Credit card fraud Detection

By Group 8:
2440007226 - Eduardo Brilliandy
2440026585 - Felix Museng
2440007062 - Gian Reinfred Athevan
2440012390 - Hizkia Christian Purnomo

Introduction

Credit card fraud is one of the most frequent crimes that happens in today's society. Stolen personal information, stolen cards, or credit card skimming are common today. This research attempts to reduce these type of crime by identifying fraudulent transactions.

Introduction

This experiment attempts to detect which transaction is fraudulent or a real transaction using 3 Tree based classification algorithms and determine which algorithm is most suitable. This experiment also attempts to determine the best strategy to deal with a heavily imbalanced dataset.

Data

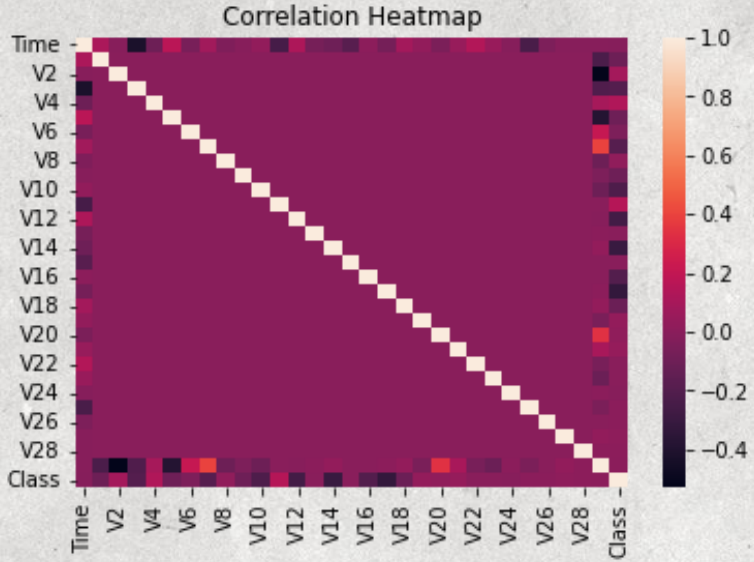
The dataset that is used is a Credit Card Fraud Detection dataset from Kaggle. The dataset contains Anonymized credit card transactions labeled as fraud or genuine transactions.

The dataset contains 284807 rows of data and 30 features. The dataset contains no missing values and all of the data are already numerical values, therefore Data Cleaning is no longer needed.

First the features in the dataset, is scaled using Standard Scaler then processed using PCA (Principal component analysis) Dimensionality reduction, except for the Time and Transaction Amount Features.

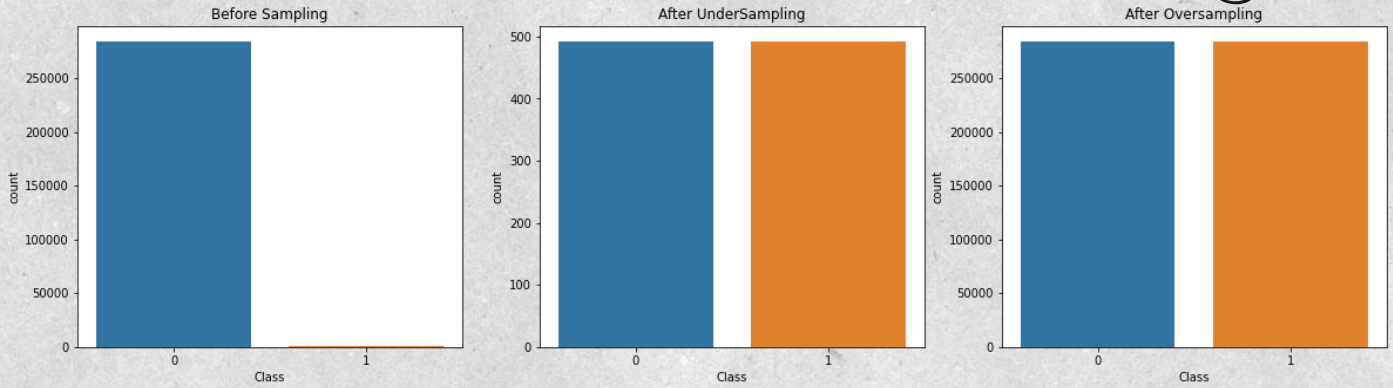
The features are also selected to use only the most relevant features, the feature are selected using the Select K Best method. This method selects the features that contribute most to the target class based on the variation of the features.

