Lesson Plan

Branch: Computer Engineering
Semester: VIII **Year: 2022-23**

Course Title: **Applied Data Science**(CSDC 8013)	SEE: 3 Hours – Theory
Total Contact Hours: 35 Hours	Duration of SEE: 3 Hrs
SEE Marks: 80 (Theory) + 20 (IA)	
Lesson Plan Author: Prof.Ashwini Pansare	Date: 10 th January 2023
Checked By: Prof. Merly T	Date:

Course Outcomes and Assessment Plan

Prerequisite: Engineering Mathematics, Machine Learning, Data Structures &Algorithms	
Course Objectives:	
1	To introduce students to the basic concepts of data science.
2	To acquire an in-depth understanding of data exploration and data visualization.
3	To be familiar with various anomaly detection techniques.
4	To understand the data science techniques for different applications.
Course Outcomes:	
1	To gain fundamental knowledge of the data science process.
2	To apply data exploration and visualization techniques.
3	To apply anomaly detection techniques.
4	To gain an in-depth understanding of time-series forecasting.
5	Apply different methodologies and evaluation strategies.
6	To apply data science techniques to real world applications.

Syllabus:

Module		Content	Hrs
1		**Introduction to Data Science**	**05**
	1.1	Introduction to Data Science, Data Science Process	
	1.2	Motivation to use Data Science Techniques: Volume, Dimensions and Complexity, Data Science Tasks and Examples	
	1.3	Overview of Data Preparation, Modeling, Difference between data science and data analytics	
2		**Data Exploration**	**09**

	2.1	Types of data, Properties of data **Descriptive Statistics:** Univariate Exploration: Measure of Central Tendency, Measure of Spread, Symmetry, Skewness: Karl Pearson Coeff.of skewness, Bowley's Coefficient, Kurtosis	
		Multivariate Exploration: Central Data Point, Correlation, Different forms of correlation, Karl Pearson Correlation Coefficient for bivariate distribution	
	2.2	Inferential Statistics: Overview of Various forms of distributions: Normal, Poisson, Test Hypothesis, Central limit theorem, Confidence Interval, Z-test, t-test, Type-I, Type-II Errors, ANOVA	
3		**Methodology and Data Visualization**	**06**
	3.1	Methodology: Overview of model building, Cross Validation, K-fold cross validation, leave-1 out, Bootstrapping	
	3.2	Data Visualization Univariate Visualization: Histogram, Quartile, Distribution Chart Multivariate Visualization: Scatter Plot, Scatter Matrix, Bubble chart, Density Chart Roadmap for Data Exploration	
		Self-Learning Topics: Visualizing high dimensional data: Parallel chart, Deviation chart,Andrews Curves.	
4		**Anomaly Detection**	**06**
	4.1	Outliers, Causes of Outliers, Anomaly detection techniques, Outlier Detection using Statistics	
	4.2	Outlier Detection using Distance based method, Outlier detection using density-based methods, SMOTE	
5		**Time Series Forecasting**	**05**
	5.1	Taxonomy of Time Series Forecasting methods, Time Series Decomposition	
	5.2	Smoothening Methods: Average method, Moving Average smoothing, Time series analysis using linear regression, ARIMA Model, Performance Evaluation: Mean Absolute Error, Root Mean Square Error, Mean Absolute Percentage Error, Mean Absolute Scaled Error	
		Self Learning Topics: Evaluation parameters for Classification, regression and clustering.	
6		**Applications of Data Science**	**04**

	6.1	Predictive Modeling: House price prediction, Fraud Detection Clustering: Customer Segmentation Time series forecasting: Weather Forecasting Recommendation engines: Product recommendation	
		Total	35

Textbooks:	
1	Vijay Kotu, Bala Deshpande. “Data Science Concepts and Practice”, Elsevier, M.K. Publishers.
2	Steven Skiena, “Data Science Design Manual”, Springer International Publishing AG
3	Samir Madhavan. “Mastering Python for Data Science”, PACKT Publishing
4	Dr. P. N. Arora, Sumeet Arora, S. Arora, Ameet Arora, “Comprehensive Statistical Methods”, S.Chand Publications, New Delhi.
References:	
1.	Jake VanderPlas. “Python Data Science Handbook”, O’reilly Publications.
2.	Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor, “Recommender Systems Handbook”, Springer.
3	S.C. Gupta, V. K. Kapoor “Fundamentals of Mathematical Statistics”, S. Chand and Sons, New Delhi.
4	B .L .Agrawal. “Basic Statistics”, New Age Publications, Delhi.

