APPROACH NOTE

TEAM MEMBERS:
Khushi Agrawal
Ishita Agarwal
BRANCH:
Electronics and Telecommunication.

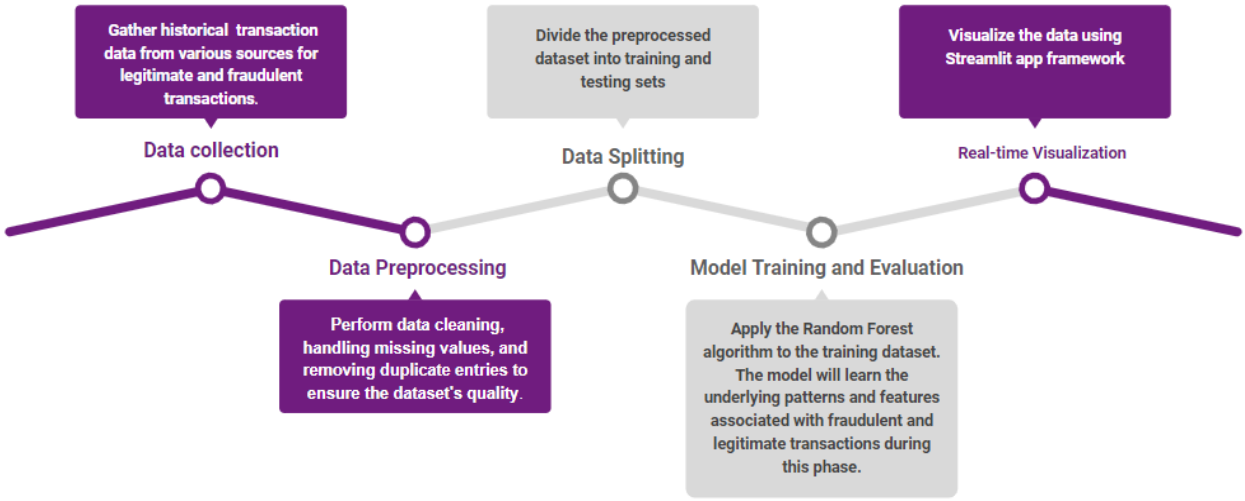
Theme: SPATIAL DATA ANALYTICS PLATFORM

Approach Taken: Detecting Financial Fraud: Modeling and Actionable Insights

This approach aims to address the pressing issue by leveraging data-driven approaches, advanced analytics, and machine learning models.
We have Developed a Machine Learning model for predicting fraudulent transactions for a financial company and then those insights can be used from the model to develop an actionable plan. We have taken the dataset from Kaggle.
The provided data has the financial transaction data as well as the target variable **isFraud**, which is the actual fraud status of the transaction and **isFlaggedFraud** is the indicator which the simulation is used to flag the transaction using some **threshold value**.

High level architecture :

1. Data Collection : Gather historical credit card transaction data from various sources, including both legitimate and fraudulent transactions. This dataset will serve as the basis for training and evaluating the model.
2. Data Preprocessing : Perform data cleaning, handling missing values, and removing duplicate entries to ensure the dataset's quality.
3. Data Splitting : Divide the preprocessed dataset into training and testing set
4. Model Training : Apply the Random Forest algorithm to the training dataset. The model will learn the underlying patterns and features associated with fraudulent and legitimate transactions during this phase.
5. Model Evaluation : Evaluate the trained model's performance on the testing dataset using appropriate metrics such as accuracy, precision, recall.



- The dataset is split into 80:20 ratio as training as testing.The dataset contains step that maps a unit of time in the real world. In this case 1 step is 1 hour of time. Total steps 744 (30 days simulation). type - **CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.** We are using the **Random Forest Classifier**. Decision Tree can also be used here.
- From the results it is concluded that **TRANSFER** and **CASH_OUT** are two most used modes of transaction and we can see that transfer and cash_out are also the only way in which fraud happens.
- Heatmaps are even plotted for representing data graphically where values are depicted by color.
  Average precision score is 0.7687.
  Machine learning can be used for the detection of fraud transactions. Predictive models produce good precision scores and are capable of detection of fraud transactions

- One of the most challenging issues faced by the Financial services company is Evolving Fraud Patterns. Fraudsters constantly adapt and develop new techniques to evade detection. So here we are visualizing the data with the help of **Streamlit**, **matplotlib library** in python to understand the fraud patterns.