Scenario based questions –ML set1

1.Supervisied Regression problem,

Output- To predict house price-numerical

Input-square footage,number of bedrooms,and location

Step1:

Data collection: gathering data with respect tocorresponding features and target variable price

Step2:

Data preprocessing:Handling missing values-Imputation or interpolation.

Encoding-convert categorical like loction into numerical using like oneHot encoding or Label encoding.

Scale/normalize feature-to improve model performance

Step3:

Feature Engineering:extract relevant features by using select Kbest,rfe factor for better performance.

Step4:

Model Selection-Linear Regression

Decision Tree

Random Forest

Gradient Boosting

Support Vector Machines

Step5:

Model Training and Evaluation-Split the dataset: Split the dataset into training and testing sets

(80% training and20% for testing)

Train the model: Train the selected model on the training set.

Evaluate the model:Evaluate the model on the testing test using metrics likeMean Absolute Error(MAE),Mean squared Error(MSE),andR-squared.

Step6: Hyperparameter Tuning-using to tune the Model's Hyperparameters to optimise its

Performanance like Gridsearch,randomsearch,or Bayesian optimization

Step7:Model Deployment-Deploy the model using trained Model in a production-ready environment

Where it can receive input features and predict house prices.

Step8:Model Monitering and Maintenance-contineously moniter and add new data

And retain the model as necessary to maintain its accuracy.

Update the Model:Update the model to adapt to changes in the market or data distribution.

By following these steps,the real estate company can develop a robust and accurate house price

2.Detection –Bank Fraudulent Transactions:

Supervisied Classifiction –to predict to classify transactions as either legitimate or fradulent

Step1;Data collection:

Gather transactional data,including:

Customer demographic

Transaction History(amount,time,location,type)

Device information(IPaddress,device ID)

Behavirol data Ilogin frequency,transaction frequency)

Ensure the dataset includes both legitimate and fraudulent transactions.

Step2: Data preprocessing:

Handle missing values

Encode Categorical variables

Scale/normalize features

Feature Engineering: Extract relevant features from

Transaction frequency and amount patterns

Time of day/week/month

Location-based features (e.g.,country city)

Step3:Exploratory Data Analysis(EDA)

Analysis the distribution of legitimate and fraudulent transactions.

Identify patterns and correlations in the data.

Step4: Model Selection-Supervised Learning:

LogisticRegression

Decision Trees

Random Forest

Support Vector Mechines

Unsupervised learning:

Local Outlier Factor

Isolation Forest

Autoencoders

Hybrid approaches:

Combining supervised and unsupervised techniques

Step5: Model Training and Evaluation

Split the data-train and test set

Train the model usin train set

Evaluate the model-test set-metrics like

Precision

Recall

F1-score

ROC-AUC

Confusion marix

Step6: Hyperparameter Tuning-Tune the model's hyperparameter to optimize its performance

Step7: Model Deployment-Deploy to trained model in a production-ready environment where it can analyse transactions in real time

Stea8: Model Monitoring and Maintenance

3.Supermarket Customer Shopping Patterns:

Clustering Problem-goal is to group customers based on their shopping patterns abd preferences

Step:1:Data Collection,

Tranaction history(productspurchased,quantity,frequency)

Demographic data(age,location,income)

Loyalty program data(if available)

Online behaviour(if available)

Ste2:Data Preprocessing: Handle missing,Data transformation-OneHot Encoding

Feature Engineering:

Purchase frequency

Average basket size

Product categories purchased

Step4:Clustering

K-means

Hierarchical clustering

DBSCAN

RFM(Recency,Frequency,Monetary)analysis

Determine the optimal number-elbow or silhouette score

Step5: Cluster Analysis and Profiing

Demographic data

Shopping behavior

Product prefrences

Create customer profiles for each cluster

Step6: Personalized Promotions

Develop targeted promotions for each customer segment,

Discounts on preferred products

Loyalty rewards

Personalised marketing messages

Product recommendations

Step7:Implementation and Evaluation

Sales lift

Customer retention

Customer satisfaction

Refine the Strategy basedbon customer feedback and performance data.

4.Employees Salary: Regression-goal is to predict a continuous output variable (salary)

Based oninput features(years of experience, job title,education level)

Step1:Data Collection-

Years of experience

Job title

Education level

Salary

Step2: DataPreprocessing-

Handle missing

Encoding

Scale/normalize

Step3:Feature Engineering

Step4:Model Selection choose

Linear Regression

Decision Trees

Random Forest

Gradient Boosting

Support Vector Mechines

Step5: Model Training and Evaluation

Split the data

Train the model

Evaluate the model

Mean Absolute Error(MAE)

MSE

R-Squared

Step6: Hyperparamerter tuning

Step7:Model Development

5.Email provider :Binary Classification-goal is toclassify either spam or not spam based on their contents and sender details.

Step1: Data Collection

Email content(subject and body)

Sender details(email address,IP address)

Scenario based questions –ML set1

Label(spam Or not spam)

Step2: Data preprocessing

Tokenization:split the email content into individual words or tokens

Stopword removal:Remove common words like "the","and" etc.

Stemming or Lemmatization:Reduce words totheir baseform

Feature Extraction:Extract from email.

Word frequency

Keyword presence

Sender reputation

IP address reputation

Step3:Feature Engineering-Sentiment analysis

Topic modelling

Keyword extraction

Ste5:Model Selection-suitable binary classification

Navie Bayes

Logistic Regression

Decision Trees

Randon Forest

SVMs

Step5: Model Training and Evaluation

Split the dataset

Train the model

Evaluate the model

Accuracy

Precision

Recall

F1-Score

ROC-AUC

Step6:Hyperparameter Tuning

Step7:Model Deployment