Introduction

Online payments have become the predominant method of payments worldwide due to their efficiency as well as the convenience of the transactions. However with the growth of the digital payment methods, it has led to an increase in payment fraud that present significant challenges for the users of these services such as financial institutions and consumers etc. The worst-case scenario is these fraudulent transactions not only result in financial losses but also decreases the integrity and trust in online payment systems. This posing substantial risk to those service platforms also. So to identify and distinguish between fraudulent and non-fraudulent online payments, is crucial to be handled. This study aims to analyse fraudulent transactions using a dataset from Kaggle, which include past data on various transaction types and their outcomes. By examining them, we seek to identify underlying patterns and characteristics of fraudulent transactions to enhance the mechanism, fraud detection.

Objective of the Study

- Build a predictive model to identify fraudulent and non-fraudulent payments.

Data

Variable	Description
Step	Represents a unit of time where 1 step equals 1 hour.
Type	Type of online transaction (e.g., payment, transfer).
Amount	The amount of the transaction.
Nameorig	Identifier for the customer initiating the transaction.
Oldbalanceorg	Balance of the customer before the transaction.

Newbalanceorig	Balance of the customer after the transaction.
Namedest	Identifier for the transaction recipient.
Oldbalancedest	Initial balance of the recipient before the transaction.
Newbalancedest	Balance of the recipient after the transaction.
Isfraud	Indicator of whether the transaction is fraudulent.
Isflaggedfraud	Indicator of whether the transaction was flagged as fraudulent by the system. (Will not be used in the machine learning model building)

Significance of the Study

This study aims to enhance the detection and prevention of fraudulent activities in online payments. By identifying fraudulent transactions, financial institutions and customers can improve their confidence in online payment platforms. And also the service providers can also implement more robust security measures, and reduce economic losses.

Suggested Methodology

Logistic Regression will be used to do the classification task of detecting fraud or non-fraud transactions as the dependent variable is being of binary type. In the case of advancement of the model, the Decision Tree Classifier and Random Forest Classifier will be used.

References

Dataset – https://www.kaggle.com/datasets/jainilcoder/online-payment-fraud-detection