1. A bank wants to predict whether a loan applicant will default based on credit score, income, and past loan history. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to predict the loan applicant is default or not, it is a ***Classification*** problem.

**Steps to solve the problem:**

* **Data Collection**: Credit score, income and past loan history of customers should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier,RandomForest or logistic regression
* **Train the model:** Fit the model using labelled loan default data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict loan default by passing new applicant data to the model to be deployed

1. A retail store wants to predict the demand for different products to optimize inventory levels. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to predict the demand for each product with their sales history/purchase details, this falls under ***Regression Type***

**Steps to solve the problem:**

* **Data Collection**: Collect past sales data/purchase details of each product and demand for each product in the retail store
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** Linear Regression, SVM Regression, Decision Tree or Random Forest regression
* **Train the model:** Fit the model using input data(sales and demand data)
* **Model Performance**: Evaluate the model performance using metrics like r\_score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict Product Demand by passing future sales data

1. A factory wants to detect whether a manufactured product is defective based on sensor readings and quality control data. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to predict product is defective or not, this falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: sensor readings, Quality control data and defective status (yes or no) is collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest or logistic regression
* **Train the model:** Fit the model using labelled product defective status data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict product is defective or not by passing future sensor readings and quality control data of a product

1. A healthcare provider wants to analyze patient symptoms and classify them into different disease categories. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to Classify patient symptoms to different disease categories, it falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: Past patient symptoms and disease details should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest,XGboost or logistic regression
* **Train the model:** Fit the model using labelled disease data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict disease based on the patient symptoms data

1. An e-commerce company wants to identify and remove fake reviews posted by bots or fraudsters. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to predict review is fake or not, this falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: Past reviews posted by human and bots(fraudsters)
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest or logistic regression
* **Train the model:** Fit the model using labelled review data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict review is fake or not based on the review submitted

1. A financial firm wants to predict stock price movements based on historical price data and market indicators. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to predict stock price based on the price data and market indicators, this falls under *Regression Type*

**Steps to solve the problem:**

* **Data Collection**: Past price data and market indicators and their respective stock price details should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler.

Outliers should be handled by replacing mean, median or mode.

* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** Linear Regression, SVM Regression, Decision Tree regression or Random Forest Tree regression

**Train the model:** Fit the model using past input and output data(Price, market indicators and stock price)

* **Model Performance**: Evaluate the model performance using metrics like r score or rmse
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict stock price by passing future price data and market indicators

1. A social media platform wants to detect fake user accounts based on user activity and profile data. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to user is fake or not based on their activity and profile data, this falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: Past user activity and profile data with their respective account type (fake or real) should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest,logistic regression or XGBoost Classifier
* **Train the model:** Fit the model using labelled user data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict user is fake or not by passing current user activity and profile data

1. A marketing agency wants to segment customers into different groups based on their purchasing behavior. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to segment customers into different groups based on purchasing behavior, this falls under *Clustering Type*

**Steps to solve the problem:**

* **Data Collection**: Past Customer Purchase Data should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Algorithm Selection:** Kmeans, Meanshift or DBSCAN
* **Train the model:** Fit the model using customer purchasing data
* **Model Performance**: Evaluate the model performance using metrics like silhouette score
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Segment the customer based on the future purchasing data of customer. Check whether the clustering or segmentation of customer is predicted correctly based on the customer purchase data

1. A geospatial research team wants to analyze satellite images to classify different land types (forest, water, urban). What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need to classify different land types based on satellite images, it falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: Past Satellite images related to different land types should be collected
* **Data Pre-processing**: Normalize pixel values, rotate the image and extract image features
* **Algorithm Selection:** Decision Tree or CNN based models (like Resnet, YOLO etc.)
* **Split the Dataset:** Divide the data into training and test sets
* **Train the model:** Fit the model on labelled satellite images
* **Model Performance**: Evaluate the model performance using metrics like

Accuracy and confusion matrix

**Select Best Model**: Based on the metrics select the best model to deploy them

* **Make Predictions**: Predict/Classify the image type by passing the realtime satellite image captured

1. A streaming service wants to predict which users are likely to cancel their subscriptions. What type of ML problem is this, and what steps would you take to solve it?

**Problem type:**

As we need classify whether the user continue or cancel their subscriptions, this falls under *Classification Type*

**Steps to solve the problem:**

* **Data Collection**: Past user activity(time spent daily/weekly) and their subscription details including the subscription status should be collected
* **Data Pre-processing**: Missing values should be handled. If any standardization required that should be handled using StandardScaler. Categorical values should be converted to numerical values using encoding technique.
* **Split the dataset**: split the dataset into test and train data
* **Algorithm Selection:** DecisiontreeClassifier, RandomForest,logistic regression or XGBoost Classifier
* **Train the model:** Fit the model using labelled subscription data
* **Model Performance**: Evaluate the model performance using metrics like AUC\_ROC, F1\_score,Recall and precision
* **Select Best Model**: Based on the metrics select the best model to deploy them
* **Make Predictions**: Predict user would continue or cancel the subscription by passing user activity with subscription amount as input data