Fraud Detection

Team Name: khushirajpurohit2021

Participant name: Khushi Rajpurohit

Problem Statement: Financial institutions face challenges in detecting and preventing fraudulent transactions, which can result in significant financial losses. Develop a machine learning model that analyses transaction patterns and user behaviours to detect and flag potentially fraudulent activities in real-time.

Objective: Enhance the security of financial transactions by providing an advanced fraud detection system that identifies suspicious activities quickly and accurately.

Dataset used: I have used Fraud dataset, which includes variables such as amount, type of transaction, old and new transaction detail, "isFraud" states if that transaction is fraud(1) or not(0).

Data exploration: head(),.shape(),.info(),.describe(), .duplicated() and .isnull().sum() methods are used to get an overview of dataset, its structure and missing values.

Exploratory Data Analysis:

- Total count of each Payment type across the dataset.
- Amount of transaction.
- Percentage of amount for each Payment type.
- Percentage of fraud and Normal transaction in the dataset.
- Counting fraud transaction with respect to Payment types.

Insights from the Analysis:

- There are five types of transaction that includes CASH-OUT, CASH IN, DEBIT, TRANSFER, and PAYMENT.
- Most of the transaction was carried out by "cash _out" Payment type. And the least payment type for the transaction was "Debit". Mostly people use financial transaction for either getting the cash out or for the payment.
- The distribution of transaction amounts appears to be heavily skewed, with a significant number of transactions being of the dominant amount. This could indicate a specific type of transaction or a system-generated value.
- Fraud happens only in "cash out" and "transfer" Payment types.
- The frauds in financial transactions are very small in size as compared to the whole dataset, with a percentage of as least as 0.1.

Predictive Modeling:

- Prepared data for data modeling
    - Feature selection
    - Encodes categorical features using Label Encoder
    - Splits data into training and test set with help of train_test_split.

- o Converted the features to array and normalized them for easy prediction.
- Trained a logistic regression, KNN, decision Tree model and evaluated its performance using metrics like Log Loss, Jaccard index, accuracy score, presion, recall, f1score and finally created a Confusion matrix to visualize the Prediction in comparison to the actual labels.
- **Logistic Regression**
  - o Accuracy:
  - o F1 score:
  - o Log Loss:
  - o Jaccard Index:
- **Decision Tree:**
  - o Accuracy:
  - o F1 Score:
  - o Jaccard Index:
- KNN :
  - o Accuracy:
  - o F1 Score:
  - o Jaccard Index:

Final Interpretation:

- With the result of all models accuracy, the best performing model is Decision Tree. As a result I choose Decision Tree model for detecting Fraud Transactions.