



ARTIFICIAL INTELLIGENCE & DATA SCIENCE

AND

ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

DATA MINING AND VISUALIZATION LAB (S3CCSA02)

Laboratory Cycles and Evaluation Scheme - 2024-25 (Odd Sem)

Lab Hours/ Week	: 2	Credits:	1
CIE Marks:	50	SEE Marks:	50

Guidelines to the students:

- Students are assessed every week for 25 marks.
- The students must complete implementation of the respective cycle programs on or before mentioned deadline.
- The students must carry observation book to the lab which contains programs to be executed along with algorithms required for current lab set program.
- Students must maintain a file which includes the evaluation sheets and printouts of successfully executed programs with proper inputs and outputs and related information about each program as per the format provided.
- CIE marks for the practical course is computed by adding the average of the marks secured by the student for conducting each of the experiments with the marks secured in the test conducted and also the marks secured for the open-ended experiments at the end of the course. Two announced tests will be conducted. The syllabus and marks for each test is given below:

Activity	Syllabus	Marks
TEST	Cycle 1 + Cycle 2	$(T1(15) + T2(15))/2$
Open Ended Experiments	Cycle 3	10

Lab test schedule:

Test-1: December 1st Week (offline Mode)

Test-2: January 4th Week (offline Mode)

Demo of open-ended experiments: January 4th Week (offline Mode)

Evaluation Scheme:

- Evaluation scheme for Continuous Internal Evaluation (CIE) - 50 Marks and Semester End Examination (SEE) - 50 Marks.

Break-up of CIE marks for practical courses with Open Ended Experiments

<i>Details</i>	<i>Marks</i>
Regular Lab Work and writing lab records	(15+10) 25 marks
Lab test and Viva-voce at the end of the semester	(10+5) 15 marks
Evaluation of open ended experiment	10 marks
CIE	50 marks
SEE	50 marks
Total	100 marks

- CIE is computed by adding the average of the marks secured in weekly evaluation plus the average of marks secured in tests.

Every week the students are assessed for 35 Marks based on the following Rubrics:

- CIE is computed by adding the average of the marks secured in weekly evaluation plus the average of marks secured in tests.
 - Evaluation scheme for Test :(15 Marks)
 - Write Up - 5 Marks
 - Execution - 5 Marks
 - Viva + Modification- 5 marks
- **Evaluation scheme for weekly evaluations:(25 Marks)**

Every week the students are assessed for 25 Marks based on the following Rubrics:

File maintenance = 05 marks									
Maintaining file (05 Marks)									
Excellent		Very Good		Good		Average		Poor	
5 Marks		4 marks		3-4 Marks		2 marks		1 mark	
Implementation and execution of the programs = 15 marks									
Execution of programs for all possible inputs (10 Marks)					Program Analysis (5 Marks)				
Excellent	Very Good	Good	Average	Poor	Excellent	Very Good	Good	Average	Poor
10 marks	8-9 marks	6-7 marks	4-5 marks	1-3 mark	5 marks	4 marks	3 marks	2 marks	1 mark
Answering 100% of Questions asked (5 Marks)									
Excellent		Good		Average		Poor			
5 marks		4 marks		2-3 marks		1 mark			

- Evaluation scheme for Open Ended evaluations:(10 Marks)

Excellent	Good	Average	Poor
9-10 marks	7-8 marks	5-6 marks	1-4 marks

Evaluation Scheme for SEE:

- In SEE all question Carries equal marks.
- Student is required to execute one program from Part A,
- The questions are allotted based on lots.
- SEE is evaluated for 50 marks using following rubrics.
- Semester End Evaluation (SEE): 50 marks
 - Write Up – 10 Marks
 - Execution – 20 Marks
 - Results with modification – 10 Marks
 - Viva – 10 Marks
- On taking change of program, question from Part A will be changed and the questions are evaluated for 80% of marks only.

Course outcomes (COs):

CO 1	To understand the need for Data Mining and advantages to the business world
CO 2	To get a clear idea of various classes of Data Mining techniques, their need, scenarios (situations) and scope of their applicability
CO3	To learn the algorithms used for various type of Data Mining problems
CO 4	To understand how to explore and communicate data using data visualization techniques using data analytics tools such as PowerBI and Tableau

Mapping of Course outcomes (COs) to Program outcomes (Pos):

Program Articulation Matrix															
Course Outcomes	Program Outcomes												Program Specific Outcome		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1			2											2	
CO2			2											2	
CO3			2											2	
over all			2											2	

Degree of compliance 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

Experiments:

Lab Cycle Questions	Deadline: 23. 01.2025	<u>Mapping</u>
---------------------	-----------------------	----------------

1.

