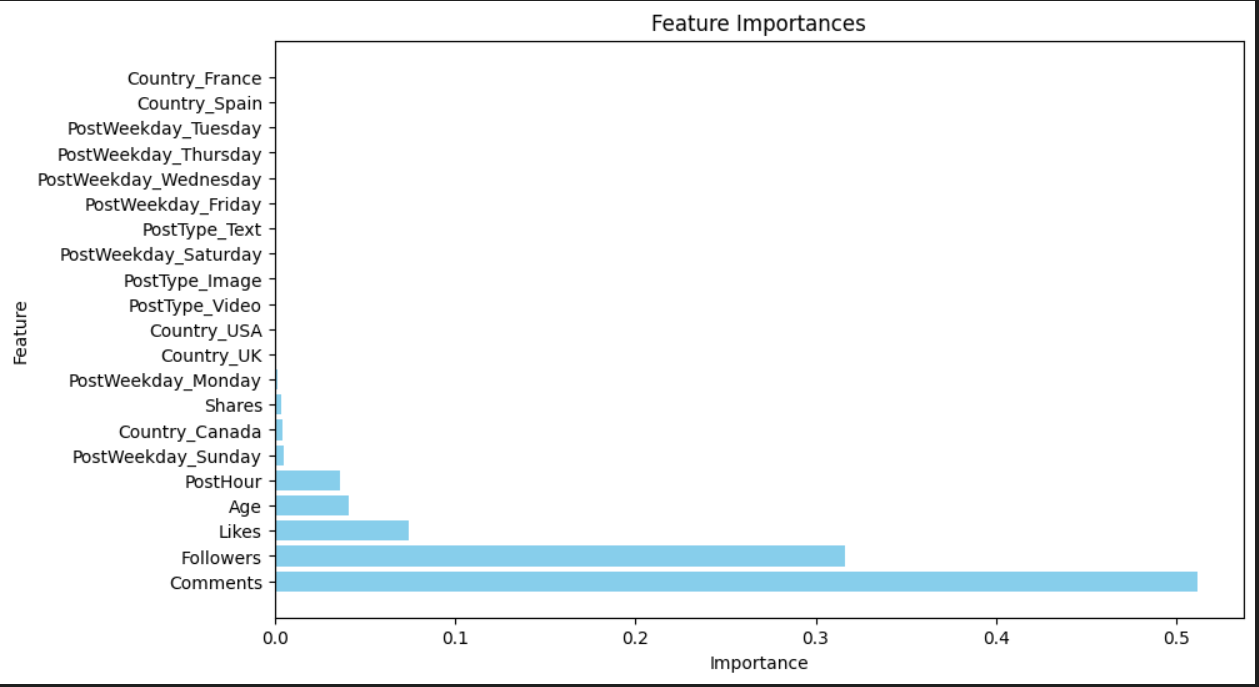
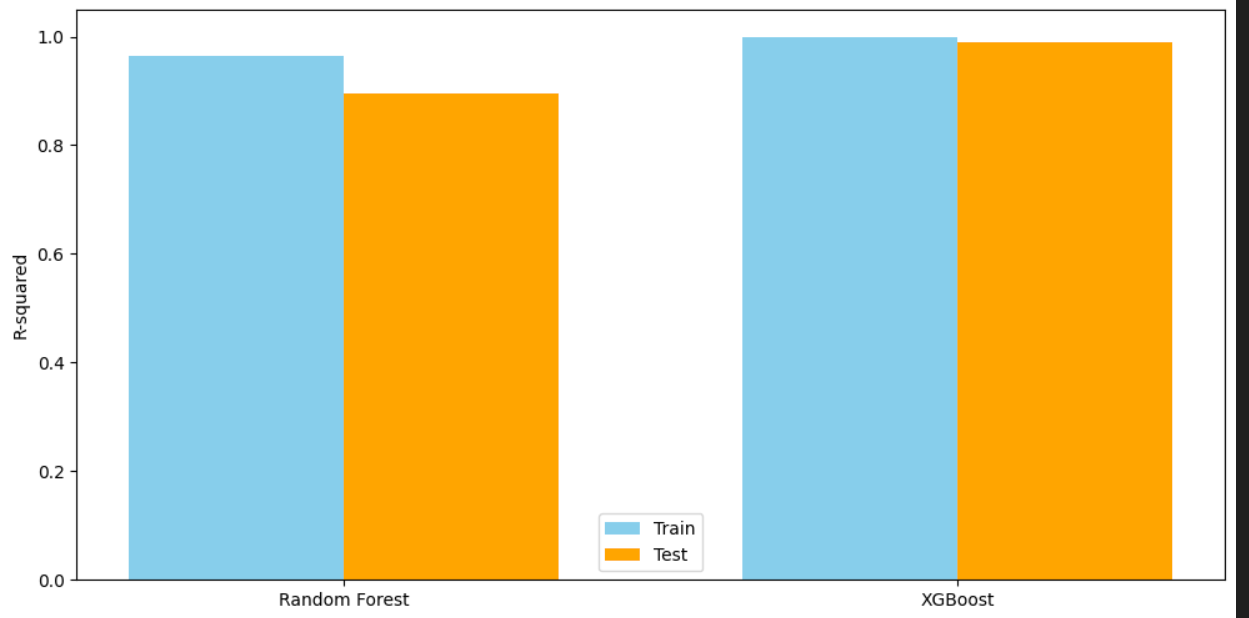
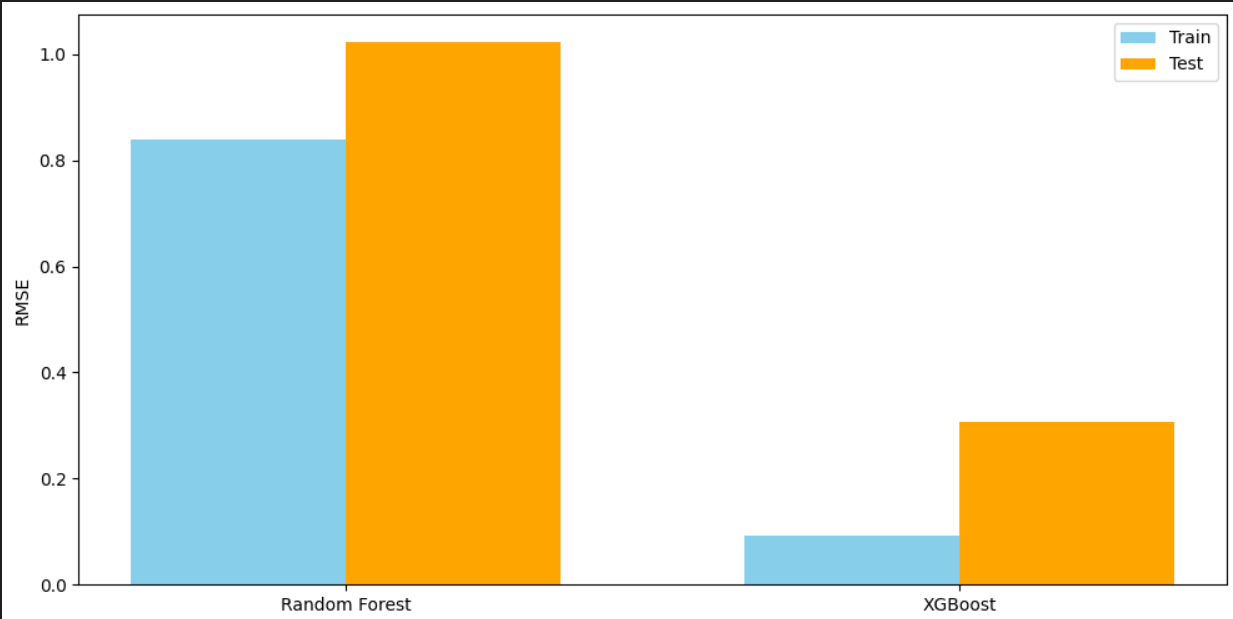
