

Решения кейса компании

СИТИМОБИЛ

команды MLшарики

Описание решаемой проблемы



Бизнес проблема

- ▶ Расширение клиентской и водительской баз



Проблема клиентов

- ▶ Безопасное перемещение
- ▶ Высокий уровень сервиса



Проблема водителей

- ▶ Получение справедливого вознаграждения за труд

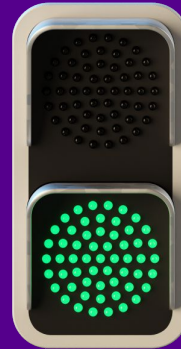
Исследование рынка

- ▶ Отсутствие динамических/соревновательных систем оценивания водителей
- ▶ Фиксированные баллы достижения определенного статуса



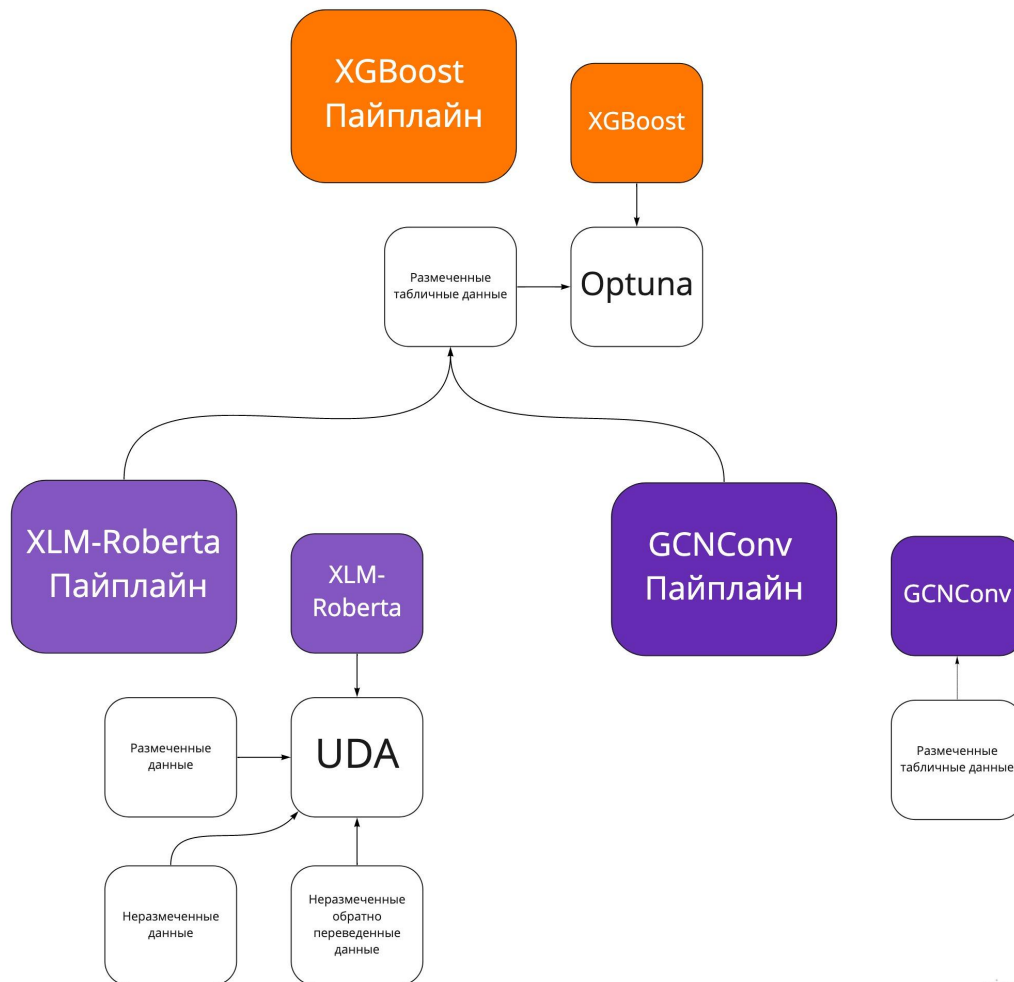
Описание Решения

- 1) Создание стека робастных моделей для распознавания случаев агрессивного вождения
 - ▶ Модель для обработки комментариев
 - ▶ Модель для обработки табличных данных
 - ▶ Модель для обработки маршрутов
- 2) Создание Метрики для оценивания качества работы водителей
- 3) Разработка возможных рангов водителей на основании метрики
- 4) Создание системы перехода между рангами



Стэк технологий

- Unsupervised Data Augmentation с XLM-Roberta
- Optuna с XGBoost
- Graph Attention Networks



XGBoost

Признаки:

1. arrived_distance
2. arrived_duration
