

Отчет по работе 2го ИНТЕНСИВА

Работу выполнили:
Крылов Савва
Пелюшенко Влад



самолет

О кейсе

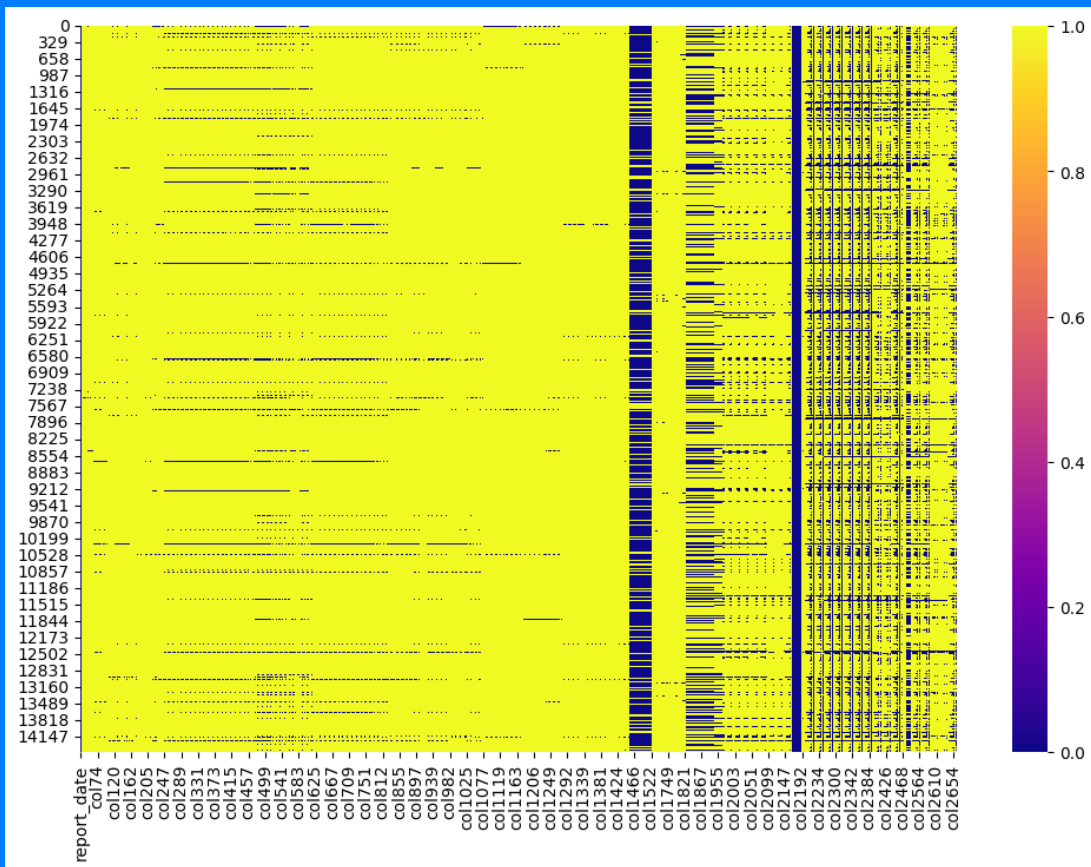
Название: Модель склонности клиента к приобретению машиноместа

Цель задания: Разработать модель на основе анализа данных о взаимодействии с клиентами, чтобы прогнозировать вероятность покупки дополнительных услуг, таких как машиноместа. Цель — выделить владельцев квартир, наиболее склонных к покупке, и провести с ними целевую коммуникацию через SMS или email.

