

# ML Boot Camp III

Карачун Михаил

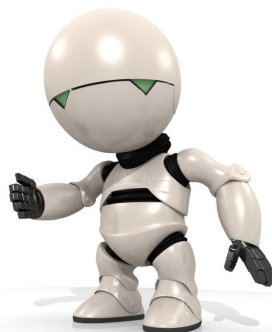
- победитель двух соревнований ML Boot Camp
- [karachun.michael@gmail.com](mailto:karachun.michael@gmail.com)





# Постановка задачи

- Онлайн игра: прогноз оттока пользователей
- Бинарная классификация
- Logloss
- 25000 объектов
- 12 признаков



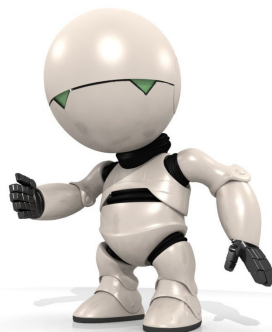
# Данные

- Статистика игроков за 14 дней
- Несбалансированные классы 0.288
- Чистые данные
- Почти все признаки непрерывные

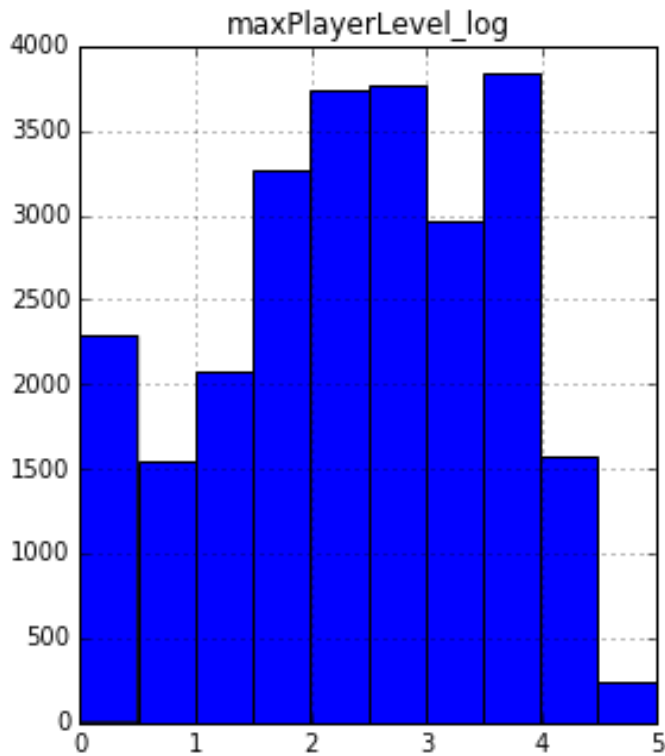
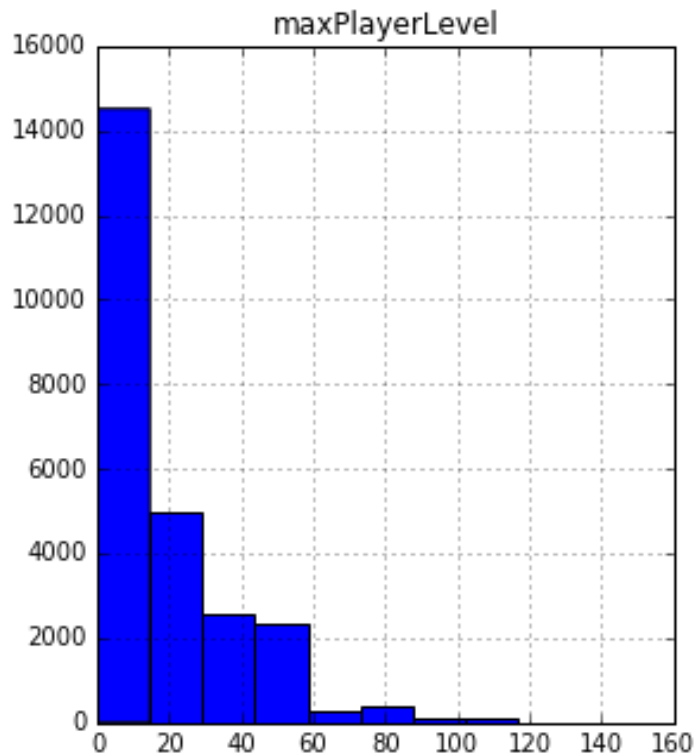


