не спостерігається, оскільки результати крос-валідації стабільні і відповідають загальним метрикам. Модель ймовірно добре буде працювати і на нових даних завдяки високим значенням метрик та стабільності в крос-валідації

Модель показує хороші результати з точки зору точності, повноти, точності і F1-оцінки. Результати крос-валідації підтверджують, що модель не перенавчена і має стабільну продуктивність на різних наборах даних. Графіки також підтримують висновок, що модель добре справляється з класифікацією.

****

**IV ВИКОРИСТАНІ БІБЛІОТЕКИ**

**'seaborn', 'sklearn.model\_selection', 'sklearn.preprocessing', 'matplotlib.pyplot', 'pandas', 'sklearn.preprocessing', 'google.colab', 'sklearn.metrics', 'numpy', 'xgboost' , 'streamlit'**

****