



# Решение Allstate

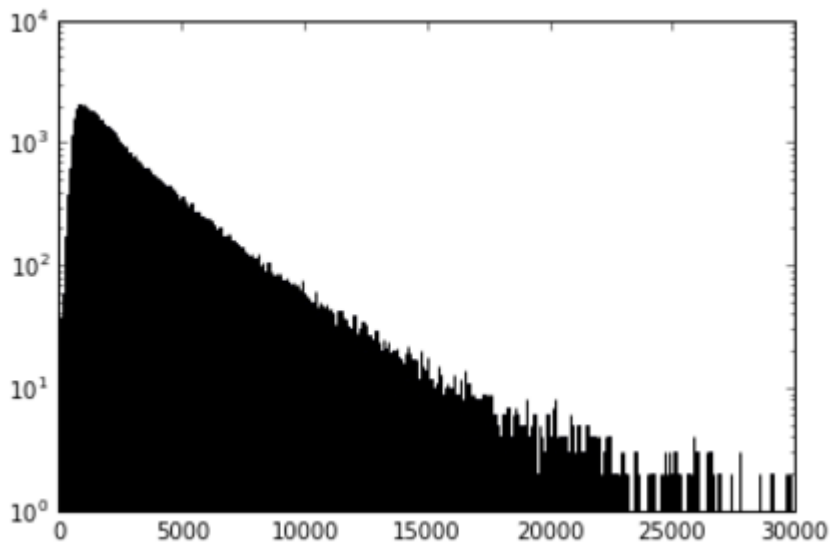
Парпулов Д.  
DMIA, 2016

# Данные

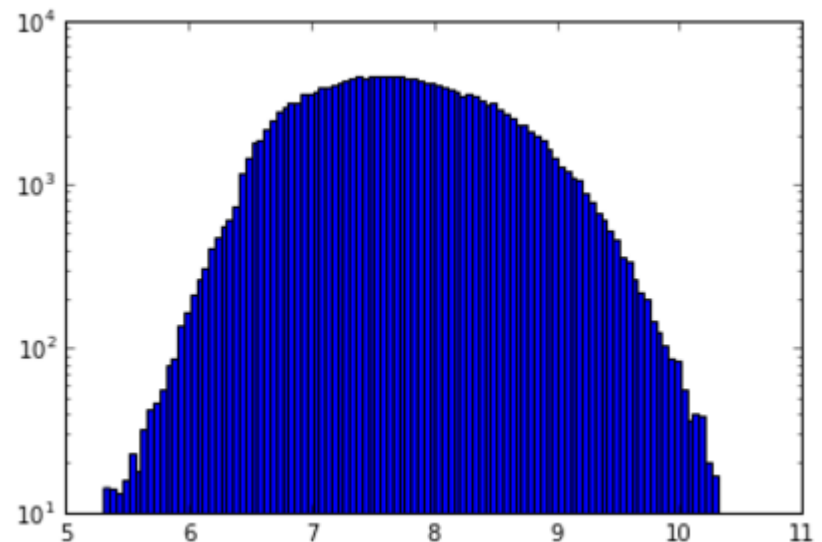
- Строки – каждый иск о страховом возмещении
- 116 категориальных переменных
- 14 вещественных переменных
- Нужно предсказать сумму страхового возмещения

## Target

loss



$\log(\text{loss}+200)$



# Baseline 1

За основу берем скрипт

<https://www.kaggle.com/mariusbo/allstate-claims-severity/xgb-lb-1106-33084>

Внутри:

- Формирование квадратичных признаков из 35 категориальных фич + LabelEncoder в 27 label
- Предобработка вещественных признаков
- Определение кастомного objective для xgboost

В результате получаем dataframe с трансформированными признаками и таргетом.

# Baseline 1

- Фиксируем разбиение train на 5 фолдов
- Обучаем xgboost OOF, получаем предсказания для каждого фолда train, *делаем также предсказание и для test*
- Для каждого фолда повторяем такую процедуру 10 раз (по различным seed) и усредняем итоговое предсказание (для train и test)
- В конце обучаемся на всех данных и предсказываем для теста
- Так же повторяем это 10 раз с разными seed и усредняем итоговое предсказание.

Результат:

1106.54618 – Public LB

1120.08163 – Private LB

# Baseline 2

За основу берем скрипт <https://www.kaggle.com/mtinti/allstate-claims-severity/keras-starter-with-bagging-1111-84364> и повторяем ровно то же самое, что и с xgboost

- Обучаем NN OOF, получаем предсказания OOF для train, *делаем предсказание и для test*
- Для каждого фолда повторяем такую процедуру 10 раз (по различным seed) и усредняем итоговое предсказание (для train и test)
- В конце обучаемся на всех данных и предсказываем для теста
- Так же повторяем это 10 раз с разными seed и усредняем итоговое предсказание.

Результат:

1112.91244 – Public LB

1124.12466 – Private LB

# Let's time to blend!

- Берем OOF предсказания для train от xgboost и NN
- Откладываем один из фолдов в качестве holdout
- Подбираем коэффициент для смешивания
- Валидируемся на холдауте

Результат( $w_{xgb}=0.57$ ):

1105.41214 – Public LB

1118.31162 – Private LB



# Больше xgboost'ов!

- Лень было что-то придумывать, времени было мало, обучим еще xgboost'ов и NN
- Усредним не только итоговые предсказания, но и OOF предсказания.
- Итого: смешано 3 xgboost'а (успело обучиться только 3) и 13 NN.
- Умножение результата на 1.01 прибавило score на ЛБ

Результат:

1101.26014 – public LB

1113.24990 – private LB

I'm lazy and I know it.



**Allstate Claims Severity**

4 days ago · Top 3%

26 entries as a solo competitor



**85<sup>th</sup>**  
of 3055



Спасибо за внимание!