Course Outcomes (CO):

On successful completion of course learner will be able to:

CSDO8013.1 Describe data science process.
CSDO8013.2 Apply data exploration and visualization techniques to a specific problem.
CSDO8013.3 Apply anomaly detection techniques to a specific problem.
CSDO8013.4 Discover essentials of time-series forecasting.
CSDO8013.5 Apply different methodologies and evaluation strategies to a specific problem

Mapping of CO and PO/PSO

Relationship of course outcomes with program outcomes: Indicate 1 (low importance), 2 (Moderate Importance) or 3 (High Importance) in respective mapping cell.

Program Specific Outcomes (PSOs)

Student will have ability to

PSO1: Develop Artificial Intelligence and Machine Learning based systems.

PSO2: Apply cyber security mechanisms to ensure the protection of Information Technology assets.

	PO1 (Engg Know)	PO2 (Ana)	PO3 (De sign)	PO4 (inve stiga)	PO5 (tools)	PO6 (engg Soci)	PO7 (Env)	PO8 (Eth)	PO9 (ind Team)	PO10 (comm.)	PO11 (PM)	PO12 (life Long)
CSC402.1	1	1										2
CSC402.2	1	1		3	1							
CSC402.3	1	1										
CSC402.4	1	2	3	3								
CSC402.5	1	3	3	3	1							2
SUM	5	8	6	9	2							
Course To PO	1	2	3	3	1							2

CO	**PSO1**	**PSO2**
CSC402.1	1	-
CSC402.2	1	-
CSC402.3	1	-
CSC402.4	2	-
CSC402.5	3	-
Course to PSO	2	-

Mapping Justification:

Course Outcome	BL	Competency	Performance Indicator	PO	Mapping
CSDO8013.1	2	1.4 Demonstrate competence in specialized engineering knowledge to the program	1.4.1 Apply theory and principles of Computer Science and engineering to solve an engineering problem	PO1	1
		2.2 Demonstrate an ability to formulate a solution plan and methodology for an engineering problem	2.2.3 Identify existing solution/methods to solve the problem, including forming justified approximations and assumptions. 2.2.4 Compare and contrast alternative solution/methods to select the best methods	PO2	1
		12.1 Demonstrate an ability to identify gaps in knowledge and a strategy to close these gaps 12.2 Demonstrate an ability to identify changing trends in engineering knowledge and practice	12.1.1 Describe the rationale for the requirement for continuing professional development 12.1.2 Identify deficiencies or gaps in knowledge and demonstrate an ability to source information to close this gap 12.2.1 Identify historic points of technological advance in engineering that required practitioners to seek education in order to stay current 12.2.2 Recognize the need and be able to clearly explain why it is vitally important to keep current regarding new developments in your field	PO12	2
		1.1 Demonstrate competence in mathematical modelling, and engineering fundamentals.	1.1.1 Develop mathematical concepts required for ML and AI algorithms. 1.1.2 Devise the concepts of modelling for the said systems.	PSO1	1

			1.1.3 Apply theory and principles of Computer Science and engineering.		
CSDO8013.2	4	1.4 Demonstrate competence in specialized engineering knowledge to the program	1.4.1 Apply theory and principles of Computer Science and engineering to solve an engineering problem	PO1	1
		2.3 Demonstrate an ability to formulate and interpret a model 2.4 Demonstrate an ability to execute a solution process and analyze results	2.3.1 Able to apply computer engineering principles to formulate modules of a system with required applicability and performance. 2.4.1 Applies engineering mathematics to implement the solution.	PO2	1
		4.1 Demonstrate an ability to conduct investigations of technical issues consistent with their level of knowledge and understanding 4.2 Demonstrate an ability to design experiments to solve open-ended problems 4.3 Demonstrate an ability to analyse data and reach a valid conclusion	4.1.1 Define a problem for purposes of investigation, its scope and importance 4.1.2 Able to choose appropriate procedure/algorithm, dataset and test cases. 4.1.3 Able to choose appropriate hardware/software tools to conduct the experiment. 4.2.1 Design and develop appropriate procedures/methodologies based on the study objectives 4.3.1 Use appropriate procedures, tools and techniques to and analyze collect data 4.3.2 Critically analyze data for trends and correlations, stating possible errors and limitations 4.3.3 Represent data (in tabular and/or graphical forms) so as to facilitate analysis and explanation of the data, and drawing of conclusions 4.3.4 Synthesize information and knowledge about the problem from the raw data to reach appropriate conclusions	PO4	3

