22MBABA304

USN

Third Semester MBA Degree Examination, Dec.2023/Jan.2024
Exploratory Data Analysis for Business

Time: 3 hrs.   Max. Marks: 100

Note: 1. Answer any FOUR full questions from Q.No.1 to Q.No.7.
2. Question No. 8 is compulsory.
3. M : Marks , L: Bloom's level , C: Course outcomes.

			M	L	C
Q.1	a.	Write the applications in Data Mining.	3	L1	CO1
	b.	Using the example of email span filtering, describe the typical steps involved in a data mining process.	7	L2	CO1
	c.	Explain the importance of data visualization in Exploratory Data Analysis (EDA) and provide examples of tools used for visualizing relationship between two variables.	10	L3	CO1
Q.2	a.	Define Prediction Error.	3	L1	CO2
	b.	Explain the holdout method for Cross – Validation and its limitations.	7	L2	CO2
	c.	Evaluate different cross validation techniques to split the sample into multiple training and test data sets.	10	L4	CO2
Q.3	a.	What is Linear Regression?	3	L1	CO3
	b.	Explain the difference between Point estimates and Interval estimates in linear regression.	7	L2	CO3
	c.	Discuss the advantages and disadvantages of different variables selection techniques like forward selection and backward selection.	10	L3	CO3
Q.4	a.	Define Singular Value Decomposition (SVD) and its role in Principal Components Analysis (PCA).	3	L1	CO4
	b.	Illustrate Bayes classification rule with an example.	7	L2	CO4
	c.	Explain the process and steps in construction of a decision tree.	10	L5	CO4
Q.5	a.	Define the concept of shrinkage in regression.	3	L1	CO5
	b.	Evaluate the performance of metrics logistic regression.	7	L4	CO5
	c.	Write a short note on Discriminant Analysis and Logistic Regression for classification.	10	L1	CO5

Q.6	a.	Differentiate between Linear and Non – linear Support Vector Machine (SVM).	3	L1	CO6
	b.	Explain the Logistic Regression with suitable examples.	7	L2	CO6
	c.	Compare and contrast the squared loss for Ridge Regression and Lasso Regression methods.	10	L4	CO6
Q.7	a.	Define Exploratory Data Analysis (EDA) and its significance in data mining.	3	L1	CO1
	b.	Give an overview of Support Vector Machines (SVM) and their applications.	7	L2	CO6
	c.	With the help of dimensionality diagram and example, discuss Best fitting , Over fitting , Under fitting in classification of data.	10	L3	CO4
Q.8		Explain the following : 1. Assumption of linear regression. 2. List the steps involved in Building a linear regression model. 3. Any one application of linear regression in detail.	20	L3	CO4

* * * * *