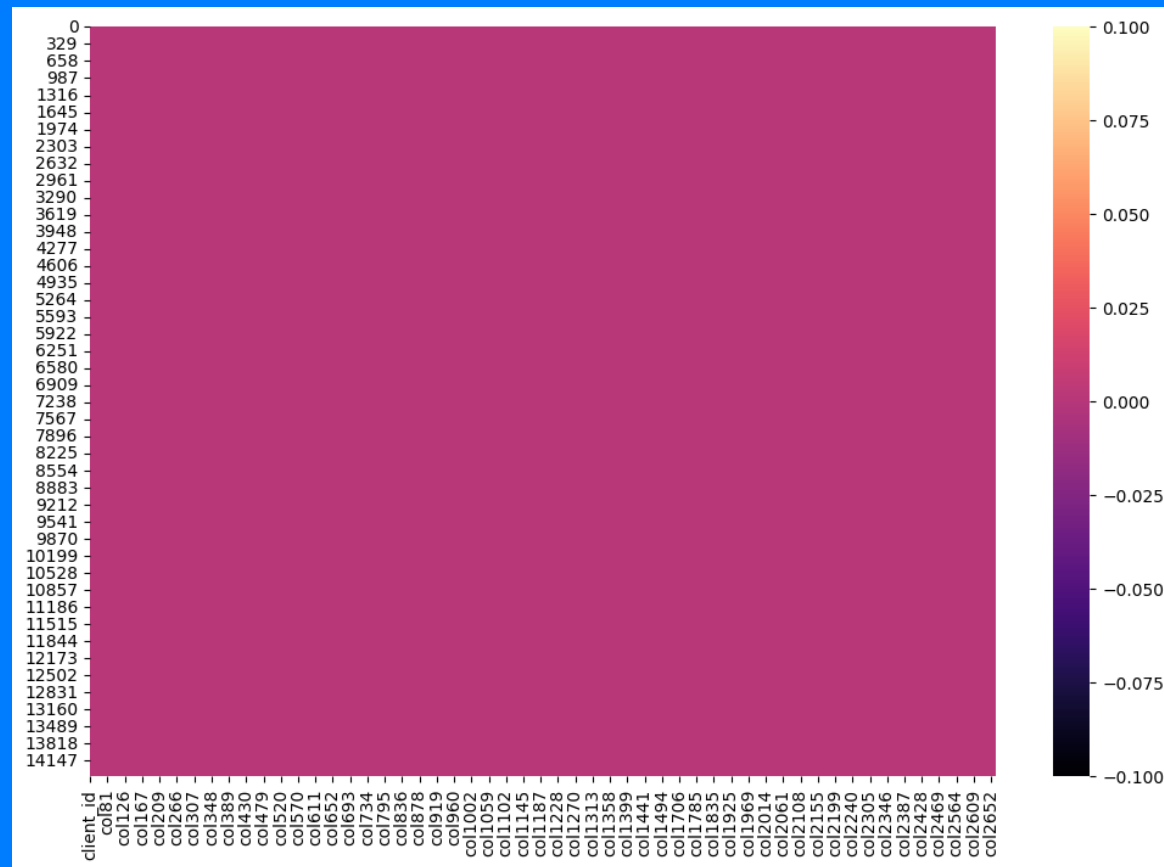


Графики о датасете

Тепловая карта изначально



Тепловая карта после чистки



самолет

Модели

Xgboost

Это мощный алгоритм градиентного бустинга, который используется для решения задач классификации и регрессии. Он отличается высокой производительностью и эффективностью.

RandomForestClassifier

Это ансамблевый метод классификации, основанный на построении множества деревьев решений и объединении их прогнозов для повышения точности. Он входит в состав библиотеки Scikit-learn и используется для решения задач классификации.



Обработка данных

Удаление колонок
заполненных меньше чем на
50%

Заполнение нулями NaN
значений.

Заполнение пустыми строками
NaN значений в нечисловых

Удаление специфических
данных:

Hash - непонятные символы

Links- ненужные данные

Удаление дубликатов

Предобработка данных для обучения

1. Масштабирование данных

Чтобы все числовые признаки находились в одном масштабе, что важно для алгоритмов, зависящих от расстояния (например, линейной регрессии или SVM).

