15. The daily temperature of a city over a week are as folllows

    Monday=30

    Tuesday=28

    Wednesday=31

    Thursday=33

    Friday=29

    Saturday=32

    Sunday=34

    1. Compute mean and standard deviation
    2. Normalize using Z score normalization
    3. Identify which day has the most normalized temperature

[CO1,Evaluate]

## Part B

**Answer any 2 questions from the bunch of CO2. Each question carries 6 mark.** **(6x2=12)**

16. State the advantages of supervised learning algorithms [CO2,Analyse]
17. State the advantages of unsupervised learning algorithms [CO2,Analyse]
18. List three real-world applications of Unsupervised learning with brief explanations [CO2,Evaluate]

## Part B

**Answer any 2 questions from the bunch of CO3. Each question carries 6 mark.** **(6x2=12)**

19. Discuss different distance and similarity measures. [CO3,Apply]
20. Compare and contrast Hierarchical and Density-based clustering methods. [CO3,Apply]
21. Compare and contrast partitioning and grid-based clustering methods. [CO3,Apply]

## Part C

**Answer any 1 question from the bunch of CO4. Each question carries 10 mark.** **(10x1=10)**

22. Find the association rule for the following Database D using Apriori algorithm.

    Min support = 2 Min. confidence = 50%

    Database D

| TID | Items |
|---|---|
| T1 | I1, I2, I5 |
| T2 | I2, I4 |
| T3 | I2,I3 |
| T4 | I1,I2,I4 |
| T5 | I1,I3 |
| T6 | I2,I3 |
| T7 | I1,I3 |
| T8 | I1,I2,I3,I5 |
| T9 | I1,I2,I3 |

[CO4,Evaluate]

TB253232L

Reg. No :.....

Name :.....

**BACHELOR'S DEGREE EXAMINATION, OCTOBER 2025**

**2024 ADMISSIONS REGULAR**

**SEMESTER III - MAJOR (COMPUTER APPLICATIONS )**

**CA3DSEA01B24 - Principles of Data Science**

**Time : 2 Hours** **Maximum Marks : 70**

**Part A**

**Answer all questions from the bunch of CO1. Each question carries 2 mark.** **(2x3=6)**

1. Find the mean and median of the following data 23,34,54,65,20,100,31. [CO1,Evaluate]
2. Define Outlier [CO1,Remember]
3. Define an Attribute. [CO1,Remember]

**Part A**

**Answer all questions from the bunch of CO2. Each question carries 2 mark.** **(2x3=6)**

4. Define Non Parametric method in statistics. [CO2,Understand]
5. Give one example each of a parametric and a non-parametric algorithm [CO2,Apply]
6. Which type of model generally requires more training data: parametric or non-parametric? [CO2,Apply]

**Part A**

**Answer all questions from the bunch of CO3. Each question carries 2 mark.** **(2x3=6)**

7. Define the term "hyperplane" in SVM. [CO3,Understand]
8. What does the "k" in k-means clustering represent? [CO3,Analyse]
9. Name any two clustering algorithms [CO3,Remember]

**Part A**

**Answer all questions from the bunch of CO4. Each question carries 2 mark.** **(2x3=6)**

10. What are support and confidence in association rule mining? [CO4,Understand]
11. State Apriori Property [CO4,Remember]
12. Write the formula for confidence in association mining. [CO4,Remember]

**Part B**

**Answer any 2 questions from the bunch of CO1. Each question carries 6 mark.** **(6x2=12)**

13. The following contingency table summarizes supermarket transaction data, where hot dogs refers to the transaction containing hot dogs, No hot dog refers to the transaction that do not contain hot dogs. Hamburgers refers to the transaction containing hamburgers , No hamburgers refers to the transaction that do not contain hamburgers. Generate an Expected Frequency table.

| | Hot Dogs | No Hot Dogs | Total |
|---|---|---|---|
| Hamburgers | 2000 | 500 | 2500 |
| No Hamburgers | 1000 | 1500 | 2500 |
| Total | 3000 | 2000 | 5000 |

[CO1,Evaluate]

14. Given the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Use smoothing by bin means to smooth the above data, using a bin depth of 3, Illustrate your steps. [CO1,Evaluate]