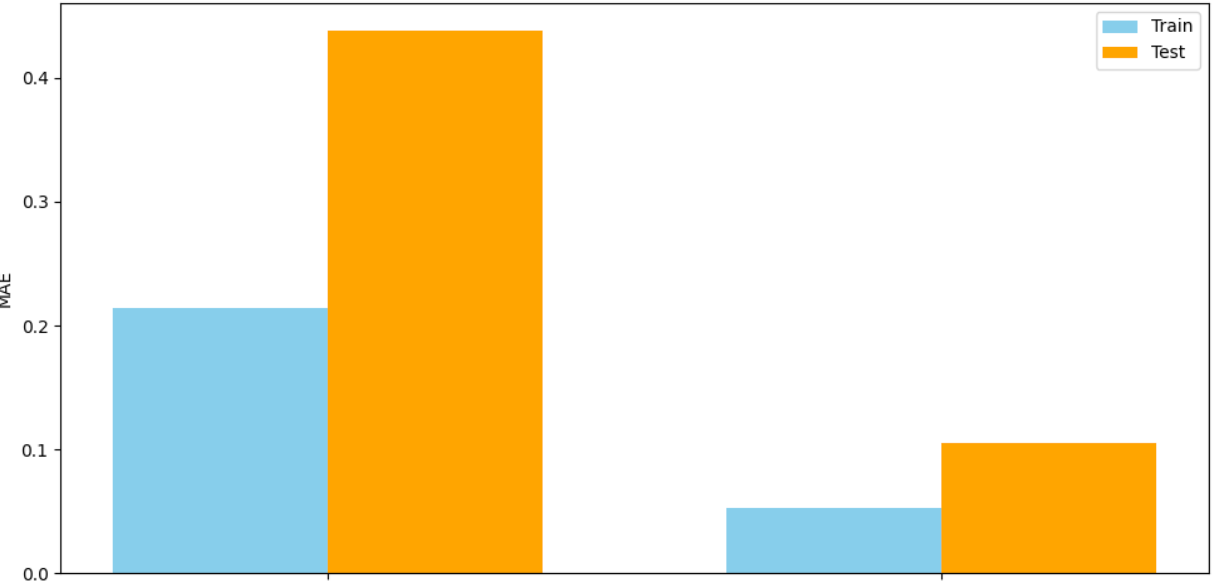
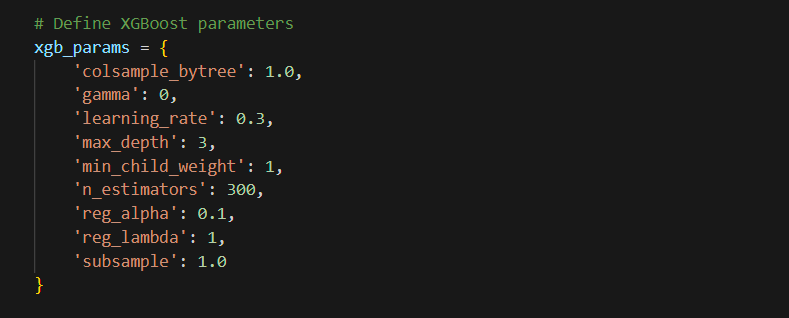
Завдання 1 Ви хороший детектив?

