Reg. No.												

Question Paper Code	11900

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2023

Sixth Semester

Computer Science and Engineering

20CSEL605 - PREDICTIVE MODELLING

(Regulations 2020)

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks, K-Level, CO
1.	List the advantages of data mining.	2,K1,CO1
2.	Define SEMMA approach.	2,K1,CO1
3.	Define dev set.	2,K1,CO2
4.	Compare and contrast between balanced and imbalanced dataset.	2,K2,CO2
5.	Illustrate the steps to construct the decision tree.	2,K2,CO3
6.	Make use of the below table calculate the conditional probability.	2,K2,CO3

	FEMALE	MALE	TOTAL
TEACHER	8	12	20
STUDENT	32	48	80
TOTAL	40	60	100

(i) P(Teacher\Male)
(ii) P(Teacher\Female)

7.	Compare and contrast between classification tree and regression tree.	2,K2,CO4
8.	Define ROC curve.	2,K1,CO4
9.	What is meant by PMML?	2,K1,CO5
10.	Define model composition.	2,K2,CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11.	a)	Explain in detail about the Methods and process of data cleaning.	13,K2,CO1
		OR	
	b)	(i) Explain in detail about the stages in SEMMA approach.	6,K2,CO1
		(ii) Explain in detail about data partitioning.	7,K2,CO1

12. a) Illustrate the Artificial Neural Network in detail. *13,K2,CO2*

OR

b) Explain in detail about the balanced and imbalanced dataset with examples. *13,K2,CO2*

13. a) Illustrate the steps involved in constructing the decision tree and the measures used for splitting. *13,K2,CO3*

OR

b) Apply the KNN algorithm to predict the grade for the new student Kumar (8, 2). The parameters with the scale of 1 to 10. *13,K3,CO3*

Name	Academic Score	Extra Curricular	Grade
Ria	8	8	Outstanding
Raj	9	1	Academically sound
Surya	4	8	Sporty
Kamal	2	1	Below Average

14. a) Explain in detail about the Support Vector Machine algorithm with suitable examples. *13,K2,CO4*

OR

b) Illustrate the pseudo code for CART algorithm for constructing the decision tree. *13,K2,CO4*

15. a) Illustrate the PMML structure to mining the models using XML. *13,K2,CO5*

OR

b) Explain in detail about the Normalization and discretization in PMML. *13,K2,CO5*

PART - C (1 × 15 = 15 Marks)

16. a) Explain the architecture of WEKA with neat diagram. *15,K2,CO*

OR

b) Illustrate the vectors operation in detail with suitable examples in R. *15,K2,CO6*