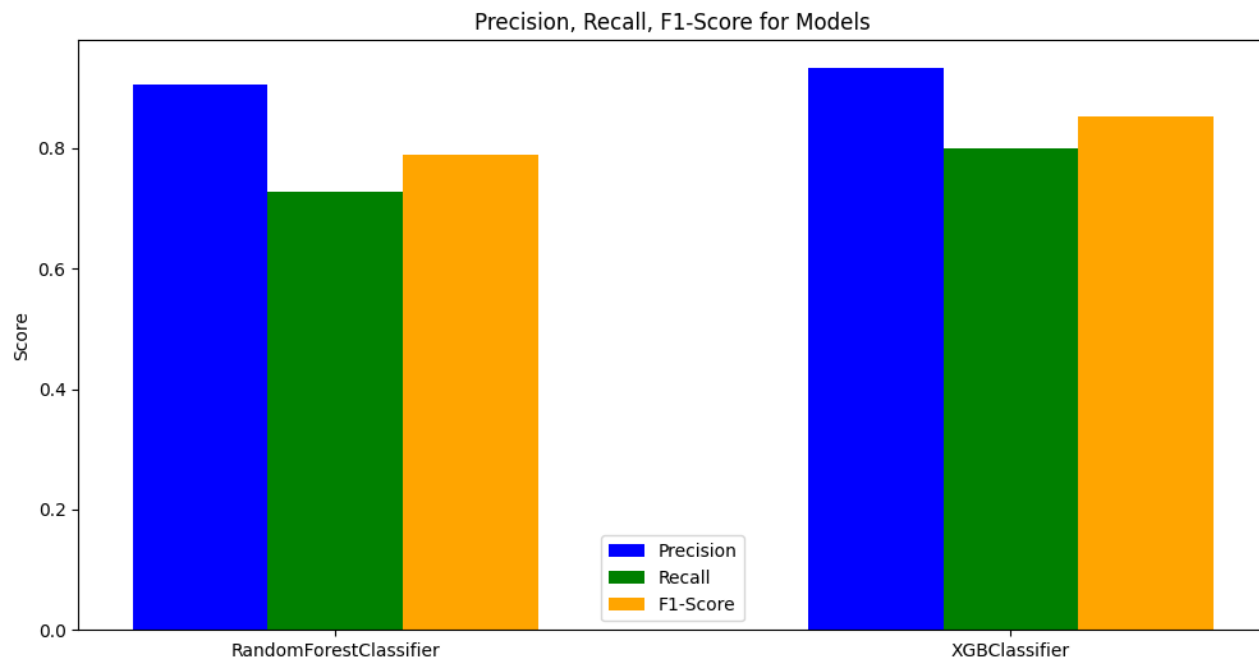
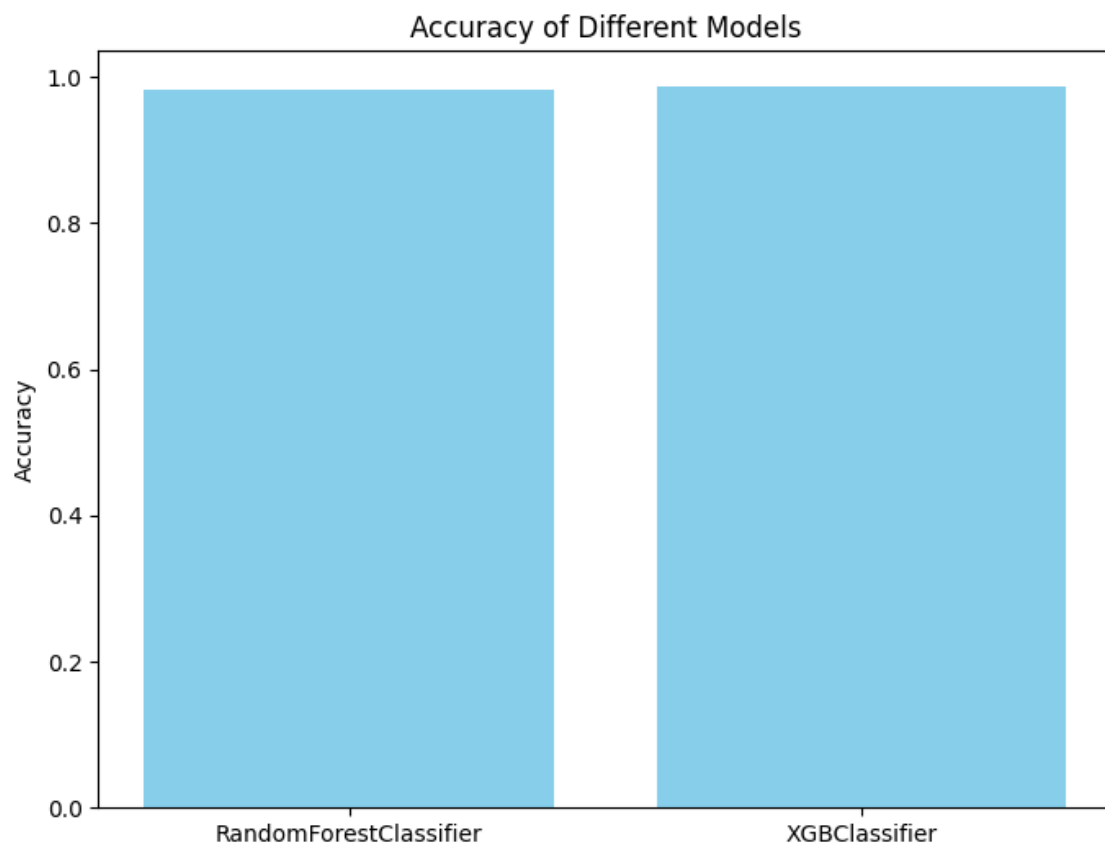
2. Разделение данных на обучающую и тестовую выборки

