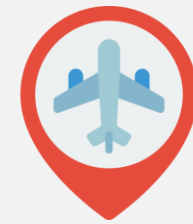


Сервис ранжирования предложений **Auto Avia Offer**



Команда
«fit_predict»
Трек № 2



Описание задачи

Решение

Проблема

Агентам приходится вручную подбирать релевантные варианты для перелета клиентам

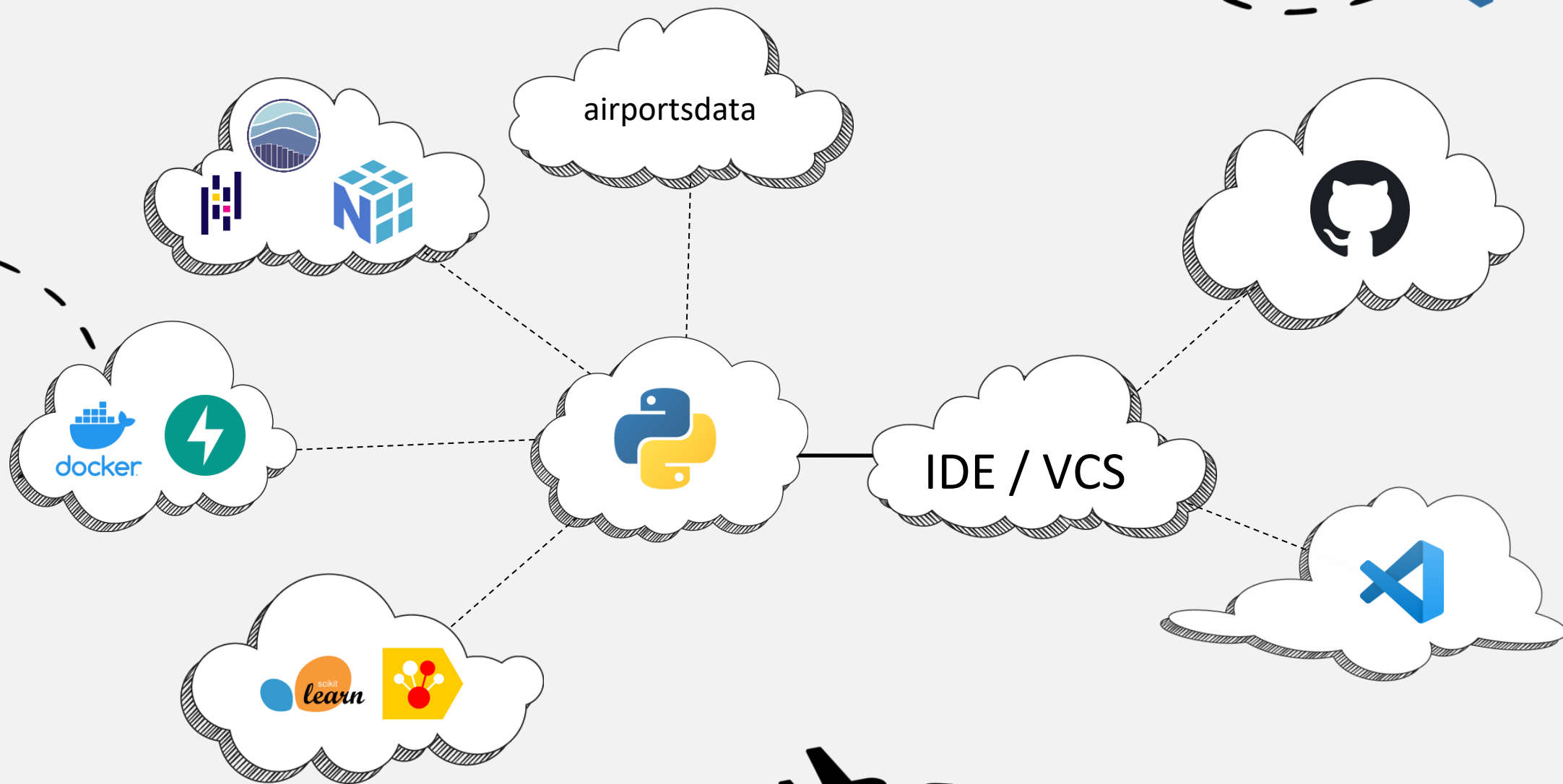
Автоматизировать процесс подбора вариантов перелета в ответ на соответствующую заявку

