VIT Vidyalankar Institute of Technology
Accredited A+ by NAAC

Version 08/06-06

Consolidated Academic Administration Plan for the Course
Data Warehousing and Mining(DWM) Professional Elective (AIML and DA track)
Sem. V – Program CMPN 2025-2026 –Odd Semester
Faculty -Dr. Vidya Chitre (Cluster Mentor) & Faculty – Dr. Kavita P Shirsat.

The academic resources available in VIT –

VMIS (ERP)	V-Refer and V-Live	VIT Library	VAC & MOOC Courses
Institute & Department Vision and Mission	Former IA question papers and solutions (prepared by faculty)	Former IA question papers solutions - hardcopy	Value Added Courses (VAC) are conducted throughout the semester & in the semester break - Enrol for the VACs
Program Educational Objectives (PEO)	MU end semester examination question papers and solutions (prepared by faculty)	MU end semester exam question paper & solutions - by faculty, hardcopy	
Program Specific Outcome (PSO)	Class notes and Digital Content for the subject (scanned / typed by faculty)	All text books, reference books, e -books mentioned in the syllabus & AAP	Online courses from NPTEL, Coursera etc. are pursued throughout the semester - Register for the course & get certified
Program Outcome (PO)	Comprehensive question bank, EQ, GQ, PPT, Class Test papers	Technical journals and magazines for reference	
Departmental Knowledge Map	Academic Administration Plan & Beyond Syllabus Activity report	VIT library has many resources e:g :- IEEE, Nimbus, xplore, EBSCO etc.	Watch former lectures captured in LMS at VIT

1.a Course Objectives (Write in detail – as per NBA guidelines)

Cognitive	What do you want students to know?	Students should understand how to analyze data and choose relevant models and algorithms for respective applications.	
Affective	What do you want students to think / care about?	Students should think about designing data warehousing schemas and applying the right data mining algorithms effectively.	
Behavioural	What do you want students to be able to do?	Students should cultivate a research interest in innovations in data mining.	

Advice to Students:

Attend every class!!! Missing even one class can have a substantial effect on your ability to understand the course. Be prepared to think and concentrate, in the class and outside. I will try to make the class very interactive. Participate in the class discussions. Ask questions when you don't understand something. Keep up with the class readings. Start assignments and homework early. Meet me in office hour to discuss ideas, solutions or to check if, what you understand is correct.
The v-Refer Link : http://vidyalankarlive.com/vrefer/index.php/apps/files/?dir=/vRefer/CMPN/SEM%20V/2025-26/Verticals/PC-PEC/Data%20Warehousing%20and%20Data%20Mining_PECE02T/KS&fileid=1080659.

Collaboration Policy:

We encourage discussion between students regarding the course material. However, no discussion of any sort is allowed with anyone on the assignment and homework for the class. If you find solution to some problems in a book or on the internet, you may use their idea for the solution; provided you acknowledge the source (name and page in the book or the website, if the idea is found on the internet). Even though you are allowed to use ideas from another source, you must write the solution in your own words. If you are unsure whether or not certain kinds of collaboration is possible, please ask the teacher.

1.b **Course Outcome (CO) Statements and Module-Wise Mapping (follow NBA guideline)**

CO No.	Statements	Related Module/s
CO1	Recall and explain the fundamental concepts and architecture of data warehousing, data lakes, and data mining.	All
CO2	Design dimensional data warehouse schemas (star, snowflake) and describe the ETL (Extract, Transform, Load) process involved in building and maintaining a data warehouse.	1
CO3	Perform data preprocessing and visualization tasks to prepare datasets for mining, including data cleaning, transformation, and exploratory analysis.	3
CO4	Analyze and compare various data mining algorithms (classification, clustering, association) for solving real-world problems, selecting the most appropriate approach.	3,4 ,6
CO5	Evaluate the effectiveness of implemented data mining solutions, interpret results, and propose improvements or optimizations for enhanced decision-making.	All

1.c **Course Outcome (CO) Statements and Module-Wise Mapping (follow NBA guideline)**

	Mapped to Learning Outcomes
CO1	Recall and explain the fundamental concepts and architecture of data warehousing, data lakes, and data mining.
CO2	Design dimensional data warehouse schemas (star, snowflake) and describe the ETL (Extract, Transform, Load) process involved in building and maintaining a data warehouse.
CO3	Perform data preprocessing and visualization tasks to prepare datasets for mining, including data cleaning, transformation, and exploratory analysis.
CO4	Analyze and compare various data mining algorithms (classification, clustering, association) for solving real-world problems, selecting the most appropriate approach.
CO5	Evaluate the effectiveness of implemented data mining solutions, interpret results, and propose improvements or optimizations for enhanced decision-making.

1.d **Mapping of COs with POs (mark S: Strong, M: Moderate, W: Weak, Dash '–': not mapped) (List of POs is available in V-refer)**

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
	Knowledge A	Analysis	Design	Investigation	Modern Tools	Society	Environment &sustainability	Ethics T	Teamwork C	Communication	Project Mgt	Life long learning
CO 1	07/39	08/39	-	-	01/39	-	-	-	-	-	-	07/39

CO2	03/39	02/39	-	-	03/39	-	-	-	-	-	-	03/39
CO3	05/39	04/39	-	01/39	-	-	-	-	-	-	-	05/39
CO4	-	02/39	-	02/39	-	-	-	-	-	-	-	-
CO5	-	01/39	-	01/39	-	-	-	-	-	-	-	-

1.e **Mapping of COs with PSOs (mark S: Strong, M: Moderate, W: Weak, Dash '–': not mapped)**

	PSO 1	PSO 2	PSO 3
CO1	15/39	01/39	-
CO2	05/39	03/39	-
CO3	10/39	-	-
CO4	04/39	-	-
CO5	02/39	-	-

1.f **Teaching and Examination Scheme (As specified by the autonomous syllabus) for the Course**

Verticals	BSC/ESC	Program Courses	Multidisciplinary Courses	Skill Courses	HSSM	Experiential Learning	Liberal Learning	
Tick suitable category						√		

Subject Code	Subject Name	**Teaching Scheme**			Credits Assigned			
		Theory	Practical	Tutorial	Theory	TW/Practical	Tutorial	Total
PECE02T	Data Warehousing & Mining	2	-	-	2	-	-	2
PECE02P	Data Warehousing & Mining Lab	-	2	-	-	1	-	1

	Subject Name			**Examination Scheme**			
		Theory		**Total**	**Practical**		**Total**

