Porto seguro safe driver prediction

# Executive Summary

The average claim will set a car insurance company back £3000. Car accidents can be notoriously difficult to predict due to the inherent randomness of many accidents. Despite this we have been able to locate trends that can provide additional insight into individuals who may be involved in an accident. Based on the data you provided, we have harnessed several sophisticated and ground breaking modelling techniques including neural networks and gradient boosted models that have allowed us to identify high risk customers. Our unique approach involves creating an ensemble of several different model types in order to explain as much of the data as possible. We found that regional, individual and car information were useful in predicting claims however the calculated features within the data did not provide any significant prediction capabilities. Our model could potentially save Porto Seguro in excess of $200,000 a year going forward. With this new information Porto Seguro can alter prices according to the probability of claims and also refuse to take on customers that are likely to make them a loss.

# Purpose of our model

The purpose of our model is to predict which potential insurance policy holders have a higher chance of making an insurance claim in the events of an accident. Different models were considered and analysed to help us decide which model fit your purpose best. Our model will be able to help your organisation to categorise which policy holders are high risk and low risk claimants depending on the factors presented and also historical trends. By examining different variables you will be able to see which factors affect the likelihood of your policy holders in making a claim, being able to categorise customers in risk groups this will help your organisation’s ability to make alterations in policy renewal prices objectively, based on the data our model analyses.

In terms of the accuracy our model will be able to show you trends in which customers who made claims in the past. The analysis of past trends can be used to help your policy analytics team to be able to decide an optimal policy price in consideration to risk factors to maximise profit but also provide a competitive insurance prices which will attract new customers.

# Exploratory Data Analysis

To start with, we were carefully cleaning the dataset to avoid any further problems. Among all the records, we removed the columns that contain more than half of the missing values. Also, the columns with some calculations are not statistically significant in modelling. Therefore, they were also excluded. Then after this stage, there are 30 variables remained and used in this report.

# Why should they split with large consulting firms and potentially license our model

We believe that your current contractual agreement with is not delivering you a very cost efficient solution in terms of providing advice with regards to setting policy prices based on their analysis. By cutting out an intermediary data analytics team which your current consultancy firm does, you will be able to save on external outsourcing costs in which are usually passed down to you. With the implementation and direct support, we will be able to educate your staff so that they will be able to independently run our model on your current data processing systems and workstations.

With regards to protecting our model, we are also looking for an agreement with your organisation to fund our business plan to license our model, in return you will gain editing rights to our source code which will give you the right to alter the code if you wish to, but also distribution rights. By agreeing to license our model we will instantaneously offer a discounted price for our services and will contractually agree to remain committed to working your company only, and refuse to work with any others insurance policy companies in Brazil.

# Monetary benefit of our model

As well as having access to our model and right to distribute, one major benefit our implementing our statistical model is the monetary gain you will be able to make through savings.

# Conclusion