PREDICTING MOTOR INSURANCE CANCELLATIONS

Adam Nelson; MSc in Big Data

UNIVERSITY of STIRLING

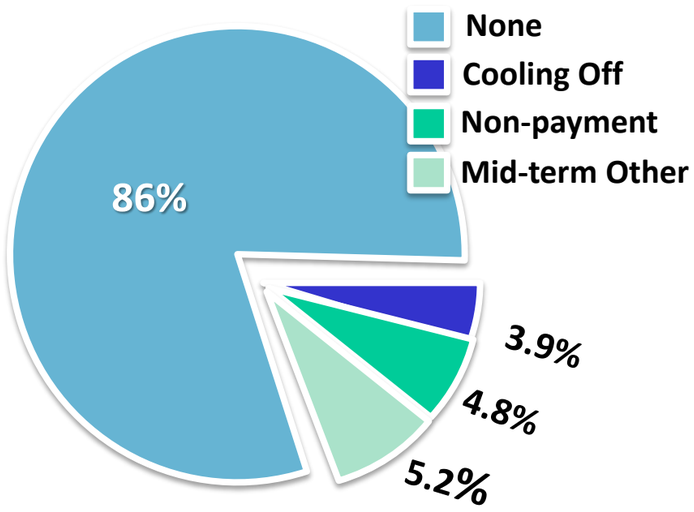
Introduction

Commission from insurances represent a major component of Sainsbury's Bank's income. The objective of this project is to construct a machine learning classifier that categorises customers based on their likelihood of cancelling a motor insurance policy at any point during coverage period.

This model will ultimately be utilised as part of an extended project by Sainsbury's Bank data team to build an insurance pricing optimiser.

Data Set

The raw dataset comprises 127316 rows and 167 columns consisting of customer data. The target variable is cancellation reason.



Overarching goal is that there will be a single output from the model – the probability that a customer will cancel – which will be further subdivided into cancellation types.

Process

PRE-PROCESSING

- Null values filled
- Binning/recoding of features
- Continuous variables curbed or scaled.
- Resampling techniques

MODEL BUILDING

- Random Forest
- Logistic Regression
- Support Vector Machines
- Multilayer Perceptrons

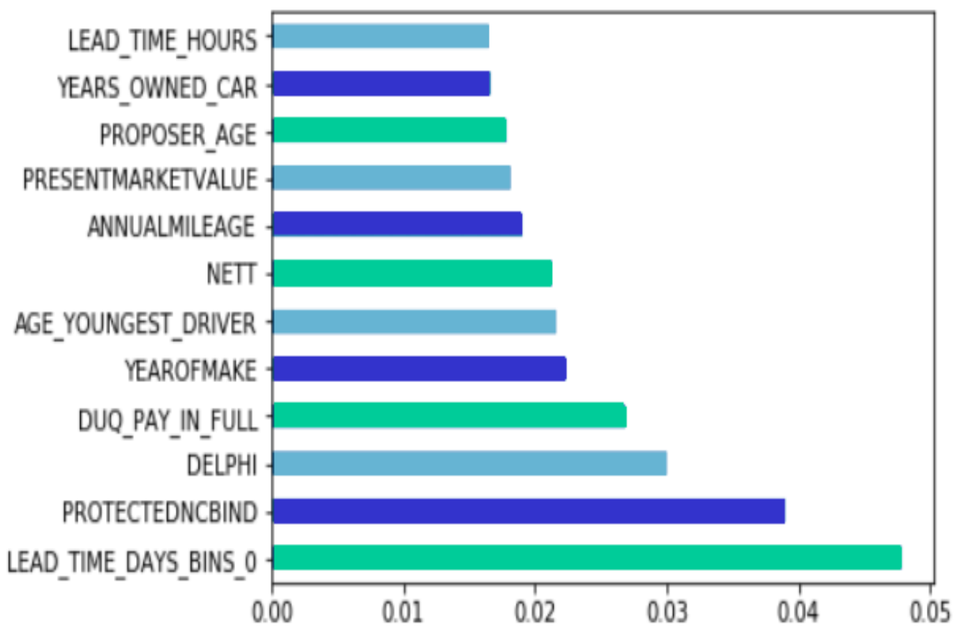
MODEL EVALUATION

- How does the model perform on data not previously exposed to?
- Metrics: AUC, recall, precision, confusion matrix

