DATA WAREHOUSING WITH MINING TECHNIQUES **Course Code : 316321**

Programme Name/s	**: Artificial Intelligence/ Artificial Intelligence and Machine Learning/ Cloud Computing and Big Data/ Data Sciences/ Information Technology/ Computer Science & Information Technology**
Programme Code	**: AI/ AN/ BD/ DS/ IF/ IH**
Semester	**: Sixth**
Course Title	**: DATA WAREHOUSING WITH MINING TECHNIQUES**
Course Code	**: 316321**

I. RATIONALE

Data warehousing provides the structure and storage needed to handle large datasets, while data mining enables the extraction of useful knowledge from those datasets. Together, they empower businesses to make smarter, data-driven decisions, optimize operations, and gain a deeper understanding of their customers and markets. This course aims to equip students with the practical skills to leverage data warehousing and mining techniques.

II. INDUSTRY / EMPLOYER EXPECTED OUTCOME

Apply mining tools to extract information from data warehouse.

III. COURSE LEVEL LEARNING OUTCOMES (COS)

Students will be able to achieve & demonstrate the following COs on completion of course based learning

- CO1 - Explain the architecture, models, and processes involved in data warehousing and its distinction from operational databases.
- CO2 - Apply OLAP operations for data analysis by designing multidimensional data models.
- CO3 - Apply data mining algorithms to discover frequent item-sets and association rules.
- CO4 - Apply various classification algorithms on a data set.
- CO5 - Apply various clustering algorithms on a data set.

IV. TEACHING-LEARNING & ASSESSMENT SCHEME

Course Code	Course Title	Abbr	Course Category/s	Learning Scheme					Credits	Assessment Scheme												
				Actual Contact Hrs./Week			SLH	NLH		Paper Duration	Theory				Based on LL & TL				Based on SL		Total Marks
															Practical						
				CL	TL	LL					FA-TH	SA-TH	Total		FA-PR		SA-PR		SLA		
											Max	Max	Max	Min	Max	Min	Max	Min	Max	Min	
316321	DATA WAREHOUSING WITH MINING TECHNIQUES	DWM	DSE	3	-	2	1	6	3	3	30	70	100	40	25	10	25#	10	25	10	175

Total IKS Hrs for Sem. : Hrs
Abbreviations: CL- ClassRoom Learning , TL- Tutorial Learning, LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, IKS - Indian Knowledge System, SLA - Self Learning Assessment
Legends: @ Internal Assessment, # External Assessment, *# On Line Examination , @$ Internal Online Examination
Note :

1. FA-TH represents average of two class tests of 30 marks each conducted during the semester.
2. If candidate is not securing minimum passing marks in FA-PR of any course then the candidate shall be declared as "Detained" in that semester.
3. If candidate is not securing minimum passing marks in SLA of any course then the candidate shall be declared as fail and will have to repeat and resubmit SLA work.
4. Notional Learning hours for the semester are (CL+LL+TL+SL)hrs.* 15 Weeks
5. 1 credit is equivalent to 30 Notional hrs.
6. * Self learning hours shall not be reflected in the Time Table.
7. * Self learning includes micro project / assignment / other activities.

V. THEORY LEARNING OUTCOMES AND ALIGNED COURSE CONTENT

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
1	TLO 1.1 Describe the need of Data warehousing. TLO 1.2 Differentiate between Operational Database Systems and Data Warehouses. TLO 1.3 Differentiate the working of different data warehouse models. TLO 1.4 Differentiate the concept of data pond, data lake and data ocean.	**Unit - I Basics of Data Warehousing** 1.1 Introduction to Data Warehouse 1.2 Need of Data Warehousing 1.3 Differences between Operational Database Systems and Data Warehouses 1.4 A Multi-Tiered Architecture of Data Warehouse 1.5 Data Warehouse Models: Enterprise Warehouse, Data Mart, And virtual Warehouse 1.6 Extraction, Transformation and Loading (ETL) 1.7 Metadata Repository 1.8 Concept of data pond, data lake, data ocean	Case Study Presentations Lecture Using Chalk-Board
2	TLO 2.1 Extract data from multidimensional data models. TLO 2.2 Design schemas for multidimensional data model. TLO 2.3 Illustrate the relationship of dimensions and measures. TLO 2.4 Perform OLAP operations.	**Unit - II Data Warehouse Modelling – Data Cube and Online Analytical Processing (OLAP)** 2.1 Data Cube: A Multidimensional Data Model 2.2 Stars, Snowflakes, and Fact Constellations: Schemas for Multidimensional Data Models 2.3 Dimensions: The Role of Concept Hierarchies 2.4 Measures: Categorization and Computation 2.5 OLAP Operations - Roll-up, Drill-down, Slice and Dice	Presentations Lecture Using Chalk-Board