# Пример признаков

- Максимальный уровень игрока
- Количество набранных очков
- Количество попыток
- Количество дней игры



# Первичная обработка



# Первая модель

- Xgboost
- Hyperopt
- StratifiedKFold (n\_folds=3)

**Public:0.383**



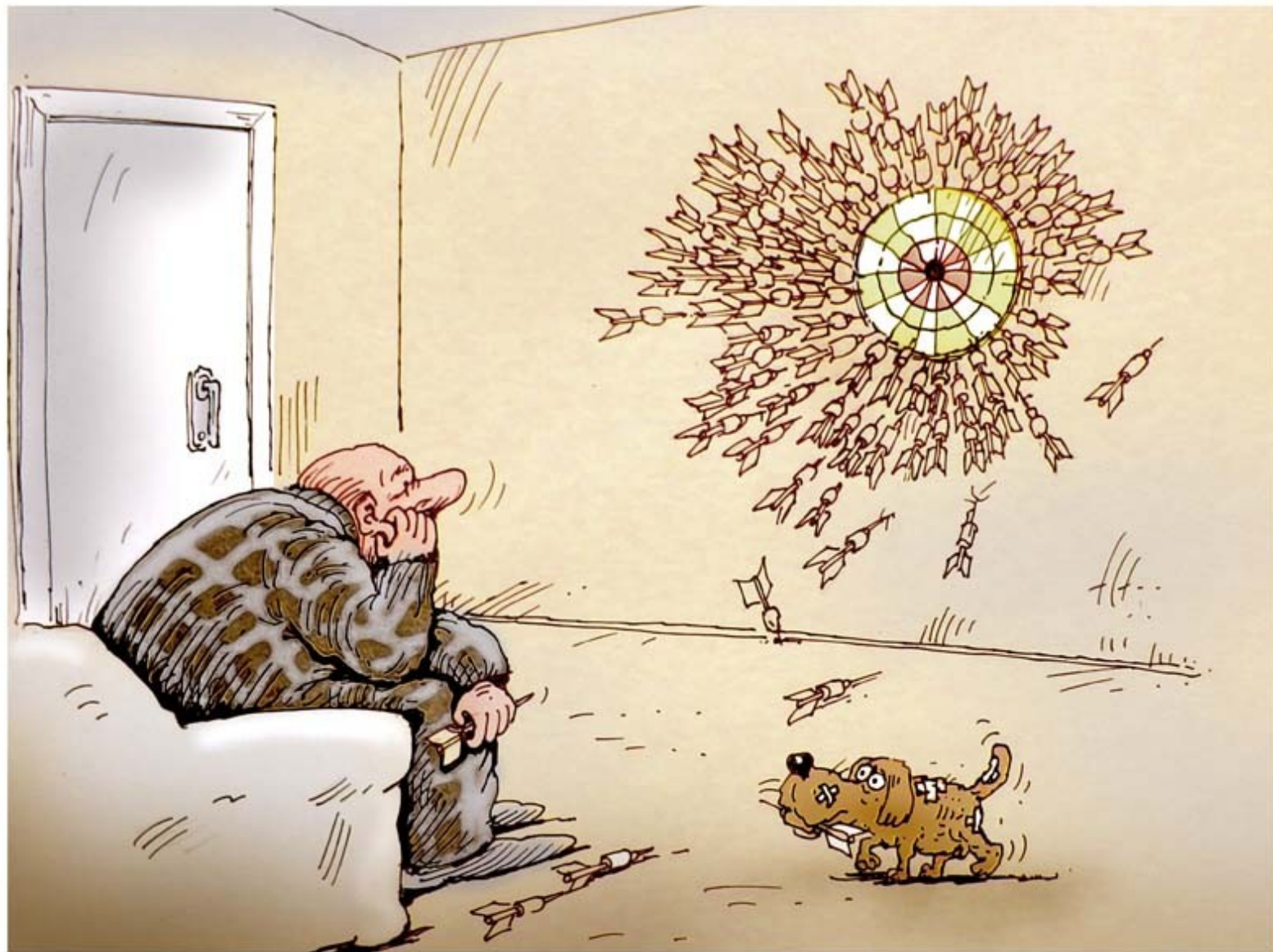
# Генерация признаков

- Признаки + log
- Сочетания
- RFECV

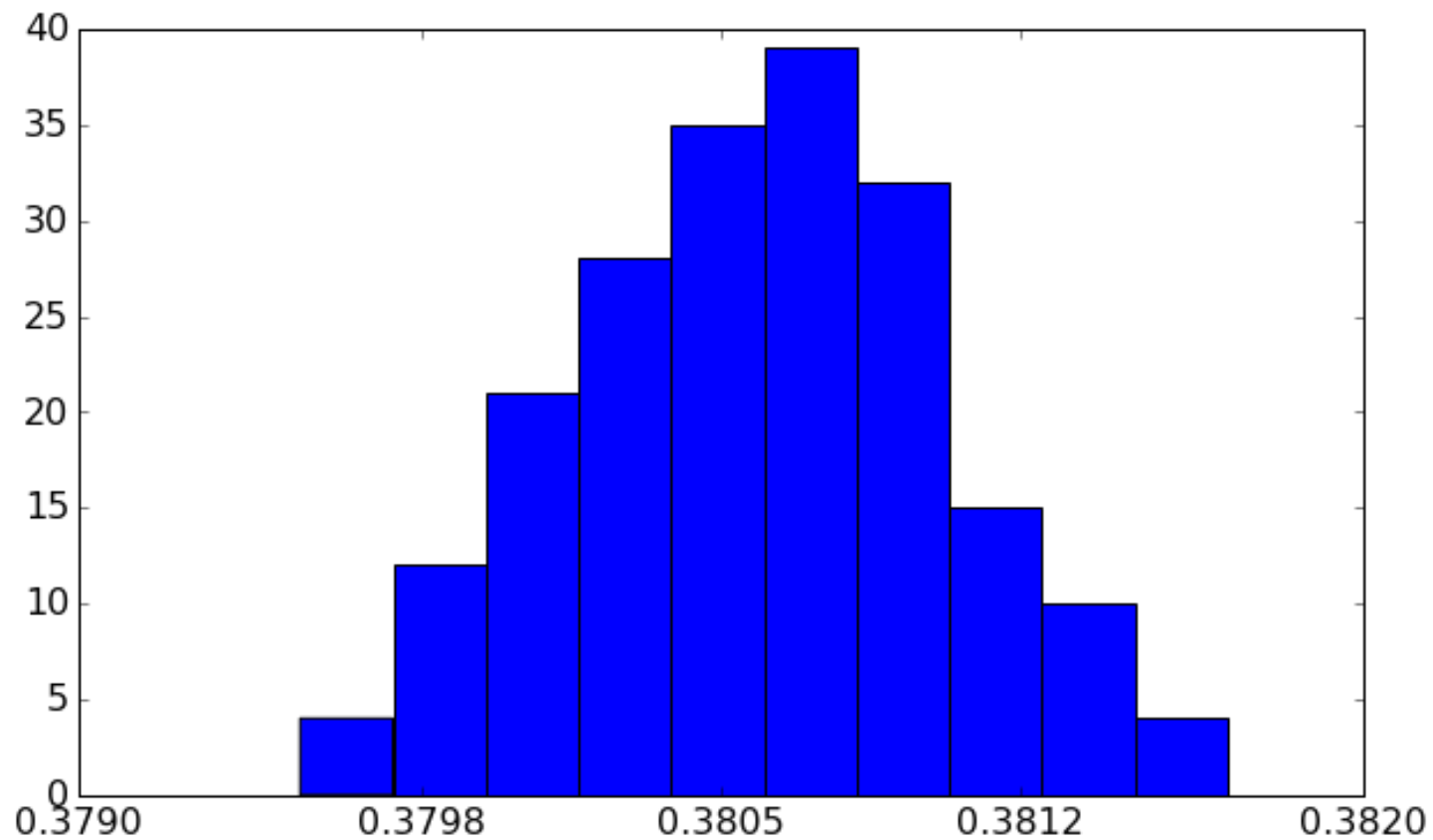
**Public:0.382**







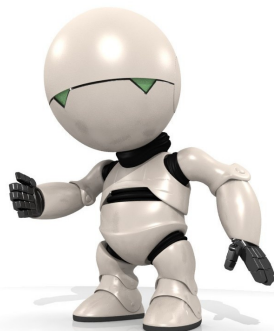