Subject Code		ISA	MSE	ESE	**(Theory)**	ISA	ESE	ORAL	**(Practical)**
PECE02T	Data Warehousing& Mining	15	20	40	75				
PECE02P	Data Warehousing& Mining Lab		-		25	25			50

Subject Code	Subject Name	**MSE-1***		
		Q, No	Module wise % Distribution	Relevant to Bloom Taxonomy
PECE02T	Data Warehousing& Mining	Q1 a	Module 1 10%	L2
		Q1b	Module 1 10%	L3
		Q2a	Module 2 10%	L3
		Q2b	Module 3 10%	L2

Subject Code	Subject Name	**MSE-2***		
		Q, No	Module wise % Distribution	Relevant to Bloom Taxonomy
PECE02T	Data Warehousing& Mining	Q1 a	Module 3 10%	L2
		Q1b	Module 4 10%	L2
		Q2a	Module 4 10%	L3
		Q2b	Module 5 10%	L3

Subject Code	Subject Name	**ESE#**		
		Q, No	Distribution	Relevant to Bloom Taxonomy
PECE02T	Data Warehousing& Mining	Q1 a	Module 1 8%	L2
		Q1b	Module 2 8%	L3
		Q2a	Module 3 8%	L2
		Q2b	Module 4 8%	L2
		Q3 a	Module 5 8%	L2
		Q3b	Module 5 8%	L3
		Q4a	Module 6 8%	L3
		Q4b	Module 6 8%	L3

*** Recommended distribution: -** 30 Marks from Assignments, 40 marks based on assignments with slightly enhance difficulty /complex, 30 marks from thought provoking

 # Recommended distribution: - 30 Marks from Assignments, 40 marks based on assignments/MSE with slightly enhance difficulty /complex, 30 marks from thought provoking

1.g Faculty-Wise Distribution of all Lecture-Practical-Tutorial Hours for the Course

Divisions	Lecture (Hrs.)	Practical (Hrs.)				Tutorial (Hrs.)			
		Batch 1	Batch 2	Batch 3	Batch 4	Batch 1	Batch 2	Batch 3	Batch 4
A	KS	KS	KS	KS	KS	-	-	-	-
B	KS	KS	KS	SNA	KS	-	-	-	-
C	KS	KS	KS	SNA	KS				

1.h Office Hours (Faculty will be available in office in this duration for solving students' query)

Division	Day	Time (at least 1 Hr. / Division)	Venue (Office Room No.)
A	Monday	3:45 pm to 4:45 pm	M-209
B	Wednesday	3:45 pm to 4:45 pm	M-209
C	Tuesday	3:45 pm to 4:45 pm	M-209

2.a Syllabus: Module Wise Teaching Hours and % Weightage in autonomous syllabus Question Paper

Module No.	Module Title and Brief Details	Teaching Hrs. for each module	% Weightage in autonomous syllabus Question Papers			Performance Indicator Mapping
			ISA	MSE	ESE	
1	Introduction to data warehouse and ETL Process	8	18.18	25	10	*P.I.- 1.4.2 , 2.1.2,,2.1.4, 1.4.3 , 2.2.3,5.2.3, 1.4.2 , 5.2.1 5.3.2*
Learning Outcome-1.1	*Explain key data warehousing concepts and architectures(P.I.- 1.4.2 , 2.1.2 and 2.1.4).(CO1)*					
Learning Outcome-1.2	*Apply dimensional modeling techniques for designing warehouse schemas. (PI 1.4.3 , 2.2.3 and 5.2.3) (CO2)*					
Learning Outcome-1.3	*Describe the role and steps of the ETL process. (PI 1.4.2 , 5.2.1 and 5.3.2) (CO2)*					
2	Introduction to Data Lakes	3	9.09	12.5	10	*1.4.1, 2.1.2, 2.1.4 , 5.2.4, 2.1.5 ,1.4.2*
Learning Outcome-2.1	*Explain the purpose and structure of data lakes. (PI 1.4.1 , and 2.1.2) (CO1)*					

Module No.	Module Title and Brief Details	Teaching Hrs. for each module	% Weightage in autonomous syllabus Question Papers			Performance Indicator Mapping
			ISA	MSE	ESE	
Learning Outcome-2.2	*Compare and contrast data lakes with data warehouses. (PI 2.1.4 and 5.2.4) (CO1)*					
Learning Outcome-2.3	*Identify key challenges and functionalities of data lakes in large-scale data storage. (PI 2.1.5 and 1.4.2) (CO2)*					
3	Data Exploration and Data Preprocessing	5	18.18	25	15%	*1.4.2 , 2.1.2, 1.4.3 , 2.4.3, 4.3.4 , .1.3*
Learning Outcome-3.1	*Explain the KDD process and role of preprocessing. (PI 1.4.2 , and 2.1.2) (CO1)*					
Learning Outcome-3.2	*Apply data cleaning, integration, and transformation methods. (PI 1.4.3 , and 2.4.3) (CO3)*					
Learning Outcome-3.3	*Perform exploratory data analysis using visualization techniques. (PI 4.3.4 , and 1.1.3) (CO3)*					
4	Classification	5	18.18	25	15%	*1.4.2 , 2.1.2 1.4.3 , 2.4.3, 4.3.5 , 2.4.4*
25%Learning Outcome-4.1	*Explain the principles of decision tree and Naïve Bayes classifiers. (PI 1.4.2 , and 2.1.2) (CO1)*					
Learning Outcome-4.2	*Apply classification and regression techniques on sample datasets. (PI 1.4.3 , and 2.4.3) (CO3)*					
Learning Outcome-4.3	*Analyze the strengths and weaknesses of various classifiers. (PI 4.3.5 , and 2.4.4) (CO4)*					
5	Clustering	4	18.18	12.5	25%	*1.4.2 , 2.1.2, 1.4.3 ,2.4.3, 4.3.4 , 2.4.4*
Learning Outcome-5.1	*Explain partition-based and hierarchical clustering methods. (PI 1.4.2 , and 2.1.2) (CO1)*					
Learning Outcome-5.2	*Apply K-means and hierarchical clustering on datasets. (PI 1.4.3 , and 2.4.3) (CO3)*					
Learning Outcome-5.3	*Analyze clustering results and interpret cluster quality. (PI 4.3.4 , and 2.4.4) (CO4)*					
6	Mining frequent patterns and associations	5	18.18	0	25%	*1.4.2 , 2.1.2, 1.4.3 , 2.4.3, 4.3.5 , 2.4.5*
Learning Outcome-6.1	*Explain frequent itemsets, closed itemsets, and association rule concepts. (PI 1.4.2 , and 2.1.2) (CO1)*					
Learning Outcome-6.2	*Apply Apriori and pattern-growth algorithms to discover frequent patterns. (PI 1.4.3 , and 2.4.3) (CO3)*					

Module No.	Module Title and Brief Details	Teaching Hrs. for each module	% Weightage in autonomous syllabus Question Papers			Performance Indicator Mapping
			ISA	MSE	ESE	
Learning Outcome-6.3	*Evaluate the performance and effectiveness of frequent pattern mining techniques. (PI 4.3.5 , and 2.4.5) (CO5)*					
	* Insert rows for more modules in the Course **Total**	**30**	**15**	**60**	**80**	

Note: - As an attachment Annexure is required for assessment criteria of learning outcomes.

