



Факультет компьютерных наук

Машинное обучение и  
высоконагруженные системы.

Москва 2024

# ***Предсказательная модель для игры Dota 2***

Командный проект



Выполнили студенты  
образовательной  
программы  
«Машинное обучение и  
высоконагруженные  
системы»

Горбач Илья  
Мустафин Фарид  
Кучиев Александр

Куратор:

Макарова Мария



## *Краткий ликбез по игре Dota 2:*

*Dota 2 – многопользовательская командная компьютерная игра в жанре MOBA.*

Игра изображает сражение на карте особого вида; в каждом матче участвуют две команды по пять игроков, управляющих разными «героями» — персонажами с различными наборами способностей и характеристиками.

*Цель игры – уничтожить особое строение врага - «Ancient».*



Рис. Игровой процесс.

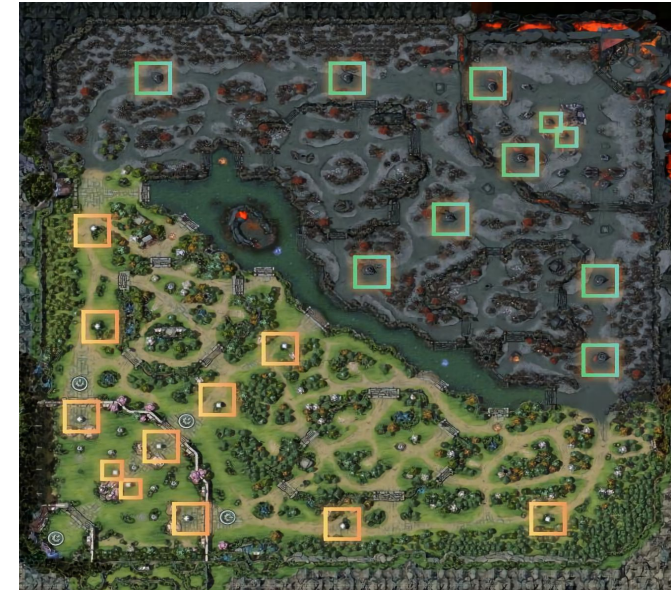


Рис. Карта игры

*Dota 2 является крайне популярной игрой – ежегодно проводятся десятки турниров с призовым фондом в десятки-сотни тысяч долларов*

Ежегодно проводится главный чемпионат по игре – The International. Призовой фонд турнира в 2021 году составил 40 млн. долларов.

*Игра популярна также среди обычных игроков - средний онлайн в игре составляет порядка 400-500 тысяч игроков.*





The International 2019



Российская команда Team Spirit – 2-х кратный чемпион The International.

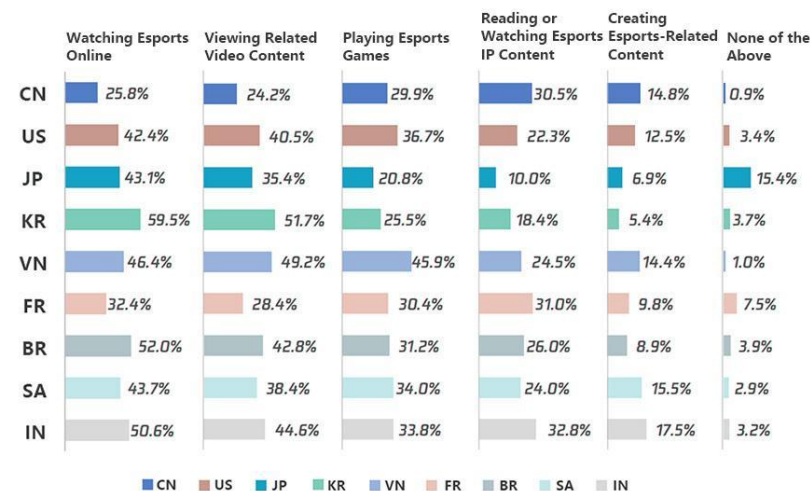


## Актуальность работы:

По прогнозам, в 2023 году глобальная аудитория киберспорта увеличится до 574 миллионов человек. Из них 283 миллиона будут являться основными энтузиастами киберспорта, а остальные 291 миллион будут случайными зрителями.



