**Topic Title**

**Bachelor of Technology**

**Computer Science and Engineering**

Submitted By

NAME (ROLL NUMBER)

MONTH 2025



**Techno Main**

**EM-4/1, Sector-V, Salt Lake**

**Kolkata- 700091**

**West Bengal**

**India**

**TABLE OF CONTENTS**

1. Introduction
2. Details

i

Ii

Iii

Iv

……..

……..

1. Conclusion
2. References

**Please use Times New Roman, 12 and Justified Alignment, Numbering should be Automatic**

1. **Introduction**

This report analyses the **Credit Card Fraud Detection Dataset** from **Kaggle**, comprising **284,807** anonymized transactions labelled as fraudulent or legitimate. The objective is to apply descriptive statistical techniques—including measures of central tendency (mean, median), dispersion (variance, standard deviation), skewness, kurtosis, and quartile analysis—to three numerical features:

* Time (transaction timestamp),
* Amount (transaction value),
* V1 (a PCA-transformed feature).

The analysis aims to characterize the data distribution, identify anomalies (e.g., outliers), and summarize insights to enhance understanding of transaction patterns, particularly for fraud detection. Key focuses include assessing asymmetry (skewness), tail behaviour (kurtosis), and deviations in transaction amounts or transformed features that may signal fraudulent activity.

1. **Methodology**
   1. **Data Preprocessing:**
      1. Load the dataset (284,807 transactions) and inspect its structure.
      2. Check for and handle missing values (if present) to ensure data integrity.
      3. Select three numerical columns for analysis: Time, Amount, and V1 (PCA-transformed feature).
   2. **Descriptive Statistics:**
      1. **Central Tendency:** Compute mean, median, and mode to identify central values.
      2. **Dispersion:** Calculate range, variance, standard deviation, and interquartile range (IQR) to assess spread.
      3. **Skewness & Kurtosis:** Analyze symmetry (skewness) and tail behavior (kurtosis) of data distributions.
   3. **Outlier Detection:**
      1. Compute percentiles (25th, 75th) and use the IQR method to detect outliers.
      2. Visualize outliers and distributions using boxplots, histograms, and density plots.
   4. **Tools & Libraries:**
      1. **Python libraries:** Pandas (data handling), NumPy (statistical computations), Matplotlib/Seaborn (visualization).
2. **Results & Analysis**
   1. **Data Preprocessing**
      1. The dataset is loaded and the first 5 rows are displayed.
      2. Missing values are checked and handled.
      3. Three numerical columns, e.g., **id**, **V1 and V2** are selected for further analysis.
   2. **Measures of Central Tendency**

For each selected column, we calculate:

* + 1. **Mean**: The average value.
    2. **Median**: The middle value when sorted.
    3. **Mode**: The most frequently occurring value.

These measures help understand the central location of data points.

* 1. **Measures of Dispersion**

For each selected column, we calculate:

* + 1. **Range**: The difference between the maximum and minimum values.
    2. **Variance**: The spread of data around the mean.
    3. **Standard Deviation**: The average deviation from the mean.
    4. **Interquartile Range (IQR):** The range between the 25th and 75th percentiles.
    5. **Visualization**
       1. **Histogram and Box Plot**
       2. **Interpretation**
  1. **hmgmhm**

1. **Conclusion**
2. **References**