2.b Prerequisite Courses

No.	Semester	Name of the Course	Topic/s
1	IV	DBMS	DB design & fundamental concepts

2.c Relevance to Future Courses

No.	Semester	Name of the Course
1	VI	Machine Learning
2	VIII	Deep learning (Professional Elective)
3	VII, VIII	Projects

2.d

See :- Identify real life scenarios/examples which uses the knowledge of the subject ,(Discussion on how to prepare examples and case studies e.g. "Boeing Plane": C Programming Language – Intro to Computer Science – Harvard's CS50 (2018) – Bing video)

Real Life Scenario	Concept Used
1. Classifying medical images 2. Detecting tumours from MRI scans 3. Categorizing emails into spam or non-spam/ Identifying abnormalities in X-rays or MRI scans.	Classification Algorithm
1. Weather/ Stock Market Prediction 2. Predicting patient recovery time	Regression Algorithm
1. Customer or patient segmentation 2. To identify anomalous network traffic behaviour 3. detect cybersecurity threats. 4. Credit card fraud detection.	Clustering Algorithm
1. Market Basket Analysis 2. Identifying co-occurring symptoms in medical records 3. Analysing DNA sequences to discover genetic markers associated with specific diseases 4. Examining social media interactions to identify influential users	Association Mining Algorithm

3 Past Results – Division-Wise

Details	Target – MAY 2025	DEC 2024	MAY 2024	DEC 2023
Course Passing % – Average of 2 Divisions	100%	99%	100%	100%
Marks Obtained by Course Topper (mark/100)	80	90	90	90

	Division A		Division B	
Year	Initials of Teacher	% Result	Initials of Teacher	% Result
May 2024	KS	99	KS	100
May 2023	KS	100	KS	100
May 2022	KS	100	VDD	100

4 All the Learning Resources – Books and E-Resources

4.a List of Textbooks (T – Symbol for Textbooks) to be Referred by Students

Sr. No	Textbook Titles	Author/s	Publisher	Edition	Module Nos.	Available in our Library
1	Data Mining Concepts and Techniques	Han, Kamber	Elsevier	3nd edition	3,4,5	Yes
2	Data Warehousing: Fundamentals for IT Professionals.	Paulraj Ponniah,	Wiley India	3nd edition	1	Yes
3	Data Lake Architecture	Bill Inmon,	Technics Publication	1st edition	2	Yes
4	The Data Warehouse Toolkit	Margy Ross and Ralph Kimball,	Willey	3rd edition	1	Yes

4.b List of Reference Books (R – Symbol for Reference Books) to be Referred by Students

Sr. No	Reference Book Titles	Author/s	Publisher	Edition	Module Nos.	Available in our Library
1	Data Mining Concepts and Techniques	Han, Kamber	Elsevier	3nd edition	3,4,5	Yes
2	Data Warehousing: Fundamentals for IT Professionals.	Paulraj Ponniah,	Wiley India	3nd edition	1	Yes
3	Data Lake Architecture	Bill Inmon,	Technics Publication	1st edition	2	Yes

4.c List of E - Books (E – Symbol for E-Books) to be Referred by Students

Sr. No	E- Book Titles	Author/s	Publisher	Edition	Module Nos.	Available in our Library
1	Building the Data Warehouse	W. H. Inmon	Wiley Computer Publishing	3rd edition,	1	Yes
2	Data Mining	Ian H. Witten, Eibe Frank and Mark A. Hall	Morgan kaufmann publisher	3nd edition	3,4,5	Yes
3	Introduction to Data Mining	Pang-Ning Tan, Vipin Kumar, Michael Steinbach	Pearson Education	2nd edition	3,4,5	yes

4.d **Reading latest / top rated research papers (at least 5 papers)**

Name of Paper	Name of Authors (Background)	Published in		Problem Statement	Available in our Library
		Date	Journal		
Gender-Based Analysis of User Reactions to Facebook Posts	Yassine El Moudene,Jaafar Idrais	25 December 2024	Big Data Mining and Analytics IEEE Explore	Visualize the correlation of reactions on Online Social Networks (OSNs)with emotional expressions	IEEE Xplore
Modelling Audiograms for People with Dementia Who Experience Hearing Loss Using Multiple Linear Regression Method	Abeer Elkhouly Advanced Communication Engineering (ACE) Centre of Excellence Universiti Malaysia Perlis	18 November 2020	IEEE Explore	Create a mathematical model of audiograms of individuals with dementia who experience hearing loss. by using multiple regression polynomial.	IEEE Xplore
Stock Price Forecasting Based on Improved Support Vector Regression	Yuhan Fang Hong Kong University of Science and Technology School of Economics Hong Kong,China	20 September 2021	IEEE Xplore	Stock Market Analysis Using Linear Regression	IEEE Xplore
Understanding Residents' Behavior for Smart City Management by Sequential and Periodic Pattern Mining	Cong Ma; Huy Quan Vu Department of Information Systems and Business Analytics Deakin University Burwood, VIC, Australia	February 2024	IEEE Transactions on Computational Social Systems	Smart City :Human Activity Analysis Based on Pattern Mining	IEEE Xplore
Time Series Data Mining: A Case Study With Big Data Analytics Approach	Fang Wang; Menggang Li; Yiduo Mei; Wenrui Li	14 January 2020	IEEE Access	Time series Stock Market prediction to accurately predict the trend.	IEEE

4.e **Based on research paper an identify the current Problem statement**

Problem Statement			Used in				
	Quiz	Assignment	Lab	Mini Project	Poster Presentation	Test	Any Other
1. Visualize the correlation of reactions on Online Social Networks (OSNs)with emotional expressions			✓				
2. Create a mathematical model of audiograms of individuals with dementia who experience hearing loss. by using multiple regression polynomial.				✓			
3. Stock Market Analysis Using Linear Regression		✓					
4. Smart City :Human Activity Analysis Based on Pattern Mining				✓			
5, Time series Stock Market prediction to accurately predict the trend.				✓			

