CLAIM SEVERITY

FOR ALLSTATE

TING SIT

JULY 2020

OBJECTIVE

- Predict claim severity in absolute dollar amount (known as "loss" in this project)

- Benefits:
  - Reduce human errors on calculations, which improve company's profit line
  - Reduce labor hours with calculations being automated, increase productity
  - Improve consumer experience with faster/more accurate claim reimbursement

- Model Evaluation requirements:
  - Mean Absolute Error (MAE)

MODEL CONSIDERATION

- Regression problem: target variables = “loss”
- 4 models: SGD Regression, Lightbgm, Xgboost, Catboost
- Features:
  - 10 continuous variables
  - 1033 starting categorical variables, narrowed to 102 through features reduction

RESULTS

Model	MAE	Time
Dummy	1783	1s
SGD Regression	1266	4s
XGBoost	1149	44s
LightGBM	1129	14s
CatBoost	1122	43s

- Catboost has the lowest MAE
  - Recommend as the winning model

CONCLUSION

1. CatBoost has the smallest MAE and is the winning model
2. Model tend to underpredict for high value claims, and will need special audit.
3. Overprediction on small values claims happen in a fair frequency. Need further investigation on the drivers

NEXT STEPS

- Further improvement would require knowing the definition of the features for better feature engineering
- Align expectations with clients on the MAE that is "good enough"
- MAE tend to underestimate the impacts from over/under predictions in claims, and could results in financial burden as well as customer's dissatisfactions
  - Need to introduce other KPIs to ensure control to the model outputs.