Experiment to be conducted using WEKA tool:

1	outlook	temperature	humidity	windy	play
2					
3	sunny	85	85	FALSE	no
4	sunny	80	90	TRUE	no
5	overcast	83	86	FALSE	yes
6	rainy	70	96	FALSE	yes
7	rainy	68	80	FALSE	yes
8	rainy	65	70	TRUE	no
9	overcast	64	65	TRUE	yes
10	sunny	72	95	FALSE	no
11	sunny	69	70	FALSE	yes
12	rainy	75	80	FALSE	yes
13	sunny	75	70	TRUE	yes
14	overcast	72	90	TRUE	yes
15	overcast	81	75	FALSE	yes
16	rainy	71	91	TRUE	no

1. Preprocess and Classify panels
2. Draw the histogram to show how the values of the *play* class occurs for each value of the *outlook* attribute
3. Derive minimum and maximum values, mean, and standard deviation
4. Perform operations such as filter, delete, invert, Pattern, Undo, Edit, search, Select, Conversions etc
5. Build the decision tree and analyze the weather data.
6. Examine the Output , classification error and Kappa statistics
7. Visualize threshold curve
8. Apply Logistic Regression model to classify
9. Measure the log likelihood of the clusters of training data. (Consider large data set.)

CO:1, 2

PO:3, PSO:2

2.

Consider the following data set

No.	eid	ename	salary	exp	address
	Numeric	Nominal	Numeric	Numeric	Nominal
1	101.0	raj	10000.0	4.0	pdtr
2	102.0	ramu	15000.0	5.0	pdtr
3	103.0	anil	12000.0	3.0	kdp
4	104.0	sunil	13000.0	3.0	kdp
5	105.0	rajiv	16000.0	6.0	kdp
6	106.0	sunitha	15000.0	5.0	nlr
7	107.0	kavitha	12000.0	3.0	nlr
8	108.0	suresh	11000.0	5.0	gtr
9	109.0	ravi	12000.0	3.0	gtr
10	110.0	ramana	11000.0	5.0	gtr
11	111.0	ram	12000.0	3.0	kdp
12	112.0	kavya	13000.0	4.0	kdp
13	113.0	navya	14000.0	5.0	kdp

Use the data sources, like ARFF, XML ARFF files. Do the following

- i) Classify , Invoke MultiLayerPerception
- ii) Build neural network GUI as below
 - a) Beginning the process of editing the network to add a second hidden layer
 - b) The finished network with two hidden layers
- iii) Apply Lazy classifier, multi instance classifier
- iv) Apply any MetaLearning Algorithm