DATA WAREHOUSING WITH MINING TECHNIQUES **Course Code : 316321**

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
3	TLO 3.1 Describe Frequent Item sets, Closed Item sets, and Association Rules in market basket analysis. TLO 3.2 Explain Apriori algorithm. TLO 3.3 Explain the technique of Mining Frequent Item sets Using Vertical Data Format. TLO 3.4 Describe the concept of Mining Closed and Max Patterns.	**Unit - III Basics of Data Mining** 3.1 Market Basket Analysis 3.2 Frequent Item sets, Closed Item sets, and Association Rules 3.3 Apriori Algorithm: Finding Frequent Item sets by Confined Candidate Generation 3.4 Mining Frequent Item sets Using Vertical Data Format 3.5 Mining Closed and Max Patterns	Video Demonstrations Case Study Presentations Lecture Using Chalk-Board
4	TLO 4.1 Elaborate classification by learning. TLO 4.2 Explain Attribute Selection Measures for tree induction. TLO 4.3 Explain Bayes Classification Method. TLO 4.4 Apply Rule-Based Classification using IF-THEN for given data.	**Unit - IV Classification Techniques** 4.1 Introduction to Classification 4.2 Decision Tree -Decision Tree Induction, Attribute Selection Measures, Tree Pruning 4.3 Bayes Classification Methods - Bayes' Theorem, Naïve Bayesian Classification 4.4 Rule-Based Classification -Using IF-THEN Rules for Classification	Video Demonstrations Presentations Lecture Using Chalk-Board Hands-on Case Study
5	TLO 5.1 Describe features and applications of cluster analysis. TLO 5.2 Explain the given Partitioning Method for cluster analysis. TLO 5.3 Differentiate between Agglomerative and Divisive Hierarchical Clustering.	**Unit - V Cluster Analysis** 5.1 Introduction to Clustering 5.2 Cluster Analysis – Features and Applications of cluster analysis 5.3 Partitioning Methods - k-Means, k-Medoids 5.4 Hierarchical Methods- Agglomerative versus Divisive Hierarchical Clustering	Video Demonstrations Presentations Lecture Using Chalk-Board

VI. LABORATORY LEARNING OUTCOME AND ALIGNED PRACTICAL / TUTORIAL EXPERIENCES.

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 1.1 Install software for data mining.	1	Install Python along with scikit-learn, pandas, matplotlib, and DBMS	2	CO1
LLO 2.1 Implement ETL process.	2	*Implementing Extraction, Transformation and Loading process of Data Warehouse	2	CO1
LLO 3.1 Perform data mining operations such as Access specific data points, slice the cube, and aggregate multidimensional data.	3	*Write a program in python to access specific data points, slice the cube, and aggregate data of a dataset along different dimensions	2	CO2
LLO 4.1 Create star schema using SQL.	4	*Create star schema of 5 tables (one fact table and 4 dimension tables)	2	CO2
LLO 5.1 Create snowflake schema.	5	*Create snowflake schema using fact table, dimension tables and sub-dimension table	2	CO2
LLO 6.1 Create fact constellation schema.	6	Create fact constellation schema of 5 tables of student database	2	CO2

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 7.1 Implement Slice, Dice operations of OLAP.	7	Write a program in python to perform Slice, Dice operations of OLAP	2	CO2
LLO 8.1 Implement Drill-down, Roll-up operations of OLAP.	8	*Write a program in python to perform Drill-down, Roll-up operations of OLAP	2	CO2
LLO 9.1 Implement the Apriori Algorithm to solve given problem.	9	*Write a program in python to apply the Apriori Algorithm for Market Basket Analysis	2	CO3
LLO 10.1 Implement Naïve Bayesian classification algorithm.	10	*Write a program in python to apply Naïve Bayesian Classification algorithm for given STUDENT/CUSTOMER data set	2	CO4
LLO 11.1 Implement K-means algorithm.	11	*Write a program in python to apply K-means algorithm for STUDENT/CUSTOMER dataset	2	CO5
LLO 12.1 Implement K- Medoids algorithm.	12	Write a program in python to apply K- Medoids algorithm for STUDENT/CUSTOMER dataset	2	CO5
LLO 13.1 Implement Naïve Bayesian Classification to Image data set.	13	Write a program in python to apply Naïve Bayesian Classification to classify images (Use any dataset from Kaggle)	2	CO4
LLO 14.1 Implement K-medoid Clustering to Image data set.	14	Write a program in python to apply K-medoid Clustering to classify images (Use any dataset from Kaggle)	2	CO5
LLO 15.1 Implement K-means Clustering to Image data set.	15	Write a program in python to apply K-means Clustering to classify images (Use any dataset from Kaggle)	2	CO5

Note : Out of above suggestive LLOs -

- '*' Marked Practicals (LLOs) Are mandatory.
- Minimum 80% of above list of lab experiment are to be performed.
- Judicial mix of LLOs are to be performed to achieve desired outcomes.

VII. SUGGESTED MICRO PROJECT / ASSIGNMENT/ ACTIVITIES FOR SPECIFIC LEARNING / SKILLS DEVELOPMENT (SELF LEARNING)

Micro project

- Perform various operations of data warehousing and data mining for any data set from kaggle.com using jupyter notebook.
- Implement Partitioning Methods like K-Means clustering or k-Medoids using C/CPP/JAVA to group similar data points.
- Perform data mining operations on image data set.