# Первые проблемы



# Больше моделей

- «Плохие» тоже полезны
- Тюнинг в сочетании с остальными
- Среднее всех моделей

**Public:0.381**

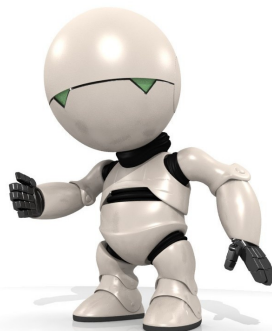
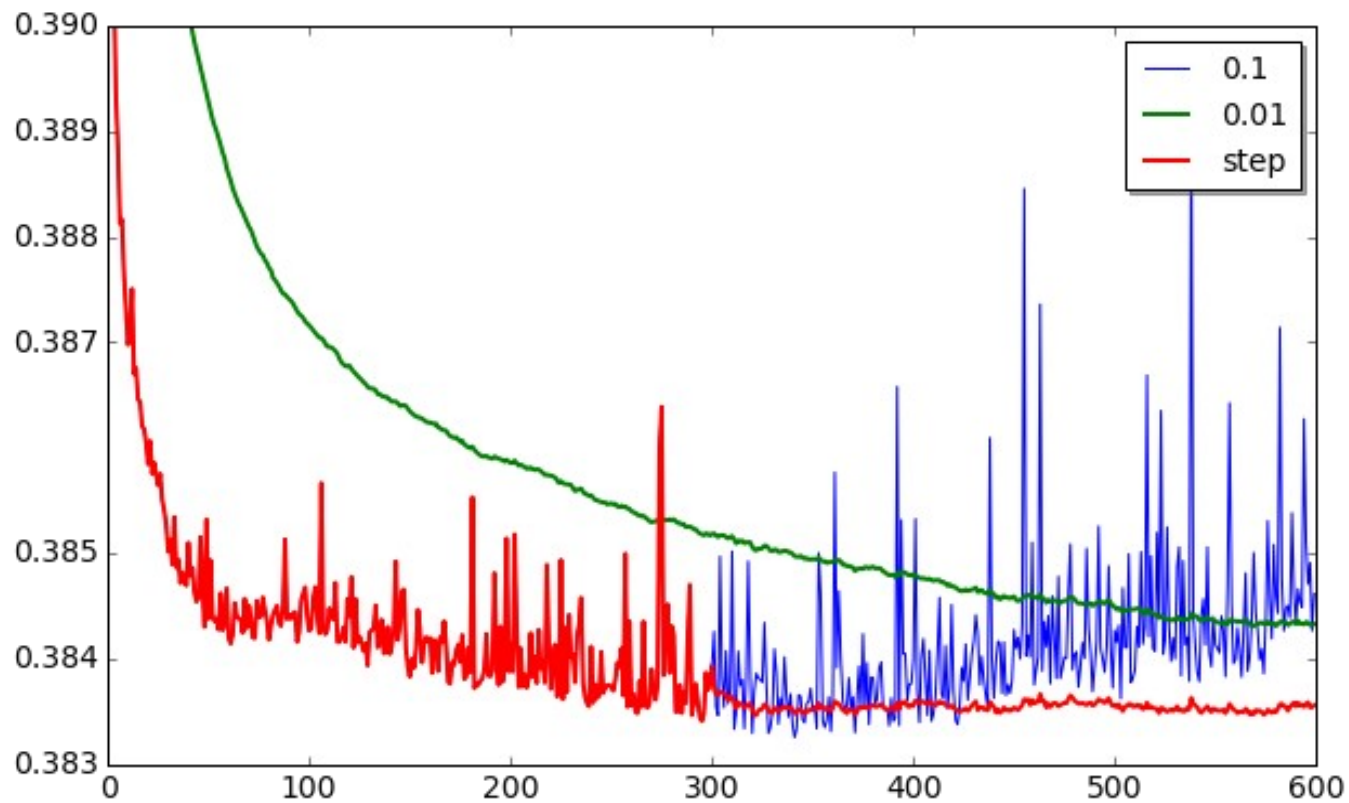