3. distance
4. duration
5. from_latitude
6. from_longitude
7. to_latitude
8. to_longitude
9. mark
10. client_rate_ride
11. xlm_roberta
12. gnn

Отбор признаков -
Жадный перебор

Тюнинг - Optuna

СИТИМОБИЛ

XLM-Roberta

Подходы:

1. Обратный перевод комментариев
2. UDA

GNN

Признаки:

1. Расстояние между точками
2. Время между точками
3. Скорость между точками
4. Угол поворота

Метрика

DriverScore =

$(time_dedication + mean_driver_rate + (1 - mean_aggr) + achievements) / 4 * 1000$

- ▶ $time_dedication = \min(driver_working_time / top_twenty_percent_working_time, 1)$
- ▶ $mean_driver_rate = driver_rates.mean() / 5$
- ▶ $mean_aggr = aggr_array.mean()$
- ▶ $achievements = (achievement_results / best_achievement_results).mean()$

Система рейтинга

- 1) Каждую неделю создается список таксистов отсортированный по метрике *DriverScore*
- 2) Далее водители разделяются на следующие ранги:



Ранг 1



Ранг 2



Ранг 3



Ранг 4



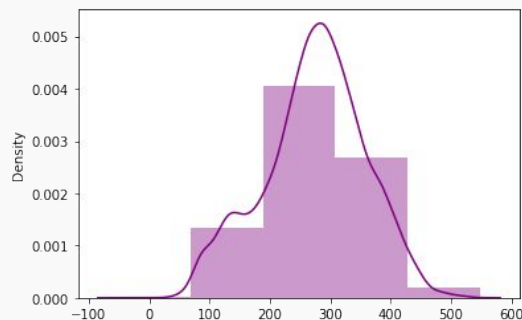
Ранг 5

- 3) По *DriverScore* нижнего человека с округлением до сотен каждой группы устанавливаются пороги прохождения в ранги

