*Project Report on Credit-Card Fraud Prediction Model*

Summery

Electronic transactions made easy to user easily transfer their capital to one another. Credit card is one of way to transfer capital electronically with millions of users worldwide, therefore credit card fraud is serious concern now a days. Detecting fraudulent transactions is crucial to minimize financial losses and maintain customer trust. In this project, we developed a credit card fraud prediction model using machine learning techniques. The objective was to build a model capable of predict fraudulent transactions based on the historical transaction data.

Data Collection

Collected dataset containing historical credit card transactions. The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions

Data Preprocessing

Before training the model, we performed several preprocessing steps:

**1. Load and overview Data-set:** Credit card data set is loaded by **pd.read\_csv(r‘file path’)** and overview on data-set is done by looking at features and functions like **info(), describe(), isna()**

**2.** **Exploratory Data Analysis:** Analyse and understand the data-set to identify patterns, relationships, and trends in the data by using Descriptive Statistics and Visualizations of various charts.

**3. Data Cleaning:** This might include standardization, handling the missing values and outliers in the data.

**4. Dealing with Imbalanced data:** Data set is highly imbalanced in ‘Class’ feature, therefore we use undersampling and oversampling method to balance data to train model.

Model Selection

**1. Model Selection:** To predict Fraudulent transactions we use ML algorithms, including

* Logistic Regression
* Decision Tree Classifier
* Random Forest Classifier

Model selection in both under-sampling and oversampling method, the most appropriate model that can be used for this project.

Model Training and Evaluation

**1. Model Training:** Split the data into train & test sets and use the train set to estimate the best model parameters.

**2. Model Validation:** Evaluate the performance of the model on data that was not used during the training process. The goal is to estimate the model's ability to generalize to new, unseen data and to identify any issues with the model, such as overfitting.

Model Deployment

**1. Model Deployment:** Model deployment is the process of making a trained machine learning model available for use in a production environment.

Design Choices and Performance Evaluation

After thorough experimentation and evaluation, we found that the Random Forest Classifier in oversampling model getting overfitted, therefore we select Decision Tree model in under-sampling which shows relevant scores to each other. The final selected model achieved the following performance metrics on the test set:

* Accuracy: 88.94%
* Precision: 91.75%
* Recall: 87.75%
* F1-score: 89.44%

Future Work with Model

Model saved as **“credit\_card\_model”**

To use this model we use this code **model = joblib.load("credit\_card\_model")**

Predicting credit card fraud is an ongoing challenge due to the evolving nature of fraudulent activities and the need to stay ahead of increasingly sophisticated fraudsters. Following are some points for future work in credit card fraud prediction:

* **Feature Engineering:** Identifying new features or enhancing existing features to learn patterns that indicates of fraudulent transactions.
* **Real-Time Detection**: Building low memory taking model that can quickly process features transaction data and identify transaction class in real-time.
* **New Data Exploration**: Collecting new data related to fraudulent transaction so that we can train model on updated and modified dataset.
* **Model Deployment**: Streamlining the deployment and integration of fraud detection models into existing banking systems is crucial for operational efficiency. Future work could explore automated deployment pipelines, model monitoring systems, and integration with fraud management workflows to ensure seamless implementation and maintenance of fraud detection solution

Conclusion

In this project, we successfully developed a credit card fraud prediction model capable of accurately identifying fraudulent transactions. The model's high recall indicates its effectiveness in detecting fraudulent activity, thereby enabling financial institutions to take timely action to mitigate losses. Future work could involve deploying the model in a real-world environment and continuously updating it with new transaction data to improve its performance over time.