		5.1 Demonstrate an ability to identify/create modern engineering tools, techniques and resources	5.1.1 Identify modern engineering tools, techniques and resources for engineering activities	PO5	1
		1.1 Demonstrate competence in mathematical modelling, and engineering fundamentals.	1.1.1 Develop mathematical concepts required for ML and AI algorithms. 1.1.3 Apply theory and principles of Computer Science and engineering.	PSO1	1
CSDO8013.3	4	1.4Demonstrate competence in specialized engineering knowledge to the program	1.4.1Apply theory and principles of Computer Science and engineering to solve an engineering problem	PO1	1
		2.2 Demonstrate an ability to formulate a solution plan and methodology for an engineering problem	2.2.3 Identify existing solution/methods to solve the problem, including forming justified approximations and assumptions. 2.2.4 Compare and contrast alternative solution/methods to select the best methods	PO2	1
		1.1Demonstrate competence in mathematical modelling, and engineering fundamentals.	1.1.3 Apply theory and principles of Computer Science and engineering.	PSO1	1

CSDO8013.4	3	1.4 Demonstrate competence in specialized engineering knowledge to the program	1.4.1 Apply theory and principles of Computer Science and engineering to solve an engineering problem	PO1	1
		2.1 Demonstrate an ability to identify and formulate complex engineering problem 2.2 Demonstrate an ability to formulate a solution plan and methodology for an engineering problem	2.1.1 Articulate problem statements and identify objectives 2.1.2 Identify processes/modules of a computer-based system and parameters to solve a problem 2.1.3 Identify an algorithm that applies to a given problem 2.2.3 Identify existing solution/methods to solve the problem, including forming justified approximations and assumptions. 2.2.4 Compare and contrast alternative solution/methods to select the best methods	PO2	2
		3.1 Demonstrate an ability to define a complex/ open-ended problem in engineering terms	3.1.1 Able to define a precise problem statement with objectives and scope. 3.1.2 Able to identify and document system requirements from stake- holders. 3.1.3 Able to review state-of-the-art literature to synthesize system requirements.	PO3	3
		3.2 Demonstrate an ability to generate a diverse set of alternative design solutions	3.2.1 Able to explore design alternatives.		
		3.3 Demonstrate an ability to select optimal design scheme for further development	3.3.1 Able to perform systematic evaluation of the degree to which several design concepts meet the criteria.		

		4.1 Demonstrate an ability to conduct investigations of technical issues consistent with their level of knowledge and understanding 4.2 Demonstrate an ability to design experiments to solve open-ended problems 4.3 Demonstrate an ability to analyze data and reach a valid conclusion	4.1.1 Define a problem for purposes of investigation, its scope and importance 4.1.2 Able to choose appropriate procedure/algorithm, dataset and test cases. 4.1.3 Able to choose appropriate hardware/software tools to conduct the experiment. 4.2.1 Design and develop appropriate procedures/methodologies based on the study objectives 4.3.1 Use appropriate procedures, tools and techniques to and analyze collect data 4.3.2 Critically analyze data for trends and correlations, stating possible errors and limitations 4.3.3 Represent data (in tabular and/or graphical forms) so as to facilitate analysis and explanation of the data, and drawing of conclusions 4.3.4 Synthesize information and knowledge about the problem from the raw data to reach appropriate conclusions	PO4	3
		1.1 Demonstrate competence in mathematical modelling, and engineering fundamentals. 1.3 Demonstrate an ability to identify solutions/methods to solve the problem.	1.1.3 Apply theory and principles of Computer Science and engineering. 1.3.1Identify processes/modules of an Artificial Intelligence and Machine Learning system and parameters to solve a problem 1.3.2 Apply various methods and evaluate their performance. 1.3.3 Choose an appropriate method for the given problem	PSO 1	2
CSC402.5	4	1.4 Demonstrate competence in specialized engineering knowledge to the program	1.4.1 Apply theory and principles of Computer Science and engineering to solve an engineering problem	PO1	1