4.f **Identify Companies / Industries which use the knowledge of the subject and thus may provide Internships and final Placements**

Name of the Company	To be / Contacted for		
	Student Internship	Student Final Placement	Faculty Internship
1. Oracle Data Mining 2. IBM 3. Amazon	Student Internship		

4.g **Identify suitable relevant TOP Guest Speakers from Industry,**
Example: - (CS50 Lecture by Mark Zuckerberg - 7 December 2005 - YouTube)

Name of the Identified Guest Speaker	Designation	Name of the Company

Mr. Hemant Tendolkar	Enterprise Architect, AWS Business Unit	TCS

4.h **Identify relevant technical competitions to participate [Competitions -Paper Presentations, Projects, Hackathons, IVs etc..]**

Name of the Relevant Technical Competition Identified to participate	Organized by	Date of the Event
Kaggle Competitions	Kaggle	Entire year
AI Song Contest 2025 human–AI co-creativity in the songwriting process.	Dutch public broadcaster VPRO	June 2025
Smart India Hackathon	Government of India, Pune University, and Various Deemed university	As per the scheduled announced

4.i **Identify faculty in TOP schools / Universities who are teaching same / similar subject and develop rapport e.g. Exchange Lecture Material (Assignments / Tests / Project etc..), Joint Paper Publication**

University	Name of the Course	Name of Faculty	Type of Collaboration		
			Exchange of Lecture Material	Joint Publication/ Research	Other
IIT Kharagpur	Data Mining	Prof. Pabitra Mitra	Lecture notes		
NITT Tiruchirappalli	Data Warehousing and Data Mining	Dr. E Sivasankar	Chapter 1 PPT		

4.j **Module Best Available in – Title of the best resource [from *4.a* to *4.d* in this AAP] and other details as necessary**

Module No.	Title of the Module	Textbook	Mention the Tile					
			Reference Book	E-books	Journal	E-Journal	Available in our Library	V-refer
1	Introduction to Data Warehouse and ETL Process	Dataware housing and OLAP operation	A Review: Analysis on Data Warehousing and Data Mining IEEE				1	Introduction to Data Warehouse and ETL Process
2	Introduction to Data Lakes		IEEE Transactions on Knowledge and Data Engineering		Introduction to Data Lakes	https://www.altexsoft.com/blog/data-lake-architecture/	2	Introduction to Data Lakes
3	Data Exploration and Data Preprocessing	https://youtu.be/NSxEiohAH5o	Data Mining and Knowledge Discovery				3	Data Exploration and Data Preprocessing

			Springer					
4	Classification	https://archive.nptel.ac.in/courses/111/106/111106164/	International Journal of Data Mining, Modelling and Management. Inder Science				4	Classification
5	Clustering	https://www.youtube.com/watch?v=ZTP6b3LaVGg	Statistical Analysis and Data Mining Wiley				5	Clustering
6	Mining frequent patterns and associations	Association Mining	Frequent Pattern Mining Algorithms for Finding Associated Science Direct				6	Mining frequent patterns and associations

4.k **Referred to any top-rated university in that subject for content**

University	Name of the Course	Name of Faculty	Date of Delivery of the Course	Remarks
MIT	Prediction Machine learning and statistics	Prof. Cynthia Rudin	Winter 2021	
Buffalo	Data Mining	Sargur Srihari	Winter 2021	
IIT Kharagpur	Data Mining	Prof. Pabitra Mitra	Feb 2021	https://youtu.be/PiGtQK7_fg4

4.l **Faculty received any certification related to this subject. List of Certifications Identified / Done**

Course	Certifying Agency	No. of Hours	Level of the Course		Certification		Remarks
			Introductory	Advance Skill Development	Done on	Proposed to be on	
Deep learning and AI	Coursera	2022		Advance	20/2/22		completed

Crash course on Machine Learning	Great Learning	47	Basic course		30 January 2024		completed
Application of ML in Sustainable Technologies	NIIT Rourkela	1 week		Advance	January 2024		completed

4.m **Completed subject wise/cluster wise training with cluster mentor.**
List of relevant Refresher Course Identified / Done

Course	Certifying Agency (As suggested by DAB/Cluster Mentor/Industry/University other than MU)	Certification		Remarks
		Done on	Proposed to be on	
Pedagogy	Coursera	2021	Data mining	Completed
	Udemy	2022	Machine learning	Completed
PBL	Coursera	2019	Optimization	Completed
	Udemy	2019	Data mining	Completed
Sub. Content Training	Coursera	2022	Machine learning	Completed
	Udemy Great Learning	2022 2024	Machine learning Machine Learning	Completed

4.n **Best Practices Identified and adopted**

No.	Item	Best Practices Identified		
		IIT Ropar and Kharagpur	MIT	NIT Rourkela
1	Microsite			
2	Video Lectures	✓	✓	
3	Assignments	✓		
4	Mini Project	✓	✓	
5	Assessment Metric			
6	Quizzes	✓	✓	✓
7	Labs/ Practical (PBL)	✓		✓
8	Tests			
9	Peer Assessment			
10	Any Other			

4.o **Web Links for Online Notes/YouTube/VIT Digital Content/VIT Lecture Capture/NPTEL Videos**

Students can view lectures by VIT professors, captured through LMS 'Lecture Capture' in VIT campus for previous years.

No.	Websites / Links	Module Nos.

1	https://youtu.be/PiGtQK7_fg4	All
2	https://www.coursera.org/specializations/data-mining	3 to 6

s

4.p **Recommended MOOC Courses like Coursera / NPTEL / MIT-OCW / edX/VAC etc.**

Sr. No.	MOOC Course Link	Course conducted by – Person / University / Institute / Industry	Course Duration	Certificate (Y / N)
1	https://www.mygreatlearning.com/academy/courses/11331425/94723	Great Leraning	42 hrs	Y
2	https://www.coursera.org/specializations/data-mining	Coursera	3 months	Y

5 **Consolidated Course Lesson Plan**

	From (date/month/year)	From (date/month/year)	Total Number of Weeks
Semester Duration	07/07/2025	28/010/2025	14

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
1	1	1	Chapter 1: Introduction to Data Warehouse and ETL Process Introduction to Data Warehouse and Data Mart, Data Warehouse architecture.		L2	CO1		Fundamentals Paulraj Ponnianh Chapter 15 page no 373–405
	2	1	Data Warehouse vs Data Marts Information Package Diagram		L2	CO1		Data Mining Concepts and Techniques, Han, Kamber page no 135-164
2	3	1	Data Warehouse Schemas; Star Schema, Snowflake Schema, Fact Constellation Schema.		L2	CO1	Yes	Fundamental Paulraj Ponnianh