Цель

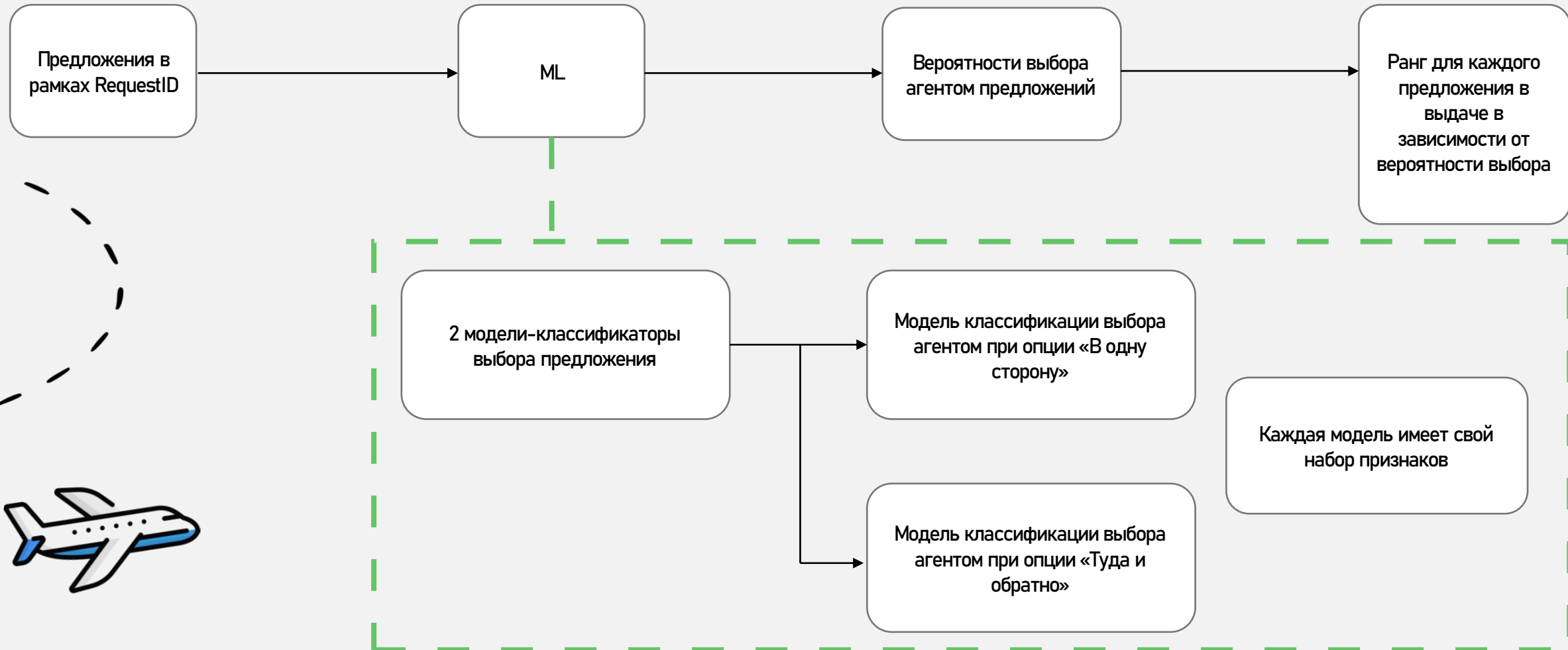
Реализовать сервис машинного обучения, который можно внедрить в существующую платформу



