



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

BUSINESS DATA MINING, WAREHOUSING AND VISULIZATION								
IV Semester: MBA								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
CMBE61	Elective	L	T	P	C	CIA	SEE	Total
		4	-	-	4	40	60	100
Contact Classes: 45	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 45			
Prerequisite: Business Analytics								
SDGs Mapped: SDG 9 (Industry, Innovation and Infrastructure), SDG 12 (Responsible Consumption & Production)								

I. COURSE OVERVIEW:

This course provides foundational knowledge and hands-on skills in data mining, data preprocessing, warehouse management, and business data visualization. Students will explore techniques to clean, transform, and reduce data, apply logistic regression for classification problems, and design dashboards for effective decision-making. Emphasis is placed on practical applications, case studies, and IT tools for solving real-world business problems.

II. OBJECTIVES:

The students will try to learn:

- I. The fundamental concepts of data mining, data warehouses, and societal implications.
- II. The preprocessing techniques for cleaning, transforming, and reducing data.
- III. The various warehousing operations, efficiency measurement, and workforce management.
- IV. The various classification methods using logistic regression and performance evaluation.
- V. The ability to visualize business data through dashboards and monitoring systems.

III. COURSE OUTCOMES:

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

- CO1 Explain the concepts of data mining, types of data, and major issues in mining.
- CO2 Apply preprocessing techniques such as cleaning, reduction, and transformation to real-world datasets.
- CO3 Analyze warehouse operations and evaluate efficiency, workforce design, and risk management.
- CO4 Build and evaluate logistic regression models for classification problems using business datasets.
- CO5 Design and develop dashboards for data visualization and business activity monitoring.
- CO6 Integrate data mining, warehouse management, and visualization techniques for effective business decision-making.

IV. COURSE CONTENT:

MODULE-I: INTRODUCTION TO DATA MINING, WAREHOUSE AND VISULIZATION: (08)

Data Mining; Kinds of data that can be mined-Database Data, Data Warehouses, Transactional Data, Other Kinds of Data; Major Issues in Data Mining-Mining Methodology, User Interaction, Efficiency and Scalability, Diversity of Database Types, Data Mining and Society.

MODULE –II: DATA PRE-PROCESSING: (10)

Data Preprocessing: An Overview-Reasons to process the data, Major Tasks in Data Preprocessing; Data Cleaning-Missing Values, Noisy Data, Data Cleaning as a Process; Data Reduction-Principal Component Analysis, Histograms, Clustering, Sampling, Data Cube Aggregation; Data Transformation and Data Discretization-Data Transformation by Normalization, Discretization by Binning, Discretization by Histogram Analysis.

Case Study: Handling Missing Values in Melbourne Housing Price Data.

MODULE –III: MANAGING WAREHOUSE EFFICIENCY: (09)

Order picking - Picking methods-pick path - Measuring Warehouse Efficiency - Warehouse Workforce design and development - cross docking.

Warehousing Operations: warehousing operations- inbound process, outbound processes, Functions of Warehouse-break-bulk, cross docking, order mixing, Risk management

MODULE –IV: CLASSIFICATION PROBLEMS – I-LOGISTICS REGRESSION: (10)

Overview of Classification; Binary Logistic Regression; Classification-Encoding Categorical Features, Building Logistic Regression Model, Printing Model Summary, Predicting on Test Data; Measuring Accuracies-Creating Confusion Matrix, Receiver Operating Characteristic (ROC) and Area Under the Curve; Finding Optimal Classification Cut-off -Youden's Index and Cost-Based Approach.

Case Study: Predicting Employee Attrition on HR Attrition Dataset.

MODULE –V: DATA VISUALISATION FOR BUSINESS: (08)

Introduction to Data Visualization for businesses, Visualization of Numerical and Non-Numerical Data Creation of Dashboards using IT Tools, Business Activity Monitoring through Dashboard.

V. TEXT BOOKS:

1. Ikvinderpal “Data Mining & Warehousing”, 1st edition, 2020.
2. **Max Bramer and Springer “Principles of Data Mining 2020” 8th edition, 2020.**
3. Szabo, Gungar Polatkan, Oscar Boykin, Chalkiopoulos, “Social Media Data Mining and Analytics”, 3rd edition. Wiley, 2019.
4. Pang-Ning Tan, “Introduction to Data Mining, Global Edition”, 4th edition, May 2019.
5. Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, “Data Mining: Practical Machine Learning Tools and Techniques”, 4th edition. Elsevier, 2017.

VI. REFERENCE BOOKS:

1. Luis Torgo, “Data Mining with R: Learning with Case Studies”, 2nd edition, CRC Press, 2011.
2. Jiawei Han, Jian Pei, Micheline Kamber, “Data Mining: Concepts and Techniques”, 3rd edition, Elsevier, 2010.
3. Joseph B. Pigus, “Data Mining with Neural Networks”, 2nd edition, TMH, 2017.
4. Robert Layton, “Learning Data Mining with Python”, 2nd edition, PACKT Publishing, 2015.

VII. WEB REFERENCES:

1. https://www.amazon.in/dp/B07YG4QSZR/ref=cm_sw_r_apan_glt_2FGRNQAEBE4AEV5JQRB7?_encoding=UTF8&psc=1

VIII. E-TEXT BOOKS:

1. https://www.amazon.in/dp/B00UVBJSQA/ref=cm_sw_r_apan_glt_1Y1H0P2MRK1KM0SFG1AJ?_encoding=UTF8&psc=1.
2. https://www.amazon.in/dp/B075GB7FT5/ref=cm_sw_r_apan_glt_G8RNA0P6W9YSQZ2N0P16?_encoding=UTF8&psc=1.