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
	4	1	Factless Fact Table, Update to the dimension tables.		L3	CO1	Yes	Chapter 15 page no 373–405 Data Mining Concepts and Techniques Han, Kamber, page no 135-164
3	5	1	OLAP operations		L3	CO1	Yes	Data Mining Concepts and Techniques,Han, Kamber page no - 327-434
	6	1	Rollup, drill down, Slice, Dice, pivot		L3	CO1	Yes	Data Mining Concepts and Techniques , Han, Kamber page no - 327-434
4	7	1	Basic steps of the ETL process, different extraction methods, transformations		L1	CO1	Yes	Data Mining Concepts and Techniques , Han, Kamber

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
								page no - 248-265
	8	1	Different loading techniques		L1	CO1	Yes	Data Mining Concepts and Techniques, Han, Kamber page no - 248-265
5	9	2	Introduction to Data Lakes Definition, key attributes of the data lake		L1	CO2	Yes	Data Lake Architecture Bill Inmon, Page no 30-37
	10	2	challenges, functionalities, architecture, Curating data lakes		L1	CO2	Yes	Data Lake Architecture Bill Inmon, Page no 47-63
6	11	2	Data Lake vs. Data Warehouse		L1	CO2	Yes	Data Lake Architecture Bill Inmon, Page no 63-77
	12	3	**Data Exploration and Data Preprocessing**		L1	CO3	Yes	Data Mining Concept

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
			The KDD process, Data mining system architecture, Data Exploration: Types of Attributes					s and Techniques, Han, Kamber page no 83-117
7	13	3	Statistical Description of Data		L2	CO3	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117
	14	3	Data Visualization: box plots, line & bar charts, and scatter plots.		L2	CO3	Yes	Data Mining Concepts and Techniques, Han, Kamber page no - 327-434,
8	15	3	Data Preprocessing: Descriptive data		L2	CO3	Yes	Data Mining Concepts and Technique Han, Kamber page no - 327-434
	16	3	summarization, Cleaning, Integration &		L2	CO3	Yes	Data Mining Concepts and Technique Han,Kamber page no - 248-265
9	17	3	transformation, Data reduction.		L2	CO3	Yes	Data Mining Concepts and Techniques

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
								, Han, Kamber page no - 248-265
	18	4	Introduction to data mining techniques, Classification:		L2	CO4	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117
10	19	4	Decision Tree Induction		L3	CO4	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117
	20	4	Naïve Bayesian Classification.		L3	CO4	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117
11	21	4	Simple Regression		L3	CO4	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
	22	4	Multiple Regression		L3	CO4	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 83-117
12	23	5	Clustering: Partition based:		L3	CO5	Yes	Data Mining Concepts and Techniques, Han, Kamber page no 444-487
	24	5	K-means Clustering (Single and double)		L3	CO5	Yes	Data Mining Concepts and Techniques, Han, Kamber
13	25	5	Hierarchical Methods (Agglomerative, Divisive)		L3	CO5	Yes	Data Mining Concepts and Techniques, Han, Kamber
	26	6	Basic Concepts: Market Basket Analysis, Frequent Itemset, Closed Itemset, and Association Rules; The Apriori Algorithm: Finding Frequent Itemset Using Candidate Generation, Generating Association Rules from requent Itemset,		L2	CO6	Yes	Data Mining Concepts and Techniques, Han, Kamber

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	Mapping Bloom Taxonomy level	COs Mapped	Recommended Prior Viewing / Reading	
							Lecture No. (on LMS)	Chapter No./ Books/ Web Site
14	27	6	Improving the Efficiency of Apriori, A pattern growth approach for mining		L3	CO6	Yes	Data Mining Concepts and Techniques, Han, Kamber
	28	6	Frequent Itemset. Mining Methods: Frequent Itemset, Mining Frequent Itemset using vertical data formats.		L1	CO6	Yes	Data Mining Concepts and Techniques, Han, Kamber
15	29		Revision			CO6	Yes	
	30		Revision			CO6	Yes	

6 Rubric for Grading and Marking of Term Work (inform students at the beginning of semester)

- Activity/ies should be designed as per reference of credit structure.
- If the subject is of 2 credit, activity/ assignment should be design for 2 hours with appropriate complexity and engaging time.

Theory (ISA=15)												Total
Class Participation	Activity-1	Activity-2	Activity-3	Activity-4	Activity-5	Activity-6	Activity-7	Activity-8	Activity-9	Activity-10	Activity-11	
04	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	Assignment	15

	1 marks	1 marks	1 marks (Mooc Course)	1 marks Class test 1marks	1 marks	1 marks	1 marks	1 marks	Poster 1 marks	IEEE paper Review 1 marks	1 marks	

Practical (ISA=25)												
Class Participation	Activity-1	Activity-2	Activity-3	Activity-4	Activity-5	Activity-6	Activity-7	Activity-8	Activity-9	Activity-10	Activity-11	Total
10	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Experiment 1 marks	Mini project 5 marks	Experiment 1 marks	Experiment 1 marks	25

Class Participation	MSE-1	MSE-2	ESE	Total
15 marks	Q1 -10 Marks Q2 -10 Marks Total 20 marks	Q1 -10 Marks Q2 -10 Marks Total 20 marks	Q1 -10 Marks Q2 -10 Marks Q3 -10 Marks Q4 -10 Marks Total- 40 marks	Total -75 marks

7 Assignments / Tutorials Details

Assignment/ Tutorial No.	Title of the Assignments / Tutorials	CO Map	Mapping Bloom Taxonomy Level	Assignment/ Tutorials given to Students on	Assignments to be submitted back on
1	Assignment 1	01	L2	Week 2	Week 3
2	Assignment 2	01	L2	Week 3	Week 4
3	Assignment 3	02	L3	Week 4	Week 5
4	Assignment 4	02	L3	Week 5	Week 6

5	Assignment 5	02	L3	Week 6	Week 7
6	Assignment 6	03	L3	Week 7	Week 8
7	Assignment 7	03	L3	Week 8	Week 9
8	Assignment 8	03	L3	Week 9	Week 10
9	Assignment 9	03	L4	Week 10	Week 11
10	Assignment 10	04	L4	Week 11	Week 12
11	Assignment 11	04	L4	Week 12	Week 5

