

PORTO SEGURO MODEL PROPOSAL

Authors: 150004311, 150003431, 180025784, 180028634, 180024570

WORD COUNT: 1026

Contents

Porto Seguro's Safe Driver Prediction:	1
Executive Summary:	1
Problem statement:	1
Exploratory data analysis:	1
Proposed Model and Methodology:	2
Impact of Model Implementation:	3
References	4

Porto Seguro's Safe Driver Prediction:

Executive Summary:

The car insurance industry supports thousands of policy holders. The industry relies on statistical analysis to understand the key factors that influence the likelihood that a policy holder will make a claim as this will influence the premium rates charged and the overall profitability of the business being written. On average insurance companies incur a cost of £3000 per claim (IMF, 2012). Despite car accidents being problematic to predict, the model that we have developed projects a reduction in sums paid through claims by Porto Seguro more than £250,000 annually. This model was developed using data provided and the application of various machine learning techniques including logistic regression, neural networks and gradient boosting. Through the application of these techniques, we have developed several models to identify high risk policy holders. Analysis undertaken has identified that from these models the optimal model that Porto Seguro should utilise is the XGBoost model. This model has identified that specific car, region and individual informative features optimise the identification of risk of claims. In particular, ps_car_13, ps_reg_03 and ps_ind_05 are useful predictors which should be critically analysed before offering an insurance contract to a potential policy holder. Our generated model should be used to support the determination insurance rates, or identification of potential customers that would likely make frequent claims impacting the profitability of the business.

Problem statement:

The objective was to develop a model that would aid in identifying the risk of claims based on a series of specific factors that would be available as part of an insurance application. The application of the model would enable Porto Seguro to offer more competitive insurance rates to those at little risk of making a claim increasing the potential client base, as well as, increasing the rates for potential high-risk individuals leading to a reduction in loss due to claims per annum.

Exploratory data analysis:

An examination of the data highlights that there are an approximate 595,000 data sets with 59 variables. The variable "*target*" is of importance as it identifies whether a policy holder has made a claim and is the factor that our model aims to predict. Identifying which policy holders are likely to make a future claim is complex. From the data provided 3.6% of policy holders are recorded as having made a claim. Of the provided data there are 846,458 individual data items missing. These are particularly prevalent within variables ps_car_03_cat, ps_car_05_cat, ps_reg_03 which have 70%, 45% and 18% of their data missing, respectively. The percentage of missing values across the different variables is summarised in Figure 1 below.

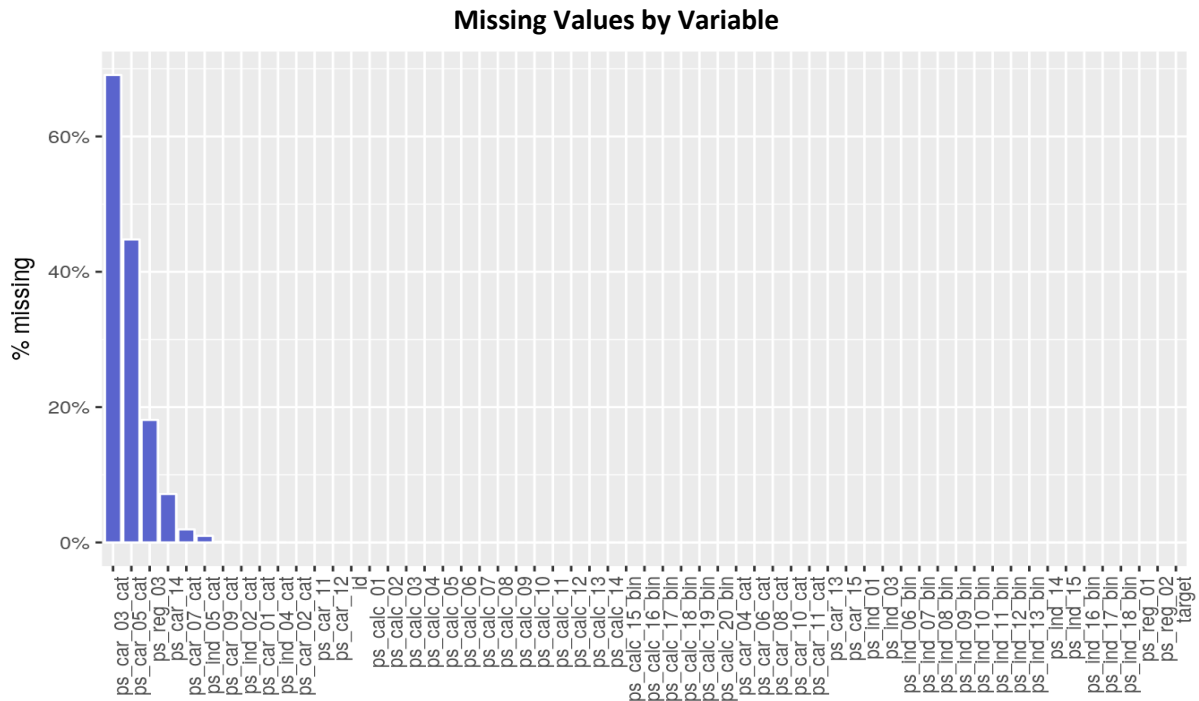


Figure 1: Representation of missing data by variable

Proposed Model and Methodology:

Based on the data provided, the key objective was to identify which of the variables supported the greatest predictive value.