Distribution of eSports Engagement Among  
Internet Users in Different Countries



## ТОП-5 киберспортивных игр по количеству часов просмотра в 2022-м году:

League of Legends — 611,77 млн часов

CS:GO — 448,4 млн часов

Mobile Legends: Bang Bang — 331,4 млн часов

Dota 2 — 313,6 млн часов

VALORANT — 283,4 млн часов



## Суть проекта

*Реализация support-player system закрывающая основные потребности игроков:*

*Dota Pro Tracker + Dota Buff + Stratz*

- Выбор наиболее подходящих героев относительно своего и вражеского пика
- Выбор покупки наиболее оптимальных предметов (итем билд) в зависимости игровой ситуации
- Выбор наиболее оптимальной прокачки персонажа в течение всей игры
- Вероятность победы команды исходя из внутриигровой ситуации



## Описание задачи:

1. Сбор, обработка и анализ данных для построения модели машинного обучения для игры Dota
2. Построение модели машинного обучения для предсказания победы в текущий момент лайв матча
3. Разработка телеграмм – бота для вывода лайв статистики и предсказаний победы для выбранного матча
4. Разработка мобильного приложения для полного вывода статистики и предсказаний победы





## *Состав нашей команды:*



**Горбач Илья** – разработчик архитектуры бд и алгоритмов для работы с данными

@paral1ax



**Александр Кучиев** – аналитик, разработчик и тестировщик модели машинного обучения

@ascetto



**Мустафин Фарид** – аналитик и разработчик телеграмм

бота  
@lanarich



**Мария Макарова**  
*Куратор проекта*

@mariagolddd



## Описание данных:

Главными входными данными являются завершенные матчи Dota 2

- Общие поля для матча:

"match_id": 123,	"tower_status_dire": 0,
"barracks_status_dire": 0,	"tower_status_radiant": 0,
"barracks_status_radiant": 0,	"version": 0,
"duration": 0,	"replay_salt": 0,
"engine": 0,	"series_id": 0,
"first_blood_time": 0,	"series_type": 0,
"game_mode": 0,	"radiant_team": { },
"human_players": 0,	"dire_team": { },
"leagueid": 0,	"league": { },
"lobby_type": 0,	"skill": 0,
"radiant_gold_adv": [0],	"radiant_win": true,
"radiant_score": 0,	"radiant_xp_adv": [0]

- Общие поля для игрока:

- Предметы
- Золото в минуту
- Опыт в минуту
- Убийства
- Смерти
- Помощи
- Позиция
- Статус отключения от игры
- Добивания крипов
- И тд