Analysis of Assignment / Tutorial Questions and Related Resources

Assignment / Tutorial No.	Week No.	Type* (√)			Module No.	Based on #			Question Type (√)	
		OT	CS	DTP		Textbook	Reference Book	Other Learning Resource	Real Life Assignments	Thought Provoking
1	Week 2		√		1		√		√	
2	Week 3		√		2			√	√	
3	Week 4		√		2		√		√	√
4	Week 5		√		3			√	√	√
5	Week 6		√		3			√	√	√
6	Week 7		√		3			√		√
7	Week 8		√		4		√		√	√
8	Week 9		√		4			√	√	
9	Week 10		√		5			√	√	
10	Week 11		√		5		√			√
11	Week 12		√		5			√	√	

* Tick (√) the Type of the Assignment: Online Tools (OT); Collaborative Assignments (CS); Design /Thought provoking (DTP)

Write number for textbook, reference book, other learning resource from this AAP – *from Points 4.a to 4.d*

8 In Semester Assessment (ISE) / Other Class Test / Open Book Test (OBT)/Take Home Test (THT) Details

Tests	Test Dates	Module No.	CO Map	MSE Question Paper Pattern	Policy
ISE	Week 4	3	CO1		
Minute Paper	Week 5	4	CO4		
Pop Quiz	Week 8	5	CO3	10 questions from IIT Ropar assignments	
Class Test	Week 10	6	CO4		

*** Failures of IA test (IA1+IA2) shall appear for IA test in the next semester. There is no provision for re-test in the same semester.**

9. Practical Activities

Practical No.	Module No.	Title of the **Experiments**	Type of Experiment		Topics to be highlighted	CO Map
			PBL	Newly Added		
1	1	Build Data Warehouse/Data Mart for a given problem statement i) Identifying the source tables and populating sample data ii) Design dimensional data model i.e. Star schema, Snowflake schema, and Fact Constellation schema(if applicable)			Data Warehouse schema design	CO1
2	1	You are a data analyst for **RetailMart**, a large retail chain that stores its sales data in a **star schema** data warehouse. The **fact table** Sales_Fact records daily sales transactions and is connected to the following **dimension tables**: Date_Dim(Date_ID, Day, Month, Quarter, Year) Store_Dim(Store_ID, Store_Name, City, State, Region) Product_Dim(Product_ID, Product_Name, Category, Sub_Category) Customer_Dim(Customer_ID, Gender, Age_Group, Income_Bracket) The **fact table** has the following fields: Sales_Fact(Sales_ID, Date_ID, Store_ID, Product_ID, Customer_ID, Units_Sold, Revenue) **a)** Describe the **ROLL-UP** operation needed to produce this summary from the base-level daily transaction data. **b)** Later, they want to **DRILL-DOWN** from region to individual cities within the region to identify underperforming cities. What would this operation look like? A marketing analyst wants to understand customer buying behavior for female customers **in the age group 25–35**, focused only on **Electronics** category. **a)** Define a **SLICE** operation to restrict the data to this age group and gender. **b)** Then apply a **DICE** operation to compare **revenue by category and region** only for this sliced group.		New	OLAP	CO1

		c)The CEO wants a pivoted sales report that shows quarters as columns and product categories as rows, with revenue values filled in. Describe how you would achieve this using OLAP cube or SQL-based pivot**.**				
3	3	Implement and apply various data visualization techniques on a given dataset to explore and understand the characteristics of the data. Use appropriate univariate, bivariate, and multivariate plots to analyze distributions, trends, and relationships. Interpret the findings and summarize the insights gained from each visualization			Data Visualization	CO3
4	4	A retail company wants to predict the monthly sales revenue of its stores based on past data. Initially, the company believes that the revenue depends mainly on the store's advertising budget. Later, they also consider other influencing factors such as store location rating, number of staff, and customer footfall. Implement linear regression models (both single-variable and multiple-variable) to**:** 1. Predict sales using only the advertising budget (simple linear regression). 2. Predict sales using multiple features such as advertising budget, location rating, staff count, and footfall (multiple linear regression). 3. Compare the performance of both models. 4. Visualize and interpret the results."		PBL	Regression analysis	CO4
5	4	A healthcare company wants to identify which patient attributes are most influential in predicting whether a patient will be readmitted within 30 days. You are provided with a dataset containing patient demographics, medical history, treatment type, and past admission records. Use Decision Tree-based attribute relevance analysis to: 1. Determine the most important features affecting patient readmission. 2. Visualize the decision tree. 3. Interpret and rank the attributes by their relevance.			Classificatio n	CO4

		4. Suggest how this analysis can support hospital resource planning and patient car **Dataset Link:** https://www.kaggle.com/datasets/uciml/diabetes				
6	5	A telecom company wants to segment its customer base to design targeted marketing strategies. You are given a dataset containing information about customer usage patterns such as call duration, data consumption, monthly charges, and number of customer service calls. Your task is to implement the K-Means clustering algorithm to group customers with similar behavior. Analyze and interpret the resulting clusters to identify distinct customer segments and recommend suitable service plans for each cluster. 1. Preprocess the data (scaling, normalization if needed). 2. Apply the **K-Means algorithm** to find optimal customer segments. 3. Use **Elbow Method** to determine the number of clusters. 4. Visualize clusters using 2D or 3D plots. 5. Interpret cluster characteristics and suggest targeted strategies. **Dataset Link:** https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom		New	Clustering	CO5
7	5	An e-commerce company wants to segment its customers based on their purchasing behavior to better tailor product recommendations. The company provides you with customer data including average order value, frequency of purchases, total spending, and product category preferences. Your task is to implement the Agglomerative Hierarchical Clustering algorithm to group customers into meaningful segments. Analyze the resulting clusters to understand different customer profiles and recommend targeted marketing strategies for each group		New	Clustering	CO5

		1. Preprocess the data (encode categorical features, normalize numeric values). 2. Apply **Agglomerative Clustering** using distance metrics (Euclidean, Ward). 3. Use a **dendrogram** to decide the optimal number of clusters. 4. Label and visualize the clusters in 2D space. 5. Interpret the characteristics of each cluster. **Dataset Link** https://github.com/AakritiKinra/Agglomerative-Hierarchical-clustering/blob/main/Wholesale%20customers%20data.csv				
8	6	A supermarket wants to understand customer purchasing behavior to improve product placement and offer personalized promotions. You're provided with a dataset of transaction records, where each row lists the items bought together. Your task is to implement the Apriori algorithm to identify frequent itemsets and generate association rules. Use the results to suggest product bundling strategies **Dataset Link** https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store		New	Association Analysis	CO6
9	1,2,3,4,5,6	**Mini-Project** *on Data Mining Techniques: Classification, Clustering, and Association" on a real-world dataset using data mining tools.* **Each student/group must:** 1. Select a real-world dataset (e.g., retail, health, education, e-commerce). 2. Perform Data Preprocessing o Handle missing values, encoding, scaling, etc. 3. Apply Classification Algorithm (e.g., Decision Tree, Naive Bayes) 4. Apply Clustering Algorithm (e.g., K-Means, Agglomerative) 5. Apply Association Rule Mining (e.g., Apriori)		New	ALL	CO6

