1. A telecom company notices that 30% of customers who made complaints in the last three months have left. However, 70% of customers who didn't complain remained. How would you modify your feature engineering strategy to improve churn prediction?

   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   CUSTOMER CHURN PREDICTION
   SUPERVISED LEARNING – CLASSIFICATION (DEFAULT / NOT DEFAULT)
   Data collection
   Collect the customer churn data
   Data preprocessing
   Identify and replace all the missing values, and use one hot encoding and label encoding to convert all the categorical variables.
   Input output split
   Feature selection
   Train and test split
   Model creation – try with all the classification algorithms to find the best model.
   Evaluation f1score, roc-auc score, accuracy, precision, recall
   Confusion matrix, classification report

2. A warehouse manager observes that demand for certain products fluctuates seasonally, but your model is not capturing these patterns. How would you improve the model's predictive power?

   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   SUPERVISED LEARNING – TIME SERIES REGRESSION

   Data collection
   Collect historical product demand data over multiple seasons (daily / weekly / monthly).

   Data preprocessing
   Handle missing values, remove outliers, and normalize the data if required.

   Feature engineering
   Create time-based features such as month, week, quarter, holidays, and seasonal indicators.
   Add lag features and rolling averages to capture past demand patterns.

   Train and test split

Split data based on time (e.g., past data for training and recent data for testing).

Model creation
Use models that capture seasonality such as SARIMA, Holt-Winters, Prophet, or machine learning models with seasonal features.

Model tuning
Optimize hyperparameters to better capture seasonal trends.

Evaluation
Evaluate using MAE, RMSE, MAPE, and compare predicted vs actual demand.

Model improvement
Continuously retrain the model with new data to adapt to changing seasonal patterns.

3. Your bank's fraud detection model has a high false positive rate, flagging many legitimate transactions as suspicious. What steps would you take to reduce false alarms while still detecting money laundering effectively?

Problem Identification:
Stage 1: Domain selection – Machine Learning
Stage 2: Learning selection – Supervised Learning
Stage 3: Regression/ Classification – Classification

CREDIT CARD FRAUD PREDICTION
SUPERVISED LEARNING – CLASSIFICATION (DEFAULT / NOT DEFAULT)
Data collection
Collect credit card transaction details data
Data preprocessing
Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
Input output split
Feature selection
Train and test split
Model creation – try with all the classification algorithms to find the best model.
Model tuning -Tune hyperparameters to improve precision and reduce false alarms.
Evaluation f1score, roc-auc score, accuracy, precision, recall
Confusion matrix, classification report

4. An airline wants to use weather data in its flight delay prediction model, but 20% of past records lack weather conditions. How would you handle missing weather data without compromising model accuracy?

   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   FLIGHT DELAY PREDICTION
   SUPERVISED LEARNING – CLASSIFICATION (DEFAULT / NOT DEFAULT)
   Data collection
   Collect customer churn data
   Data preprocessing
   Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
   Input output split
   Feature selection
   Train and test split
   Model creation – try with all the classification algorithms to find the best model.
   Model tuning -Tune hyperparameters to improve precision and reduce false weather predictions.
   Evaluation f1score, roc-auc score, accuracy, precision, recall
   Confusion matrix, classification report

5. A student with high quiz scores but low engagement time is placed in an advanced learning path by your clustering model. However, they struggle with the content. How would you refine your clustering approach?
   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   STUDENT LEARNING PATH PERSONALIZATION
   UNSUPERVISED LEARNING - CLUSTERING
   Data collection
   Collect student data such as quiz scores, engagement time, completion rates, interaction frequency, and learning pace.
   Data preprocessing
   Handle missing values, remove outliers, normalize features, and encode categorical variables.
   Feature selection & weighting
   Re-evaluate features used for clustering and assign appropriate weights so engagement metrics are not overshadowed by quiz scores.

