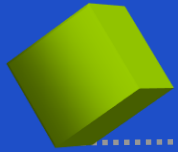


Анализ трековых данных воздушных судов

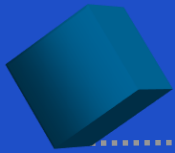
Команда RSP



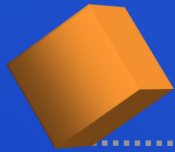
Команда



Петров Родион Андреевич – team leader



Суворов Максим Дмитриевич – data scientist



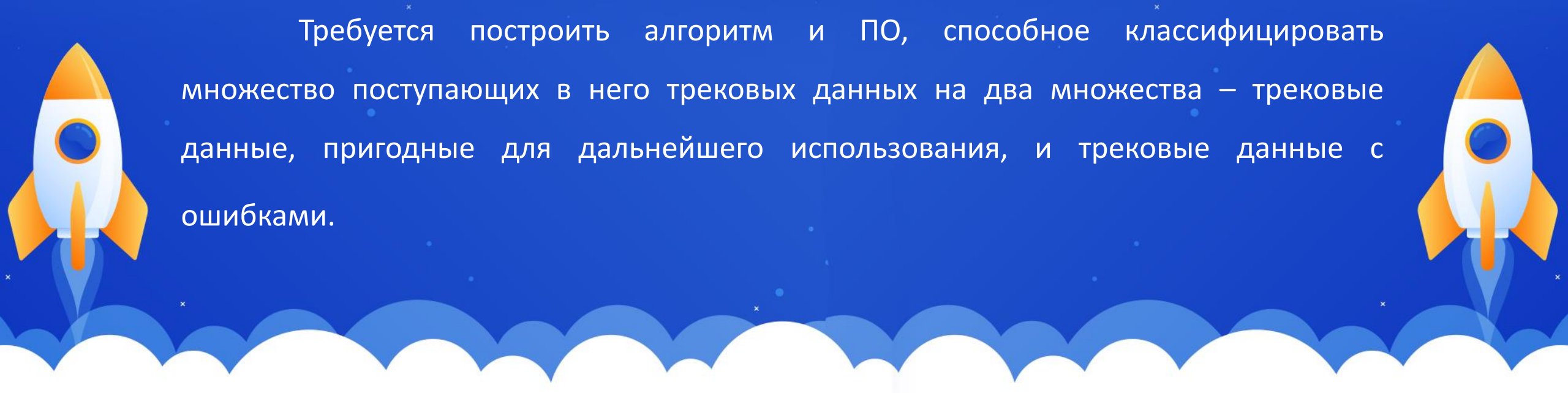
Ракитянский Владислав Максимович – UI/UX designer



Проблема

Поставлена задача определения треков воздушных судов, содержащих ошибки. Данная проблема является предварительным этапом проведения исследований, а также разработки интеллектуальных систем поддержки системы ОрВД.

Требуется построить алгоритм и ПО, способное классифицировать множество поступающих в него трековых данных на два множества – трековые данные, пригодные для дальнейшего использования, и трековые данные с ошибками.



Описание решения

1

Входные данные в виде текстового файла следующего формата:
время, идентификатор трека, широта, долгота, высота, код
ответчика, позывной.

2

Предварительная обработка данных для подачи в алгоритм ML.

3

Получение вероятностного прогноза о принадлежности трека к
• одному из классов (хороший трек или трек с ошибкой) •

