

# **AAIA'17 Data Mining Challenge**

## **Helping AI to Play Hearthstone**

Евгений Патеха (3 место)

# Конкурс на fedcsis.org

## Helping AI to Play Hearthstone

- Задача – предсказание вероятности победы в игре для оценки качества ходов игрока-компьютера

- Объем данных:

- Train – **2 000 000 + 1 250 000**

- Test – **750 000**

- Метрика – **AUC**

public lb – **5%**, private lb – **95%**

	private	public	dif
1 iwannabetheverybest	0,80185	0,8041	-0,0022
2 hieuvq	0,79922	0,8016	-0,0024
3 johnpateha	0,79895	0,8001	-0,0011
4 vz	0,79733	0,7997	-0,0024
5 jj	0,79707	0,7997	-0,0026
6 karek	0,79685	0,8000	-0,0032
7 podludek	0,79657		
8 akumpan	0,79654	0,7995	-0,0030
9 iran-amin	0,79637		
10 basakesin	0,79617	0,7988	-0,0026

# Данные

- Hearthstone – компьютерная игра с использованием тематических карт
- Базовые данные – текущий ход, количество кристаллов, здоровье игроков, id героев, количество карт
- Данные о картах на столе у игрока и оппонента (только minions) - id карт, атака, здоровье, другие способности
- Данные о картах в руке у игрока - id карт, типы карт (minions, spell, weapon), атака, здоровье и др.

# Валидация

- Данные содержали какой-то лик и градиентный бустинг переобучался. Возможно в данных было по несколько различных стадий одной игры.
- Единственные признаки в данных по которым можно определить такие ситуации – id героев игрока и оппонента (9 разных)
- Решение – формирование фолдов для кросс-валидации с исключением в обучающей и отложенной выборках одинаковых сочетаний id героев
  - Каждая пара id игрока – id оппонента относится целиком в один фолд (81 сочетание разбито на 9 фолдов)
  - В каждом фолде по одному id каждого вида для игрока и оппонента
- Принятие решения об улучшении модели только если улучшился результат на большинстве фолдов

# Основные модели

**lightGBM – 140 признаков. Public lb 0.7987**

- Базовые и новые признаки плюс признаки – количество карт определенных id на столе или на руках (больше всего добавили признаки spell-карт).

**lightGBM – 58 признаков. Public lb 0.7983**

- Укороченный вариант 1 модели – признаки отбирались добавлением по одному. Признаки id карт не добавлялись. Плюс новые признаки - число карт с определенным cost.

**mxnet – 140 признаков. Public lb 0.7980**

- Нейронная сеть с 2 скрытыми уровнями. Признаки аналогичны 1й модели. На локальной валидации результаты были близки к 1-2 моделям, на lb чуть хуже

**Финальные результат - среднее с весами 30-30-40. Public lb 0.8001**

# Мелкие улучшения

## Отдельная модель для 1-2 ходов

- На 1-2 ходу, за исключением редких случаев победа еще не predetermined. Отдельная модель исключает использование закономерностей, определенных на последующих ходах. Результатом заменял результаты 1й модели для 1-2 ходов.

## Сближение трейна и теста

- Корректировка теста - `op.deck_count - 1`
- Корректировка трейна
  - `if (turn==1) op.hand_count:=op.hand_count-2`
  - `if (turn==2) op.hand_count:=op.hand_count-1`
  - `if (turn<11 & pl.crystals_all<turn) pl.crystals_all:=turn`
- Обучение только на первых 16 ходах

# Новые признаки

- Разница между суммарной атакой игрока и суммарным здоровьем оппонента (сумма по картам в игре + показатели героя) и наоборот
- Отношение разницы в здоровье игрока и оппонента к здоровью игрока
- Сумма здоровья minions в руке
- Прирост суммарной силы атаки карт в игре относительно базового уровня
- Число миньонов в игре, готовых к атаке
- Число карт с определенным id (отдельно в игре и в руке)
- Число карт с определенным cost (отдельно для карт в игре, minions и spell)

## Что не получилось

- Использование внешних данных по свойствам игровых карт
- Разворот карт игрока-оппонента и увеличение вдвое обучающей выборки
- Кодирование сочетаний признаков вероятностью победы
- logreg, KNN, xgboost
- Не удалось построить хорошую модель 2 уровня – прирост относительно усреднения был небольшой и на lb не проявлялся



# Вычислительные ресурсы. ПО

- Ноутбук - 2 ядра, 12 GB
- Google Cloud – 8-12 ядер 40-60 GB, GPU
- R пакеты `data.table`, `Matrix`, `jsonlite`, `lightGBM`, `mxnet`, `xgboost`, `glmnet`, `largeVis`

Спасибо за внимание!