

JAMIA HAMDARD
Department of Computer Science & Engineering
School of Engineering Science & Technology
New Delhi – 110062
Supplementary Even Semester Examination October 2023

Paper Name: Data Mining and Prediction by Machines
Time: 3 Hours

Paper Code: BTCSEAI 410
Maximum Marks: 75

COURSE OUTCOMES:

The students will be able to:

- CO-1:** Understand the process of formulating business objectives, data selection/collection, preparation and process to successfully design, build, evaluate and implement predictive models for a various business application.
- CO-2:** Create a Model for Data mining.
- CO-3:** Evaluate predictive modelling to analyses time series
- CO-4:** Select appropriate predictive modeling approaches to identify cases to progress with.
- CO-5:** Apply predictive modeling approaches using a suitable package such as Python/Tableau

Follow proper numbering for answering the questions.

Section-A				
All questions are compulsory in this Section				
Ques.		Marks	CO	BL
1.	What is Data Mining and What are the different applications of Data Mining?	5	2	1,2
2.	Write Short notes on Data Cleaning, Data Integration & Data Transformation?	5	1	1,2
3.	What is Predictive Modelling? What are the common applications of Data Mining and Prediction?	5	3	3,4
4.	What is data transformation? Why it is essential in the form of KDD? Give example.	5		1,2
5.	Let there be 10 balls (6 white and 4 red balls) in a box and let it be required to pick up the red balls from them. Suppose we pick up 7 balls as the red balls of which only 2 are actually red balls. What are the values of precision and recall in picking red ball?	5	2	5,6
6.	Explain the model of Neural Network?	5	4	2
7.	What is Web mining? What are the different purposes of web mining.	5	5	1,2
Section-B				
Attempt any FOUR Questions from this Section				
1	What is Association Rule Mining? What is the benefit of Rule Mining.	10	1	1,2
2	Describe the data classification process with a neat diagram. How does the Naive Bayesian classification work? Explain.	10	2	2,5,6
3	What measures are used to find best split in Decision Tree Induction algorithm? How Can we improve the scalability in Decision Tree Induction algorithm?	10	3	2

4	The marks for 15 students on mid-term and end-term examinations in Data Analytics course are: <table><tr><td>Mid-Term</td><td>80</td><td>75</td><td>96</td><td>62</td><td>82</td><td>88</td><td>51</td><td>82</td><td>77</td><td>90</td><td>60</td><td>81</td><td>34</td><td>55</td><td>96</td></tr><tr><td>End-Term</td><td>75</td><td>88</td><td>85</td><td>75</td><td>75</td><td>74</td><td>62</td><td>91</td><td>77</td><td>85</td><td>48</td><td>67</td><td>51</td><td>25</td><td>75</td></tr></table>	Mid-Term	80	75	96	62	82	88	51	82	77	90	60	81	34	55	96	End-Term	75	88	85	75	75	74	62	91	77	85	48	67	51	25	75	10	3,4	1,2
Mid-Term	80	75	96	62	82	88	51	82	77	90	60	81	34	55	96																					
End-Term	75	88	85	75	75	74	62	91	77	85	48	67	51	25	75																					
	a) Obtain the simple linear regression analysis to predict the score on the end-term examination from the mid-term examination score. b) It is suggested that if the regression is significant, then there is no need to have final examination. How you test the significance level of your regression analysis?																																			
5	For the given confusion Matrix below for 3 classes. Find Sensitivity & Specificity metrics to estimate Predictive Accuracy of Classification Method? <table><tr><th rowspan="2">Predicted Class</th><th colspan="3">True Class</th></tr><tr><th>1</th><th>2</th><th>3</th></tr><tr><td>1</td><td>5</td><td>5</td><td>4</td></tr><tr><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>3</td><td>0</td><td>0</td><td>2</td></tr></table>	Predicted Class	True Class			1	2	3	1	5	5	4	2	3	4	5	3	0	0	2	10	5	4,5													
Predicted Class	True Class																																			
	1	2	3																																	
1	5	5	4																																	
2	3	4	5																																	
3	0	0	2																																	
6	Write Short notes on any two: a) Evaluating the performance of a classifier? b) Explain Minimum Support and Confidence Threshold c) Support Vector Machine		2,4	1,2																																

BL- Bloom's Taxonomy Levels (1-Remembering, 2-Understanding, 3-Applying, 4-Analysing, 5-Evaluating, 6-Creating), CO- Course Outcomes