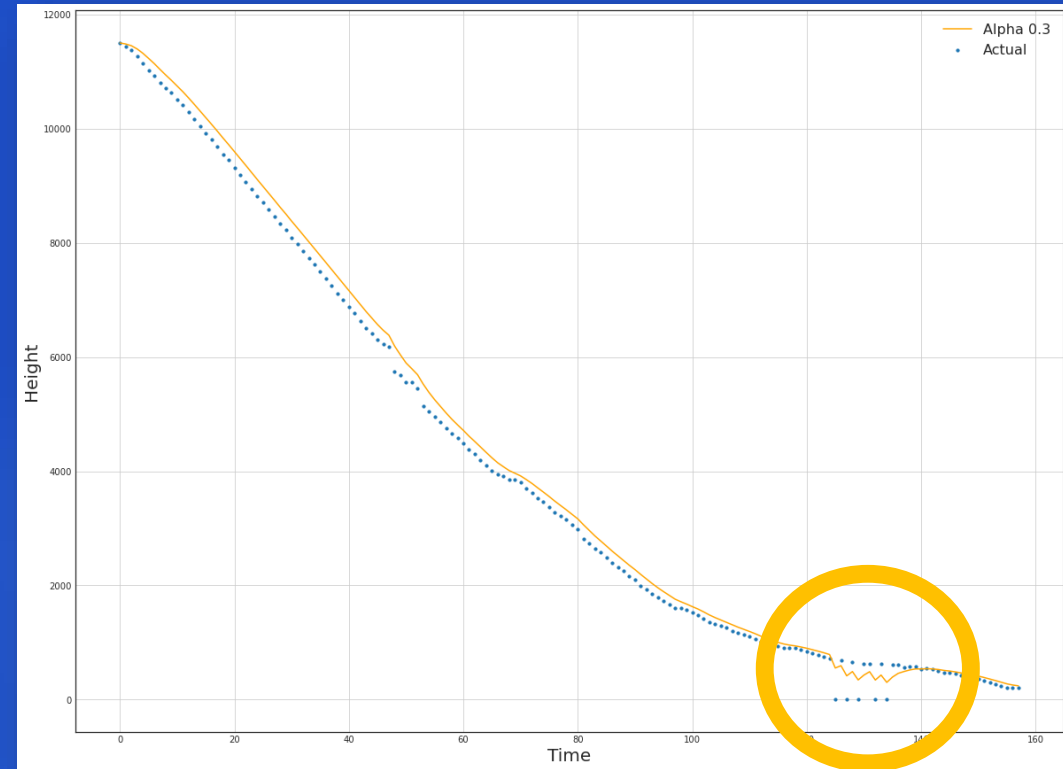


Особенности решения

Предварительная обработка данных основана на вычислении экспоненциального скользящего среднего для координат каждого трека.

Детектирование аномальных показаний происходит путем поточечного сравнения предсказанного значения с исходным с помощью различных метрик (MAE, MSE).

Идея метода скользящего среднего основана на предсказании будущего значения на основе предыдущего, поэтому резкие скачки или разрывы в данных увеличивают показания метрик.



Особенности решения

В результате обработки всех треков создается dataframe со следующими параметрами: ID трека, MAE по координатам, MSE по координатам, среднее значение координат, среднее отклонение координат, размер трековых данных.

Полученный dataframe передается в обученную модель на основе алгоритма «случайный лес». Данный метод показал наилучшие результаты по сравнению с линейными моделями.

Выходными данными является список идентификаторов трека с заданными вероятностями принадлежности к заданному классу.



Бизнес ценность

Решение данной проблемы позволяет уменьшить трудозатраты на разметку трековых данных с целью проведения исследований, а также построения сглаживания и фильтрации траекторий.



Стек технологий

Преимущества использованного стека:

- Невысокая требовательность к вычислительным мощностям
- Интерпретируемость результатов
- Модульность ПО

Использованные средства разработки:

- PyCharm
- Jupyter Notebook

Использованные библиотеки:

- Numpy
- Pandas
- Sklearn
- Matplotlib
- PySimpleGUI

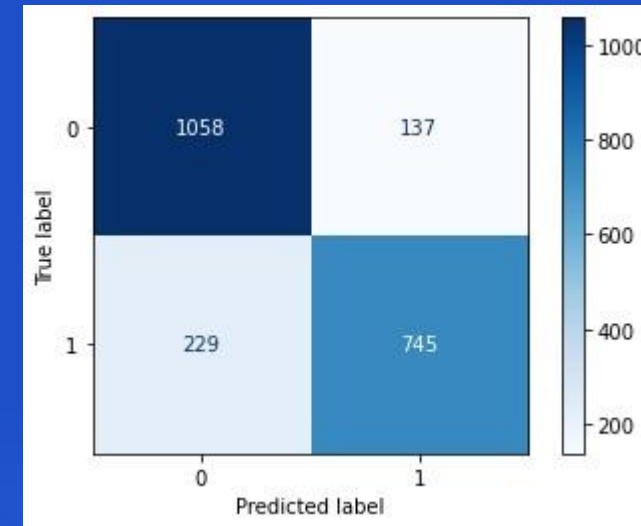
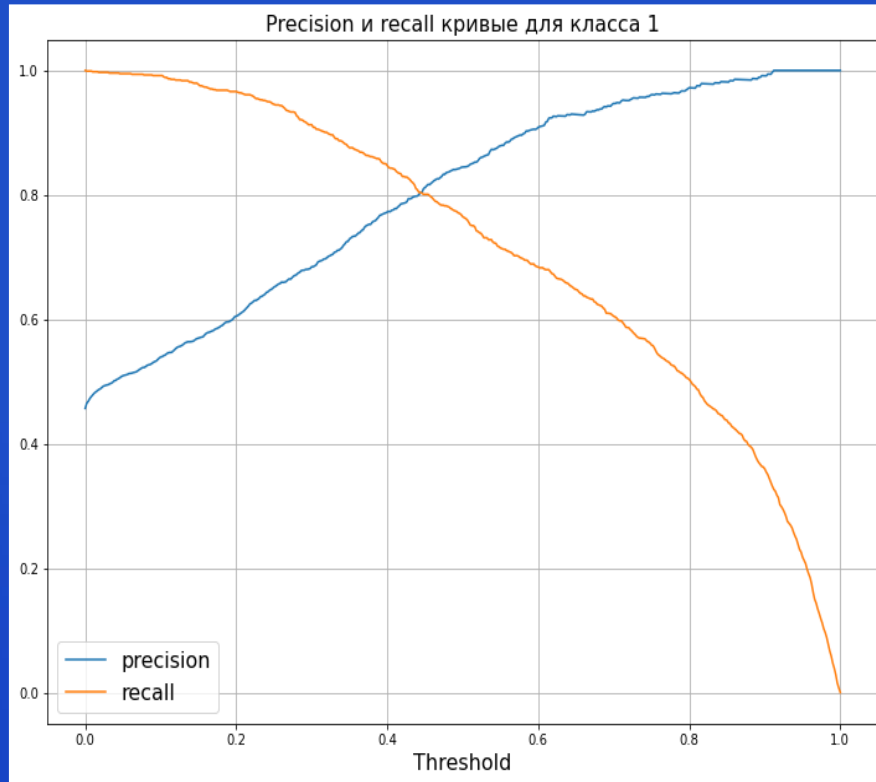


Потенциал решения

- Улучшение модели на основе исследования паттернов плохих треков
- Оптимизация предварительной обработки данных
- Визуализация треков в виде трехмерных проекций
- Улучшения пользовательского интерфейса



Демонстрация решения



	precision	recall	f1-score	support
0	0.82	0.89	0.85	1195
1	0.84	0.76	0.80	974
accuracy			0.83	2169
macro avg	0.83	0.83	0.83	2169
weighted avg	0.83	0.83	0.83	2169

Программное обеспечение

Track Checker

File for check:

In file: C:/Users/Vlad/Desktop/lol_ya_zhe_typoy/TESTdf.txt

good tracks id

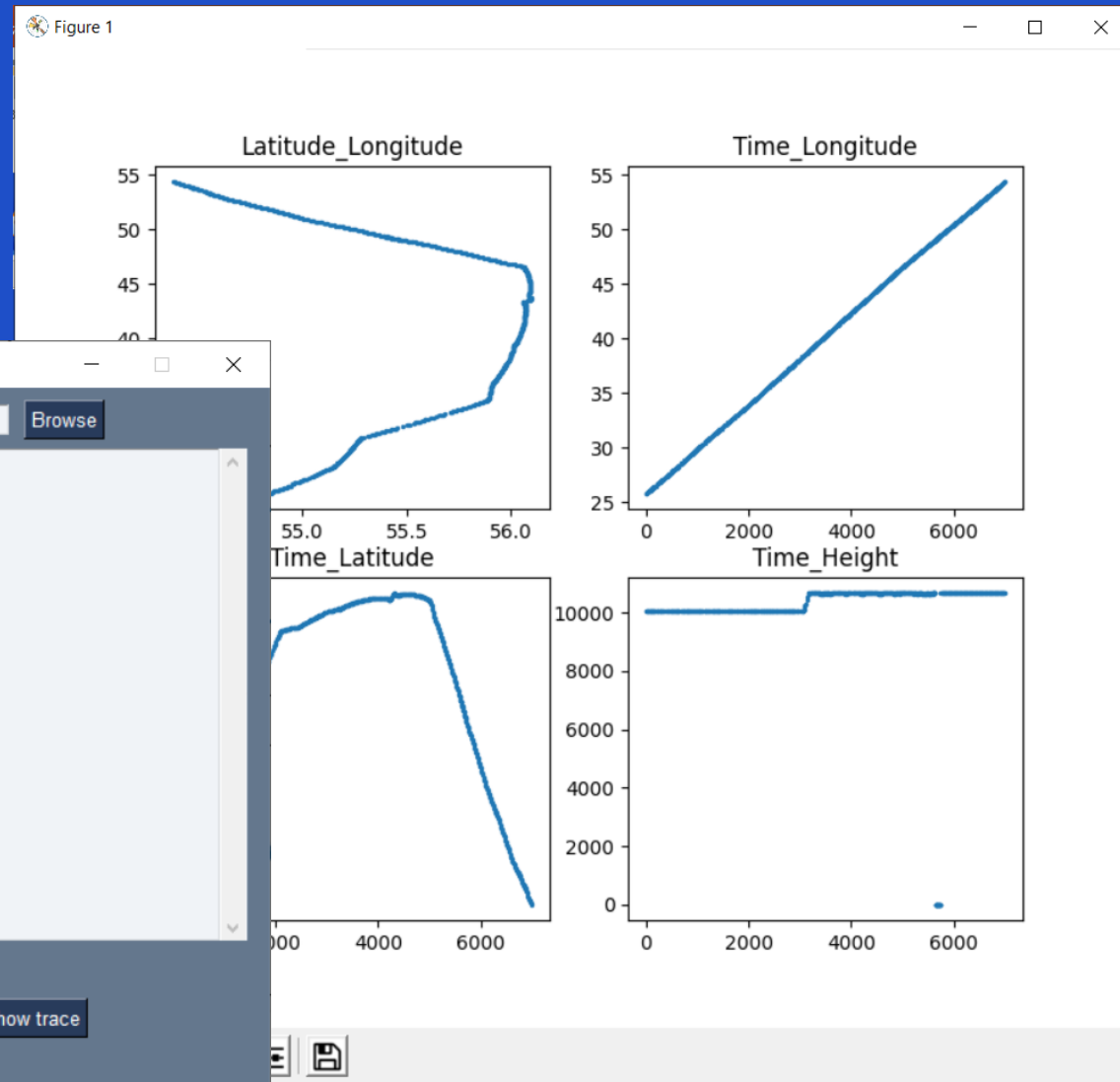
1 3 2

bad tracks id

5 4

id_to_class_probabilities.txt created!

Enter id:



Спасибо за внимание!