Модель	Библиотека	Признаки	CV	Public
Xgboost1	Xgboost	70	0.378	0.382
Xgboost2	Xgboost	24	0.376	0.382
GB1	sklearn	70	0.3802	0.382
GB2	sklearn	24	0.376	0.382
LightGBM	LightGBM	70	0.376	0.382
NeuralNet	Keras	24	0.3812	0.383
LogisticRegression	sklearn	161	0.374	0.383
Среднее				0.3811

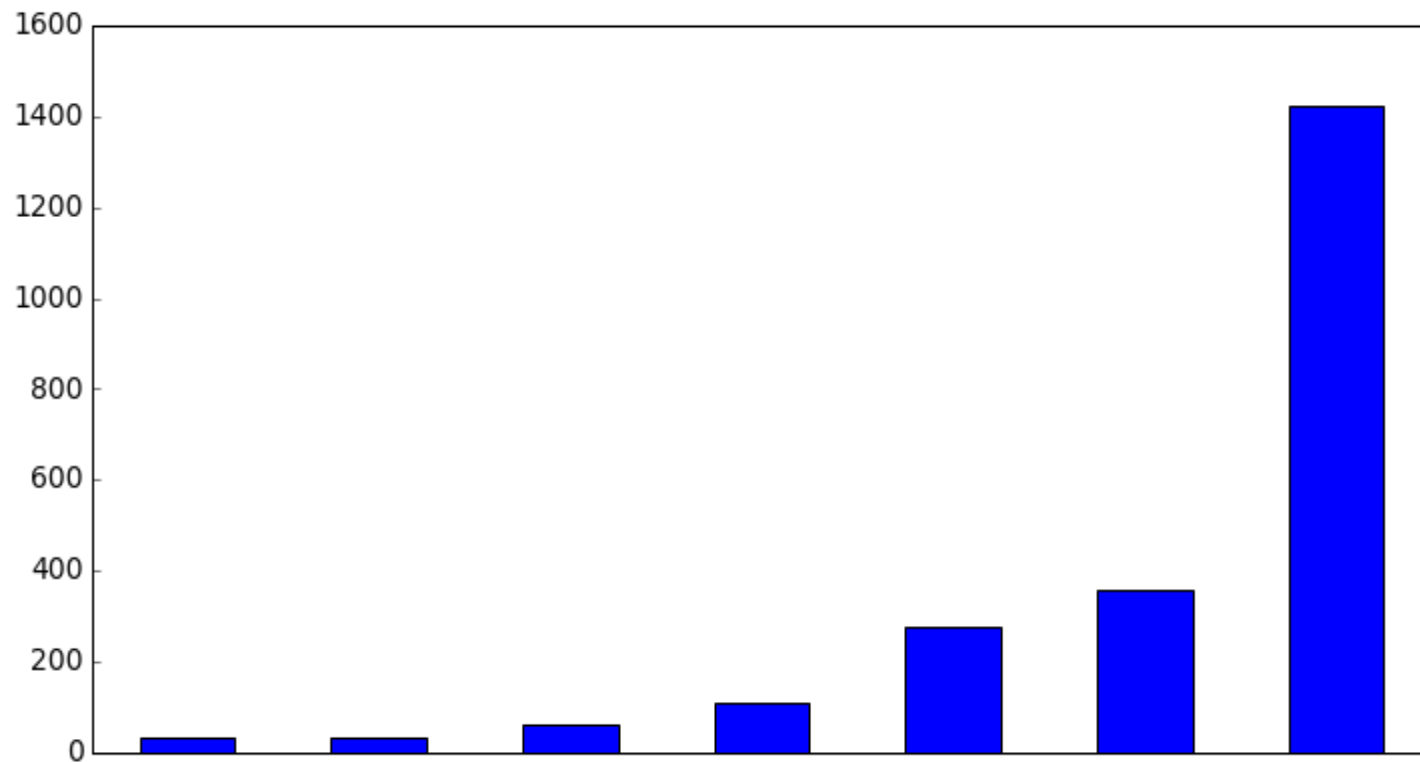
# Регрессия

- Признаки с одинаковым распределением
- Различная нормализация
- Разбиение на бины