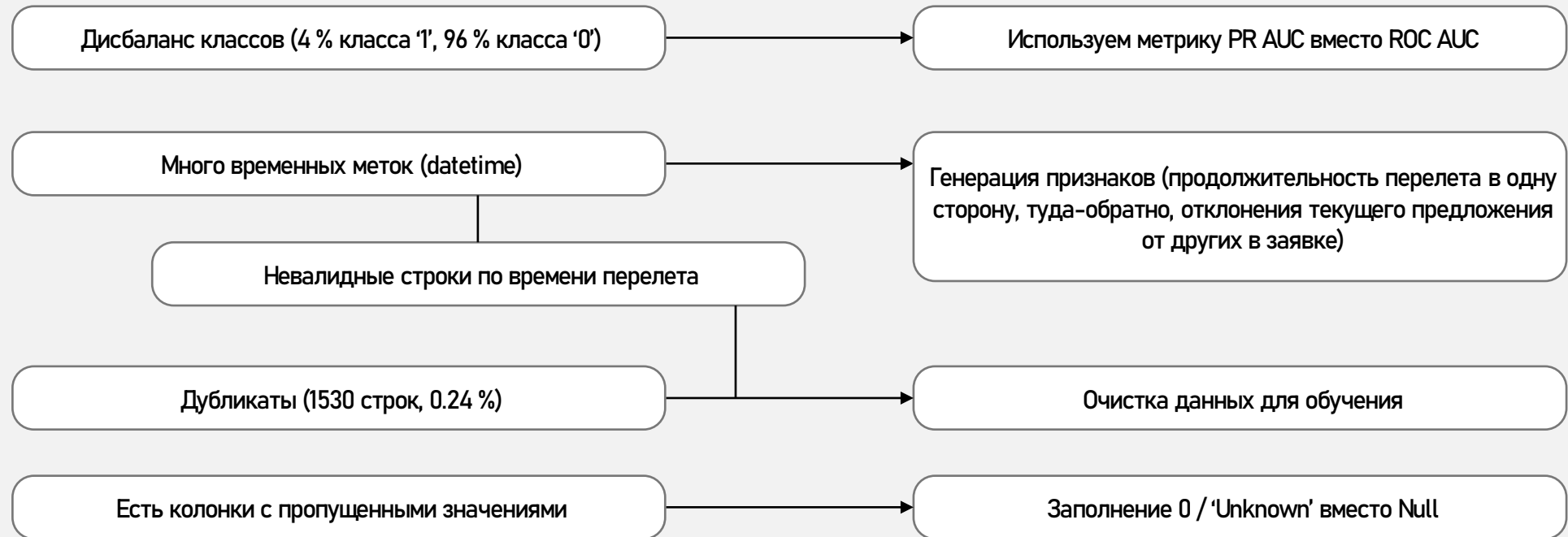
Используемые библиотеки и ПО



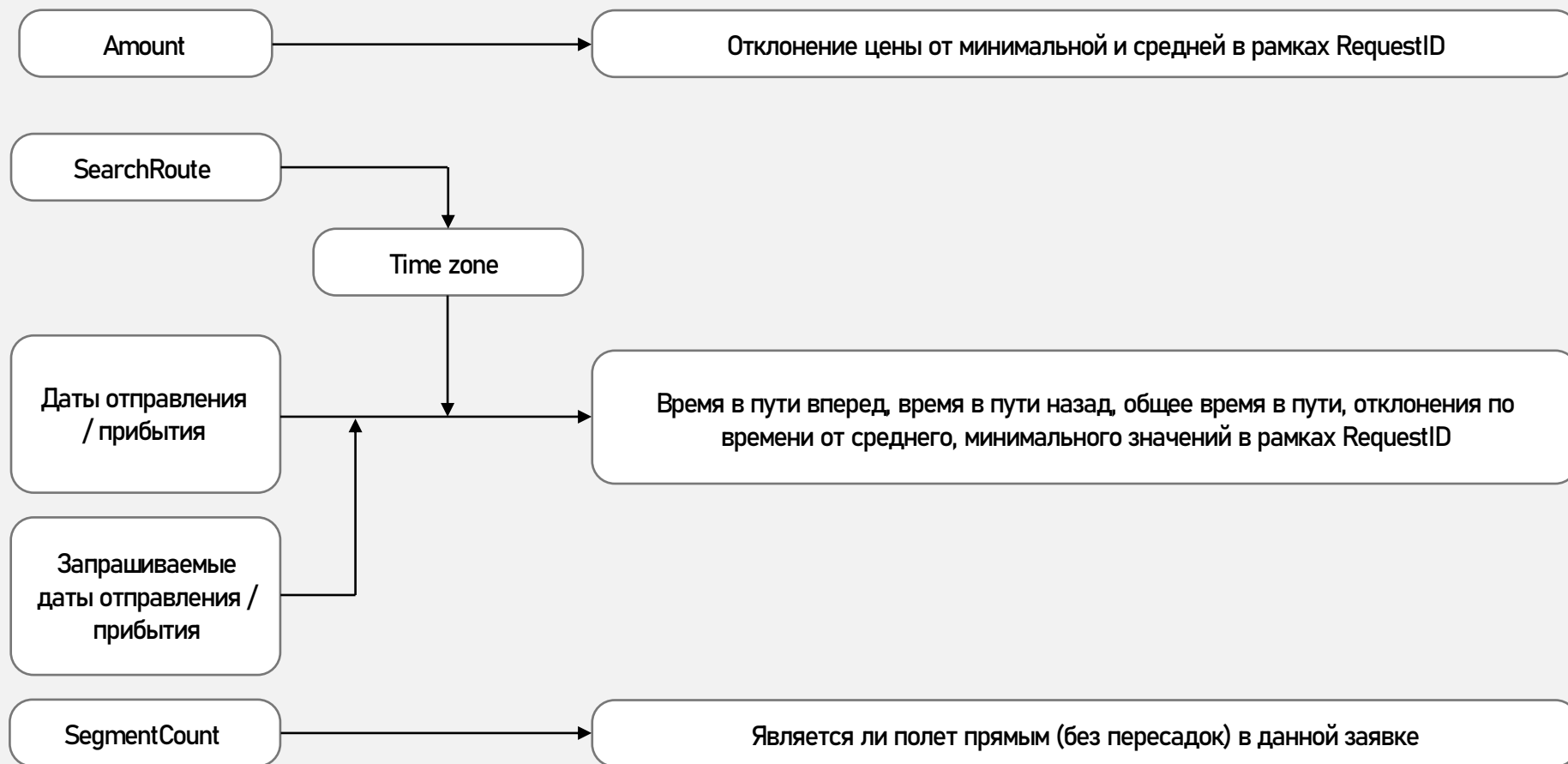
Архитектура решения



Работа с данными