		2.1 Demonstrate an ability to identify and formulate complex engineering problem 2.2 Demonstrate an ability to formulate a solution plan and methodology for an engineering problem 2.3 Demonstrate an ability to formulate and interpret a model 2.4 Demonstrate an ability to execute a solution process and analyze results	2.1.1 Articulate problem statements and identify objectives 2.1.2 Identify processes/modules of a computer-based system and parameters to solve a problem 2.1.3 Identify an algorithm that applies to a given problem 2.2.3 Identify existing solution/methods to solve the problem, including forming justified approximations and assumptions. 2.2.4 Compare and contrast alternative solution/methods to select the best methods 2.3.1 Able to apply computer engineering principles to formulate modules of a system with required applicability and performance. 2.4.1 Applies engineering mathematics to implement the solution. 2.4.2 Analyze and interpret the results using contemporary tools.	PO2	3
		3.1 Demonstrate an ability to define a complex/ open-ended problem in engineering terms	3.1.1 Able to define a precise problem statement with objectives and scope. 3.1.2 Able to identify and document system requirements from stake-holders. 3.1.3 Able to review state-of-the-art literature to synthesize system requirements.	PO3	3
		3.2 Demonstrate an ability to generate a diverse set of alternative design	3.2.1 Able to explore design alternatives.		

	solutions			
	3.3 Demonstrate an ability to select optimal design scheme for further development	3.3.1 Able to perform systematic evaluation of the degree to which several design concepts meet the criteria.		
	4.1 Demonstrate an ability to conduct investigations of technical issues consistent with their level of knowledge and understanding	4.1.1 Define a problem for purposes of investigation, its scope and importance 4.1.2 Able to choose appropriate procedure/algorithm, dataset and test cases. 4.1.3 Able to choose appropriate hardware/software tools to conduct the experiment.	PO4	3
	4.2 Demonstrate an ability to design experiments to solve open-ended problems	4.2.1 Design and develop appropriate procedures/methodologies based on the study objectives		
	4.3 Demonstrate an ability to analyze data and reach a valid conclusion	4.3.1 Use appropriate procedures, tools and techniques to and analyze collect data 4.3.2 Critically analyze data for trends and correlations, stating possible errors and limitations 4.3.3 Represent data (in tabular and/or graphical forms) so as to facilitate analysis and explanation of the data, and drawing of conclusions 4.3.4 Synthesize information and knowledge about the problem from the raw data to reach appropriate conclusions		
	5.1 Demonstrate an ability to identify/create modern engineering tools, techniques and resources	5.1.1 Identify modern engineering tools, techniques and resources for engineering activities	PO5	1
	12.1 Demonstrate an ability to identify gaps in knowledge and a strategy to close these gaps	12.1.1 Describe the rationale for the requirement for continuing professional development	PO12	2

			12.1.2 Identify deficiencies or gaps in knowledge and demonstrate an ability to source information to close this gap		
		12.2 Demonstrate an ability to identify changing trends in engineering knowledge and practice	12.2.1 Identify historic points of technological advance in engineering that required practitioners to seek education in order to stay current 12.2.2 Recognize the need and be able to clearly explain why it is vitally important to keep current regarding new developments in your field		
		1.1 Demonstrate competence in mathematical modelling, and engineering fundamentals.	.1.1 Develop mathematical concepts required for ML and AI algorithms. .1.2 Devise the concepts of modelling for the said systems. 1.1.3 Apply theory and principles of Computer Science and engineering.	PSO 1	3
		1.2 Demonstrate an ability to identify the applicability of AI and ML solutions to a problem.	1.2.1 Articulate problem statements and identify objectives 1.2.2 Analyze the problem for applicability of AI and ML solutions. Identify an algorithm that applies to a given problem		
		1.3 Demonstrate an ability to identify solutions/methods to solve the problem.	1.3.1 Identify processes/modules of an Artificial Intelligence and Machine Learning system and parameters to solve a problem 1.3.2 Apply various methods and evaluate their performance. 1.3.3 Choose an appropriate method for the given problem		
		1.4 Demonstrate an ability to execute the solution.	1.4.1 Use the chosen method to implement the solution. 1.4.2 Analyse and interpret the results using contemporary tools.		