Чтобы проверить качество модели на новых данных.



Результаты

Обе модели были обучены,
для них были подобраны
гиперпараметры с
помощью GridSearchSv



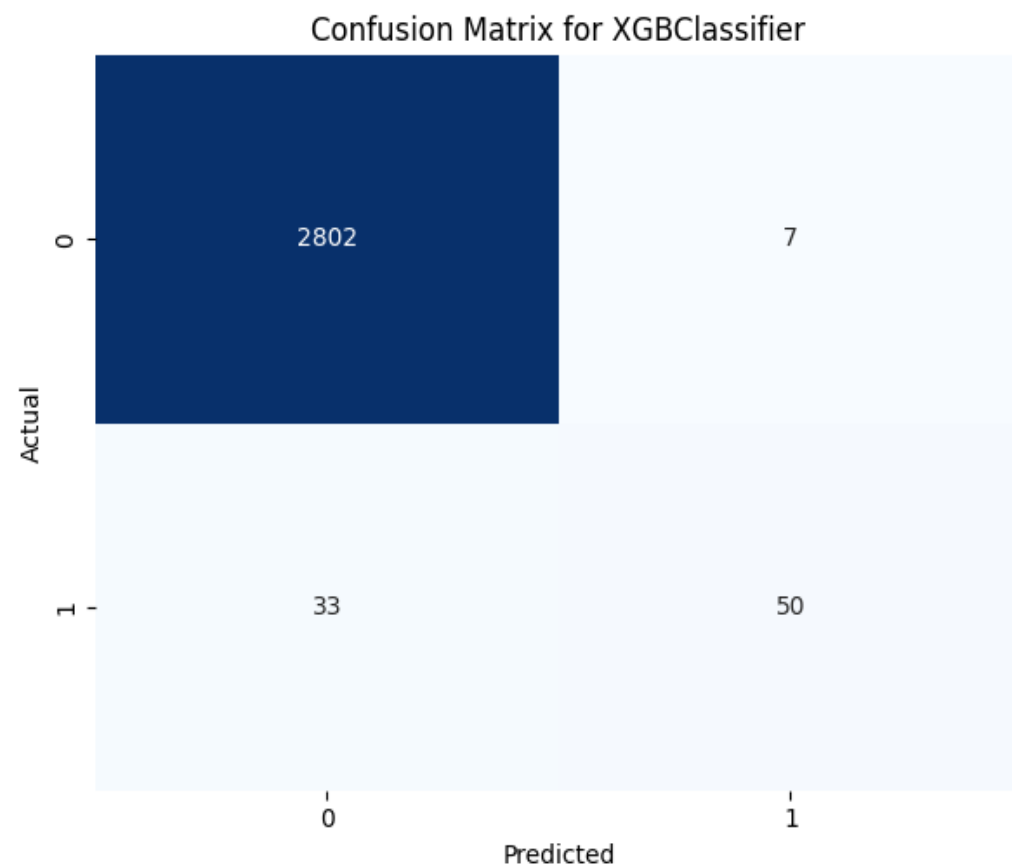
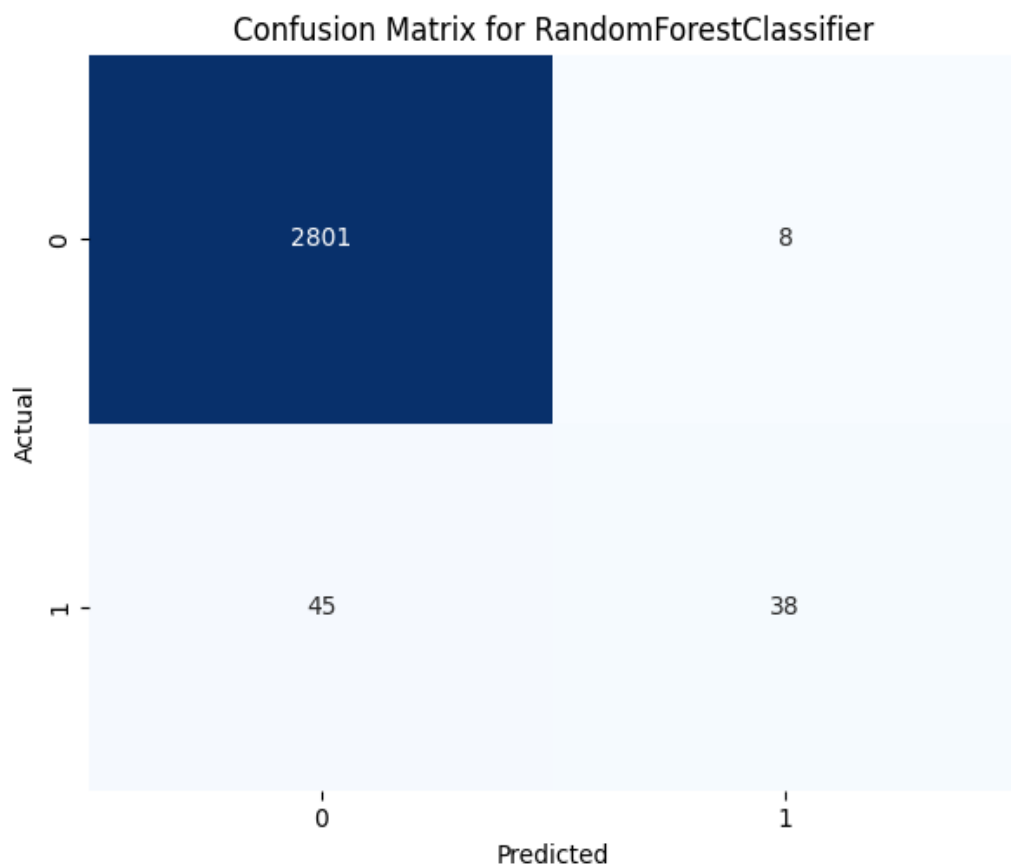
Графики разных метрик