CO:1,2

PO:3; PSO:2

	<div>v) Optimize base classifier's performance</div> <div>vi) Use clustering algorithm such as Cobweb, and Hierarchical Cluster</div> <div>vii)Select attribute by specifying an evaluator and a search method</div>																																																																																																													
3.	<div>Consider the data set</div> <div><table><tr><th colspan="6">Relation: employee</th></tr><tr><th>No.</th><th>age</th><th>income</th><th>stud</th><th>creditrates</th><th>buyscomp</th></tr><tr><td></td><td>Nominal</td><td>Nominal</td><td>Nominal</td><td>Nominal</td><td>Nominal</td></tr><tr><td>1</td><td>L20</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>2</td><td>20-40</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>3</td><td>G40</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>4</td><td>L20</td><td>low</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>5</td><td>G40</td><td>high</td><td>no</td><td>excellent</td><td>yes</td></tr><tr><td>6</td><td>L20</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>7</td><td>20-40</td><td>high</td><td>yes</td><td>excellent</td><td>no</td></tr><tr><td>8</td><td>G40</td><td>low</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>9</td><td>L20</td><td>high</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>10</td><td>G40</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>11</td><td>L20</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr><tr><td>12</td><td>G40</td><td>high</td><td>yes</td><td>excellent</td><td>no</td></tr><tr><td>13</td><td>20-40</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>14</td><td>L20</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>15</td><td>G40</td><td>high</td><td>yes</td><td>excellent</td><td>yes</td></tr></table><div><div>i) Load ARFF file and explore knowledge flow interface</div><div>ii) configure the data source , check the status area after executing the configuration</div><div>iii) Perform operations such as Attribute Selection, Filter, Classify, Data Sink, Visualization and Evaluation</div><div>iv) Apply incremental learning and analyze the result</div><div>v) do clustering : use generator properties, two clustering schemes, and result panel</div><div>vi) Generate classification Matrix and Construct Decision tree</div><div>i) Perform Linear Regression and Analyze , Validate and Visualize the data</div></div></div>	Relation: employee						No.	age	income	stud	creditrates	buyscomp		Nominal	Nominal	Nominal	Nominal	Nominal	1	L20	high	no	fair	yes	2	20-40	low	yes	fair	yes	3	G40	medium	yes	fair	yes	4	L20	low	no	fair	no	5	G40	high	no	excellent	yes	6	L20	low	yes	fair	yes	7	20-40	high	yes	excellent	no	8	G40	low	no	fair	yes	9	L20	high	yes	excellent	yes	10	G40	high	no	fair	yes	11	L20	low	yes	excellent	no	12	G40	high	yes	excellent	no	13	20-40	medium	yes	excellent	yes	14	L20	medium	yes	fair	yes	15	G40	high	yes	excellent	yes	<div>CO:1,2,3</div> <div>PO:3, PSO:2</div>
Relation: employee																																																																																																														
No.	age	income	stud	creditrates	buyscomp																																																																																																									
	Nominal	Nominal	Nominal	Nominal	Nominal																																																																																																									
1	L20	high	no	fair	yes																																																																																																									
2	20-40	low	yes	fair	yes																																																																																																									
3	G40	medium	yes	fair	yes																																																																																																									
4	L20	low	no	fair	no																																																																																																									
5	G40	high	no	excellent	yes																																																																																																									
6	L20	low	yes	fair	yes																																																																																																									
7	20-40	high	yes	excellent	no																																																																																																									
8	G40	low	no	fair	yes																																																																																																									
9	L20	high	yes	excellent	yes																																																																																																									
10	G40	high	no	fair	yes																																																																																																									
11	L20	low	yes	excellent	no																																																																																																									
12	G40	high	yes	excellent	no																																																																																																									
13	20-40	medium	yes	excellent	yes																																																																																																									
14	L20	medium	yes	fair	yes																																																																																																									
15	G40	high	yes	excellent	yes																																																																																																									
4.	<div>Consider glass data set.</div> <div><div>i) How many attributes are there in the dataset? What are their names? What is the class attribute? Run the classification algorithm IBk (weka.classifiers.lazy.IBk). Use cross-validation to test its performance, leaving the number of folds at the default value of 10.</div><div>ii) What is the accuracy of IBk (given in the Classifier Output box)? Run IBk again, but increase the number of neighboring instances to k = 5 by entering this value in the KNN field. Use cross-validation as the evaluation method.</div><div>iii) What is the accuracy of IBk with five neighboring instances (k = 5)?</div><div>iv) Obtain best accuracy higher than the accuracy obtained on the full dataset. Verify ,Is this best accuracy an unbiased estimate of accuracy on future data?</div><div>v) Record the cross-validated accuracy estimate of IBk for 10 different percentages of class noise and neighborhood sizes</div><div>vi) Analyze, What is the effect of increasing the amount of class noise?</div><div>vii) Analyze, What is the effect of altering the value of k?</div><div>viii) Verify the amount of training data</div></div>	<div>CO:3</div> <div>PO:3, PSO:2</div>																																																																																																												
5.	<div>Consider Titanic Data Set. Perform the following using Tableau platform:</div> <div><div>i. Perform Calculations:</div><div>i) Calculate the survival rate.</div><div>Hint: you can use the following formula</div><div>SUM(IIF([Survived] = 1, 1, 0)) / COUNT([PassengerId])</div></div>	<div>CO:3,4</div> <div>PO:3, PSO:2</div>																																																																																																												

