

Banking Transaction Fraud Detection

The Blocker Fraud Company is a company specialized in detecting fraud in financial transactions made through mobile devices. The company has a service called "Blocker Fraud" which guarantees the blocking of fraudulent transactions.

The business model of the company is of the Service type with the monetization made by the performance of the service provided, in other words, the user pays a fixed fee on the success in detecting fraud in the customer's transactions.

However, the Blocker Fraud Company is expanding in Brazil and to acquire customers more quickly, it has adopted a very aggressive strategy. The strategy works as follows:

- The company will receive 25% of the value of each transaction truly detected as fraud;
- The company will receive 5% of the value of each transaction detected as fraud, but the transaction is truly legitimate.
- The company will return 100% of the value to the customer, for each transaction detected as legitimate, however the transaction is truly a fraud.

With this aggressive strategy, the company assumes the risks of failing to detect fraud and is remunerated for assertive fraud detection.

For the client, it is an excellent business to hire the Blocker Fraud Company. Although the fee charged is very high on success, 25%, the company reduces its costs with fraudulent transactions detected correctly and even the damage caused by an error in the anti-fraud service will be covered by the Blocker Fraud Company itself.

For the company, in addition to getting many customers with this risky strategy to guarantee reimbursement in the event of a failure to detect customer fraud, it depends only on the precision and accuracy of the models built by its Data Scientists, in other words, how much the more accurate the "Blocker Fraud" model, the greater the company's revenue. However, if the model has low accuracy, the company could have a huge loss.

Case Study Task :

1. **Data Description:** In this first section the data will be collected and studied. The missing values will be threatened or removed. Finally, a initial data description will carried out to know the data. Therefore some calculations of descriptive statistics will be made, such as kurtosis, skewness, media, fashion, median and standard desviation.
2. **Feature Engineering:** In this section, a mind map will be created to assist the creation of the hypothesis and the creation of new features. These assumptions will help in exploratory data analysis and may improve the model scores.

3. Data Filtering: Data filtering is used to remove columns or rows that are not part of the business. For example, columns with customer ID, hash code or rows with age that does not consist of human age.
4. Exploratory Data Analysis: The exploratory data analysis section consists of univariate analysis, bivariate analysis and multivariate analysis to assist in understanding of the database. The hypothesis created in step 02 will be tested in the bivariate analysis.
5. Data Preparation: In this fifth section, the data will be prepared for machine learning modeling. Therefore, they will be transformed to improve the learning of the machine learning model, thus they can be encoded, oversampled, subsampled or rescaled.
6. Feature Selection: After the data preparation in this section algorithms, like Boruta, will select the best columns to be used for the training of the machine learning model. This reduces the dimensionality of the database and decreases the chances of overfitting.