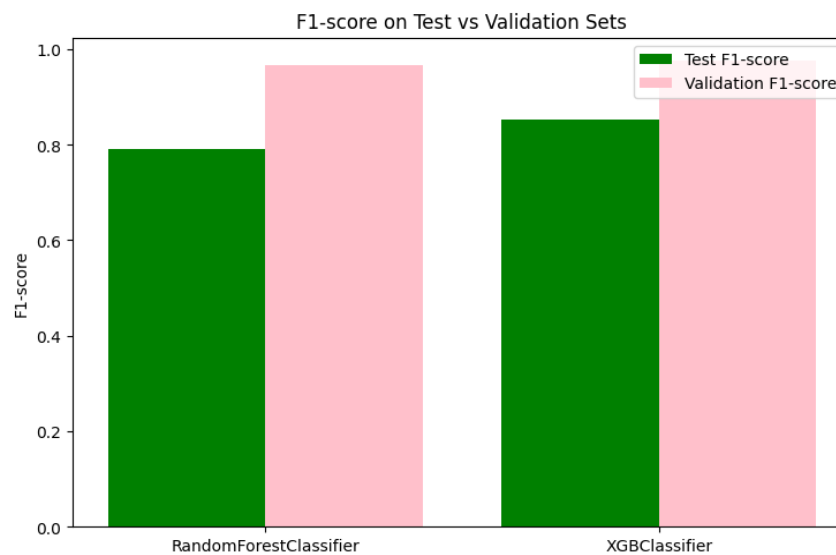
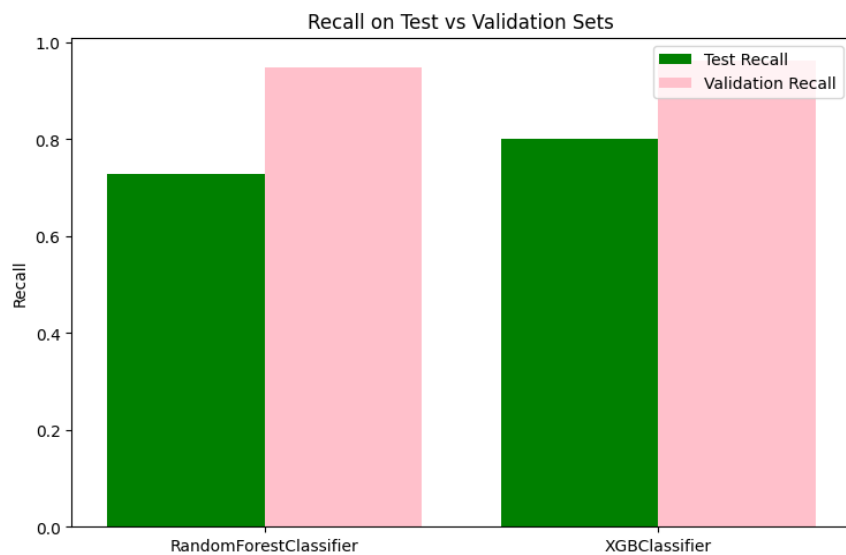