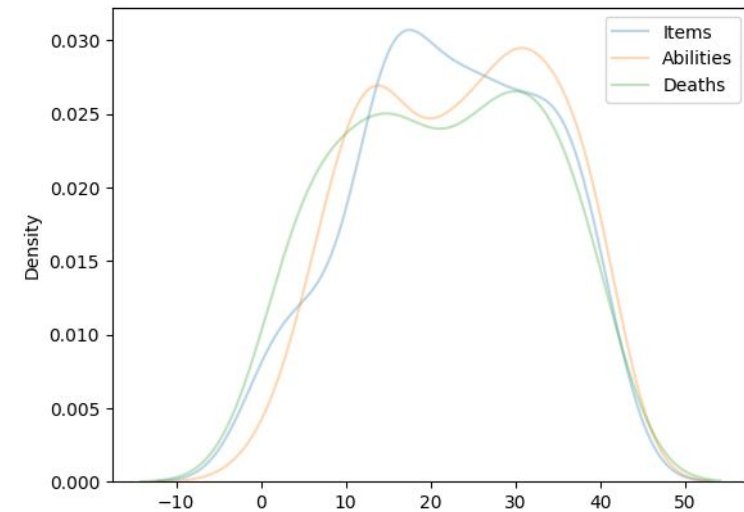
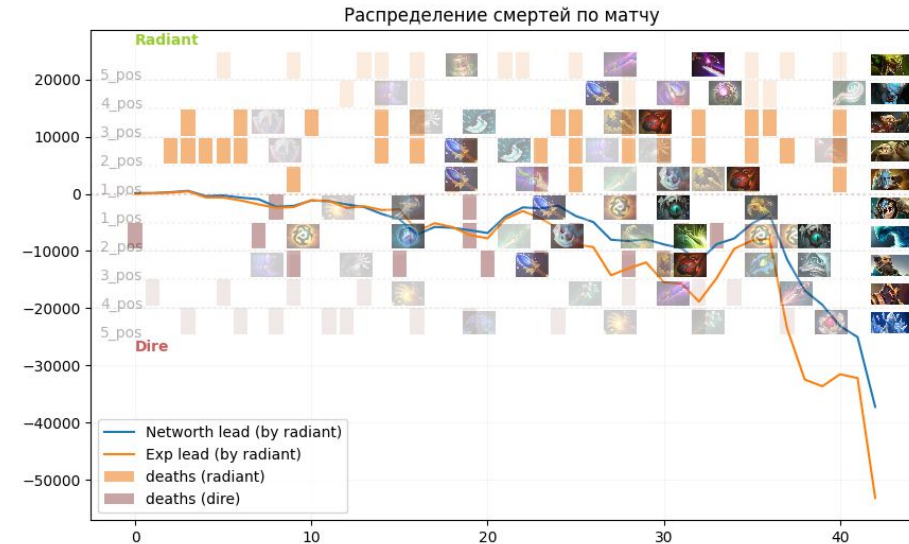
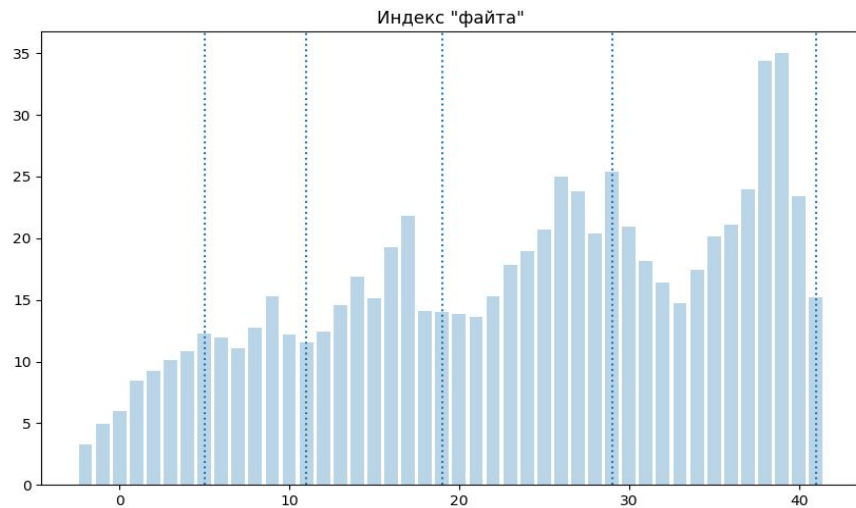
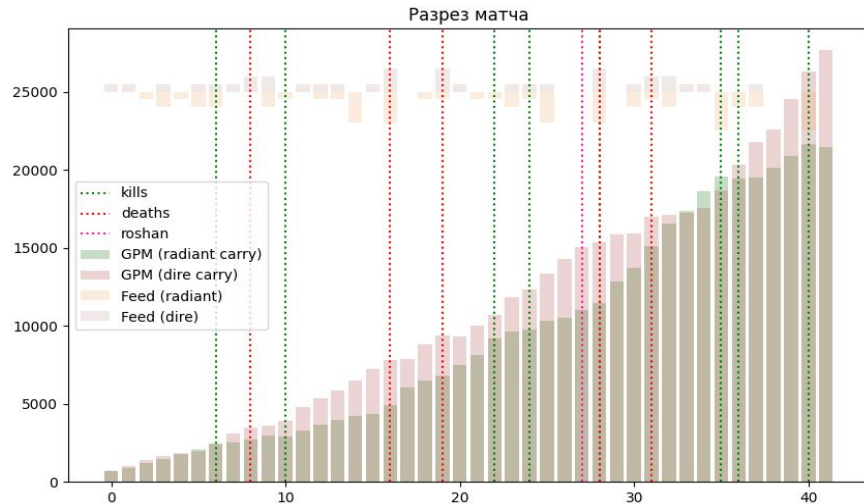