Конструирование признаков



Выбор и обучение модели



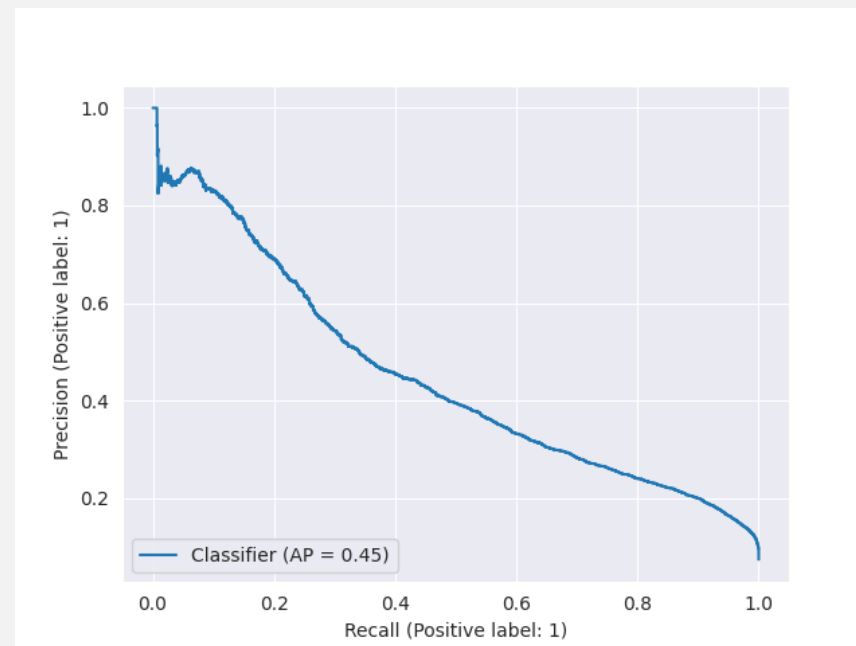
20 % исходной выборки для теста + стратификация



