|  |  |
| --- | --- |
| **Name: MAHENDRA SINGH** | **Roll No.: 22/AI/29** |

**ASSIGNMENT**

**POORNIMA COLLEGE OF ENGINEERING, JAIPUR**

**III B.TECH. (VI Sem.) SEC- D**

**Code: 6CAI6-02**

**Subject Name–Machine Learning**

**(BRANCH: ADVANCE COMPUTING (AI))**

**Max. Time: 2 hrs. Max. Marks: 20 Marks**

**INSTRUCTIONS: UPLOAD THE SOLUTION ON YOUR GITHUB REPOSITORY and MENTIONED THE URL OF THE REPOSITORY ON TCSION**

**ASSIGNMENT QUESTION 1: CO3**

**Fault Prediction Using Supervised Machine Learning**

**Problem Context**

**You are an engineer working for a power distribution company responsible for maintaining and ensuring the reliability of the electrical grid. Your task is to develop a system for detecting and classifying electrical faults in the grid. Electrical faults can lead to disruptions, damage equipment, and pose safety hazards. The company is interested in a predictive maintenance system that can identify and classify different types of electrical faults to facilitate timely intervention.**

**Fault Prediction Dataset:** [**https://www.kaggle.com/code/pythonafroz/fault-prediction-usingdecision-tree-algorithm**](https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decision-tree-algorithm)

|  |  |  |
| --- | --- | --- |
| S/r No. | Question | Marks |
| Q1. | Name any 4 libraries required for the implementation of the problem statement using python | 2 |
| Ans1. |  Pandas – for data manipulation and analysis   NumPy – for numerical computations   Matplotlib – for data visualization   sklearn – for machine learning algorithms and utilities |  |
| Q2. | Go through the above Kaggle link and answer the following: About this dataset file:  [https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decisiontree-algorithm](https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decision-tree-algorithm)   1. Total no. of columns in the dataset: 6 2. Write and count input columns: 5 ['Phase Current', 'Voltage', 'Frequency', 'Power Factor', 'Temperature'] 3. Write and count the output column: 1 ['Fault Type'] | 3 |
| Q3. | What is the purpose this library used for the given problem statement:   |  | | --- | | from sklearn.preprocessing import LabelEncoder | | 1 |
| Ans3. | LabelEncoder is used to convert categorical labels into numerical values so that they can be used in machine learning algorithms which require numerical input. |  |
| Q4. | What is the purpose this library used for the given problem statement:   |  | | --- | | from sklearn.model\_selection import train\_test\_split | | 1 |
| Ans4. | train\_test\_split is used to divide the dataset into training and testing sets. This helps evaluate the performance of the machine learning model on unseen data. |  |
| Q4. | List all the algorithms through which you can able to find Electrical Faults Detection and Classification | 3 |
| Ans4. |  Decision Tree   Random Forest   K-Nearest Neighbors (KNN)   Support Vector Machine (SVM)   Logistic Regression |  |
| Q5. | How to read the Classification\_Report generated by several mod els in the given problem statement:  [https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decisiontree-algorithm](https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decision-tree-algorithm) | 5 |
| Ans5. | The classification\_report includes the following metrics:   * **Precision**: True Positives / (True Positives + False Positives) * **Recall**: True Positives / (True Positives + False Negatives) * **F1-score**: Harmonic mean of precision and recall * **Support**: Number of actual occurrences of each class   These metrics help evaluate how well the model performs for each class. |  |
| Q6. | From the mentioned link: [https://www.kaggle.com/code/pythonafroz/faultprediction-using-decision-tree-algorithm](https://www.kaggle.com/code/pythonafroz/fault-prediction-using-decision-tree-algorithm)  Do one sight analysis and figure out which algorithms work well on the given dataset. And on what basis are Model comparisons done over there? | 5 |
| Ans6. | From the analysis at the Kaggle link, **Decision Tree** performs well with high accuracy and clarity in interpretation.   * **Basis of model comparison**:   + Accuracy Score   + Confusion Matrix   + Classification Report (Precision, Recall, F1-score) |  |

**ASSIGNMENT QUESTION 2: CO4**

**Customer Segmentation using Unsupervised Problem Context:**

|  |  |
| --- | --- |
| **Name: MAHENDRA SINGH** | **Roll No.: 22/AI/29** |

**You are a data scientist working for a retail company that wants to improve its marketing strategies by better understanding customer behaviour. One approach is to segment customers into distinct groups based on their purchasing habits. This will allow the company to tailor marketing campaigns to specific groups, ultimately increasing sales and customer satisfaction.**

**Task:**

**Design and explain the customer segmentation model using any one of the unsupervised algorithms.**

1. Data Collection: Obtain a dataset containing customer purchase history, including details such as purchase frequency, amount spent, types of products purchased, etc. You may use publicly available datasets or simulate data for this assignment Include the first 10 rows of the dataset that you are going to consider. 5 marks

1. Data Preprocessing: Clean the dataset and perform necessary preprocessing steps such as normalization, handling missing values, and feature engineering. 5 marks
2. Unsupervised Learning (Clustering): Apply an unsupervised learning algorithm (e.g., Kmeans clustering, hierarchical clustering) to segment customers into distinct groups based on their purchasing behaviour. 5 marks
3. Evaluate the clustering results using appropriate evaluation metrics. 5 marks

