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?

   Identify the Problem type: Classification

   Logic:

- **Collect data:** Collect historical customer data, including call duration, complaints, and contract length
- **Preprocessing Data:** Handle missing values and encode categorical variables
- **Choose an algorithm:** Logistic Regression, Decision Tree, or Random Forest.
- **Evaluate Model:** Calculate the accuracy, precision, recall, and F1-score.
- **Make Prediction:** Deploy the model to identify customers at risk of churning.

---

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?

   Identify the Problem type: Regression

   Logic:

- **Collect data:** Collect past sales and inventory data for different products.
- **Preprocessing Data:** Handle missing values and normalize numerical values.
- **Choose algorithm:** Linear Regression, XG Boost, or Random Forest Regression.
- **Evaluate Model:** Evaluate using RMSE and Mean Absolute Percentage Error (MAPE).
- **Make Prediction:** Deploy the model for real-time demand forecasting.

---

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?

   Identify the Problem type: Classification

   Logic:

- **Collect data:** Collect the data for the transaction's history
- **Preprocessing Data:** Handle the missing values
- **Choose algorithm:** Use random forest, logistic regression, decision tree
- **Train model:** Train the model using the training data
- **Evaluate Model:** Use recall and f1 score.

- **Make Prediction:** Deploy the model to flag suspicious transactions in real-time

---

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?

   Identify the Problem type: Classification

   Logic:

- **Collect data:** Collect flight data, including departure times, weather conditions, and past delays.
- **Preprocessing Data:** Handle the missing values and normalize the numerical features.
- **Choose algorithm:** Use the Logistic Regression, Decision Tree, or XGBoost.
- **Train model:** Train the model using past flight delay records.
- **Evaluate Model:** Evaluate using accuracy, precision, and recall.
- **Make Prediction:** Deploy the model to predict flight delays in advance

---

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?

   Identify the Problem type: clustering

   Logic:

- **Collect data:** Collect the student score and engagement time as input data
- **Preprocessing Data:** Normalize the numerical values and encode categorical variables
- **Choose algorithm:** Use k-means clustering or DBSCAN
- **Evaluate Model:** Use Elbow method for the quiz score
- **Make Prediction:** Now the give the input value for the engagement time

---

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?

   Identify the Problem type: Regression

   Logic:

- **Collect data:** Get the data about the properties

- **Preprocessing Data:** Handle missing values and encode categorical variables.

- **Choose algorithm:** Linear Regression, Random Forest, or Gradient Boosting.
- **Train model:** Train the model using the training data
- **Evaluate Model:** Evaluate using RMSE and R^2 score
- **Make Prediction:** Redeploy the model and test the new values

---

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?

   Identify the Problem type: Anomaly Detection

   Logic:
- **Collect data:** Collect network traffic logs with normal and anomalous patterns.
- **Preprocessing Data:** Normalize numerical values and extract important features.
- **Choose algorithm:** Isolation Forest, One-Class SVM, or Autoencoders
- **Evaluate Model:** Evaluate using anomaly detection metrics like False Positive Rate
- **Make Prediction:** Deploy the model for real-time threat detection.

---

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?

   Identify the Problem type: classification

   Logic:
- **Collect data:** Get the extra data flu and covid-19
- **Preprocessing Data:** Handle the missing symptoms that which all are not overlap.
- **Choose algorithm:** Decision Tree, Random Forest, or Support Vector Machine (SVM).
- **Train model:** Train the model using test data and now splite the data for training in first 25%
- **Evaluate Model:** Evaluate using accuracy, precision, recall, and confusion matrix
- **Make Prediction:** Now give the different input from the new collected data.

---

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?

Identify the Problem type: Regression

Logic:

- **Collect data:** Get the response data from ad engagement
- **Preprocessing Data:** Normalize numerical values and encode categorical variables
- **Choose algorithm:** Gradient Boosting, Random Forest Regression, or XGBoost.

- **Evaluate Model:** Evaluate using RMSE and R^2 score
- **Make Prediction:** Now predict the future response

---

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?

Identify the Problem type: Classification

Logic:

- **Collect data:** Get the data from the recruiter job description, and his user credential.
- **Preprocessing Data:** Convert job descriptions into numerical format using TF-IDF or Word2Vec.
- **Choose algorithm:** Logistic Regression, Random Forest, or Neural Networks
  Train model: Train the model using training data
- **Evaluate Model:** Evaluate using accuracy, precision, recall, and AUC-ROC score
- **Make Prediction:** Give the input to find the given details of the fake job detection.