Визначити, як розраховується EngagementScore на основі різних характеристик користувачів і постів. Дослідити, які фактори впливають на цей показник.

EngagementScore вираховується за допомогою декьльков фичь з датасету а саме з колонок Comments Followers Likes Age PostHour Shares PostType (Розташовані в порядку перший має найбільшу вагу останній найменшу)

Прогнозування непередбачуваного: оцініть, чи можна точно передбачити EngagementScore для новостворених постів.

Я експерементував з багатьма моделями і багатьма підходами для подачі даних в модель, у фінал вийшли ці моделі RandomForest та XGBoost. В кінцевому всече вигравав XGBoost на метриці R квадрат він показав на тестовому сеті 0.9906690775974784 коли 0.8958117083373824 метрики як RMSE і MAE XGBoost теж лідирував. На мій погляд XGBoost може передбачити EngagementScore. Скріншоти додаю нижче.



Завдання 2 Ви хороший хакер?

У другому завданні я так само випробував багато варінтів, спочатку я думав, що це завдання запитання-відповідь, але о 2-й годині ночі четверга мені спало на думку, що, найімовірніше, потрібно дивитися під іншим кутом на це завдання, бо це завдання все ж таки для SentenceTransformer. Я зупинився на моделі 'distilroberta-base-v2', але так само експериментував з іншими версіями BERT, за основу сайту взяв Django, а базу даних, куди записуються дані, - SQLite, так само я скористався логікою RAG, яка пов'язувала LLM із базою даних, чесно кажучи, я не довів проєкт до ідеалу і розумію, що можна його поліпшити за багатьма параметрами, але основні функції він виконує.