

Project Introduction

Project : French Motor Claims

Author : Bhishan Poudel, Ph.D Physics

Goal : Implement Frequency modelling, Severity modelling and Pure Premium Modelling

Tools : pandas, scikit-learn, xgboost, pygam

References:

- <https://www.kaggle.com/floser/french-motor-claims-datasets-fremtpl2freq>
- https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.TweedieRegressor.html

Project Notebooks

Notebook	Rendered	Description	Author
a01_data_cleaning.ipynb	ipynb , rendered	ohc, kbin, logscaling	Bhishan Poudel
b01_freq_modelling.ipynb	ipynb , rendered	Poisson	Bhishan Poudel
b02_severity_modelling.ipynb	ipynb , rendered	Gamma	Bhishan Poudel
b03_pure_premium_modelling.ipynb	ipynb , rendered	Poisson*Gamma and Tweedie	Bhishan Poudel
b04_tweedie_vs_freqSev.ipynb	ipynb , rendered	comparison	Bhishan Poudel
b05_lorentz_curves_comparison.ipynb	ipynb , rendered	Lorentz Curve	Bhishan Poudel
c01_xgboost_tweedie.ipynb	ipynb , rendered	'objective':'reg:tweedie'	Bhishan Poudel
d01_gam_linear.ipynb	ipynb , rendered	n_splines=10, grid_search	Bhishan Poudel

Data

- [openml french motor freq](#) (Multiple features.)
- [openml french motor severity](#) (Two features Id policy and Claim Amount.)

Data Cleaning

Some of the features are chosen for modelling.

```
one hot encoding = ["VehBrand", "VehPower", "VehGas", "Region", "Area"]
kbins discretizer = ["VehAge", "DrivAge"]
log and scaling = ["Density"]
pass through = ["BonusMalus"]
```

Results

Module	Distribution	y_train	sample_weight	train
sklearn	Frequency Modelling (Poisson Distribution)	df_train['Frequency']	df_train['Exposure']	0.051
sklearn	Severity Modelling (Gamma Distribution)	df_train.loc[mask_train, 'AvgClaimAmount']	df_train.loc[mask_train, 'ClaimNb']	-
sklearn	Pure Premium Modelling (TweedieRegressor)	df_train['PurePremium']	df_train['Exposure']	2.01802
xgboost	Xgboost Tweedie Regression	dtrain.set_base_margin(np.log(df_train['Exposure']).to_numpy())	dtest.set_base_margin(np.log(df_test['Exposure']).to_numpy())	-
pygam	GAM Linear Model	df_train["AvgClaimAmount"].values	N/A	-