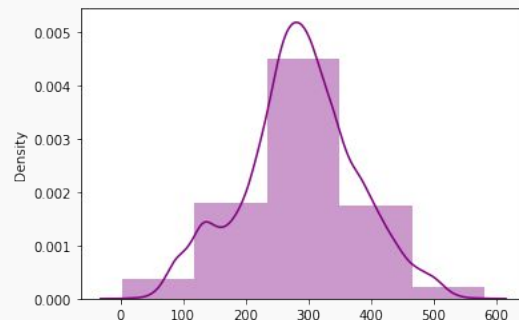
Распределения DriverScore

Рассчитав DriverScore
по 4 месяцам мы
получили следующие
распределения

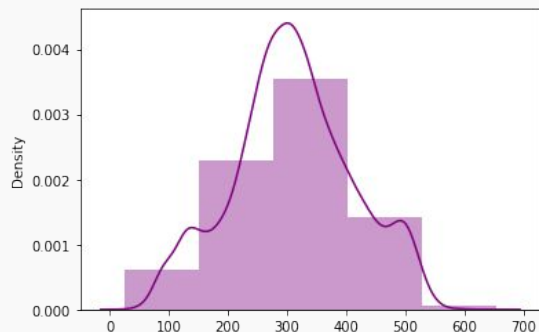
Март



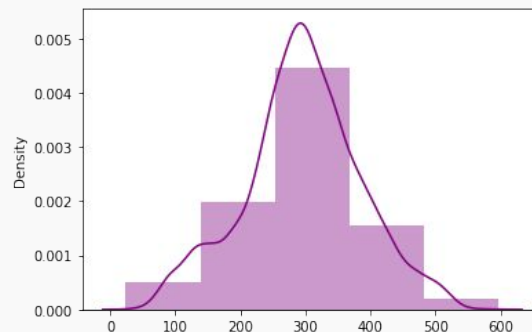
Апрель



Май

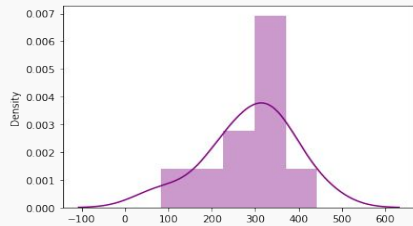


Июнь



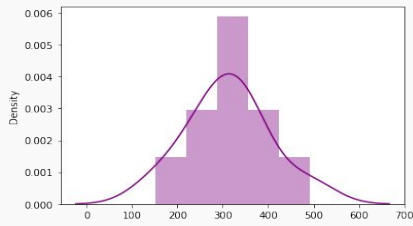
Пример работы системы на случайной выборке

Март



id_driver	driver_score
e36fc7379e0b25cc5c2f1c8bec53fb8	82.062854
bbaded8e02a7bcaa2509bcd049a9539d	208.017691
47798fe48fae5304debcdefdc0291e06	227.665771
e55257c1e5553b1ba3014ae7f83cc783	253.177015
9a5915145597716199146e872b18d36a	311.428018
efbc74dfaa05538bd9d71ea48ef58c011	312.752967
13d88dfae97d09078df664b5f251807	336.582348
ec11fcba8654b37c5719bedf367a82c6	353.612482
964d12da2622477f5882b83d0f86be28	353.753380
2197dcfc473a3cc1f453546be7096f80	442.949938

Апрель



id_driver	driver_score
6fbac7e3d172cd66effe7e0e00166e5	152.114643
34cc8a00b6986adff3d743ec240bc888	231.863038
f774467164b43ecbe3458c27bd670575	245.177894
cffb3ac3238b0ce50b9072ffaf874cf6	294.718034
7b538fc404e26e7dcca954602cb1fc60	301.766376
da46a65230c4ad3e3d86efc020c5c2226	329.729136
a2e4a57a9d2e18689aa1b80e1c319af4	330.236056
6794843abf2ba8cd25e13378039c3386	355.867290
ba11e44ed384b234ba5f72388ada61e9	387.198013
55088dce4cb15a9c3af0423e8f0c35d6	491.546008



Полученные разделители:

- 154.24027035
- 226.41768716
- 298.59510396
- 370.77252076

Плюсы системы:

- 👍 Мотивация водителей к улучшению уровня сервиса
- 👍 Повышение лояльности водителей к сервису Ситимобил
- 👍 Отсутствие поощрения чрезмерной работы



Как улучшить решение?

1) Повысить точность модели

а) Мотивировать пассажиров ставить отметки и комментарии (повысить полноту данных)

б) Лучше обучить модели и/или подобрать более подходящие

2) Сделать метрику более справедливой, отталкиваясь от практической ценности её параметров



Наша команда



Кирилл Захаров
kil4gen@gmail.com

1. Работы с табличными данными
2. Реализация метрики



Лев Новицкий
levnovitskiy@gmail.com

1. NLP
2. Создание рейтинга



Семён Беляев
b.semeka.13@gmail.com

Работа с GNN

Решения кейса компании

СИТИМОБИЛ

команды MLшарики