Feature engineering
Create new features such as consistency of performance, time spent per topic, and improvement trends over time.
Clustering algorithm selection
Experiment with different clustering algorithms (K-Means, Hierarchical Clustering, DBSCAN) to find better student groupings.
Optimal cluster selection
Use methods like the Elbow method or Silhouette score to choose the optimal number of clusters.
Cluster validation
Analyze cluster characteristics to ensure students with low engagement are not incorrectly grouped into advanced paths.
Hybrid approach
Combine clustering with rule-based conditions or supervised models to refine learning path assignment.
Evaluation & feedback
Monitor student performance after placement and adjust clusters using feedback and outcomes.

6. Your model predicts unusually low rental prices for high-end properties in luxury neighborhoods. What factors could be missing from the dataset, and how would you improve your model?
Problem Identification:
Stage 1: Domain selection – Machine Learning
Stage 2: Learning selection – Supervised Learning
Stage 3: Regression/ Classification – Regression

HOUSE PRICE PREDICTION
SUPERVISED LEARNING = REGRESSION
Data collection
Collect house rental price, squarefeet area, location, amenities details
Data preprocessing
Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
Input output split
Feature selection
Train and test split
Model creation – try with all the regression algorithms to find the best model.
Model tuning -Tune hyperparameters to improve precision and reduce false prices.
Evaluation R2score, MSE, MAE

7. Your anomaly detection system raises an alert about a network traffic pattern, but IT specialists confirm it's a routine software update. How would you improve the system to reduce such false positives?
   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   ANAMOLY DETECTION SYSTEM
   SUPERVISED LEARNING – CLASSIFICATION (ANAMOLY/NOT)
   Data collection
   Collect the data of the software update and patches app details, anamoly detection details
   Data preprocessing
   Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
   Input output split
   Feature selection
   Train and test split
   Model creation – try with all the classification algorithms to find the best model.
   Model tuning -Tune hyperparameters to improve precision and reduce false alarms.
   Evaluation f1score, roc-auc score, accuracy, precision, recall
   Confusion matrix, classification report

8. A hospital's classification model struggles to differentiate between flu and COVID-19 because many symptoms overlap. What techniques can improve the model's ability to distinguish between these conditions?
   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   SUPERVISED LEARNING – CLASSIFICATION (COVID POSITIVE/NEGATIVE)
   Data collection
   Collect covid 19 symptoms data,
   Data preprocessing
   Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
   Input output split
   Feature selection
   Train and test split
   Model creation – try with all the classification algorithms to find the best model.
   Model tuning -Tune hyperparameters to improve precision and reduce false positives.

Evaluation f1score, roc-auc score, accuracy, precision, recall, Confusion matrix, classification report

9. Your ad engagement prediction model performs well on historical data but fails to generalize to new ad formats. What strategies can you use to improve its adaptability?

   Problem Identification:
   Stage 1: Domain selection – Machine Learning
   Stage 2: Learning selection – Supervised Learning
   Stage 3: Regression/ Classification – Classification

   SUPERVISED LEARNING – CLASSIFICATION (AD ENGAGED YES/NO)
   Data collection
   Collect the ad details and people preference whether clicked / not details , user interest shown / not details
   Data preprocessing
   Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
   Input output split
   Feature selection
   Train and test split
   Model creation – try with all the classification algorithms to find the best model.
   Model tuning -Tune hyperparameters to improve precision and reduce false positives.
   Evaluation f1score, roc-auc score, accuracy, precision, recall
   Confusion matrix, classification report

10. A recruiter tries to bypass the fake job detection model by slightly modifying fraudulent job descriptions. How would you make the model more robust against such adversarial tactics?

    Problem Identification:
    Stage 1: Domain selection – Machine Learning
    Stage 2: Learning selection – Supervised Learning
    Stage 3: Regression/ Classification – Classification

    SUPERVISED LEARNING – CLASSIFICATION (FAKE/REAL)
    Data collection
    Collect customer churn data
    Data preprocessing
    Identify and replace all the missing values and use one hot encoding and label encoding to convert all the categorical variables.
    Input output split
    Feature selection

Train and test split
Model creation – try with all the classification algorithms to find the best model.
Model tuning -Tune hyperparameters to improve precision and reduce false alarms.
Evaluation f1score, roc-auc score, accuracy, precision, recall
Confusion matrix, classification report