GANPAT UNIVERSITY									
FACULTY OF MANAGEMENT STUDIES									
Programme	Masters of Business Administration				Branch/Spec.	Business Analytics			
Semester	III				Version	1.0.0.0			
Effective from Academic Year		2021-22			Effective for the batch Admitted in			June 2020	
Subject code	**IIIA04DAM**	Subject Name			**Data Mining and Data Warehousing**				
Teaching scheme					Examination scheme (Marks)				
(Per week)	Lecture (DT)		Practical (Lab.)	Total		CE	SEE	Total	
	L	TU	P	TW					
Credit	4				04	Theory	60	40	100
Hours	4				60	Practical			

Course Objectives:

CO 1: To Understand the concept of data mining.
Co 2: To understand data processing and classification.
Co 3: To learn clustering and association analysis.
Co 4: To understand data warehousing and data architecture.
Co 5: To learn data design and information delivery.
Co 6:- To learn testing and implementation.

Theory Syllabus

Unit	Content	Hrs
1	Introduction to Data Mining Concept, Definitions and Need of Big Data, Data Mining, Business Intelligence. Data Mining Process, Business Intelligence techniques. Introduction to Data Mining Tasks (Classification, Clustering, Association Analysis, Anomaly Detection). Concept, Definitions of model, descriptive models, predictive modeling. Real-world data mining applications - Big Data Analytics in Business Environment, Fraud Detection and Prevention with Data Mining Techniques.	10
2	Data Preprocessing and Classification Understanding of Data,Types of attributes, properties of attribute values, types of data, data quality, Sampling, Data Normalization, Data Cleaning, Similarity Measures, Feature Selection/Instance Selection with importance. Classification: Decision-Tree Based Approach, Rule-based Approach, Instance-based classifiers, Support Vector Machines, Ensemble Learning, Model Selection and Evaluation, Applications: B2B customer buying stage prediction, Recommender Systems.	10
3	Clustering and Association Analyisis Partitional and Hierarchical Clustering Methods, Graph-based Methods, Density-based Methods, Cluster Validation, Applications: Customer Profiling, Market Segmentation. Association Analysis: Apriori Algorithm and its Extensions, Association Pattern Evaluation, Sequential Patterns and Frequent Subgraph Mining, Applications: B2B Customer Buying Path Analysis, Medical Informatics, Telecommunication alarm diagnosis. Anomaly Detection: Statistical-based and Density-based Methods , Ethics of data mining, privacy	10
4	Data Warehousing and Data Architecture Concept of Data Warehousing, Applications and Drivers of Business Intelligence , Data Shadow Systems, Data Warehouse Project, The Project Team, Project Management Considerations, Defining the Business Requirements, Dimensional Analysis, Information Packages, Requirements Gathering	10

Note: Version 1.0.0.0 (First Digit= New syllabus/Revision in Full Syllabus, Second Digit=Revision in Teaching Scheme,Third Digit=Revision in Exam Scheme, Forth Digit= Content Revision)

L=Lecture, TU=Tutorial, P= Practical/Lab., TW= Term work, DT= Direct Teaching, Lab.= Laboratory work

	Methods, Requirements as the Driving Force for Data Warehousing, Data Design,Data Storage Specifications, Information Delivery Strategy. Data Architecture: Concept, features, Architectural Framework, Technical Architecture, Infrastructure Supporting Architecture, Concept of Meta data, Importance, Metadata Types,	
5	Data Design And Information delivery: Concept of data design, Principles of Dimensional Modeling, The STAR Schema, STAR Schema Keys- Advantages, Updates , Dimensions, The Snowflake Schema, Aggregate Fact Tables, Families of STARS, Data Extraction, Data Transformation, Data Loading, ETL Summary, Data Quality – Challenges and Tools. Information Delivery: Concept, Tools, Users, Features and Functions, OLAP Models, OLAP Implementation, Building Web-Enabled Data Warehouse, Web-Based Information Delivery, OLAP and the Web.	10
6	Testing and Implementation Data volume, Test techniques, ETL, The Physical Design Process and Considerations, Physical Storage, Indexing the Data Warehouse, Performance Enhancement Techniques, Data Warehouse Deployment, Major Deployment Activities, Considerations for a Pilot, Security, Backup and Recovery, Growth and Maintenance, Monitoring the Data Warehouse, User Training and Support, Managing the Data Warehouse, Scoping, Estimation, proposal writing and Risk planning.	10
Reference Books:		
1	Data Mining: The Textbook by Charu C. Aggarwal	
2	Data Science for Business by Foster Provost and Tom Fawcett, O'Reilly	
3	Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Addison Wesley	
4	Data Mining and Analysis: Fundamental Concepts and Algorithms by Mohammed J. Zaki and Wagner Meira	
5	The Data Warehouse Lifecycle Toolkit, Ralph Kimball	
6	Data Warehouse Project Management , Sid Adelman, Larissa T. Moss	
7	The Data Warehouse Lab: A step-by-step guide using SSIS and SSAS, Amin Jalali	
8	Data Warehouse for Project Managers, Laura Reeves	
9	The Profit Impact of Business Intelligence, Steve Williams , Nancy Williams	

Note: Version 1.0.0.0 (First Digit= New syllabus/Revision in Full Syllabus, Second Digit=Revision in Teaching Scheme,Third Digit=Revision in Exam Scheme, Forth Digit= Content Revision)

L=Lecture, TU=Tutorial, P= Practical/Lab., TW= Term work, DT= Direct Teaching, Lab.= Laboratory work