			1.4.3 Identify the limitations of the solution and sources/causes.		

Course Outcomes	***Direct Method (80%)***					***Indirect Method (20%)***
	Unit Tests		Assignments		End Sem Exam	Course exit survey
	1	2	1	2		
CSDC8013.1	20%	--	20%	--	60%	100%
CSDC8013.2	15%	5%	10%	10%	60%	100%
CSDC8013.3	-	20%	--	20%	60%	100%
CSDC8013.4	--	20%	--	20%	60%	100%
CSDC8013.5	--	20%	--	20%	60%	100%

CO Assessment Tools:

CSDC8013.1: Direct Methods(80%): Unit Test 1 + Assignment 1+SEE(T)

CO1dm = 0.2T +0.2Assignment+0.6SEE(T)

InDirect Methods(20%): Course exit survey

CO1idm

CSDC8013.1 = 0.8*CO1dm + 0.2* CO1idm

CSDC8013.2: Direct Methods(80%): Unit Test 1 + Assignment 1+SEE(T)

CO1dm = 0.2T +0.2Assignment+0.6SEE(T)

InDirect Methods(20%): Course exit survey

CO1idm

CSDC8013.2 = 0.8*CO1dm + 0.2* CO1idm

CSDC8013.3: Direct Methods(80%): Unit Test 1 + Assignment 1+SEE(T)

CO1dm = 0.2T +0.2Assignment+0.6SEE(T)

InDirect Methods(20%): Course exit survey

CO1idm

CSDC8013.3 = 0.8*CO1dm + 0.2* CO1idm

CSDC8013.4: Direct Methods(80%): Unit Test 1 + Assignment 1+SEE(T)

CO1dm = 0.2T +0.2Assignment+0.6SEE(T)

InDirect Methods(20%): Course exit survey

CO1idm

CSDC8013.4 = 0.8*CO1dm + 0.2* CO1idm

CSDC8013.5: Direct Methods(80%): Unit Test 1 + Assignment 1+SEE(T)

CO1dm = 0.2T +0.2Assignment+0.6SEE(T)

InDirect Methods(20%): Course exit survey

CO1idm

CSDC8013.5 = 0.8*CO1dm + 0.2* CO1idm

Rubrics for Assignment

Indicator	Excellent	Good	Average	Below average
Timeline (2)	submitted on time or early (2)	Submitted next day (1)	Submitted in same week (0.5)	Submitted in next week (0)
Organization (2)	Well organized, neat and clear handwriting, neat diagrams with all labels. (2)	Organized to some extent, diagrams and handwriting is neat with some missing labels (1)	Organization not appropriate, diagrams are incomplete with some missing labels (0.5)	Poorly organized, diagrams incomplete (0)
Level of content (3)	All points are covered and answered accurately (3)	Some important points are omitted / addressed minimally (2)	Many important points are missing and the ones which are written are addresses in brief. (2-1)	Many important points are missing and the answers are not accurate. (1-0.5)
Knowledge about the topic (3)	All Concepts of a topic are clear and knows the application to real world problems (3)	All Concepts of a topic are mostly clear, lacks understanding about the application to real world problems (2)	Concepts of a topic are not understood clearly, lacks understanding about the application to real world problems (2-1)	Poor understanding of concepts and application to real world problems. (1-0.5)

Evaluation Scheme

CIE Scheme
Internal Assessment: 20 (Average of two tests)

Internal Assessment Scheme

Module		Lecture Hours	No. of questions in			No. of questions in SEE
			Test 1	Test 2	Test 3*	
1	Introduction to data science	5	06 Marks	-	--	
2	Data Exploration	9	14 Marks	06 Marks	--	
3	Methodology and data visualization	6	-	06 Marks	--	
4	Anamoly detection	6	-	08 Marks		
5	Applications of data science	5		05 Marks		

Lecture Plan:

