

Scenario Answer ML set2:

1. Predicting loan default:

Problem type-Classification.

Collection of Data: Gather customer financial history, credit scores, and loan repayment records.

Preprocess Data: Handle Missing values, normalize numerical features and encode categorical variables.

Split the data: Divide the dataset into training and testing sets.

Choose algorithm: Use Logistic Regression, Decision Tree or Random Forest

Train Model: Fit the model using labelled loan default data

Evaluate performance: Use AUC-ROC, precision, Recall, F1-score

Predictions: predict loan default for new applicants

2. Forecasting Demand for a Retail store:

Problem type: Regression

Collect Data: gather past sales data, seasonal trends, and product demand

Preprocessing data: Handling Missing values, normalize numerical data, and remove Outliers

Split the data: Divide the dataset into training and testing sets'

Choose Algorithm: Use Linear Regression, Random Forest Regression, or XGBoost.

Train the model- Fit the model using historical demand data.

Evaluate performance- Use RMSE and R2 score

Make performance- Forecast demand for upcoming sales periods

3. Detecting Defective Products in Manufacturing:

Problem type- Classification

Collect Data:Gather sensor readings,production details and defect labels

Preprocessing Data:Handle Missing values,normalize numerical values, and encode categorical features.

Split the data-Divide the data into training and testing sets

Choose Algorithm-Use Decision Trees, Support Vector Machines or Neural Networks

Train the model-Fit the model using labelled defect data.

Evaluate Performance-Use accuracy,Precision,recall, and F1-score

Deploy Model-Detect defective products in real time.

#### 4.Classifying Medical Diagnoses:

Scenario:A healthcare provider wants to classify patient symptoms into different disease categories.

Problem type: Classification

Collect data:Gather dataset patients records with symptoms and diagnoses

Preprocess data: Handle missing values,normalize medical test results and encode categorical features

Split dataset; Training and test split

Use algorithm: Logical .

Train model-Fit

Evaluate model-use accuracy,confusion matrix and f1-score

Deployment-Deploy the model

Make predictions-predict disease category based on patients symptoms.

#### 5.Identifying Fake Online Reviews-

Scenario-An e-commerce company wants to detect fake reviews posted by bots or fraudsters.

Problem type-Classification

Collect Data-Gather dataset of real and fake reviews

Preprocessed data-Tokenize text,remove stopwords and vectorise using TF-IDF

Feature Engineering-Identifying suspicious patterns like repetitive words unnatural phrasing,and review frequency.

Split data-Training and test sets

Choose Algorithm-Use NaiveBayes,LogisticRegression,or Transformer models

Train Model-Fit the model on labelled review data

Evaluate performance-Use accuracy,F1-score,and confusion matrix

Make Predictions-Detect fake reviews in real-time

## 6.Predicting Stock Market trends

Scenario-A financial firm wants to predict stock price movement based on historical price

Data and market indicators.

Identify the problem type:Regression

Collect the data-Gather historical stock prices trading volumes, and economic indicators

Preprocesses data-Handling Missing values ,Normalizing/scaling

Feature Engineering-Create new featureslike moving averages,relative strength index(RSI),or Bollinger Bands to capture trends and patterns

Time series Analysis:Apply differencing or Logging

Split the data-Training and test sets

Choose algorithm-Use Random Forest Regression, LSTMs,or Gradient Boosting

Train the model-Fit the model on historical stock data

Evaluate performance-Use RMSE and directional accuracy.

Make Prediction-Forecast future stock price movements

## 7.Detecting Fake Social Media Accounts-

Scenario-A social media platform wants to identify and remove fake user accounts

Problem type-Classification

Collect data-Gather account details, activity logs and engagement patterns

Preprocessed data-Handle missing values,engineer features like average post frequency and follower ratio.

Split data-Divide Training and test set

Choose algorithm-Use Random Forest ,Support Vector Machine or XGBoost

Train the model-Fit the model using labelled real and fake account data

Evaluate the model-aaaaaaaause precision,recall and F1-scorerering

Make predictions-identify and flag fake accounts

## 8.Optimizing Ad Targeting for Online Marketing

Scenario-A digital marketing company wants to show the most relevant ads to users based on their browsing behaviour.

Problem type-Clustering

Collect the data-Gather user click behaviour browsing history and demographic

Preprocess data-

Choose algorithm-Use K-Means or Hierarchical Cluatering

Determine Optimal Clusters-Use the Elbow Method

Train the modal-Apply clustering algorithm tosegment users

Analyse clusters-Identifi=y user groups(e.g.,” TechEnthusiasts,” “FasionLovers”)

Optimize Ads-Deliver targetd ads based on cluster preference

## 9.Classifying Land Cover in Satellitw images-

Scenario-A geospatial research team wants to classify different land types(forest,water,urban)usingsatellite images

Problem type-Classification

Collect data-Use satellite images labelled with land types

Preprocessed data- Normalize pixel values ,remove noise, and extract image features

Split dataset-Divide Training and test set

Choose Algorithm-Use Decision Trees,Support Vector Machines or CNN-based models

Train model-Fit the model on labelled satellite images

Evaluate performance-Use accuracy and confusion matrix

Make predictions-Classify new satellite images into land cover types

## 10. Predicting Customer Churn for a Subscription service

Scenario-A streaming service wants to predict which users are likely to cancel their subscriptions.

Problem type-Classification

Collect data-Gather historical data on user behaviour, including subscription details, usage patterns, demographic information and any other relevant factors that influence churn

Preprocess data-Handling missing values encoding and Normalizing variables

Feature Engineering-Create features like average watch time per session and last login frequency, number of device used.

Split dataset-Train and test set

Choose algorithm-Use LogisticRegression, Random Forest, or Gradient Boosting

Train Model-Fit the model using past churn data

Evaluate performance-Use AUC-ROC, precision, and Recall

Make predictions-Identify customers likely to churn and apply retention strategies