ONLINE PAYMENTS FRAUD DETECTION USING WITH MACHINE LEARNING:

To build an application that can detect the legitimacy of the transaction in real-time and increase the security to prevent fraud.

By

(Marri yashmitha)

(Manthina raja rishika)

(Kutagulla safa)

Guided by

Prof. Ms swetha raj

A Dissertation Submitted to
SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY, An Autonomous Institution affiliated to 'JNTU Ananthapur' in Partial Fulfilment of the Bachelor of Technology branch of ***Computer science and Engineering***

May 2024

SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY

R.V.S. Nagar Tirupathi Road, Andhra Pradesh– 517127

Data exploration and data preprocessing:

Fraud detection is an important aspect of any financial system. It helps protect businesses, individuals, and financial institutions from the negative effects of fraudulent transactions. However, detecting fraud is not a simple task. It requires a sophisticated system that can analyze a large amount of data in real-time to identify anomalies and suspicious patterns. data preprocessing is a crucial step in the fraud detection process. It involves cleaning, transforming, and preparing data for analysis. The quality of the data used in *fraud detection models* significantly impacts the accuracy of the results. In this section, we will explore data preprocessing techniques for *fraud detection*.

1. Data cleaning: This involves removing or correcting any errors or inconsistencies in the data. For example, missing values, duplicate records, or outliers can significantly affect the accuracy of *fraud detection models*.

2. Data transformation: This step involves converting the data into a format that is more suitable for analysis. For example, converting categorical data into *numerical data* or *normalizing data* to ensure it is on the same scale.

3. Feature engineering: This involves selecting and creating features that are relevant to *fraud detection*. For example, the time of day a transaction occurs or

the location of the transaction can be useful features in detecting *fraudulent activity*.

4. Sampling: This involves selecting a subset of the data for analysis. Sampling can help reduce the computational power needed for *fraud detection models* and prevent overfitting.

5. Data augmentation: This involves creating *synthetic data* to supplement the existing data. For example, creating additional *fraudulent transactions* to balance *the class distribution* of the data.

Overall, data preprocessing is a critical step in the fraud detection process. It helps ensure that the data used in *fraud detection models* is accurate, relevant, and suitable for analysis. By applying these techniques, businesses and financial institutions can improve the accuracy of their *fraud detection models* and protect themselves from *fraudulent activity*.