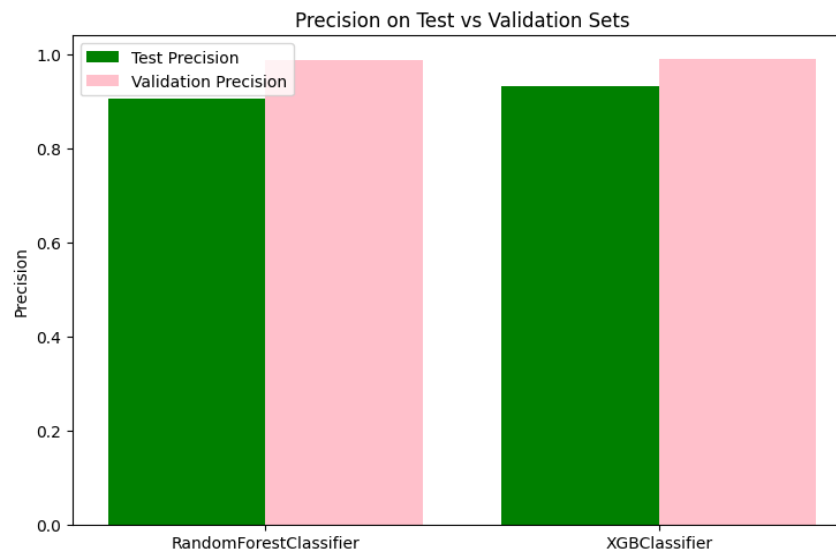
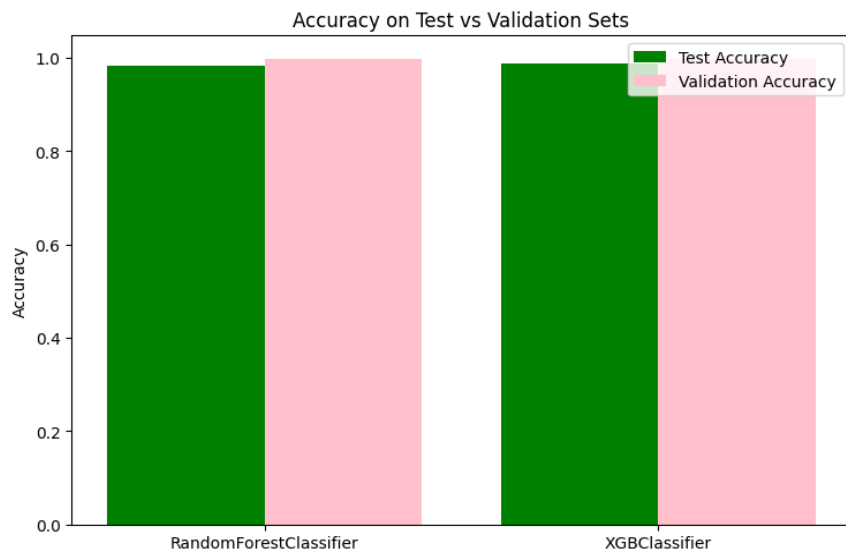
самолет