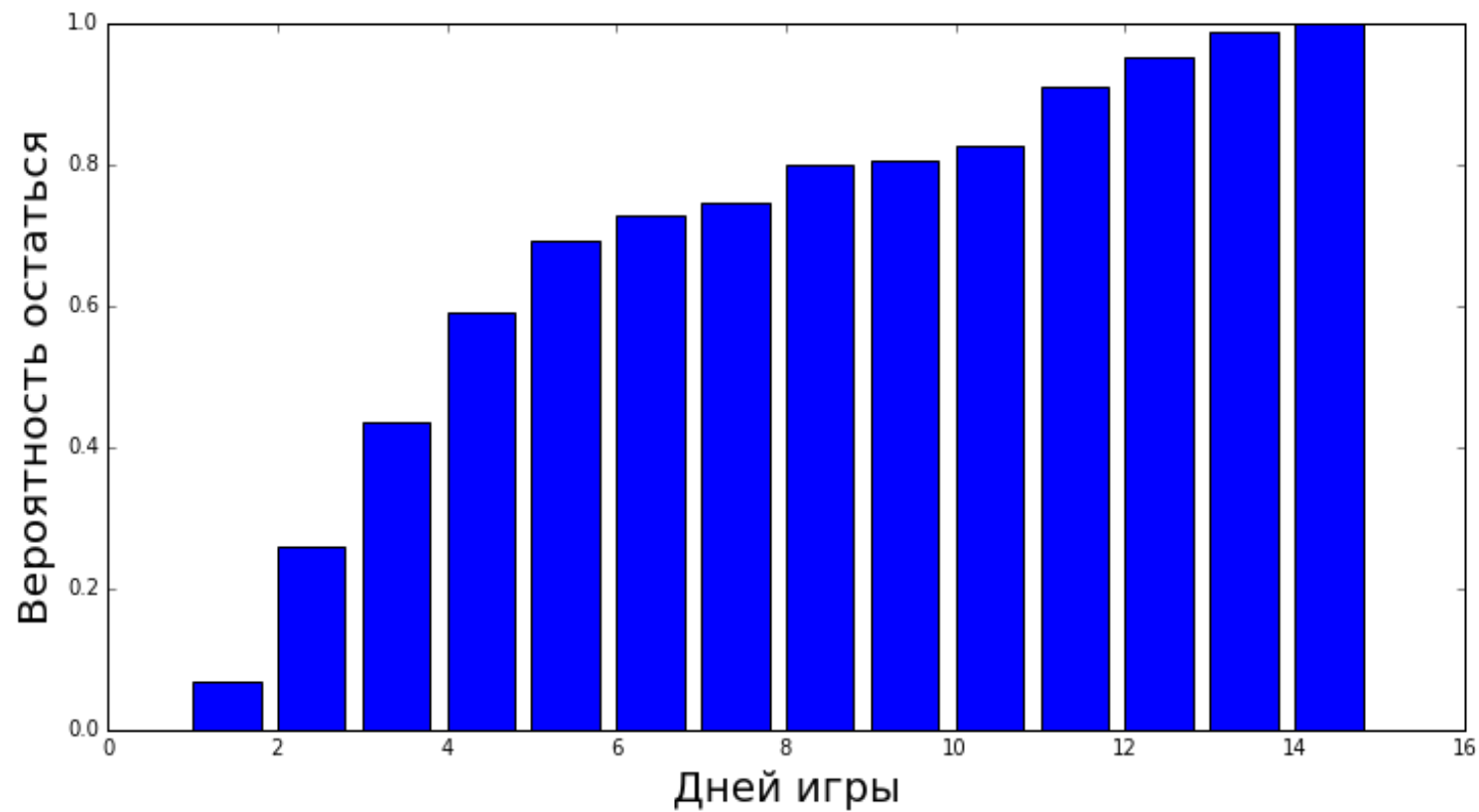
# Нейронная сеть



# Одинаковые объекты



# Дни игры



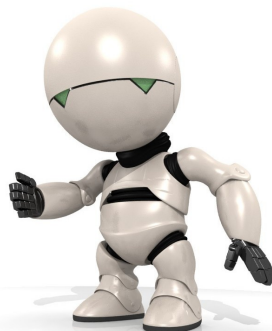
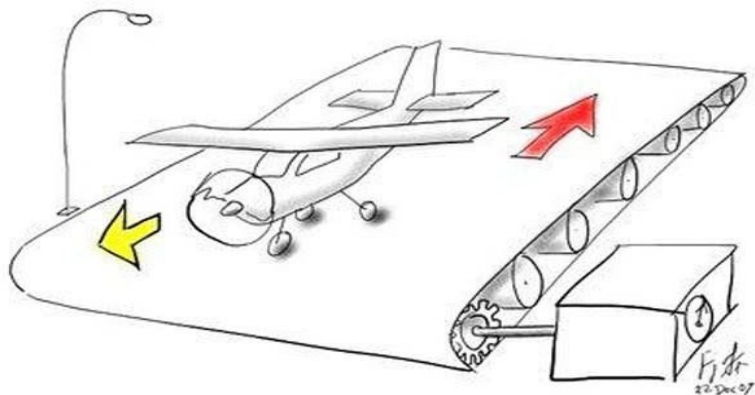




**No.**

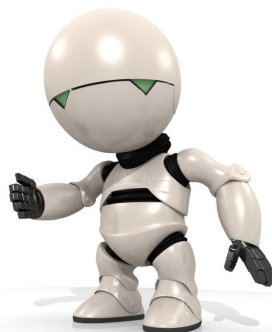
# То что не взлетело

- Стекинг
- RandomForest
- Разбиение выборки



# Итоговый результат

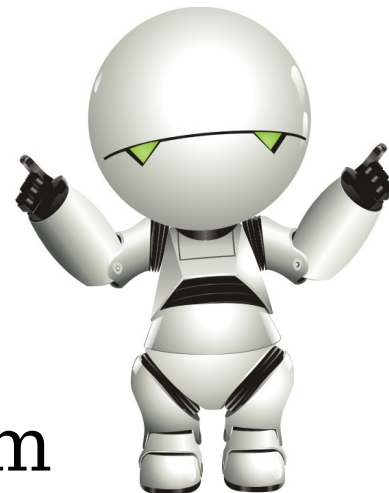
Участник	Public score	Private score
Карачун Михаил	0,3808( <b>5</b> )	0,3814( <b>1</b> )
Евгений Цацорин	0,3818( <b>37</b> )	0,3816( <b>2</b> )
Иван Тямгин	0,3807( <b>3</b> )	0,3816( <b>3</b> )
Александр Сергей	0,3800( <b>2</b> )	0,3825( <b>32</b> )
Андрей Селиванов	0,3796( <b>1</b> )	0,3844( <b>261</b> )





# Материалы

- <https://habrahabr.ru/post/324916/>
- <https://habrahabr.ru/company/mailru/blog/321016/>
- <https://github.com/KarachunMikhail/>



karachun.michael@gmail.com