

ML Scenario Question - Answers

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.

Logic:

Problem Identification:

- Stage 1 : Domain selection – Machine learning**
- Stage 2: Learning selection – Supervised learning**
- Stage 3: Regression/ Classification - Regression**

A: House price prediction

- Requirements are clearly specified
 - Output variable also present
 - As the data is labelled this is supervised learning. Since prediction is numerical values it is a regression problem.
- Data collection – importing the csv dataset into pandas,
 - Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding,
 - Input output split , train and test split
 - Model creation apply grid search cv for experimenting with different hyper tuning parameter
 - Evaluation – Use R2 score to evaluate the accuracy and performance
 - Prediction - Test the model on new data.

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.

Logic:

Problem Identification:

- Stage 1 : Domain selection – Machine learning**
- Stage 2: Learning selection – Supervised learning**
- Stage 3: Regression/ Classification - Classification**

A: Credit Card prediction

- Requirements are clearly specified
- Output variable also present
- As the data is labelled this is supervised learning. Since prediction is categorical values it is a Classification problem.

- Data collection – importing the csv dataset into pandas,
- Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding,
- Input output split , train and test split
- Feature selection

- **Model creation apply grid search cv for experimenting with different hyper tuning parameter**
 - **Evaluation – Use Confusion Matrix to evaluate the accuracy and performance**
 - **Prediction - Test the model on new data.**
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.
- A: user based recommendation**
- **Data collection – Collect user data importing the csv dataset into pandas,**
 - **Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding, create pivot table for matrix index**
 - **Compute similarity scores**
 - **Evaluation – Use metrics like Precision, Recall, RMSE, or MAE.**
 - **Prediction – Recommend top-N items for each user.**
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.
- A: supervised learning regression**
- Logic:**
- Problem Identification:**
- Stage 1 : Domain selection – Machine learning**
 - Stage 2: Learning selection – Supervised learning**
 - Stage 3: Regression/ Classification - Regression**
- A: Employee Salary prediction**
- **Requirements are clearly specified**
 - **Output variable also present**
 - **As the data is labelled this is supervised learning. Since prediction is numerical values it is a regression problem.**
- **Data collection – importing the csv dataset into pandas,**
 - **Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding,**
 - **Input output split, train and test split**
 - **Model creation apply grid search cv for experimenting with different hyper tuning parameter**
 - **Evaluation – Use R2 score to evaluate the accuracy and performance**
 - **Prediction - Test the model on new data.**
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.

A: supervised learning classification

Problem Identification:

- Stage 1 : Domain selection – Machine learning**
- Stage 2: Learning selection – Supervised learning**
- Stage 3: Regression/ Classification - Classification**

A: Spam Classification prediction

- Requirements are clearly specified
 - Output variable also present
 - As the data is labelled this is supervised learning. Since prediction is categorical values it is a Classification problem.
- Data collection – importing the csv dataset into pandas,
 - Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding,
 - Input output split , train and test split
 - Feature selection
 - Model creation apply grid search cv for experimenting with different hyper tuning parameter
 - Evaluation – Use Confusion Matrix to evaluate the accuracy and performance
 - Prediction - Test the model on new data.

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.

A: Natural Language Processing

Logic:

Problem Identification:

- Stage 1 : Domain selection – Natural Language Processing**
- Stage 2: Learning selection – Supervised learning**
- Stage 3: Regression/ Classification - Classification**

A: House price prediction

- Requirements are clearly specified
 - Output variable also present
 - As the data is text based it is NLP. And it has labelled data this is supervised learning. Since prediction is categorical values it is a Classification problem.
- Data collection – importing the csv dataset into pandas,
 - Data preprocessing – preprocessing the data –
 - Removing all punctuations, hastags, urls, emojis, Text tokenization,
 - converting text into vectors using count vectorizer and tf-idf vectorizer
 - Input output split , train and test split
 - Model creation using ML based algorithm Naïve bayes is best for NLP.
 - Evaluation – Use Confusion Matrix to evaluate the accuracy and performance
 - Prediction - Test the model on new data.

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.

A: time series

Question depends on time parameter so time series

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.

A: Item based

Recommend movies to users based on movies they have already watched or liked (watch history) here user details not included

- **Data collection – Collect user-movie data (watch history, metadata, likes) importing the csv dataset into pandas,**
- **Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding, create pivot table for matrix index**
- **Compute similarity scores**
- **Evaluation – Use metrics like Precision, Recall, RMSE, or MAE.**
- **Prediction – Recommend top-N items for each user.**

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.

A: Time series (as it deals with time data)

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.

Logic:

Problem Identification:

- Stage 1 : Domain selection – Machine learning**
- Stage 2: Learning selection – Supervised learning**
- Stage 3: Regression/ Classification - Regression**

A: Exam Score prediction

- **Requirements are clearly specified**
- **Output variable also present**
- **As the data is labelled this is supervised learning. Since prediction is numerical values it is a regression problem.**

- **Data collection – importing the csv dataset into pandas,**
- **Data preprocessing – preprocessing the data – removing null, na values, converting categorical data into numbers using one hot or label encoding,**
- **Input output split , train and test split**
- **Model creation apply grid search cv for experimenting with different hyper tuning parameter**
- **Evaluation – Use R2 score to evaluate the accuracy and performance**
- **Prediction - Test the model on new data.**