Assignment

- Set up a data warehouse for a BI dashboard (using Tableau, Power BI, etc.).
- Perform various operation using ETL process such as extract data from various data sources, integrate that data, clean that data and transform data from one DB to another DB.
- Extract data from a data warehouse and apply data mining techniques (e.g., classification or clustering) to derive insights.

Other

- Complete course -"Hands On Machine Learning For Data Mining" from Infosys Springboard.
- Complete course -"Introduction to Data Mining" from Infosys Springboard.
- Complete course -"Data Mining with Python: Implementing classification and regression"from Infosys Springboard.

Note :

- Above is just a suggestive list of microprojects and assignments; faculty must prepare their own bank of microprojects, assignments, and activities in a similar way.
- The faculty must allocate judicial mix of tasks, considering the weaknesses and / strengths of the student in acquiring the desired skills.
- If a microproject is assigned, it is expected to be completed as a group activity.
- SLA marks shall be awarded as per the continuous assessment record.
- For courses with no SLA component the list of suggestive microprojects / assignments/ activities are optional, faculty may encourage students to perform these tasks for enhanced learning experiences.
- If the course does not have associated SLA component, above suggestive listings is applicable to Tutorials and maybe considered for FA-PR evaluations.

VIII. LABORATORY EQUIPMENT / INSTRUMENTS / TOOLS / SOFTWARE REQUIRED

Sr.No	Equipment Name with Broad Specifications	Relevant LLO Number
1	PC-i3 or above with minimum 4GB RAM PYTHON 3 with scikit-learn, pandas, matplotlib Oracle/MySQL/SQL Server MS-Excel,WEKA	All

IX. SUGGESTED WEIGHTAGE TO LEARNING EFFORTS & ASSESSMENT PURPOSE (Specification Table)

Sr.No	Unit	Unit Title	Aligned COs	Learning Hours	R-Level	U-Level	A-Level	Total Marks
1	I	Basics of Data Warehousing	CO1	12	4	10	4	18
2	II	Data Warehouse Modelling – Data Cube and Online Analytical Processing (OLAP)	CO2	10	4	8	4	16
3	III	Basics of Data Mining	CO3	8	2	8	4	14
4	IV	Classification Techniques	CO4	8	2	6	4	12
5	V	Cluster Analysis	CO5	7	0	4	6	10
Grand Total				**45**	**12**	**36**	**22**	**70**

X. ASSESSMENT METHODOLOGIES/TOOLS

Formative assessment (Assessment for Learning)

- Continuous assessment based on process and product related performance indicators.
- Each practical will be assessed considering 60% weightage to process, 40% weightage to product.
- A continuous assessment based term work.

Summative Assessment (Assessment of Learning)

- End semester examination, Lab performance, Viva voce

XI. SUGGESTED COS - POS MATRIX FORM

Course Outcomes (COs)	Programme Outcomes (POs)							Programme Specific Outcomes* (PSOs)		
	PO-1 Basic and Discipline Specific Knowledge	PO-2 Problem Analysis	PO-3 Design/ Development of Solutions	PO-4 Engineering Tools	PO-5 Engineering Practices for Society, Sustainability and Environment	PO-6 Project Management	PO-7 Life Long Learning	PSO-1	PSO-2	PSO-3
CO1	2									
CO2	2	3	3	2						
CO3	3	3	2	3	1					
CO4		2	2	3		2	1			
CO5		2	2	3		2	1			

Legends :- High:03, Medium:02,Low:01, No Mapping: -
*PSOs are to be formulated at institute level

XII. SUGGESTED LEARNING MATERIALS / BOOKS

Sr.No	Author	Title	Publisher with ISBN Number
1	Jiawei Han, Micheline Kamber, Jian Pei	Data Mining Concepts and Techniques	Morgan Kaufmann Publishers, ISBN 978-0-12-381479-1
2	Alex Berson, Stephen Smith	Data Warehousing, Data Mining and OLAP	McGraw Hill, ISBN-13 - 978-0070587410
3	Parteek Bhatia	Data Mining and Data Warehousing: Principles and Practical Techniques	Cambridge University Press, ISBN-13 978-1108727747
4	Avi Silberschatz, Henry F. Korth, S. Sudarshan	Database System Concepts (Seventh Edition)	McGraw-Hill ISBN 9780078022159

XIII . LEARNING WEBSITES & PORTALS

Sr.No	Link / Portal	Description
1	https://www.analyticsvidhya.com/blog/category/data-mining/	Data Mining blog
2	https://nptel.ac.in/courses/106105174	NPTEL Data Warehousing & Mining Video lectures
3	https://www.oracle.com/database/technologies/datawarehouse	Use for Data warehousing
4	https://www.oreilly.com/library/view/what-is-a/9781492088899/ch01.html	Data pond, Data lake, Data ocean

Note :

- Teachers are requested to check the creative common license status/financial implications of the suggested online educational resources before use by the students