

Team Details: (Team 56)

Team members: 1) Shraddha Ghagare, 2) Nidhi Shivapuji, 3) Rutuja Manakapure.

Theme Chosen: Machine learning-based Fraud Detection System

( [Hackathon_Dataset.xlsx](#))

Brief abstract of the idea:

Increasingly, instances have arisen wherein customers place orders for products via online retail platforms, only to subsequently return the items after utilizing them. This deceptive practice can be identified as 'Post-Usage Return Fraud.' The underlying strategy involves customers procuring products for specific needs, often for single-use scenarios, and then orchestrating a return by alleging product defects or expressing dissatisfaction.

Our objective revolves around devising a machine-learning framework capable of effectively determining these fraudulent activities from legitimate return requests.

High level architecture:

The model that we will be using for our problem statement is Bagging Random Forest.

Bagging Random Forest is an ensemble learning technique that combines multiple decision trees trained on different subsets of the data (Bagging). It introduces feature randomness by considering only random subsets of features at each split (Random Forest). The final prediction is made through voting (classification) or averaging (regression) the outputs of individual trees. This approach reduces overfitting, improves accuracy, and enhances model robustness. It is widely used in machine learning for its ability to handle complex data and achieve high predictive performance.

Implementation plan:

The dataset comprises order details of products from a retail company. The dataset consists of details of buyer, product, shop, order no. and order status. Process the data by removing the rupees symbol, handling NaN values and converting integers to float wherever required. We will perform data visualization operations to analyze the features of our dataset against the target values to understand the prominent features of our dataset. We will then implement label encoding to convert required columns from categorical to numerical values to ease further operations. We will then split the data into training and testing. Do feature selection using appropriate ML libraries using variance as selection criteria. We will pass our preprocessed data through the Random Forest Bagging Model. Then we will perform appropriate accuracy parameters to calculate the accuracy of the model predictions. If time permits we will implement frontend using already-built external templates. Our machine learning model will detect fraud based on the absurd rise in the number of returns from a particular customer and higher number of returns to a particular store indicating fraudulent activities happening at the store.