Dataset Problem Solving



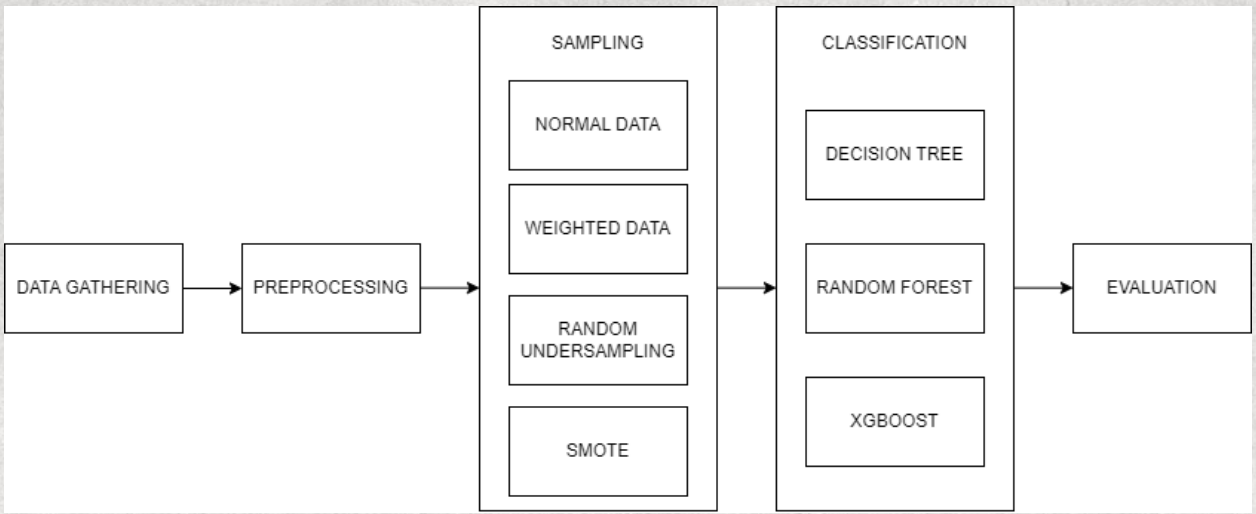
The correlation between the features is shown in the correlation heatmap in Figure 1 above. It is found that features V1 to V28 have no correlation with each other. This is expected because those features are the result of PCA.

Dataset Problem Solving



The dataset is highly unbalanced, the positive class (fraud) accounts for only 0.172% of all transactions. Therefore, undersampling and oversampling will be applied to the dataset. To undersample, Random Undersampling is used. To oversample, SMOTE (Synthetic Minority Oversampling Technique) is used. Figure 2 above shows the class distribution from before sampling, after undersampling, and after oversampling respectively. (0: not fraud, 1: fraud)

Method



Classification Algorithms:

- Decision Tree
- Random Forest
- XGBoost

Imbalanced Data Strategies:

- Normal Data
- Weighted Data
- Undersampled Data (Random Undersampling)
- Oversampled Data (SMOTE)

Evaluation

To evaluate each model, k-fold cross validation will be used with $k = 5$. This means that the data will be split into 5 subset, one subset will be used as the Testing data to test the model while the other subsets will be used as the Training data to fit the model.

The metrics that will be used to evaluate the models are Precision, Recall, F1 Score, and ROC AUC. These metrics will provide more useful information than the default accuracy because of the heavily imbalanced dataset, accuracy cannot be used as a valid evaluation metric.

Result & Analysis

Table 1 Decision Tree Results

	Normal	Weighted	Undersampled	Oversampled
Precision	0.8766	0.8795	0.5078	0.6792
Recall	0.8880	0.8687	0.9141	0.8901
F1	0.8814	0.8740	0.4890	0.7452
ROC AUC	0.8880	0.8687	0.9141	0.8901

Table 2 Random Forest Results

	Normal	Weighted	Undersampled	Oversampled
Precision	0.9763	0.9780	0.5249	0.9267
Recall	0.8923	0.8811	0.9382	0.9095
F1	0.9301	0.9239	0.5397	0.9177
ROC AUC	0.9448	0.9470	0.9797	0.9724

Table 3 XGBoost Results

	Normal	Weighted	Undersampled	Oversampled
Precision	0.9604	0.9604	0.5221	0.5498
Recall	0.8902	0.8902	0.9401	0.9462
F1	0.9221	0.9221	0.5336	0.5863
ROC AUC	0.9799	0.9799	0.9807	0.9817

Results & Analysis

- The highest Recall was achieved by the XGBoost model using Oversampling at 94.62%.
- The highest ROC AUC was achieved by the XGBoost model using Oversampling at 0.9817.
- The highest Precision was achieved by the Random Forest model using normal data at 97.63%.
- The highest F1 Score was achieved by the Random Forest model using normal data at 0.9301.
- Weighting the target class in all 3 models had little to no effects at all.
- Undersampling and Oversampling could increase the correct prediction of fraud transactions, but could perform worse in correctly predicting genuine transactions.
- Random Forest and XGBoost models performed better than the Decision Tree model.

Conclusion

In conclusion, The XGBoost model using Oversampled data performed the best if correctly labeling fraud transactions is the most important and the Random Forest model using Normal data performed the best if correctly labeling genuine and fraud transactions are equally important. Overall, the Decision Tree model performed the worst out of the 3 models.

For the strategies to deal with an imbalanced dataset, weighing the target class had little to no impact. Undersampling and Oversampling helped in correctly predicting fraud transactions but made less correct genuine transaction predictions.

Implication

Machine learning can be used as the first layer of detecting credit card scam and warns people when their credit card is compromised. The findings in this experiment can help improve future fraud detection systems in detecting fraudulent transactions. This research has shown that both Random Forest and XGBoost can be reliable ways of detecting credit card fraud in banks that provide credit cards.

Further studies could test other classification algorithms to see if higher scores could be achieved. Further studies could be done to test the effects of sampling on an imbalanced dataset by testing different ratios of sampling.

THANK YOU