



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Computer Engineering (Artificial Intelligence And Data Science)

Course / Subject Code : ME01095031

Subject Name: Data Warehousing and Data Mining

WEF Academic Year :	2024-25
Semester :	1 st Semester
Category of the Course :	PEC

Prerequisite :	Database management, programming language like (SQL,Python), data structure
Rationale :	This course in Data Warehousing data mining focuses on understanding the research methods used in the field of data mining. The course targets students who want to gain in-depth knowledge about a particular data mining topic

Course Scheme :

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	PR	C	Theory		Practical		
				ESE (E)	PA(M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Course Content :

Sr. No.	Course Content	No. of Hours	% of Weightage
1	Basics of data warehousing and data mining Data Warehousing and Business Analysis: - Data warehousing Components –Building a Data warehouse –Data Warehouse Architecture – DBMS Schemas for Decision Support – Data Extraction, Clean up, and Transformation Tools –Metadata – reporting – Query tools and Applications – Online Analytical Processing (OLAP) – OLAP and Multidimensional Data Analysis. Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process	6	15



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Computer Engineering (Artificial Intelligence And Data Science)

Course / Subject Code : ME01095031

Subject Name: Data Warehousing and Data Mining

2	<p>Basics of classification and prediction and clustering Classification and Prediction: - Classification by Decision Tree, classification by back propagation, Associative Classification – Lazy Learners Prediction – Accuracy and Error Measures – Evaluating the Accuracy of a Classifier or Predictor – Ensemble Methods – Model Selection, Classification vs. prediction</p> <p>Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods – Clustering High-Dimensional Data – Constraint-Based Cluster Analysis – Outlier Analysis.</p>	7	17
3	<p>Mining Time series Data & Data Streams Periodicity Analysis for Time-Related Sequence Data, Trend analysis, Similarity search in Time-series analysis</p> <p>Methodologies for stream data processing and stream data systems, Frequent pattern mining in stream data, Sequential Pattern Mining in Data Streams, Classification of dynamic data streams, Class Imbalance Problem</p>	7	17
4	<p>Sequential Pattern Mining Concepts, primitives, scalable methods for mining sequential patterns – GSP, SPADE, PrefixSpan, Mining Closed Sequential Patterns, Mining Multidimensional and Multilevel sequential Patterns.</p>	5	12
4	<p>Graph mining and Web mining Mining frequent subgraphs, finding clusters, hub and outliers in large graphs, Graph Partitioning; Web Mining, Mining the web page layout structure, mining web link structure, mining multimedia data on the web, Automatic classification of web documents and web usage mining.</p>	6	15



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Computer Engineering (Artificial Intelligence And Data Science)

Course / Subject Code : ME01095031

Subject Name: Data Warehousing and Data Mining

5	Big Data Mining and Analytics Big Data: Introduction to big data: distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.	10	24
Total			100

Reference Book :

1. “Data Mining Concepts and Techniques” .2nd Edition, Elsevier, 2011. J Han and M Kamber.
2. “Introduction to Data Mining”. Addison Wesley,2006, Pang Ning Tan, M Steinbach, Vipin Kumar,
3. “Sequence Data Mining”. Springer, 2007, G Dong and J Pei
4. “Data mining: Concepts, models, methods and algorithms “, M. Kantardzic, John Wiley & Sons Inc.
5. “Data Warehousing Fundamentals”, Paulraj Ponnian, John Willey.
6. “Data Mining: Introductory and Advanced Topics”,M. Dunham, Pearson Education.
7. “Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner”, G. Shmueli, N.R. Patel, P.C. Bruce, Wiley India.
8. “Mining of Massive Datasets,” Anand Rajaraman, Jeffrey David Ullman. (2011), Cambridge University Press, [ISBN: 978-1107015357].
9. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication
10. Raj Kamal, “Internet of Things: Architecture and Design Principles, Mc Graw Hill Education
11. Hanes et al “IoT Fundamentals”, Cisco Press
12. Vijay Madiseti and Arshdeep Bahga, “Internet of Things (A Hands-on-Approach)”, , Paperback, 2015.
13. A. McEwen, H. Cassimally, “Designing the Internet of Things”, Wiley, 2013.
14. Yashwant Kanetkar, “21 Internet of Things Experiments”, Kindle edition



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Computer Engineering (Artificial Intelligence And Data Science)

Course / Subject Code : ME01095031

Subject Name: Data Warehousing and Data Mining

15. Adeel Javed, “Building Arduino projects for Internet of Things”, Apress publication
16. Donald Noris, “The Internet of Things: Do it yourself Projects with Arduino, Raspberry PI and
17. BeagleBone Black” Mc Graw Hill Publication
18. Adrian McEwen & Hakim Cassimally, “Designing the Internet of things”, Willey publication
19. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication
20. Raj Kamal, “Internet of Things: Architecture and Design Principles, Mc Graw Hill Education
21. Hanes et al “IoT Fundamentals”, Cisco Press
22. Vijay Madiseti and Arshdeep Bahga, “Internet of Things (A Hands-on-Approach)”, , Paperback, 2015.
23. A. McEwen, H. Cassimally, “Designing the Internet of Things”, Wiley, 2013.
24. Yashwant Kanetkar, “21 Internet of Things Experiments”, Kindle edition
25. Adeel Javed, “Building Arduino projects for Internet of Things”, Apress publication
26. Donald Noris, “The Internet of Things: Do it yourself Projects with Arduino, Raspberry PI and
27. BeagleBone Black” Mc Graw Hill Publication
28. Adrian McEwen & Hakim Cassimally, “Designing the Internet of things”, Willey publication
29. Rahul Dubey, “An Introduction to Internet of Things: Connecting Devices, Edge Gateway, and Cloud with Applications”, Cengage India Publication

Course Outcome:

After Completion of the Course, Student will able to :

No	Course Outcomes	RBT Level*
01	Design a data mart or data warehouse for any organization	AP
02	Analyze Algorithms for sequential patterns and Time series data	AN
03	Understand Data Stream Mining models and methods	UN
04	Apply Graph mining algorithms to Web Mining	AP
05	Distinguish computing frameworks for Big Data analytics.	EL
06	Describe different methodologies used in distributed data mining	AN

*RM: Remember, UN: Understand, AP: Apply, AN: Analyse, EL: Evaluate, CR: Create

Suggested Course Practical List :

1. Creation of Data Warehouse/Data Mart for real world problem
2. Find a Frequent pattern from given data set



GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Computer Engineering (Artificial Intelligence And Data Science)

Course / Subject Code : ME01095031

Subject Name: Data Warehousing and Data Mining

3. Find a Maximal Frequent pattern from given dataset
4. Find a Closed Frequent pattern from given dataset
5. Perform time series analysis on given data set
6. Learn any data streaming tools
7. Python Web Scraping
8. Perform Network Mining using any language
9. Implementing any one Frequent Item set algorithm using Map-Reduce
10. Implement Twitter Sentiment Analysis

* * * * *