## ***Сбор данных:***

1. Анализ существующих сервисов, предоставляющих API со статистикой Dota 2 матчей.
2. Написание скрипта, для получения данных:
  1. Пробегается по всем последним сыгранным матчам выбранного ранга.
  2. С помощью Rest API сервиса OpenDota получаем id таких сыгранных матчей
  3. Кладем их в MongoDB
  4. Достаем из MongoDB id сыгранных вчера матчей (сделано, чтобы Stratz успел обработать новые матчи)
  5. С помощью GraphQL делаем запрос в сервис Stratz и оттуда получаем расширенную статистику по матчам (Так сделано, потому что OpenDota разрешает только 2к бесплатных запросов в день, а Stratz – 10к)
  6. Кладем полные данные по матчу в другую коллекцию MongoDB (60к матчей весят 500 Гб)
3. Написание скрипта, для обработки данных в приемлемый табличный вид:
  1. Каждый матч был разбит на  $n$  объектов, где  $n$  – количество минут в матче. То есть 1 матч длиною в 40 минут превращается в 40 объектов. Это сделано, чтобы избавиться от вложенных поминутных массивов
  2. Такой трюк помог увеличить выборку с 100к матчей, до 4 миллионов поминутных объектов
  3. Теперь мы имеем табличные данные без вложенных массивов. Каждый объект хранит в себе



## EDA:





## Описание модели:

### Были обучены две модели:

**CatBoostClassifier** со следующими гиперпараметрами:

- 1) iterations=200
- 2) learning\_rate=0.2
- 3) depth=5
- 4) min\_child\_samples=10,
- 5) custom\_loss=['AUC', 'Accuracy'],
- 6) boosting\_type='Ordered',
- 7) bootstrap\_type='Bayesian',
- 8) objective='CrossEntropy')

**RandomForestClassifier** со следующими гиперпараметрами:

- 1) n\_estimators=300
- 2) min\_samples\_split=39
- 3) min\_samples\_leaf=18
- 4) max\_features='sqrt'
- 5) max\_depth=20
- 6) bootstrap=True

*Гиперпараметры подбирались с помощью Optuna*

- **accuracy score** на отложенной выборке составил 0.76
- **ROC-AUC** на отложенной выборке составил 0.86

*Гиперпараметры подбирались с помощью RandomizedSearchCV*

- **accuracy score** на отложенной выборке составил 0.76
- **ROC-AUC** на отложенной выборке составил 0.84



