Question 2 Marks

1) Write about types of attributes.
2) What are some of the key features of multivariate analysis, and how can it be used in data analytics?
3) How can time series analysis be used to uncover patterns in data over time?
4) What are some applications of data analytics in real-world scenarios?
5) Why is the data analytics lifecycle important?
6) List various phases of the data analytics life cycle.
7) What do you mean by supervised learning?
8) Write the difference between precision and recall.
9) Define confusion matrix.
10) Write the difference between nominal and numerical data types.
11) Define noise and outlier in data analytics.
12) Describe the importance of data analytics in business.
13) What are the different categories of data?
14) List the name of any four python libraries used in data analysis.
15) What and why Bigdata and give the name of any three big data tools.
16) Compare Data Mining and Data Analytics with examples of each.
17) What is the significance of Correlation?
18) What is the significance of regression?
19) Write the difference between Regression and Classification.
20) Find the mean, median, and mode.
    82,77,32,88,87,32,37,47,87
    82,77,32,88,87,32,37,47
21) Write the relation between mean, median, and mode.
22) Write the difference between underfitting and overfitting.
23) What do you mean by 4V's in big data?

Questions 5 Marks

1) Illustrate and explain the steps involved in Bayesian data analysis.
2) What are the advantages and disadvantages of using structured, semi-structured, and unstructured data in data analytics projects?
3) What are the main characteristics of data that are relevant to data analytics, and how can these characteristics impact the analysis process?
4) How can data analytics be used in healthcare to improve patient outcomes and reduce costs?
5) What are the key stages of the data analytics lifecycle, and how do they contribute to the success of the project?

6) What is a prediction error? State and explain the prediction error in regression and classification with suitable examples.
7) How business industry uses data analytics to gain competitive advantages?
8) Write the difference between mean absolute error and mean squared error.
9) Compare the following.
   Simple Vs Multiple Regression
   Logistics Vs Polynomial regression.
10) Discuss different types of Data Analytics
11) Discuss Bigdata and its characteristics.

Question 10 Marks

1. Given below are the monthly income and their net savings of a sample of 10 supervisory staff belonging to a firm. Calculate the correlation coefficient.

   Monthly income(Rs.):780 360 980 250 750 820 900 620 650 390

   Net Savings: 84 51 91 60 68 62 86 58 53 47

   r= 0.78 approx.

2. The decrease in heart rate (beats/min) due to different concentrations of a drug are given below. Find the equation for the regression using the method of least square

Dose of drug(mg)	1.0	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
Decrease in heart beat	10	12	12	14	16	17	20	18	21

3. For Data

PlayTennis: training examples

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Classify the tuple(**x**')=(Outlook=*Sunny,* Temperature=*Cool,* Humidity=*High,* Wind=*Strong*)

4. Explain the role of time series analysis in the stock market. Also, explain its components with suitable examples.
5. Explain how Support Vector Machine can be used for classification and regression both.
6. Explain the knowledge discovery database with a suitable diagram.
7. Explain various phases of the Data Analytics Life Cycle with a suitable diagram.
8. How can Bayesian modelling and inference be used to make predictions about future events?