Выводы по тестовой выборке:

Отличия между моделями в производительности не существенны



Сравнение результатов на валидационном датасете с тестовым



Цифры метрик валидации

Validation Metrics for RandomForestClassifier:

Accuracy: 0.9963

Precision: 0.9884

Recall: 0.9481

F1-Score: 0.9673

Validation Metrics for XGBClassifier:

Accuracy: 0.9972

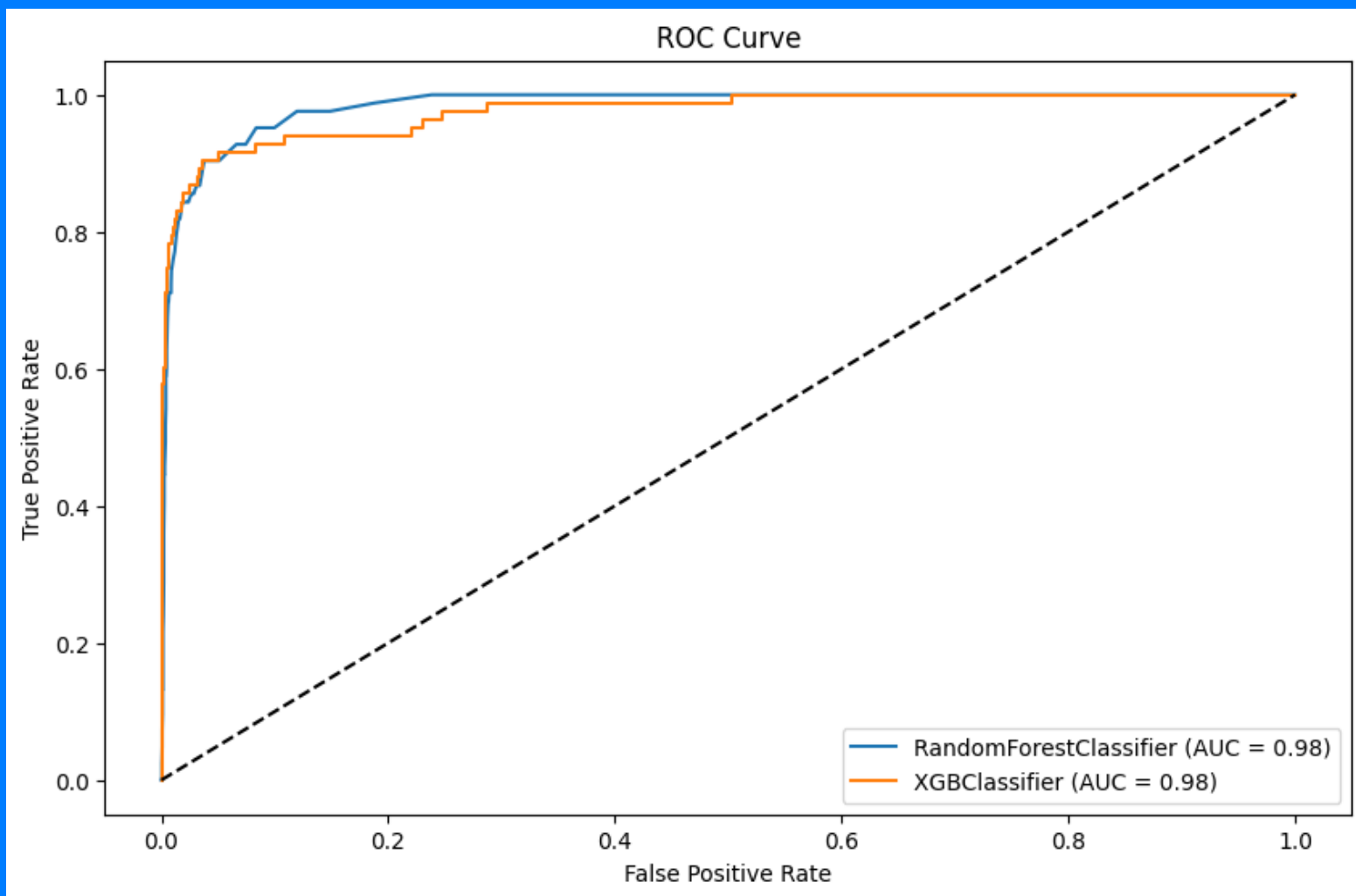
Precision: 0.9902

Recall: 0.9608

F1-Score: 0.9750



Вывод:



В процессе выполнения задачи был проведён полный цикл предобработки данных, включающий устранение пропусков, масштабирование признаков и подготовку данных для обучения модели.

