**Fraud Detection Project**

**Introduction**

This project focuses on the application of both traditional machine learning and deep learning algorithms for detecting fraudulent transactions in the financial industry. Fraudulent activities can lead to significant financial losses and damage an institution's reputation. Therefore, the development of effective fraud detection models is crucial.

We have explored various machine learning models, including Random Forest, XGBoost, Logistic Regression, K-Nearest Neighbors, Naive Bayes, and deep learning models with SimpleRNN and LSTM cells. These models produce different results, and our main objective is to identify metrics that accurately assess their performance in fraud detection. Additionally, we aim to minimize false negative transactions to reduce financial losses.

To address the highly unbalanced nature of the initial dataset, we employ the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic fraudulent transactions. Ideally, a balanced dataset with both fraudulent and genuine transactions would lead to more robust models.

**Dataset - PaySim**

Obtaining publicly available financial datasets, especially in the emerging field of mobile money transactions, is a challenge due to the sensitive nature of financial transactions. However, we used the PaySim dataset, which simulates mobile money transactions based on real transaction samples from one month of financial logs from a mobile money service in an African country. These logs were provided by a multinational company operating the mobile financial service in multiple countries worldwide.

**How to Run**

To run the code for this project, follow these steps:

1. Download the repository and the PaySim dataset from Kaggle.

2. Update the file paths within the code files to point to the downloaded dataset.

3. The quickest way to execute the code is by opening the files in Google Colab.

4. Alternatively, you can open the 'Data\_Visualisation.ipynb' file to visualize the data and the data cleaning process. The 'Algorithms.ipynb' file contains all the algorithms and their results.

**Future Research**

We envision further research and improvements in the following areas:

**Custom Loss Function**: Develop a custom loss function for both traditional machine learning and deep learning models that calculates the financial loss based on false negative transactions. This is a complex challenge, as real-world administrative costs also play a significant role in determining a bank's total loss.

**Deep Learning Model Optimization**: Explore and optimize the existing deep learning models. Experiment with deep learning models and hyperparameters such as epochs, batch sizes, and consider creating a new RNN model. Additionally, consider whether RNN is the most suitable model for this dataset, as fraud detection might not solely rely on sequential data. Other behavioural factors like user navigation and time spent on certain actions could be valuable.

**Feedback**

Your feedback is highly appreciated as it helps us improve our work and contribute to the advancement of fraud detection techniques. Thank you for your time and interest in our project!

Feel free to reach out with any questions or suggestions. 😊