**DOCUMENTATION-TASK 1**

**Objective:**

Detect fraudulent credit card transactions.

**Approach:**

Using a simple and fast model for real-time anomaly detection. You've also mentioned the importance of balancing the dataset and preserving user privacy by reducing dimensionality.

**Data Exploration and Preprocessing:**

Loaded the dataset into a Pandas DataFrame .

Conducted basic exploratory data analysis (EDA), including checking for null values and understanding the distribution of variables.

Noted the imbalance in the dataset (very few fraudulent transactions compared to legitimate ones).

Separated the data into 'legit' and 'fraud' for further analysis.

Conducted statistical analysis on transaction amounts.

Visualized the distribution of transaction amounts and times for both fraud and normal transactions.

**Feature Extraction:**

Created a balanced dataset using under-sampling, ensuring equal representation of both classes.

**Data Splitting:**

Split the data into features (X) and target (Y).

**Model Training:**

I've chosen Logistic Regression, which is a good start for its simplicity and speed. Train the model using our balanced dataset.

**Model Evaluation:**

Evaluate the model using appropriate metrics. Given the nature of the problem (fraud detection), focus on metrics like Precision, Recall, F1-Score, and ROC-AUC, rather than just accuracy.

**Feature Importance Analysis:**

Analyse which features are most important for predicting fraud. This can provide insights into the nature of fraudulent transactions.

**Model Improvement:**

Depending on the initial performance, consider trying other algorithms like Random Forest, Gradient Boosting, or even neural networks for potentially better performance.

Implement techniques like cross-validation to ensure the model's robustness.

**Handling Class Imbalance (Advanced Techniques):**

Explore advanced techniques to handle class imbalance, such as SMOTE (Synthetic Minority Over-sampling Technique) or using anomaly detection methods.

**Deployment Considerations:**

If deploying this model in a real-world scenario, consider the need for continuous learning as fraudsters' tactics evolve.

I’ve ensured the model can be updated quickly with new data without significant downtime.

**Privacy and Security:**

I’ve ensured that the model complies with privacy laws and regulations, especially when dealing with sensitive financial data.