INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

KNOWLEDGE DISCOVERY AND DATA MINING								
IV Semester: CSE (Artificial Intelligence & Machine Learning) | CSE (DS)								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
ACSE16	**Core**	**L**	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
		3	0	0	3	40	60	100
Contact Classes: 48	**Tutorial Classes: NIL**	**Practical Classes: NIL**				**Total Classes: 48**		
Prerequisite: Python Programming								

I. COURSE OVERVIEW:

Data mining refers to extracting or mining knowledge from large amounts of data. It emphasizes various techniques and algorithms used to explore, analyze and leverage data and turn it into valuable and actionable information. It includes data warehousing and data mining functionalities such as association mining, classification, clustering and outlier analysis. The techniques are used to tackle data centric applications in various domains such as financial analysis, telecommunication industry, intrusion detection, and complex data mining applications in stream, web, text, spatial and other scientific applications.

II. COURSES OBJECTIVES:

The students will try to learn:

I. The representation, search procedures, and various logical reasoning systems.
II. A framework for modeling human reasoning and dealing with uncertainty through fuzzy logic.
III. Applications of game playing, designing communicating agents, and robotic technology.

III. COURSE OUTCOMES:

At the end of the course students should be able to:

CO1 Explain appropriate knowledge representation techniques to generate inferences for any search problem.
CO2 Outline knowledge that allows systematic and rigorous reasoning to enhance clarity and precision of knowledge representation.
CO3 Examine the situations that is inherently fuzzy or ambiguous while designing logical systems to deal such data.
CO4 Choose reverse engineering human capabilities and apply them to develop intelligent systems.
CO5 Build new architectures, designs for communicating agents and robotic applications.
CO6 Select design technologies of AI for enhanced learning, reasoning, and perception to simulate or approximate human intelligence.

IV. COURSE SYLLABUS:

MODULE 1: KNOWLEDGE DISCOVARY (09)

Data Mining Definition, Knowledge Discovery In Data (Kdd), Kinds Of Data Can Be Mined, Kinds Of Patterns / Data Mining Functionalities, Technologies, Applications, Issues In Data Mining. Data Objects And Attribute Types, Basic Statistical Descriptions Of Data, Data Visualization, Measuring Data Similarity And Dissimilarity.

MODULE 2: DATA PREPROCESSING (08)

Data Preprocessing: Data Quality, Major Tasks In Data Preprocessing, Data Cleaning, Data Integration And Transformation, Data Reduction, Data Discretization.

MODULE 3: DATA WAREHOUSING AND ONLINE ANALYTICAL PROCESSING (09)

Data Warehouse Concepts, Differences Between Operational Database Systems And Data Warehouses, a Multitiered Architecture; Data Warehouse Models: Enterprise Warehouse, Data Mart, And Virtual Warehouse, Extraction, Transformation, And Loading, Metadata Repository, a Multidimensional Data Model; Schemas For Multidimensional

Data Models: Stars, Snowflakes, And Fact Constellations, Dimensions, Measures, Olap Operations, a Starnet Query Model For Querying Multidimensional Databases.
Business Analysis Framework For Data Warehouse Design, Data Warehouse Design Process, Data Warehouse Implementation, Indexing Olap Data, Olap Server Architectures, Data Generalization By Attribute, Oriented Induction

MODULE 4: MINING FREQUENT PATTERNS AND CLASSIFICATION (10)

Market Basket Analysis, Frequent Item sets, Closed item sets, And Association Rules, Frequent Itemset Mining Methods; Apriori Algorithm, Generating Association Rules From Frequent item sets, Improving The Efficiency Of Apriori Pattern-Growth Approach.

Classification: Basic Concepts, Decision Tree Induction, Bayesian Belief Networks, Classification By Back Propagation, Support Vector Machines, Classification Using Frequent Patterns, Lazy Learners, Other Classification Methods, Model Evaluation And Selection, Techniques To Improve Classification Accuracy.

MODULE 5: CLUSTERING AND RESEARCH FRONTIERS (09)

Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid Based Methods, Evaluation Of Clustering.

Mining Complex Types Of Data: Mining Sequence Data: Time-Series, Symbolic Sequences, And Biological Sequences, Mining Graphs And Networks

V.TEXTBOOKS:

1. Jiawei Han, MichelineKamber, "Data Mining,Concepts and Techniques", Morgan Kaufmann Publishers, Elsevier, 3rd Edition, 2012.
2. Alex Berson, Stephen J.Smith, "Data warehousing Data mining and OLAP", Tata McGraw-Hill, 2nd Edition, 2007.

VI.REFERENCE BOOKS:

1. Arum K Pujari, "Data Mining Techniques", Universities Press, 3rd Edition, 2005.
2. PualrajPonnaiah, "Data Warehousing Fundamentals", Wiley, Student Edition, 2004.
3. Ralph Kimball, "The Data Warehouse Life Cycle Toolkit", Wiley, Student Edition, 2006.
4. VikramPudi, P Radha Krishna, "Data Mining", Oxford University, 1st Edition, 2007.

VII. ELECTRONICS RESOURCES:

a. https://www.anderson.ucla.edu
b. https://www.smartzworld.com
c. http://iiscs.wssu.edu

VIII. MATERIALS ONLINE

1. Course Template
2. Tutorial Question Bank
3. Definition and Terminology
4. Tech-Talk topics
5. Assignments
6. Model question paper - I
7. Model question paper - II
8. Lecture notes
9. Early learning readiness videos (ELRV)
10. PowerPoint Presentations