Reg. No. | | | | | | | | |

G. VENKATASWAMY NAIDU COLLEGE (AUTONOMOUS), KOVILPATTI – 628 502.

UG DEGREE END SEMESTER EXAMINATIONS - APRIL 2025.

(For those admitted in June 2021 and later)

PROGRAMME AND BRANCH: B.Sc., COMPUTER SCIENCE

SEM	CATEGORY	COMPONENT	COURSE CODE	COURSE TITLE
VI	PART - III	CORE	U21CS611	DATAMINING USING PYTHON

Date & Session: 29.04.2025/FN **Time: 3 hours** **Maximum: 75 Marks**

Course Outcome	Bloom's K-level	Q. No.	SECTION – A (10 X 1 = 10 Marks) Answer ALL Questions.
CO1	K1	1.	What is the OneR algorithm? a) A complex machine learning algorithm for deep learning. b) A simple and easy-to-understand classification algorithm. c) An algorithm used for dimensionality reduction. d) An algorithm for clustering data points.
CO1	K2	2.	Select the following metric which s is NOT typically used to rank rules in affinity analysis? a) Support b) Confidence c) Lift d) Accuracy
CO2	K1	3.	Which of the following is a common distance metric used in nearest neighbor algorithms? a) Euclidean distance b) Manhattan distance c) Minkowski distance d) All of the above
CO2	K2	4.	Select common preprocessing step that can be included in a Pipeline. a) Scaling the features of the data. b) Selecting relevant features. c) Imputing missing values. d) All of the above
CO3	K1	5.	Which library in Python is commonly used for loading and manipulating datasets? a) NumPy b) SciPy c) Pandas d) Matplotli
CO3	K2	6.	Select What is an ensemble method in machine learning? a) A single machine learning algorithm. b) A technique that combines multiple models to improve performance. c) A method for selecting the best features. d) A way to visualize model predictions
CO4	K1	7.	Which of the following is NOT a common algorithm for affinity analysis? a) Apriori algorithm b) FP-growth algorithm c) K-means clustering d) Eclat algorithm
CO4	K2	8.	Select what is a common data format for representing sparse data in affinity analysis? a) CSV b) Market basket analysis c) Transaction database d) Decision tree
CO5	K1	9.	What are common patterns observed in features within datasets? a) Linear relationships b) Non-linear relationships c) Correlations d) All of the above

CO5	K2	10.	Select what is the purpose of the Transformer API in scikit-learn? a) To define custom preprocessing steps for machine learning pipelines b) To visualize the results of machine learning models. c) To optimize hyperparameters of machine learning models. d) To perform feature selection.
Course Outcome	**Bloom's K-level**	**Q. No.**	**SECTION – B (5 X 5 = 25 Marks)** **Answer ALL Questions choosing either (a) or (b)**
CO1	K3	11a.	Determine the method how will you load datasets with NumPy. **(OR)**
CO1	K3	11b.	Write short notes on the process of feature selection in Datamining.
CO2	K3	12a.	Write the concept of K-Nearest Neighbors in detail. **(OR)**
CO2	K3	12b.	Manipulate the advantages of pipelines in machine learning workflows.
CO3	K4	13a.	Examine the process of loading and cleaning a dataset using pandas. **(OR)**
CO3	K4	13b.	Analyze the bias-variance trade-off and how it relates to ensemble methods.
CO4	K4	14a.	Comment on the concept of association rules. **(OR)**
CO4	K4	14b.	Illustrate the steps involved in the Apriori algorithm.
CO5	K5	15a.	Evaluate how does Principal Component Analysis work, what is its advantages? **(OR)**
CO5	K5	15b.	Criticize how will you represent feature Pattern in Models.

Course Outcome	**Bloom's K-level**	**Q. No.**	**SECTION – C (5 X 8 = 40 Marks)** **Answer ALL Questions choosing either (a) or (b)**
CO1	K3	16a.	Estimate how can you evaluate the performance of clustering algorithm. **(OR)**
CO1	K3	16b.	Write short notes on real-world application of data mining.
CO2	K4	17a.	Analyze the importance of feature scaling in machine learning. **(OR)**
CO2	K4	17b.	Comment on different data Classification techniques and their implications.
CO3	K4	18a.	Analyze the performance of a machine learning model. **(OR)**
CO3	K4	18b.	Examine the advantages and disadvantages of using random forests.
CO4	K5	19a.	Criticize about Association Algorithm. **(OR)**
CO4	K5	19b.	Criticize about how you can Load the datasets using Pandas.
CO5	K5	20a.	Analyze the Importance of Feature Creation Method in Data mining. **(OR)**
CO5	K5	20b.	Evaluate the importance of unit testing.