

**1. A real estate company wants to develop a system that predicts house prices based on square footage, number of bedrooms, and location.**

**Q: Identify the problem type and outline the step-by-step logic to solve it.**

Ans:

**Machine Learning Type:** Supervised Learning

**Task:** Regression

**Reason:** Output (house price) continuous numerical value.

- Collect house data (size, bedrooms, location, price)
- Clean & preprocess data
- Split data into train and test
- Choose a regression model
- Train the model
- Predict house prices
- Evaluate accuracy

This is a supervised regression problem where a model learns from past housing data to predict continuous house prices.

**2. A bank wants to build a model to detect fraudulent transactions by analyzing customer spending behavior and transaction history.**

**Q: Identify the problem type and outline the step-by-step logic to solve it.**

Ans:

**Machine Learning Type:** Supervised Learning

**Task:** Classification

**Reason:** Output is categorical → *Fraud or Not Fraud*

- Collect transaction data with fraud labels
- Preprocess & extract features
- Split data into train and test
- Choose a classification model
- Train the model
- Predict fraud / not fraud
- Evaluate performance.

*This is a supervised classification problem where the model detects fraud based on transaction behavior.*

**3. A supermarket wants to segment its customers based on their shopping patterns to provide personalized promotions.**

**Q: Identify the problem type and outline the step-by-step logic to solve it.**

Ans:

**Machine Learning Type:** Unsupervised Learning

**Task:** Clustering

**Reason:** No predefined labels; customers are grouped based on similar shopping behaviour.

- Collect customer shopping data
- Preprocess & select features
- Choose clustering algorithm
- Group customers into clusters
- Analyse clusters
- Send personalized offers.

This is an unsupervised clustering problem used to group customers for personalized marketing.

4. A company wants to estimate an employee's salary based on their years of experience, job title, and education level.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Regression**

**Reason: Salary is a continuous numerical value.**

- Collect employee data (experience, job title, education, salary)
- Preprocess data (clean, encode job title & education)
- Split data into train and test
- Choose a regression model
- Train the model
- Predict salary
- Evaluate performance.

*This model estimates salary based on employee features.*

5. An email provider wants to automatically classify incoming emails as spam or not spam based on their content and sender details.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Classification**

**Reason: Output is categorical → Spam / Not Spam**

- Collect labelled email data (spam / not spam)
- Preprocess emails (clean text, extract features)
- Split data into train and test
- Choose a classification model
- Train the model
- Classify new emails
- Evaluate accuracy.

*This model automatically filters spam emails.*

6. A business wants to analyze customer reviews of its products and determine whether the sentiment is positive or negative.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Classification (Sentiment Analysis)**

**Reason: Output is categorical → Positive / Negative.**

- Collect customer reviews with sentiment labels
- Preprocess text (clean, tokenize, remove stopwords)
- Convert text to features (TF-IDF / embeddings)
- Split data into train and test
- Choose a classification model
- Train the model
- Predict sentiment & evaluate results

*This model identifies whether reviews are positive or negative.*

7. An insurance company wants to predict whether a customer is likely to file a claim in the next year based on their driving history and demographics.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Classification (Claim Prediction)**

**Reason: Output is categorical → Claim / No Claim**

- Collect customer data (driving history + demographics + claim history)
- Preprocess data (handle missing values, encode categorical variables, scale features)
- Engineer features (e.g., accident frequency, risk scores)
- Split data into train and test sets
- Choose a classification model (Logistic Regression, Random Forest, XGBoost, etc.)
- Train the model on the training set
- Predict claims & evaluate results (Accuracy, Precision, Recall, F1-score, ROC-AUC)

*This model identifies whether a customer is likely to file a claim next year.*

8. A streaming platform wants to recommend movies to users by grouping them based on their viewing preferences and watch history.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Unsupervised Learning**

**Task: Clustering (Movie Recommendation)**

**Reason: No labeled output; goal is to group users with similar preferences.**

- Collect user data (watch history, ratings, genre preferences, watch frequency)
- Preprocess data (handle missing values, normalize/scale features)
- Transform data into features (e.g., user-movie matrix, embeddings)
- Choose a clustering method (K-Means, DBSCAN, Hierarchical, etc.)
- Cluster users based on similarity in viewing patterns
- Analyse clusters to identify user groups and preferred content types
- Recommend movies to each user based on their cluster's popular content

*This model groups users with similar tastes to provide personalized movie recommendations.*

9. A hospital wants to predict the recovery time of patients after surgery based on their age, medical history, and lifestyle habits.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Regression (Recovery Time Prediction)**

**Reason: Output is continuous → number of days until recovery**

- Collect patient data (age, medical history, lifestyle habits, past recovery times)
- Preprocess data (handle missing values, encode categorical features, scale numeric features)
- Engineer features (e.g., comorbidity score, activity level metrics)
- Split data into train and test sets
- Choose a regression model (Linear Regression, Random Forest Regressor, XGBoost Regressor, etc.)
- Train the model on the training set
- Predict recovery times & evaluate results (MAE, MSE, RMSE,  $R^2$ )

*This model predicts how long a patient is likely to take to recover after surgery.*

10. A university wants to predict a student's final exam score based on study hours, attendance, and past academic performance.

**Q:** Identify the problem type and outline the step-by-step logic to solve it.

Ans:

**Machine Learning Type: Supervised Learning**

**Task: Regression (Exam Score Prediction)**

**Reason: Output is continuous → final exam score**

- Collect student data (study hours, attendance, past academic performance, etc.)
- Preprocess data (handle missing values, encode categorical features, scale numeric features)
- Engineer features (e.g., average past grades, consistency metrics)
- Split data into train and test sets
- Choose a regression model (Linear Regression, Random Forest Regressor, XGBoost Regressor, etc.)
- Train the model on the training set
- Predict exam scores & evaluate results (MAE, MSE, RMSE,  $R^2$ )

*This model predicts a student's likely final exam score based on their study habits and academic history.*