

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | |
| D | D | M | M | Y | Y | Y |

Assignment-1

Q1] Describe a real-world problem that could be addressed using ML, Explain which type of ML would be appropriate for this problem why

⇒

Real world problem: Detection of exoplanet

Exoplanet are planets that orbit star outside our solar system. Detecting and classifying exoplanet is crucial for understanding planetary system and the potential for life beyond Earth.

The appropriate ML : Supervised.

Reason :

There are existing data sets with labeled examples of light curves, where dips caused by exoplanets have been manually identified by astronomers. These labeled dataset are crucial for training supervised learning model.

Q2] Compare and contrast supervised, unsupervised and reinforcement learning. Provide example of algorithm for each type and discuss scenarios where one type might be more suitable than other

⇒ ① Supervised:

Algorithms:

- Linear regression
- Logistic regression
- Support vector Machine.
- Decision tree & Random forest
- Neural network based Algos like (CNN)

Use Cases

- Image classification
- Spam detection
- Medical Diagnosis
- Exponent Detection.

Pros

- High accuracy when trained on sufficient and representative data.
- Provides clear performance metric like accuracy, precision and recall

Cons

- Requires a large amount of labeled data.
- May not be useful in new type of data.

② Unsupervised:

Algorithms:

- K-Means clustering
- Hierarchical clustering
- Principal Component analysis

Use cases :

- Customer segmentation
- Anomaly Detection
- Dimensionality Reduction

Pros :

- Can work with unlabeled data
- Useful for finding patterns in new type of data.

Cons :

- Result can be harder to interpret.
- No clear metric for evaluation.

⑤ Reinforcement learning

Algorithms :

- Q-Learning
- Deep Q-Learning
- Policy Gradient method.

Use cases :

- Game Playing
- Robotics.
- Autonomous vehicles.

Pros → Can solve complex decision-making problem

- Suitable for environment where the optimal strategy can be learned through interaction.

Cons → Requires a large number of interactions, which can be time-consuming and computationally expensive

- Difficult to design heuristic function that directs to desired

Q3] Outline the steps involved in a ML model for house price prediction.

-
- ① Data collection: collect the data for the house and their price by field visit or use secondary data.
 - ② Preprocessing: Handle missing values, normalize the numerical features etc.
 - ③ Model selection: choosing appropriate model is very important, in this case we will choose multiple linear regression.
 - ④ Model training: Divide the labeled data in ratio (80-20) (training-testing) and start training it.
 - ⑤ Model evaluation: Once it is trained, validate its accuracy and use the remaining 20% data, we use various evaluation matrix like MAE, RMSE etc.