Springboard Capstone: BNP Paribas Cardif

Overview by Monisha Gopal

Outline

- Background
- Data
- Pre-processing
- Machine Learning

Background

BNP Paribas Cardif is an global insurance company that specializes in personal insurance.

Their clients want their claims to be processed faster. So BNP Paribas Cardif hopes to use data science to determine off the bat which claims can go through an accelerated process and which can’t.

Data

Three files provided:

- train.csv - training set with target (dependent variable)
- test.csv - test set without target
- samplesubmission.csv - sample submission with correct format

Initial Look:

- 33 features
  - 'ID', 'target', and 'v1' through 'v131'
  - Categorical & numerical features.
- Main limitation is anonymized data
  - Don’t know what features stand for

Pre-processing

- After cleaning data went from 131 features to 14
- Main feature engineering methods:
  - Weight of Evidence
  - Information Value
- Finally, had 10 variables

Machine Learning

- Model:
  - Logistic Regression
- Tried 3 Models
  - Model 1 - Top 5 most important variables
  - Model 2 - Top 8 most important variables
  - Model 3 - Top 10 most important variables
- Best Model was model 2
  - It performed the best on Kaggle’s public dataset

Kaggle Outcomes

Submission and Description	Private Score	Public Score
model3.csv 5 minutes ago by Monisha Gopal Top 10 variables	0.50107	0.50170
model2.csv 6 minutes ago by Monisha Gopal Top 8 variables	0.50112	0.50169
model1.csv 6 minutes ago by Monisha Gopal Top 5 variables	0.50927	0.50998

Thanks