**FRAUD DETECTION**

**OBJECTIVE: -** The primary objective of the fraud detection model is to identify and prevent fraudulent activities within an organization's operations.

**DEPENDENCIES: -** Importing the required libraries such as Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, etc.

**LOADING THE DATASET: -** The dataset has been loaded using the pandas “pd.read\_csv ()” command.

**DATA PRE-PROCESSING:-**

* **Handling Missing values: -** Data was consisting very few null values, which has been dropped as the values were very few as compared to our large dataset.
* **Label Encoding: -** Label Encoder has been used to handle the categorical data, in which the categorical data has been converted into numerical features.
* **Standard Scaling: -** “Standard Scaler” has been used to scale the dataset.

**DATA VISUALIZATION:-**

* **PAIR PLOT: -** Pair plot has been used to visualize the relationship among the various features of the dataset.
* **BOX PLOT: - :** Boxplots provide a visual summary of the central tendency, spread, and shape of the data distribution, allowing for quick comparisons between different groups or categories.
* **HEATMAP: -** Heat map has been used to depict the correlation among various attributes of the dataset.
* **SCATTER PLOT: -** Further, scatter plot has been used to depict the relationship between two continuous variables.
* **COUNT PLOT: -** Count plot has been used to have a look at the categorical features and overall dataset.

**MODEL BUILDING:-**

Various machine learning models such as “**Decision Tree**”, “**Artificial Neural Network**”, “**Random Forest Classifier**”, and “**XGBoost Classifier**” has been used to understand the performance of each model, so that the best model could be used for building our predictive system. Since each model has come up with **Training Accuracy as ---100%** and **Testing Accuracy ----- 99%.** So, in the end. “XGBoost Classifier” has been used for building our Predictive System.

**MODEL EVALUATION:-**

Our model has been evaluated through **confusion matrix, accuracy score, and performance matrix** which includes **F1 Score, Precision, Recall, and support**. Further the model has been evaluated for both training and test dataset. In which training has got **100%** accuracy and testing has got **99% accuracy**.

**PREDICTIVE SYSTEM:-**

We have also build our predictive system, in which “**XGBoost Classifier**” has been used for prediction. This model takes in an input, converts it into a NumPy array, then it further goes through “Standard Scaler” for scaling purpose. Then our “XGBoost Classifier” model give predictions whether the transaction is **fraud** or **legit**.