Module	Detailed Content	Hours	Planned date	Actual date	Content Delivery Method
1	**Introduction to Data Science**	5			
	Introduction to Data Science		10-01-23	10-01-23	PPT
	Data science Process		11-01-23	11-01-23	PPT
	Motivation to use Data Science Techniques: Volume, Dimensions and Complexity, Data Science Tasks and Examples		13-01-23	13-01-23	PPT
	Overview of Data Preparation, Modeling, Difference between data science and data analytics		17-01-23	17-01-23	PPT
	Applications of data science in various domains		18-01-23	18-01-23	
2	**Data Exploration**	9			
	Types of data, Properties of data, **Descriptive Statistics:** Univariate Exploration:		20-01-23	20-01-23,	PPT
	Measure of Central Tendency, Measure of Spread,		24-01-23	24-01-23	PPT
	Symmetry, Skewness: Karl Pearson Coefficient of skewness, Numerical		25-01-23	25-01-23	PPT & Board
	Bowley ‘s Coefficient, Numerical		27-01-23	27-01-23	PPT & Board
	Kurtosis		31-01-23	31-01-23	Board
	Multivariate Exploration: Central Data Point, Correlation		1-02-23	1-02-23	PPT & Board
	Different forms of correlation, Karl Pearson Correlation Coefficient for bivariate distribution		3-02-23	3-02-23	Board
	Inferential Statistics: Overview of Various forms of distributions: Normal, Poisson		7-02-23	7-02-23	PPT
	Test Hypothesis, Central limit theorem, Confidence Interval		8-02-23	8-02-23	PPT & Board
	Z-test, t-test, Type-I, Type-II Errors , Numericals on Z test and T test		10-02-23	10-02-23	PPT & Board
	ANOVA test , Numericals		13-02-23	13-02-23	PPT & Board
3	**Methodology and Data Visualization**	5			
	Overview of model building, Cross Validation, K-fold cross validation		15-02-23	15-02-23	PPT & Board
	leave-1 out cross validation, Bootstrapping		17-02-23	17-02-23	
	Data Visualization: Univariate Visualization: Histogram, Quartile,		21-02-23	21-02-23	PPT, lab performance

	Distribution Chart				
	Scatter Plot, Scatter Matrix, Bubble chart,		22-02-23	22-02-23	PPT, lab performance
	Density Chart Roadmap for Data Exploration		24-02-23	24-02-23	PPT, lab performance
4	**Anomaly Detection**	**5**			
	Outliers, Causes of Outliers, what if they aren't handled?		7-03-2023	7-03-2023	PPT, lab performance
	Anomaly detection techniques, Outlier Detection using Statistics		8-03-23	8-03-23	PPT , lab performance
	Outlier Detection using Distance based method		10-03-23	10-03-23	PPT , lab performance
	Outlier detection using density-based methods,		14-03-23	14-03-23	PPT, lab Performance
	SMOTE		15-0-23	15-0-23	PPT , lab performance
5	**Time Series Forecasting**	6			
	Taxonomy of Time Series Forecasting methods, Time Series Decomposition		17-03-23	17-03-23	PPT & Board
	Smoothening Methods: Average method, Moving Average smoothing, Time series analysis using linear regression,		21-03-23	21-03-23	PPT & Board
	ARIMA Model and its version		24-03-23 & 28-03-23	24-03-23 & 28-03-23	PPT & Board
	Performance Evaluation: Mean Absolute Error, Root Mean Square Error, Mean Absolute Percentage Error, Mean Absolute Scaled Error		29-03-23	29-03-23	PPT & Board
6	**Applications of Data Science**	4			
	Predictive Modeling: House price prediction, Fraud Detection		05-04-23	05-04-23	PPT & Board
	Clustering: Customer Segmentation		11-04-23	11-04-23	PPT & Board
	Time series forecasting: Weather Forecasting		12-04-23	12-04-23	PPT, Board, Lab performance
	Recommendation engines: Product recommendation		15-04-23	15-04-23	PPT & Board

SUB: APPLIED DATA SCIENCE
YEAR:22-23 CLASS: BE COMPS A

Assignment 1

CSDO8013.1 Explain data science process.

CSDO8013.2 Apply data exploration and visualization techniques to a specific problem

<table>
<tr><td></td><td>Term AI, ML and data science is often used Interchangeably. Compare the similarity and difference among them.</td><td>CSDO8013.1</td></tr>
<tr><td></td><td>Describe in detail the data science process</td><td>CSDO8013.1</td></tr>
<tr><td></td><td>Tata Soaps manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign, the mean sales per week per shop was 140 dozen. After the campaign, a sample of 26 shops was taken and the mean sales was found to be 147 dozen with standard deviation 16. Can you consider the advertisement effective?</td><td>CSO8013.2</td></tr>
<tr><td></td><td>List down the cases when one should apply Bowley’s coefficient of skewness over Karl Pearson’s coefficient</td><td>CSO8013.2</td></tr>
<tr><td></td><td>Under which circumstances one has to apply i) z test ii) T test iii) Anova test</td><td>CSO8013.2</td></tr>
<tr><td></td><td>What is kurtosis and its types? Find the standard deviation and kurtosis of the following set of data pertaining to kilowatt (kwh) of electricity consumed by 100 people in a city

