

Proposal: Fraud Detection for Transactions

Author: Kausik Chattapadhyay

Date: 03/14/2024

DSC-680 Applied Data Science

Topic: Development of a machine learning model to predict whether a financial transaction is fraudulent or legitimate.

Business Problem: The Fraud Company aims to enhance its fraud detection service in the Brazilian market. With a revenue-sharing model heavily dependent on correct fraud detection, the company faces significant financial risk if its models are not precise. This project aims to mitigate this risk by building a robust machine learning model capable of accurately finding fraudulent transactions.

Datasets: The project will use financial transaction data provided by the Fraud Company. This dataset includes transaction details such as transaction type, amount, method, and whether the transaction was flagged as fraudulent or legitimate.

<https://www.kaggle.com/datasets/ealaxi/paysim1>

Methods: The project will follow a structured approach, including data preprocessing, feature engineering, exploratory data analysis, model training, and evaluation. Various machine learning algorithms such as logistic regression, k-nearest neighbors, support vector machines, random forest, XGBoost, and LightGBM will be explored for model development. Additionally, techniques like data filtering, feature selection, and hyperparameter tuning will be employed to enhance model performance.

Research Questions:

- What is the profit expected by the fraud company when using the model?
- What is the loss expected by the Company if it classifies 100% of the transactions with the model?
- What is the revenue expected by the company classify 100% of transactions with the model?
- How reliable is the model in classifying transactions as legitimate or fraudulent?
- What is the model's Precision and Accuracy?
- Does the company give back 100% of the value for the customer in each transaction detected as legitimate, however the transaction is a fraud?
- What type of transactions and amount has higher probabilities for fraud?

Ethical Considerations: There are several ethical considerations to address in this project, including data privacy, bias in model predictions, and potential impacts on individuals falsely identified as fraudulent. It's essential to ensure that the model's predictions are fair and unbiased, and measures should be taken to protect sensitive customer information.

Challenges/Issues: One significant challenge is dealing with imbalanced data, where fraudulent transactions are relatively rare compared to legitimate ones. This imbalance can lead to biased models favoring the majority class. Additionally, ensuring the model's interpretability and explainability is crucial, especially in a financial context where decisions have significant consequences.

References: The project will rely on a combination of academic research papers, industry reports, and relevant literature on fraud detection, machine learning algorithms, and ethical considerations in data science. Some key references may include academic journals such as IEEE Transactions on Neural Networks and Learning Systems, industry reports from organizations like Febraban, and online resources like Towards Data Science.

<https://www.kaggle.com/datasets/ealaxi/paysim1>

By addressing these aspects outlined in the proposal, the Transaction Fraud Detection project aims to provide Fraud Company with a reliable and correct solution to mitigate financial risks associated with fraudulent transactions, ultimately enhancing its service and reputation in the market.