	<ul style="list-style-type: none"> ii) Calculate the average iii) Calculate the total fare <p>2. Perform Group Operations:</p> <ul style="list-style-type: none"> i) Create a group to categorize passengers by age ii) Create a group to categorize passengers by fare iii) Create a set to include only passengers who survived <p>3. Perform Set Operations:</p> <ul style="list-style-type: none"> i) Create a set to include only passengers who survived ii) Create a set to include only passengers who did not survive iii) Create a set to include only passengers who traveled in first class <p>4. Create Dashboard</p> <ul style="list-style-type: none"> i). Survival Rate by Age Group: Create a bar chart to display the survival rate by age group. ii). Average Fare by Fare Group: Create a bar chart to display the average fare by fare group. iii). Survivors by Class: Create a pie chart to display the number of survivors by class <p>5. Add Additional Visualizations</p> <p>6. Combine Dashboards</p>	
6.	<p>Consider Brazilian E-Commerce Public Dataset by Olist from Kaggle. Perform the following using Tableau platform:</p> <ul style="list-style-type: none"> i) Create a dashboard that displays the total sales, average sales price, and sales quantity for each state in Brazil, filtered by city, with custom geocoding and map layers. ii) Demonstrate Advanced Mapping technique iii) Demonstrate slicers 	CO:3,4 PO:3 PSO:2
7.	<p>Consider Sales Data set. Perform the following using powerBI platform.</p> <p>1. Create tables with</p> <ul style="list-style-type: none"> i) manufacturing, sum of sales, columns. ii) Product Category, sum of profits etc iii) Find Answers for the following: <ul style="list-style-type: none"> a) Which Manufacturer has the highest Sales? b) Which Product Category has the lowest Profit value? c) Which Channel has the highest Cost of Sales? d) Which Manufacturer has the highest Profit? e) Which Promotion Name has the highest Sales? f) Which Product Sub Category has the highest Profit? iv) Perform cross filtering between tables v) Create matrix visualization for product category , Region, sum of sales vi) Create a card visualtion for total sales, Total Profits, Avg Profits, Highest Profits, Lower Profits etc. and apply formatting <p>2. Apply appropriate visualization. Create calculations for Visualization</p> <ul style="list-style-type: none"> i) Calculate the difference between sum of Sales and Sum of Profits ii) Calcualte Profit Ratio. (Hint: Sum of Profit/Sum of Sales). Also demonstrate using in built funciton, e.g., DIVIDE, AVERAGE etc iii) Calcualte the average sales per product for each of manufacturers. (Hint: Sum of sales/Count of Product Name) iv) Calcualte the Percentage of grand total using inbuilt function <p>3. Apply Filters and Slicers</p>	CO:3,4 PO:3 PSO:2

	<p>4. Build Graphs. Draw Trend Analysis Graph. Show trends and forecasting. For example: sum of sales per year, month, quarter etc</p> <p>5. Create Interactive Dashboard</p> <p>i) Create an Interactive Report</p> <p>The Sales Manager would like you to please create the following interactive report in Power BI Desktop:</p> <p>a) Create a heading - Sales Report</p> <p>b) Create the following Card visualizations:</p> <p>i. Total Sales</p> <p>ii. Total Profit</p> <p>iii. Average Sales</p> <p>iv. Number of Products</p> <p>v. Create a Area graph displaying Sales by Year and Quarter</p> <p>vi. Create a Column graph displaying Profit by Product Category</p>	
8.	<p>Consider Sales Data set. Integrate R environment to power BI platform and perform the following</p> <p>1. Create Smart Narrative</p> <p>2. Using R packages, Build ML based Data Model</p> <p>3. Perform Visualization and build report</p>	CO:3,4 PO:3, PSO:2
9.	<p>Consider Sales Data set. Integrate R environment to power BI platform and perform the following</p> <p>1. Create Q&A Visuals</p> <p>2. Using R packages, Build ML based classification Model</p> <p>3. Perform Visualization and build report</p>	CO:3,4 PO:3 PSO:2
10.	<p>Consider Sales Data set. Integrate R environment to power BI platform and perform the following</p> <p>1. Create decomposition tree</p> <p>2. Using R packages, Build ML based Clustering Model</p> <p>3. Perform Visualization and build report</p>	CO:3,4 PO:3 PSO:2
11.	OPEN ENDED MICRO PROJECT	CO:1,2,3,4 PO:3 PSO:2

Section **Name and signatures of the
Lab Coordinator**

Signature of the HOD

AI & DS
and
AI& ML

Dr. Sumalatha Aradhya

(Dr. N R Sunitha)

INSTITUTE VISION:

To continuously strive for the total development of students by educating them in state-of-the-art-technologies and helping them imbibe professional ethics and societal commitment, so that they emerge as competent professionals to meet the global challenges.