<table><tr><td>consumption</td><td>Number of users</td></tr><tr><td>0-10</td><td>10</td></tr><tr><td>10-20</td><td>20</td></tr><tr><td>20-30</td><td>40</td></tr><tr><td>30-40</td><td>20</td></tr><tr><td>40-50</td><td>10</td></tr></table></td><td>CSO8013.2</td></tr>
</table>

SUB: APPLIED DATA SCIENCE
YEAR:22-23 CLASS: BE COMPS A

Assignment 2

CSDO8013.3 Apply anomaly detection techniques to a specific problem
CSDO8013.4 Discover essentials of time-series forecasting.

CSDO8013.5 Apply different methodologies and evaluation strategies to a specific problem

1	Cardiotocography (CTG) Data Set is given to you and you have been asked to Classify the records based on foetus state into Normal, Suspect and Pathologic State. Which outlier detection technique can be used in this case?	CSDO8013.3
2	Which Model is efficient when you are dealing with time series data? Explain all its variations and how do you choose the ideal values for its parameters	CSDO8013.4
3	Study any case study which is based on applications of time series forecasting for business decision	CSDO8013.4
4	Demonstrate with an example various performance evaluation metrics used in machine learning. in which cases they are used. List down advantages and disadvantages if applicable	CSDO8013.5

FR. CONCEICAO RODRIGUES COLLEGE OF ENGG.
Fr. Agnel Ashram, Bandstand, Bandra (W) Mumbai 400 050.

I UNIT TEST

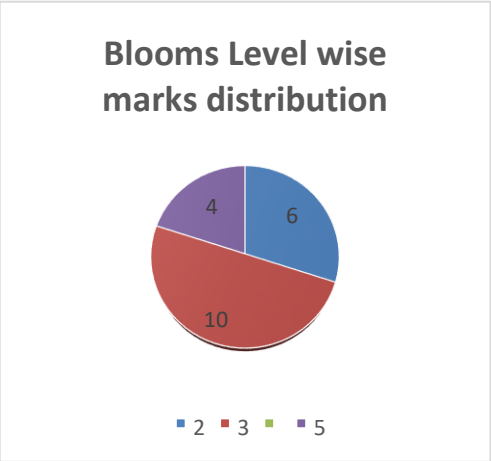
SEMESTER / BRANCH: VIII/COMPUTER-Div. A &B
SUBJECT: APPLIED DATA SCIENCE(ADS) MAX. MARKS: 20
DATE: 28/2/2023 TIMING: 2:30-3:30

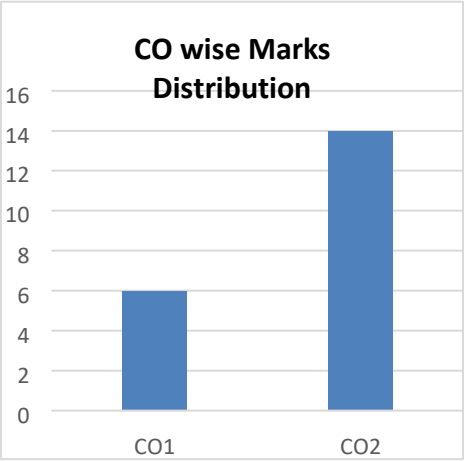
Student should be able to

CSDO8013.1	Explain data science process.
CSDO8013.2	Apply data exploration and visualization techniques to a specific problem

Q.NO	Questions	MARKS	CO	BL	PI
1.A	You wish to build a classification model. A data set provided to you consist of more than 30 percent missing values. How will you deal with this problem?	3	CSDO8013.1	2	1.4.1 2.2.3 2.2.4
1.B	Discuss why causation is different than Correlation using suitable example?	3	CSDO8013.1	2	1.4.1