Использован CatBoostClassifier (бустинг)

Поиск по сетке гиперпараметров на GPU + стратификация

PR AUC = 0.45, 0.32 (для модели #1 и #2)



Кривая Precision-Recall для модели №1 на тестовой наборе данных

Результаты прототипирования



Реализован endpoint на FastAPI для работы с данными одного RequestID + упакован в Docker

```
POST http://0.0.0.0:8001/predict_batch

{
  "691": {
    "RequestID": 4552560,
    "EmployeeId": 1789,
    "RequestDate": "2022-07-04 08:30:29.907",
    "ClientID": 45001,
    "TravellerGrade": null,
    "SearchRoute": "MOWSKD",
    "RequestDepartureDate": "2022-07-11 00:00:00.000",
    "RequestReturnDate": null,
    "FlightOption": "SU6551 SVOSKD 2022.07.11",
    "DepartureDate": "2022-07-11 00:20:00.000",
    "ArrivalDate": "2022-07-11 06:00:00.000",
    "ReturnDepartureDate": null,
    "ReturnArrivalDate": null,
    "SegmentCount": 1,
    "Amount": 19948.0,
    "class": "E",
    "IsBaggage": 1,
    "isRefundPermitted": null,
    "isExchangePermitted": null,
    "isDiscount": 1,
    "InTravelPolicy": 1
  }
}
```

Пример запроса к сервису батчем из предложений RequestID

```
POST http://0.0.0.0:8001/predict_batch

{
  "691": {
    "RequestID": 4552560,
    "EmployeeId": 1789,
    "RequestDate": "2022-07-04 08:30:29.907",
    "ClientID": 45001,
    "TravellerGrade": null,
    "SearchRoute": "MOWSKD",
    "RequestDepartureDate": "2022-07-11 00:00:00.000",
    "RequestReturnDate": null,
    "FlightOption": "SU6551 SVOSKD 2022.07.11",
    "DepartureDate": "2022-07-11 00:20:00.000",
    "ArrivalDate": "2022-07-11 06:00:00.000",
    "ReturnDepartureDate": null,
    "ReturnArrivalDate": null,
    "SegmentCount": 1,
    "Amount": 19948.0,
    "class": "E",
    "IsBaggage": 1,
    "isRefundPermitted": null,
    "isExchangePermitted": null,
    "isDiscount": 1,
    "InTravelPolicy": 1,
    "probas": 0.3491147426532159,
    "rank_level": 4
  }
}
```

Пример ответа сервиса

Выводы

Решение позволяет использовать микросервис, который будет возвращать позиции для предложений одного RequestID в выдаче пользователю

Используя скрипт можно периодически обучать новые модели силами 1 аналитика данных

Можно проводить эксперименты по добавлению новых признаков для улучшения метрики (отталкиваясь от текущего решения)

Req # 1 | Offer # 11

Req # 1 | Offer # 21

Req # 1 | Offer # 42



Req # 1 | Offer # 42

Req # 1 | Offer # 11

Req # 1 | Offer # 21

Probab score



Что можно улучшить?



Уменьшить дисбаланс классов

Довести долю класса '1' до 20-40 %

Добавить больше признаков клиента, для которого подбираются предложения

История поездок, наличие бонусных карт и тд.

Добавить информацию об авиакомпаниях

Рейтинг а/к, среднее время задержек и др.

