|  |  |
| --- | --- |
|  | Perform the following operations using Python on Iris data set.   1. Load the Dataset into pandas data frame. 2. Display information about missing values in the data 3. Display initial statistics. 4. Check the dimensions of the data frame. 5. Display data type of the variable. 6. Apply proper data type conversion 7. Convert categorical variables into quantitative variables using one hot encoding and label encoder |
|  |  |
|  | Perform the following operations using Python by creating student performance dataset.   1. Display Missing Values 2. Replace missing values using any 2 suitable 3. Identify outliers using boxplot and scatterplot 4. Handle outlier using any technique 5. Perform any 2 data normalization technique |
|  |  |
|  |  |
|  |  |
|  | Perform the following operations on iris dataset   1. Display mean, median, minimum, maximum, standard deviation 2. Provide mean, median, minimum, maximum, standard deviation for a given dataset by grouping using one of the qualitative (categorical) variable 3. Display missing values and inconsistencies. 4. Replace missing values using any 2 suitable |
|  | Perform the following operation using titanic data set.   * 1. check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram. |
|  | 1. plot a box plot for distribution of age with respect to each gender along with the information about whether they survived or not. (Column names : 'sex' and 'age') 2. Write observations on the inference from the above statistics. |
|  | Perform the following operations on iris dataset   * 1. List down the features and their types   2. Create a box plot and histogram for each feature in the dataset.   3. Compare distributions and identify outliers. |
|  |  |
|  | Create a Linear Regression Model using Python/R to predict home prices using Boston Housing Dataset. Find the performance of your model. |
|  | Create a   logistic    regression model  on social network ads.csv    to    perform    classification    on given dataset. Compute Confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall . |
|  | Create a Naïve Bayes classification model using Python on on social network ads.csv   dataset. Compute Confusion matrix to find TP, FP, TN, FN, Accuracy, Error rate, Precision, Recall on the given dataset. |
|  |  |
|  | For given text apply following preprocessing methods:   * 1. Tokenization   2. POS Tagging   3. Stop word Removal   4. Lemmatization   5. Stemming |
|  |  |
|  |  |
|  | Calculate Term Frequency and Inverse Document Frequency. Considering sentences of documents. |
|  | Write Scala program to find average temperature, average dew point and average wind speed for given weather dataset |
|  |  |
|  | Perform the following operations using Python by creating student performance dataset.   * 1. Display Missing Values   2. Replace missing values using any 2 suitable   3. Identify outliers using IQR and ZScore   4. Handle outlier using any technique   5. Perform data normalization using Min Max |

|  |  |
| --- | --- |
|  | Perform the following operations using Python by creating student performance dataset.   * + 1. Display Missing Values     2. Replace missing values using any 2 suitable     3. Identify outliers using IQR and ZScore   4.Handle outlier using any technique  5.Perform data normalization using decimal scaling |

|  |  |
| --- | --- |
|  |  |
|  | For given text apply following preprocessing methods:  1. Tokenization   1. POS Tagging 2. Stop word Removal |
|  |  |