Project Report: Predicting Credit Card Fraud

EXECUTIVE SUMMARY

AJOS Bank is currently experiencing an increase in the rate of fraud complaints in its credit card transactions for the year. The 2020 fraud rates are 0.38% (38 for every 10,000 transactions); this is double the 2019 rates. To protect its reputation, increase customer confidence and reduce financial losses, a group of fraud detection experts have been engaged to build a detection mechanism that correct identifies and prevents fraudulent transactions.

For this project four machine learning models were designed to analyze the bank's dataset, with the best selected using the confusion matrix and cost benefit analysis as the evaluation criterion. The most significant features that predict fraud cases were also identified, along with recommendations centered around these factors.

DATA DESCRIPTION

The dataset was sourced internally from the bank's database with a total of 555,719 transaction and 23 features, which include 1 target attribute (is_fraud). The values in the target column were encoded as 0 and 1, representing no fraud and fraud, respectively. It is important to note that the dataset is imbalanced, with 99.6% having no fraud cases and 0.4% having fraud cases. This was balanced to a 60:40 ratio using oversampling techniques to create more fraud cases.

Lastly, new columns that better explain the target attribute were extracted from existing ones. Some of these columns include age, hour, weekday, and month.

MODEL EVALUATION

Models were evaluated using the confusion matrix metrics as well as well as a cost benefit analysis. Each of the 4 models were trained on 2 datasets each (balanced and unbalanced) with all models tested on the original dataset. The Random Forest model scored highest on both metrics and has been selected as the model to be implemented.

Model	**Dataset**	**Accuracy**	**Testing Error**	**Sensitivity**	**Specificity**
Logistic Regression	Balanced	86.1%	13.9%	0.999	0.05
Logistic Regression	Unbalanced	99.6%	0.04%	0.996	0.00
Decision Trees	Balanced	99.4%	0.06%	0.991	1
Decision Trees	Unbalanced	99.8%	0.20%	0.999	0.711
KNN	Balanced	89.7%	10.4%	0.904	0.897
KNN	Unbalanced	99.6%	0.04%	0.722	0.826
Random Forest	**Balanced**	**99.9%**	**0.01%**	**0.999**	**1**
Random Forest	Unbalanced	99.9%	0.01%	0.999	0.647

CONCLUSIONS AND RECOMMENDATIONS:

In conclusion, we were able to build a random forest that is highly accurate with the ability to identify fraudulent transactions 99% of the time. With this recorded accuracy the model allows for us to ensure that this model will never misidentify legit transactions as fraudulent transactions. Additionally, our model will be able to cut down at our largest loss by identifying fraudulent transactions 99.9% of the time. Considering the amount that was lost to fraudulent transactions last year, this model would've been able to reduce this amount to around $1,325 which is almost 99.9% reduction from what was observed last year.

Our recommendation to the stakeholders is to implement the random forest model into production and have it as a gauge to actively monitor the legitimacy of transactions. With the high performance of this model, undetected fraudulent transactions will be below the industry standard 0.15%. We also have created a pair of guidelines we suggest for YAJOS bank to follow. Primarily, we suggest that bank maintains up to date records of their customer and merchant to provide the model with more data to retrain with in the future. Secondly, we suggest the introduction of 2FA technology ensure that larger purchases are to be authorized by users with additional conformation.