



SAHYADRI
COLLEGE OF ENGINEERING & MANAGEMENT
An Autonomous Institution
MANGALURU

Department of Information Science & Engineering

Assignment – Even Semester 2022-23

Course: Data Mining and Data Warehousing		Course Code: 18CS641
Faculty: Mr. R Pakkala & Mrs. Akhila Thejaswi R	Sem / Section: VI 'A' & 'B'	Max. Marks: 50
Date of Announcement: 05/06/2023		Date of Submission: 30/06/2023

Note:

- Answer both the Questions.
- The assignment document must contain
 - ❖ Cover Page
 - ❖ Problem Statement
 - ❖ Questionnaires
 - ❖ Snapshots of the Dataset
 - ❖ Programming Code
 - ❖ Screenshots of Execution.
- Student can write a program in Python language.

Q. No.		Questions	Marks	Blooms Level	CO No.
1	a.	Consider any real world problems of your interest and construct the dataset by preparing questionnaires' through google form and collect the response from suitable end users. The dataset must have the following requirements: a) A dataset must have minimum 10 -15 attributes. b) A dataset must have minimum of 300 instances. c) A dataset must undergo suitable preprocessing steps.	20	CL3	CO2
	b.	Apply the classification model for the input dataset obtained in the question number 1. Build the confusion matrix to evaluate the performance of the model and calculate the following performance metrics: a) Accuracy b) Error Rate c) Precision d) Recall e) F-Measure	30	CL3	CO4

Cognitive Levels of Bloom's Taxonomy

No.	CL1	CL2	CL3	CL4	CL5	CL6
Level	Remember	Understand	Apply	Analyze	Evaluate	Create

Course Outcomes

CO1	Describe the data warehouse and OLAP technology for understanding the overall data mining and knowledge discovery process.	CL2
CO2	Discuss the implementation aspects of data warehouse and illustrate the importance of data for the mining process.	CL3
CO3	Use the association rules for discovering important relationships hidden in large data sets.	CL3
CO4	Solve many diverse applications using different classification techniques.	CL3
CO5	Apply the concepts of clustering for understanding data and solving different practical problems.	CL3

Assessment Method		
Sl. No.	Assessment Component	Marks Allotted
1.	Google Form with Questionnaires	10
2.	Dataset File	10
3.	Classification Model	20
4.	Report	10
Total		50