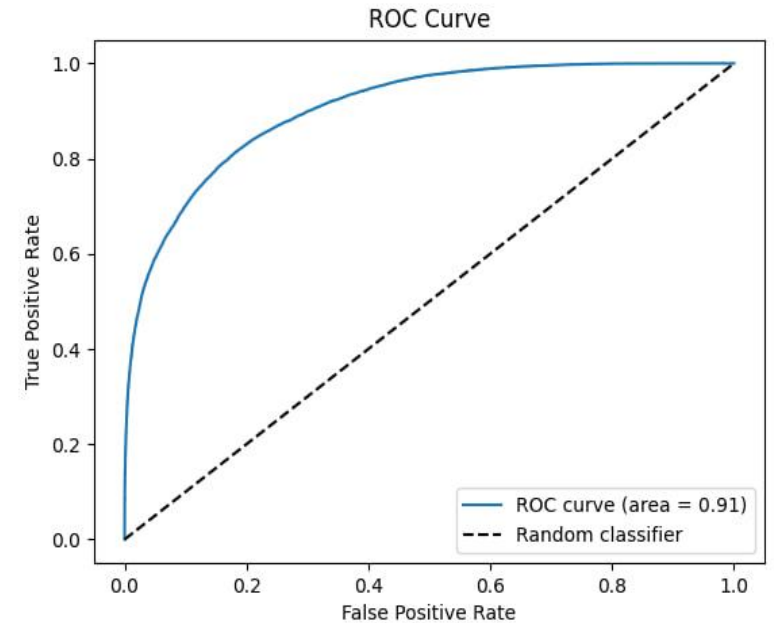
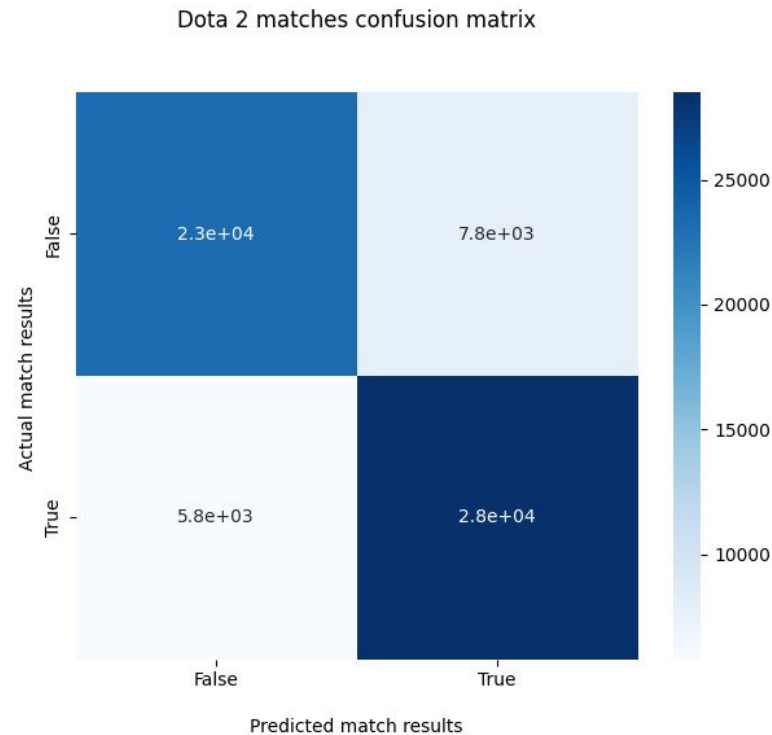


## Результаты модели:

Для проверки качества модели бинарной классификации были выбраны следующие метрики:

1. Accuracy = 0.79
2. ROC-AUC = 0.8
3. F1 - score = 0.91

```
Accuracy score: 0.791984937315352  
F1 score: 0.8074748877208392
```





## *Первичная реализация:*





## ***В процессе разработки:***

### ***Задача:***

***Предложение выбора персонажа на основе пиков (союзников и противников) и внутриигровой меты***

Сервис в реальном времени подсказывает игроку каких персонажей выбрать исходя из актуальной стадии выбора персонажа, предпочтений игрока и метовой статистики драфтов и персонажей. Предсказание осуществляется для рейтинговых ролевых игр (в дальнейшем Captains mode).

*Используемые данные и источники данных:*

- Метовая статистика (Stratz, OpenDota)
- Предпочтения игрока (приложение для сбора данных на устройстве пользователя)
- Стадия выбора персонажей (приложение для сбора данных на устройстве пользователя)

Использование приложения для сбора данных обусловлено необходимостью обхождения задержки 2 мин Dota TV.





## ***В процессе разработки:***

### **Предполагаемые модели для решения:**

- BERT4Rec
- SASRec

Данные DL рекомендательные модели используют контекст (предпочтения пользователя и метовую статистику) для определения рекомендации пользователю.



# СПАСИБО ЗА ВНИМАНИЕ!