School of Computer Science Engineering and Technology

Course- B Tech Type- Core

Course Code- Course Name- Statistical Machine learning

Year- 2024-2025 Semester- odd

Date – 01-08- 2024 Batch- ALL

<https://www.kaggle.com/datasets/camnugent/california-housing-prices>

Download the dataset from the above link.

Illustrate different functions of machine learning frameworks pandas, numpy, matplot lib in data analysis and visualization.

Question 1:

1. Given the following dataset: X={2.3,2.5,3.6,2.8,3.1,2.9,3.2,2.7,2.8,3.0}. Compute the density of the data using a histogram with 5 bins.

Question 2:

|  |  |
| --- | --- |
| (dependent variables) | (independent variables) |
| 4 | 5 |
| 5 | 6 |
| 8 | 3 |
| 2 | 8 |
| 4 | 3 |
| 2 | 7 |
| 5 | 8 |

Find Mean, Median, Mode , Variance , Standard deviation , Range , Interquartile Range (IQR), Skewness ,Kurtosis of x and y .

Question 3:

From the above data Find Mean, Median, Mode, Variance, Standard deviation, Range , Interquartile Range (IQR), Skewness ,Kurtosis of each feature.