		6. Interpret results and provide visualizations 7. Prepare a concise project report summarizing methodology, findings, and insights.				
10	6	**Problem Statement:** "A multinational retail chain generates millions of daily transactions across different product categories and locations. The management wants to uncover hidden purchase patterns to improve cross-selling strategies and inventory planning. Your task is to implement an optimized FP-Tree–based mining solution that can efficiently handle high-dimensional, sparse, and large-scale transaction data. You must construct the FP-Tree manually (not using built-in libraries), extract frequent itemsets using conditional pattern bases, and generate meaningful product association rules. Further, analyze and recommend how the insights can be used for dynamic shelf placement or personalized promotions." Tasks 1. Manually build the FP-Tree data structure from a cleaned dataset. 2. Extract conditional pattern bases and generate frequent itemsets. 3. Implement FP-Growth algorithm without using ready-made libraries. 4. Apply min-support dynamically for high/low frequency categories. 5. Compare association rules across locations or time periods. 6. Visualize the FP-Tree structure (using graphviz or custom tool). 7. Give business interpretations and strategy suggestions.	PBL	New	Association Analysis	
11	6	**Title**:User Behavior and Influence Analysis in Social Networks using Classification and Clustering. **Problem Statement:** A social media platform wants to understand how users interact with content and each other to improve personalization, moderation, and influence detection. Given a dataset of user profiles, posts, likes,	PBL		Application of clustering and classification	

		comments, and follower connections, your task is to apply clustering and classification techniques to analyze user behavior. Specifically, cluster users based on activity patterns (e.g., lurkers, influencers, content creators), and classify new users into behavior categories using supervised learning. Evaluate how these insights can improve recommendation systems or flag anomalous behavior."				

10 **Uncovering syllabus with different Activities.**

No.	Type of the Activity	Activities	Number of beneficiaries	Other Details – guest profile, feedback, mark sheet, report
1	**Experiential learning/Interaction with Outside World**	1- Guest Lectures by Industry Expert	Hemant Tendolkar TCS	Senior Analyst
		2- Workshops		
		3- Mini Project	YES	
		4- Industrial Visit		
		5- Any other activity	YES	
2	**Collaborative & Group Activity**	6- Poster Presentation	YES	
		7- Minute Papers	YES	
		8- Students Seminars	YES	
		9- Students Debates		
		10- Panel Discussion / Mock GD		
		11- Mock Interview		
		12- Any other activity		
3	**Co-Curricular Activity**	13- Informative videos (NPTEL/YouTube /TEDx/ MIT OW/edX)	YES	
		14- Lecture Capture Usage	YES	
		15- Any other activity		
4	**Tests & Assessments**	16- Class Tests/ Weekly Tests	YES	
		17- Pop Quiz	YES	
		18- Mobile App Based Quiz		
		19- Open Book Test	YES	
		20- Take Home Test	YES	
		21- Any other activity	YES	

11 AAP

No.	Programme	Course	Uploaded on V-refer	Date
1	CMPN	Dataware housing and mining		

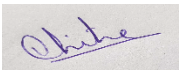
12 Lecture Guide

No.	Programme	Course	Uploaded on V-refer	Date
1	CMPN	Data warehousing and mining	Yes	06/07/2025

*** Do not delete any activity. Give details for planned events. Write 'NA' for activity Not Planned.**

Consolidated Academic Administration Plan Prepared by (mention all theory teaching faculty names with signature)

Please write below your name and sign with date of the external cluster mentor meeting

 Dr Kavita P. Shirsat			
			
Dr. Anuradha B External Industry Mentor	Dr. S. K Shinde External Academic Mentor	Dr. Vidya Chitre VIT Cluster Mentor	Dr. Ravindra Sangle. Program HOD

Annexure:

Assessment Criteria of Learning Outcomes:

Learning Outcomes: The Learner will:	**Assessment Criteria:** The Learner can:	**Evaluated under ISA/MSE/ESE/LAB**
LO 1.1 Explain key data warehousing concepts and architectures(P.I.- 1.4.2 , 2.1.2 and 2.1.4).(CO1)	• The student can define what a data warehouse is. • The student is able to describe at least 3 key characteristics of data warehousing. • The student can list and describe key components of data warehouse architecture. • The student is able to differentiate between types of data warehouse architectures. • The student can identify and match data warehouse concepts with their correct definitions or roles.	MSE ISA MSE ISA
LO 1.2: *Apply dimensional modeling techniques for designing warehouse schemas. (PI 1.4.3 , 2.2.3 and 5.2.3) (CO2)*	• The student can identify business processes and related measures (facts). • The student can determine appropriate dimension tables for context (e.g., Time, Product, Customer). • The student can ensure schema supports required level of data granularity. • The student can structure tables with proper keys and relationships for analytical querying.	MSE ISA ESE LAB
LO 1.3: *Describe the role and steps of the ETL process. (PI 1.4.2 , 5.2.1 and 5.3.2) (CO2)*	• The student can define ETL and explain its importance in data warehousing. • The student can describe the three main steps: Extract, Transform, and Load. • The student can explain how ETL ensures data quality and consistency. • The student can identify ETL challenges and common tools used.	MSE ISA ISA ESE
LO 2.1: *Explain the purpose and structure of data lakes. (PI 1.4.1 , and 2.1.2) (CO1)*	• The student can define a data lake and explain its purpose for storing large volumes of raw, unstructured, semi-structured, and structured data. • The student can describe the key characteristics of data lakes: scalability, flexibility, and schema-on-read. • The student can explain the layered structure of a data lake: ingestion, storage, processing, and consumption. • The student can differentiate data lakes from data warehouses in terms of structure, use cases, and data types.	ISA ISA MSE ISA
LO 2.2: *Compare and contrast data lakes with data warehouses. (PI 2.1.4 and 5.2.4) (CO1)*	• The student can compare data lakes and data warehouses based on data types: raw/unstructured vs. processed/structured.	MSE ISA