<table>
<tr><td>2.A</td><td>The following is the information about the settlement of an industrial dispute in a factory. Comment on the gain and losses from the point of view of workers and that of management
<table>
<tr><td></td><td>Before</td><td>After</td></tr>
<tr><td>No of workers</td><td>3000</td><td>2900</td></tr>
<tr><td>Mean wages</td><td>2200</td><td>2300</td></tr>
<tr><td>Median wages</td><td>2500</td><td>2400</td></tr>
<tr><td>Std. deviation</td><td>300</td><td>260</td></tr>
</table>
OR
Elucidate in what way measures of central tendency, variation, skewness and kurtosis are complementary to one another in understanding a frequency distribution table.</td><td>4</td><td>CSDO80
13.2</td><td>3</td><td>4.3.1
4.3.2</td></tr>
<tr><td>2.B</td><td>Calculate the coefficient of correlation for the following data.
<table>
<tr><td>x</td><td>9</td><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td></tr>
<tr><td>y</td><td>15</td><td>16</td><td>14</td><td>13</td><td>11</td><td>12</td><td>10</td><td>8</td><td>9</td></tr>
</table></td><td>3</td><td>CSDO80
13.2</td><td>3</td><td>4.3.1
4.3.2</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2 C</td><td>A principal at a certain school claims that the students in his school are above average intelligence. A random sample of thirty (50) students IQ scores have a mean score of 110. Is there sufficient evidence to support the principal’s claim? The mean population IQ is 100 with a standard deviation of 15. IQ scores are normally distributed. (Z-score at 5% level of significance is 1.645).</td><td>3</td><td>CSDO80
13.2</td><td>3</td><td>4.3.1
4.3.2</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2 D</td><td>A trucking company wishes to test the average life of the four brands of tyres. The company uses all the brands on randomly selected trucks. The records showing the lives (Thousands of miles) of tyres are as given in the table. Test the hypothesis that the average life for each band of tyres is the same. Assume standard deviation= 0.01. Apply suitable test
<table>
<tr><td>Brand 1</td><td>Brand 2</td><td>Brand 3</td><td>Brand 4</td></tr>
<tr><td>20</td><td>19</td><td>21</td><td>15</td></tr>
<tr><td>23</td><td>15</td><td>19</td><td>17</td></tr>
<tr><td>18</td><td>17</td><td>20</td><td>16</td></tr>
<tr><td>17</td><td>20</td><td>17</td><td>18</td></tr>
</table></td><td>4</td><td>CSDO80
13.2</td><td>5</td><td>4.3.1
4.3.2</td></tr>
</table>





FR. CONCEICAO RODRIGUES COLLEGE OF ENGG.
Fr. Agnel Ashram, Bandstand, Bandra (W) Mumbai 400 050.

II UNIT TEST

SEMESTER / BRANCH: VIII/COMPUTER
SUBJECT: APPLIED DATA SCIENCE(ADS)
DATE: 17/04/2023

Div. A & B
MAX. MARKS: 20
TIMING: 11:30 to 12:30

Student should be able to

CSDO8013.2	Apply data exploration and visualization techniques to a specific problem
CSDO8013.3	Apply anomaly detection techniques to a specific problem.
CSDO8013.4	Discover the essentials of time-series forecasting.

Q.NO	Questions	MARKS	CO	BL	PI
1.A	Identify the technique used to evaluate the performance of a model on unseen data. Also list all its types.	2	CSDO8013.2	3	1.4.1 2.2.3
1.B	Discuss the technique used to perform random sampling with replacement. State its advantages and disadvantages	4	CSDO8013.2	3	1.4.1 2.2.4
2.A	State the Need for anomaly detection. Specify the basic approaches to anomaly detection. Enlist the application of anomaly detection.	3	CSDO8013.3	2	1.4.1

2.B	Discuss the algorithm which will overcome the overfitting problem posed by random oversampling with an example.	3	CSDO80 13.3	3	1.4.1 2.2.3 2.2.4
3.A	Previous data of the rainfall from year 2000 to 2020 is available with us, we need to predict monthly rainfall for future years. State and explain the ideal model used for this case study.	5	CSDO80 13.4	4	2.1.3 2.2.4 3.2.3
3.B	Illustrate various smoothing methods applied on the time series data with an example.	3	CSDO80 13.4	2	2.1.3 2.3.1 3.2.3

*BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analyzing, 5 – Evaluating, 6 - Creating)