Methodologies for handling missing values include imputing mean values or mode values for the missing data, regarding missing values as a specific level of that category or eliminating the record. In the development of the proposed model we opted to regard the missing values as a separate level of each categorical feature.

The proposed model was fitted via gradient boosting, an efficient machine learning algorithm for regression and classification problems. Through one of the top gradient boosting frameworks, XGBoost, it was determined that some specific variables were more significant in terms of the model structure and value in prediction than others. In particular, one of the car-related features, *ps_car_13*, is deemed the most important feature with a weighting of 0.195. As highlighted in figure 2, three features related to the regions and the individuals, *ps_reg_03*, *ps_ind_05_cat_0* and *ps_ind_03*, are also relatively significant compared to other features.

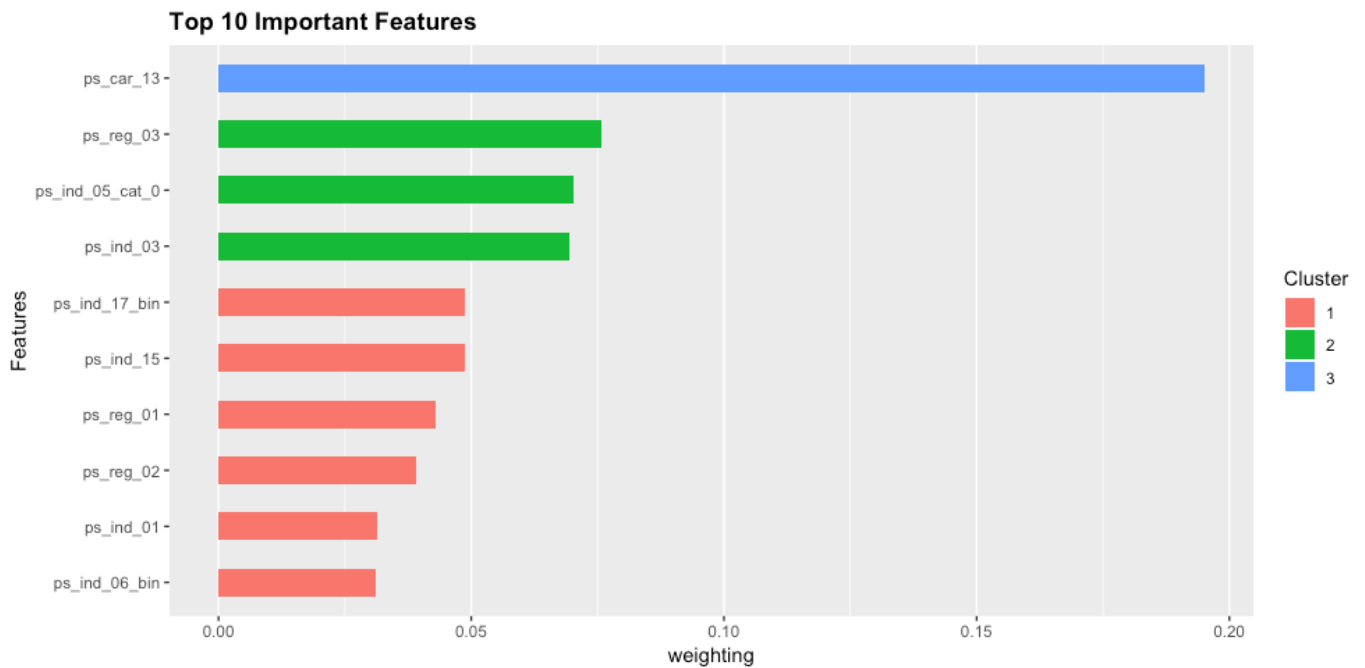


Figure 2: Impact of individual variables on predicted value

The proposed model has been tested through a third-party solution and the predictive power of the model has been determined to be comparable to the best models in the market.

Impact of Model Implementation:

The ability to identify a high-risk individual is critical to profitability for a company, such as Porto Seguro, that operates in the insurance market. By adopting the model proposed, it is projected that Porto Seguro would achieve a significant return on investment from our services. This return would be delivered through the implementation of three strategies.

1. Through utilising our model, Porto Seguro will have the opportunity to identify and withhold their services from high risk individuals that are prone to making claims. Based on available information, the use of the model would save Porto Seguro an estimated £2500 per individual claim, if the approach was to not offer insurance to individuals identified as “high risk”: the average vehicle insurance premium being £500 per annum, compared to an average claim of £3000 (IMF, 2012).
2. Porto Seguro will be able to utilise the model to increase premiums for individuals who evidence specific features. This strategy will enable more accurate targeting of premium rates, significantly reducing the aggregate loss on claim payments.
3. Within Brazil, the insurance market is growing rapidly. Given current market penetration of 3.5% (IMF, 2012), there are significant financial returns available through expansion. As the market grows it is expected to encourage an increasing number of insurance organisations to offer services. It is critical that Porto Seguro is able to offer competitive rates thereby both growing their customer base and retaining existing customers. Utilising our model, to identify both low and high-risk individuals, will enable Porto Seguro to offer rates that other competitors will not be able to match while retaining profitability. This should ensure both new customer growth and customer retention.

The combination of all three strategies can see a significant return on investment. Per annum Porto Seguro could avoid excess claim payments of more than £250,000.

References

IMF. (2012). *Brazil : Detailed Assessment of Observance of Insurance Core Principles of the International Association of Insurance Supervisors* . IMF Staff Country Reports .