Practical-1

Aim: Study of Machine learning basics

In this lab, we will go through the basics of machine learning. The student needs to make a soft copy note on the following topics:

Topics of be studied:

1. What is Machine learning

   Machine learning is a subset of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed.

2. Steps in collection of data

   - Define the problem and objectives
   - Identify data sources
   - Collect relevant data
   - Data quality check

3. Steps in importing the data in python (Through: csv, json, and other data formats)

   Through CSV:

   ```
   import pandas as pd

   # Load CSV file
   data = pd.read_csv('your_data.csv')
   ```

   Through JSON:

   ```
   import pandas as pd

   # Load JSON file
   data = pd.read_json('your_data.json')
   ```

   Other Data Formats
   Pandas supports various data formats like Excel, SQL, etc.

4. Preprocessing
    1. Remove Outliers

        ```
        # Example code using Z-score
        from scipy import stats

        z_scores = stats.zscore(data)
        filtered_data = data[(z_scores < 3).all(axis=1)]
        ```

    2. Normalize Datasets, Data encoding

        ```
        # Example code for Min-Max normalization
        from sklearn.preprocessing import MinMaxScaler

        scaler = MinMaxScaler()
        normalized_data = scaler.fit_transform(data)
        ```

    3. Handling Missing Data

        ```
        # Example code for filling missing values with mean
        data.fillna(data.mean(), inplace=True)
        ```

5. Machine Models
    1. Types of machine learning models – Supervised learning, Unsupervised learning, reinforcement learning.

        Supervised Learning: Regression, Classification
        Unsupervised Learning: Clustering, Dimensionality Reduction
        Reinforcement Learning: Learning from interaction with the environment

    2. Parameters of machine learning model (Learning rate, regularization, etc.)

        - Learning Rate
        - Regularization
        - Hyperparameters specific to each model

6. Test-train data split: using constant ration, k-fold cross validation

    - Using a Constant Ratio (e.g., 80% training, 20% testing)
    - K-Fold Cross-Validation

7. Output Inference

    - Making predictions on new data using the trained model.

8. Validation: different metrics – Confusion Matrix, Precision, Recall, F1-score

- Confusion Matrix: Describes model classification results.
- Precision: Proportion of true positives to the total predicted positives.
- Recall: Proportion of true positives to the total actual positives.
- F1-Score: Harmonic mean of precision and recall.