	• The student can explain storage and schema differences: schema-on-read (data lake) vs. schema-on-write (data warehouse). • The student can describe differences in use cases: data lakes for big data and ML, warehouses for BI and reporting. • The student can analyze performance, cost, and flexibility trade-offs between the two architectures.	ESE LAB
LO 2.3: *Identify key challenges and functionalities of data lakes in large-scale data storage. (PI 2.1.5 and 1.4.2) (CO2)*	• The student can describe core functionalities of data lakes such as storage of diverse data types, scalability, and support for advanced analytics. • The student can identify key challenges including data governance, metadata management, and data quality issues. • The student can explain the importance of access control, security, and compliance in managing large-scale data lakes. • The student can recognize performance and integration issues when working with multiple data sources in a data lake.	MSE ISA ESE LAB
LO 3.1: *Explain the KDD process and role of preprocessing. (PI 1.4.2 , and 2.1.2) (CO1)*	• The student can define Knowledge Discovery in Databases (KDD) and explain its main phases. • The student can describe the role of preprocessing as the initial and essential step in KDD. • The student can identify preprocessing tasks such as data cleaning, integration, selection, and transformation. • The student can explain how effective preprocessing improves the quality and accuracy of knowledge extracted.	MSE ISA MSE ISA
LO 3.2: *Apply data cleaning, integration, and transformation methods. (PI 1.4.3 , and 2.4.3) (CO3)*	• The student can apply data cleaning techniques such as handling missing values, removing duplicates, and correcting inconsistencies. • The student can perform data integration from multiple sources ensuring consistency and uniformity. • The student can implement data transformation methods like normalization, aggregation, and encoding. • The student can evaluate the impact of these methods on data quality and model performance.	MSE ISA MSE LAB
LO 3.3 *Perform exploratory data analysis using visualization techniques. (PI 4.3.4 , and 1.1.3) (CO3)*	• The student can use visual tools like histograms, boxplots, and scatter plots to identify data patterns and distributions. • The student can detect outliers, trends, and relationships using appropriate visualization techniques. • The student can interpret visual insights to support data understanding and decision-making.	LAB LAB LAB ISA

	• The student can apply tools such as Matplotlib, Seaborn, or Tableau for effective exploratory data analysis (EDA).	
LO 4.1: *Explain the principles of decision tree and Naïve Bayes classifiers. (PI 1.4.2 , and 2.1.2) (CO1)*	• The student can describe the working principle of Decision Tree classifiers, including attribute selection using measures like Information Gain or Gini Index. • The student can explain how Naïve Bayes uses probability and Bayes' Theorem with the assumption of feature independence. • The student can compare the strengths and limitations of both classifiers in terms of accuracy, interpretability, and efficiency. • The student can identify suitable use cases for each classifier based on data characteristics and problem requirements.	ISA ISA ISA,LAB MSE
LO 4.2: *Apply classification and regression techniques on sample datasets. (PI 1.4.3 , and 2.4.3) (CO3)*	• The student can implement classification algorithms such as Decision Tree, Naïve Bayes, or KNN on labeled datasets. • The student can apply regression techniques like Linear Regression or Polynomial Regression to predict continuous values. • The student can evaluate model performance using metrics like accuracy, precision, RMSE, and R^2. • The student can use tools like Python (scikit-learn) or R to train, test, and validate models on sample data.	LAB LAB LAB LAB
LO 4.3: *Analyze the strengths and weaknesses of various classifiers. (PI 4.3.5 , and 2.4.4) (CO4)*	• The student can compare classifiers (e.g., Decision Tree, Naïve Bayes, KNN, SVM) based on accuracy, speed, and interpretability. • The student can identify classifiers suitable for specific types of data (e.g., noisy, high-dimensional, or imbalanced). • The student can explain the limitations of each classifier, such as overfitting, assumptions, or computational cost. • The student can make informed decisions about classifier selection for different problem domain	ISA LAB LAB LAB
LO 5.1: *Explain partition-based and hierarchical clustering methods. (PI 1.4.2 , and 2.1.2) (CO1)*	• The student can define clustering and explain its objective in unsupervised learning. • The student can describe partition-based methods like K-Means and their working principles. • The student can explain hierarchical clustering methods (agglomerative and divisive) with dendrogram representation. • The student can compare both methods based on scalability, accuracy, and use cases.	ISA ISA MSE LAB
LO 5.2 *Apply K-means and hierarchical clustering on datasets. (PI 1.4.3 , and 2.4.3) (CO3)*	• The student can implement K-means clustering and determine optimal cluster count using methods like the elbow method. • The student can perform hierarchical clustering and generate dendrograms to visualize cluster formation.	LAB LAB

	• The student can preprocess and normalize data before applying clustering algorithms. • The student can evaluate clustering results using internal validation metrics like silhouette score.	LAB LAB
LO 5.3: *Analyze clustering results and interpret cluster quality. (PI 4.3.4 , and 2.4.4) (CO4)*	• The student can visualize clustering results using plots (e.g., scatter plots, dendrograms) for interpretation. • The student can evaluate cluster quality using metrics like silhouette score, Davies–Bouldin index, or intra/inter-cluster distances. • The student can interpret the meaning of formed clusters in the context of the dataset and problem domain. • The student can identify issues like overlapping clusters, poor separation, or inappropriate cluster count	ISA MSE MSE LAB
LO 6.1: *Explain frequent itemsets, closed itemsets, and association rule concepts. (PI 1.4.2 , and 2.1.2) (CO1)*	• The student can define frequent itemsets and explain their role in market basket analysis. • The student can describe closed itemsets and how they reduce redundancy in frequent pattern mining. • The student can explain association rules along with key measures: support, confidence, and lift. • The student can differentiate between frequent, closed, and maximal itemsets with examples.	ISA ISA ISA LAB
LO 6.2: *Apply Apriori and pattern-growth algorithms to discover frequent patterns. (PI 1.4.3 , and 2.4.3) (CO3)*	• The student can implement the Apriori algorithm to generate frequent itemsets and association rules from transactional data. • The student can apply pattern-growth algorithms like FP-Growth for efficient pattern discovery without candidate generation. • The student can preprocess transactional data into suitable formats for mining algorithms. • The student can compare the performance of Apriori and FP-Growth in terms of time, memory, and scalability.	LAB LAB LAB LAB
LO 6.3: *Evaluate the performance and effectiveness of frequent pattern mining techniques. (PI 4.3.5 , and 2.4.5) (CO5)*	• The student can analyze algorithm efficiency based on runtime, memory usage, and scalability (e.g., Apriori vs. FP-Growth). • The student can assess the quality of patterns using metrics like support, confidence, and lift. • The student can identify trade-offs between completeness of results and computational cost. • The student can select appropriate techniques based on dataset characteristics and business objectives	ISA ISA LAB LAB