INSTITUTE MISSION:

To develop young minds in a learning environment of high academic ambience by synergising spiritual values and technological competence.

DEPARTMENT VISION:

To work towards the vision of the institution by building a strong teaching and research environment that is capable of responding to the challenges of the 21st century.

DEPARTMENT MISSION:

To prepare under graduate, graduate and research students for productive careers in industry and academia, through comprehensive educational programs, entrepreneurship skills, research in collaboration with industry & government, dissemination by scholarly publications, co-curricular activities and professional society associations.

PROGRAM EDUCATIONAL OBJECTIVES(PEOs):

- Pursue successful careers in State/National/Multi-National companies as software developers by following sound professional and ethical practices in various cadres in key areas like networking, web design, cloud computing, big data processing, IoT, e-commerce, information security and so on.
- Work effectively in multi-disciplinary and multi-cultural teams and demonstrate good soft skills.
- Pursue higher education for a successful career in industry/academics/ research.
- Pursue life-long learning, by anticipating trends in computer science and engineering, to excel in industry/academia or own a startup for a successful career as entrepreneur.

KNOWLEDGE AND ATTITUDE PROFILE (WK) :

WK1: A systematic, theory-based understanding of the natural sciences applicable to the discipline and awareness of relevant social sciences.

WK2: Conceptually-based mathematics, numerical analysis, data analysis, statistics and formal aspects of computer and information science to support detailed analysis and modelling applicable to the discipline.

WK3: A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline.

WK4: Engineering specialist knowledge that provides theoretical frameworks and bodies of

knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline.

WK5: Knowledge, including efficient resource use, environmental impacts, whole-life cost, re-use of resources, net zero carbon, and similar concepts, that supports engineering design and operations in a practice area.

WK6: Knowledge of engineering practice (technology) in the practice areas in engineering discipline.

WK7: Knowledge of the role of engineering in society and identified issues in engineering practice in the discipline, such as the professional responsibility of an engineer to public safety and sustainable development.

WK8: Engagement with selected knowledge in the current research literature of the discipline, awareness of the power of critical thinking and creative approaches to evaluate emerging issues.

WK9: Ethics, inclusive behavior and conduct. Knowledge of professional ethics, responsibilities, and norms of engineering practice. Awareness of the need for diversity by reason of ethnicity, gender, age, physical ability etc. with mutual understanding and respect, and of inclusive attitudes.

PROGRAM OUTCOMES (POs):

PO1: Engineering Knowledge: Apply knowledge of mathematics, natural science, computing, engineering fundamentals and an engineering specialization as specified in WK1 to WK4 respectively to develop to the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development. (WK1 to WK4)

PO3: Design/Development of Solutions: Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required. (WK5)

PO4: Conduct Investigations of Complex Problems: Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions. (WK8).

PO5: Engineering Tool Usage: Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems. (WK2 and WK6)

PO6: The Engineer and The World: Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment. (WK1, WK5, and WK7).

PO7: Ethics: Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws. (WK9)

PO8: Individual and Collaborative Team work: Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

PO9: Communication: Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences

PO10: Project Management and Finance: Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

PO11: Life-Long Learning: Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change. (WK8)

PROGRAM SPECIFIC OUTCOMES (PSOs):

1. **PSO1: Computer based systems development:** Ability to apply the basic knowledge of database systems, computing, operating system, digital circuits, microcontroller, computer organization and architecture in the design of computer based systems.
2. **PSO2: Software development:** Ability to specify, design and develop projects, application software and system software by using the knowledge of data structures, analysis and design of algorithm, programming languages, software engineering practices and open source tools.
3. **PSO3: Computer communications and Internet applications:** Ability to design and develop network protocols and internet applications by incorporating the knowledge of computer networks, communication protocol engineering, cryptography and network security, distributed and cloud computing, data mining, big data analytics, ad hoc networks, storage area networks and wireless sensor networks.