Fraud Detection System

1. Introduction:

- Globally, credit card fraud is a significant financial issue.
- To reduce monetary losses and preserve consumer confidence, it is essential to identify fraudulent transactions.
- The goal of this task is to use machine learning to create a reliable model for detecting credit card fraud.

2. Dataset Loading and Preprocessing:

- **Description of the Dataset:** A dataset comprising credit card transaction records with attributes such as transaction amount, merchant information, location, demographics, and a fraud indicator is used in the project.

- **Managing Missing Values:** To deal with missing values, imputation techniques were applied.

- **Outlier Detection and Treatment:** IQR was used to identify outliers, which were then either capped or eliminated.

- **Feature Encoding:** One-hot encoding was used to transform categorical features into numerical values.

- **Feature Scaling:** Standardization was used to scale numerical features.

- **Data Splitting:** Stratified sampling was used to separate the dataset into training and testing sets.

3. **Model Development:**

- **Model Selection:** Choosing a Model Because of its robustness, interpretability, ability to handle high-dimensional data, and suitability for fraud detection, a Random Forest Classifier was selected.

  Hyperparameter Tuning: The best hyperparameters were found using grid search.

- **Cross-Validation:** Model generalization was guaranteed by k-fold cross-validation.

- **Model Fitting:** The complete training dataset was used to train the finished model.

4. Model Evaluation And Performance:

- Metrics for Evaluation: The performance of the model was assessed using:
    - **Accuracy:** The general correctness of forecasts.
    - **Precision:** The percentage of fraudulent transactions that were accurately predicted.
    - **Recall:** The percentage of real fraudulent transactions that were correctly identified.

- **F1-Score**: It is the harmonic mean of recall and precision.

- **AUC-ROC Score:** The capacity of the model to differentiate between authentic and fraudulent transactions.

5. Results Analysis:

- The testing dataset yielded good results for the Random Forest Classifier.

- These outcomes demonstrate how accurately the model classifies transactions. The model's predictions closely resemble the actual results due to its high accuracy.

- The precision shows that the model is generally right when it predicts fraud, reducing false positives and the inconvenience to real customers.

- Recall indicates that the model recognizes the real fraudulent transactions, enhancing the credit card system's security.

- The model's excellent performance in reducing false positives while identifying real fraud is confirmed by the F1-score, which strikes a balance between precision and recall.

- The model's outstanding capacity to differentiate between authentic and fraudulent transactions is indicated by the AUC-ROC score.

- Overall, the outcomes show how well the Random Forest Classifier detects credit card fraud.

- It detects a large percentage of fraudulent activity, reduces false positives, and correctly classifies transactions.

6. Conclusion:

- Using a Random Forest Classifier, this project effectively created a strong model for detecting credit card fraud.

- The model minimized financial losses and improved security by identifying fraudulent transactions with high accuracy.

- The findings demonstrate its capacity to accurately classify transactions, reduce false positives, and identify a sizable percentage of fraudulent activity.