|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Programme** | | **Computer Science & Engineering** | **Semester** | | **V** | |
| **Course** | | **Artificial Intelligence & Machine Learning** | **Max Marks** | | **30** | |
| **Course Code** | | **20CS51I** | **Duration** | | **4 hours** | |
| **Name of the course coordinator** | | **Mrs. Nagaveni Kadakol** |  | |  | |
| Note: Answer one full question from each section. | | | | | | |
| **Q.No** | **Questions** | | **CL**  **L3/L4** | **CO** | **PO** | **Marks** |
| **Section 1(Theory) 10 Marks** | | | | | | |
| 1.a) | Compare Different types of Data in data collection | | L4 | 2 | 1 | 5 |
| b) | Plot the following for drug.csv   1. Plot a bar graph for the side effects and its frequencies and determine which has the highest frequency. 2. Plot a bar graph for the effectiveness and determine which effectiveness is the highest in the data. | | L3 | 2,5 | 4 | 5 |
| 2.a) | The statistical summary of Iris dataset is as follows.    Analyse statistical metrics from above summary. | | L4 | 2,5 | 2,3,4 | 5 |
| b) | Given the Air Passengers datasets perform the following operations.   1. Read the First six months of (Jan-Jun) data into the first\_half\_df dataframe. 2. Read the next six months of data into the second\_half\_df dataframe. 3. Create a new dataframe air\_df from the dataframes first\_half\_df and second\_half\_df by inner join on their respective indices. 4. Fill any missing values with median strategy in the air\_df. 5. Create a column Year with values from 1949 to 1960 (both inclusive). 6. Find if there is any column with the number of passengers in Jan are greater than the number of passengers in Dec. Return all such rows. | | L3 | 2,5 | 2,3,4 | 5 |
| **Section 2 (Practical) 20 Marks** | | | | | | |
| 3. | A company wants to study the demographic data to make predictions about the earning potential of the population. However, the data gathered is not clean for analysis.  The company requests you, as a data scientist, to perform the following operations and gain some insights from the data for data driven competitive advantage.   * Remove data with missing values * Remove outliers * Establish the importance of the weekly working hours on earning potential * Find the features that are highly correlated with the earning potential * Find the relation between the number of years spent to get the degree and earning potential * Find the relationship between age and earning potential | | L3,  L4 | 2,5 | 2,3,4 | 20 |
| 4. | Perform and analyze univariate, bivariate, and multivariate analysis on iris.csv data set | | L3,  L4 | 2,5 | 2,3,4 | 20 |