

Allstate Claims Severity Solution

Dmitry Dryomov, 2016

Solution results

- 125/3155 (top-5%)
- 1st place score: 1109.70772
- 16th place score: 1111.38680 (Gold)
- 125 place score: 1113.64700

General ideas

- No feature engineering
- Target transformation
- Ensembling
- OOF Stacking / Averaging
- High values boosting

Features

- Pairs features interaction (selected)
- Elimination of values that are not present in train/
test

Target transformation

- $y^* = \log(\text{offset} + y)$
- $y^* = (\text{offset} + y)^c$
- Non-transformed

Ensembling

- LightGBM
- XGBoost
- NN
- *ExtraTrees, kNN tested, no improvement

Ensembling

- Building many models with the same tool
- Different folds separation
- Only basic features / with pairs
- Parameters optimisation with the fixed iterations number (without early stopping)

Stacking / Averaging

- OOF predictions of base models
- Find the best linear combination
- XGBoost on the difference ($y - p_{\text{linear_combination}}$)
- Final XGBoost + $p_{\text{linear_combination}}$

High values boosting

- All $p > \text{threshold}$: $p = \text{threshold} + (p - \text{threshold}) * a$
- $\text{threshold} \sim 2750$, $a \sim 1.034$

1st place

- Significantly more base models tried
- Significantly more parameters tuning
- Stacking with both XGBoost and NN combined
- XGBoost custom objective (and all standard tried)
- Models trained on the top/low percentiles of targets