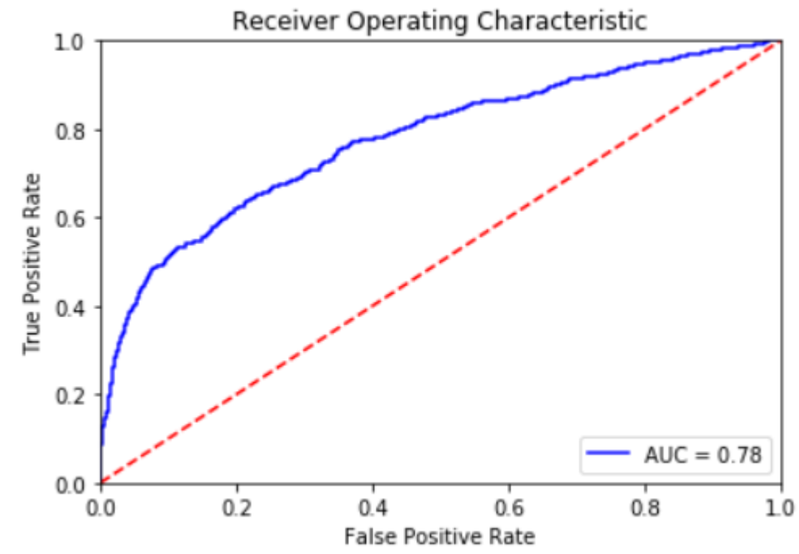
Sainsbury's Bank

jupyterhub

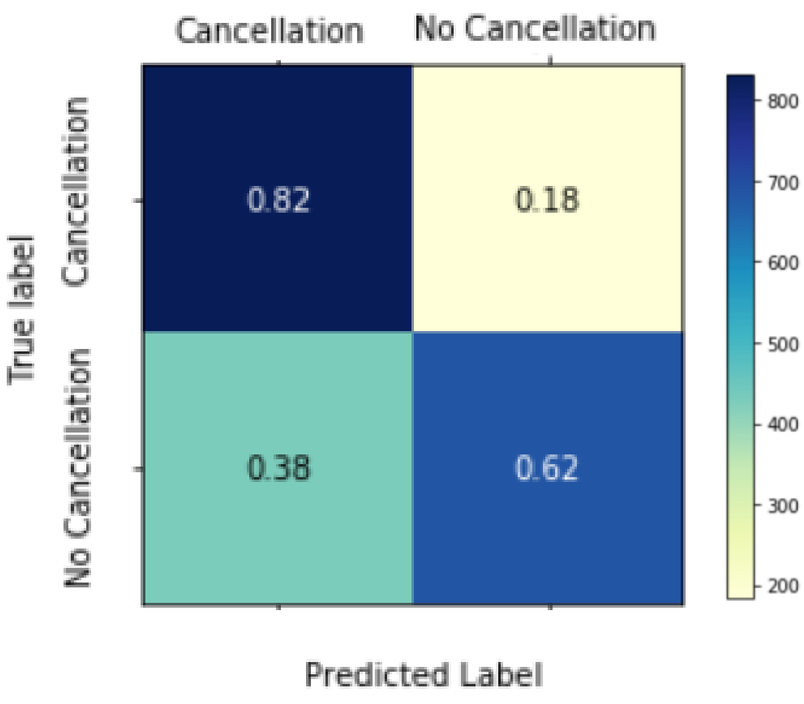
Relevant Features



Preliminary Results



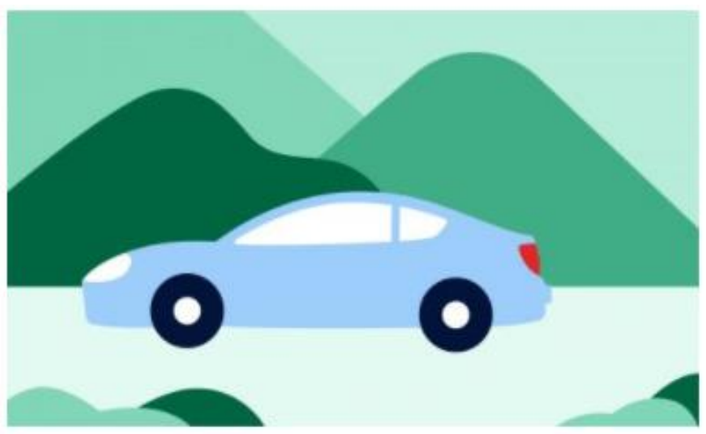
***AUC** 0.78*



Recall score: 0.6152
Precision score: 0.7894

Future Steps

- Explore other algorithms.
- Further hyperparameter tuning
- Investigate ensemble methods for resampling data.
- Extend the model to a multi-classifier, which takes into account the separate cancellation types.



stir.ac.uk

For further information contact: *Adam Nelson*

adn00015@students.stir.ac.uk

BE THE DIFFERENCE