

List of Data Science Programs

Python Programs

1. Write a program to get the largest number from a list.
2. Write a program to remove duplicates from a list.
3. Write a program to convert a tuple to a dictionary.
4. Write a program to merge two python dictionaries.
*Hint - Use `update()` method or `** operator` or `/ operator`*
5. Write a program that takes two lists and returns true if they have at least one common member.
6. Write a program to determine which one is the earlier date from the two given dates.
7. Write a program to subtract 5 days from current date.
8. Write a program to open a file and copy the contents to another file.
9. Write a program to capitalise each word in a file.
Hint - Use `str.upper()` method
10. Write a program to search a word in a file and replace with another word.
Hint - Use `str.replace("old text", "new text")` method
11. Write a program to count number of lines in a file.
12. Write a program to retrieve lines having two consecutive 1's.
13. Write programs to perform the following matrix operations using vectorisation. (Use matrices of order 3 X 3)
 - (a) addition
 - (b) subtraction
 - (c) multiplication
 - (d) scalar multiplication
 - (e) transpose
14. Write a program to create the following matrices for various 2D geometric transformations.
 - (a) translation matrix
 - (b) rotation matrix
 - (c) scaling matrix

15. Write a program to create the following matrices for various 3D geometric transformations.
 - (a) translation matrix
 - (b) rotation matrix
 - (c) scaling matrix
16. Write a program to perform SVD (Singular Value Decomposition) of a square matrix of order 3. Reconstruct the original matrix from the components.
17. The marks obtained by students in a class are given below.

22,87,5,43,56,73,55,54,11,20,51,5,79,31,27

Write a program to draw a histogram of these marks for intervals 0 - 10, 10 - 20,..., 90 - 100.

18. Write a program to draw a histogram of sepal length in the iris data set (given).
19. Write a program to draw a scatterplot that shows the relationship between rollnos and marks of students (given below) in a class.

rollnos = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15]

marks = [22,87,5,43,56,73,55,54,11,20,51,5,79,31,27]
20. Write a program to draw a scatterplot that shows the relationship between sepal length and sepal width in the iris data set (given).

R Programs

21. Given a data set of 15 food items having 4 features - ingredient, sweetness, crunchiness and food type. Write a program to predict the food type of tomato using k-NN algorithm.

data set - food.csv
22. As per a survey conducted in an institution, students are classified based on the two attributes of academic excellence(X) and other activities(Y). Write a program to identify the classification of a student with X = 5 and Y = 7 using k-NN algorithm(choose k as 3).

data set - survey.csv
23. Given a data set containing 569 examples of breast cancer biopsies, each having 32 features. Write a program to predict the result of cancer biopsies using k-NN algorithm and analyse the same.

data set - wisc_bc_data.csv
24. Given the following data on a certain set of patients seen by a doctor. Can the doctor conclude that a person having chills, fever, mild headache and without running nose has flu? Use Naive Bayes classification.

data set - symptoms.csv
25. Use Naive Bayes classification to determine whether a red domestic SUV car is a stolen car or not using the following data set.

data set - cars.csv

26. Write a program to predict the species of the iris data set(given) using Naive Bayes Classification.
27. Write a program to determine whether a person cheats using C5.0 Decision Tree Algorithm and evaluate its performance. Given a data set of 10 items, each having 5 features.
data set - person.csv
28. Write a program to determine whether play of cricket is possible depending on the weather condition using C5.0 Decision Tree Algorithm and evaluate its performance. Given a data set of 14 items, each having 5 features.
data set - cricket.csv
29. Write a program to identify risky bank loans using C5.0 Decision Tree Algorithm and evaluate its performance. Given a data set of 1000 customers, each having 17 features.
data set - credit.csv
30. Write a program to predict the Stock Index Price (the dependent variable) of a fictitious economy based on two independent variables Interest Rate and Unemployment Rate using Multiple Linear Regression Technique and evaluate its performance. Given a data set of 24 items.
data set - economy.csv
31. Write a program to predict the percentage of heart disease (the dependent variable) based on two independent variables, percentage of people biking to town and percentage of people smoking using Multiple Linear Regression Technique and evaluate its performance. Given a data set of 498 items.
data set - heart.csv
32. Write a program to predict medical expenses (the dependent variable) based on the independent variables, age, sex, bmi, children, smoker and region using Multiple Linear Regression Technique and evaluate its performance. Given a data set of 1338 items.
data set - insurance.csv
33. Write a program to partition the iris data set(given) into different clusters using k-means clustering algorithm. Choose k as 3. Check clustering result against species class label.
34. Given a data set containing 569 examples of breast cancer biopsies, each having 32 features. Write a program to partition this data set into different clusters using k-means clustering algorithm. Choose k as 2. Check clustering result against diagnosis class label.
data set - wisc_bc_data.csv