

<b>COURSE TITLE</b>	<b>DATA SCIENCE</b>
<b>COURSE CODE</b>	<b>01ML0401</b>
<b>COURSE CREDITS</b>	<b>2</b>

**Objective:**

- 1 The course aim to provide students with practical exposure to techniques used for discovering patterns, trends, and relationships within large datasets. Through hands-on experiments using tools such as Python, SQL, Tableau, and Power BI, students learn how to preprocess data, perform exploratory analysis, implement data mining algorithms, and visualize insights effectively.

**Course Outcomes:** After completion of this course, student will be able to:

- 1 Apply data acquisition and preprocessing techniques using tools such as Python and SQL to prepare datasets for data mining tasks.
- 2 To Analyze datasets using exploratory data analysis and visualization techniques to identify patterns, correlations, and anomalies.
- 3 Apply and analyze data mining algorithms such as classification, clustering, and association rule mining to extract meaningful knowledge from data.
- 4 Evaluate the performance of data mining models using appropriate evaluation metrics and interpret the results for decision-making.
- 5 Develop interactive data visualizations and analytical dashboards using tools such as Tableau and Power BI to communicate insights effectively.
- 6 Design and implement a complete data mining workflow integrating data extraction, preprocessing, modeling, and visualization for solving real-world problems.

**Pre-requisite of course:**Python programming, Basic statistics and probability

**Teaching and Examination Scheme**

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
0	0	4	0	0	0	50	50

Contents : Unit	Topics	Contact Hours
<b>Total Hours</b>		

**Suggested List of Experiments:**

Contents : Unit	Topics	Contact Hours
1	<b>Experiment 1</b> To analyze the structure of real-world datasets by identifying feature types, dataset dimensions, and statistical characteristics for further data mining tasks.	4

**Suggested List of Experiments:**

<b>Contents : Unit</b>	<b>Topics</b>	<b>Contact Hours</b>
2	<b>Experiment 2</b> To prepare raw datasets for analysis by performing data cleaning, handling missing values, removing inconsistencies, and transforming data into suitable formats for mining.	4
3	<b>Experiment 3</b> To retrieve, filter, and aggregate structured data using SQL queries in order to support data mining and analytical operations.	6
4	<b>Experiment 4</b> To analyze datasets using statistical summaries and exploratory techniques to discover patterns, relationships, and anomalies within the data.	4
5	<b>Experiment 5</b> To visualize datasets using graphical techniques in order to interpret trends, distributions, and relationships among variables.	4
6	<b>Experiment 6</b> To create interactive visualizations and dashboards using Tableau for effective data exploration and decision-making.	4
7	<b>Experiment 7</b> To develop analytical dashboards that present key performance indicators and insights derived from data mining processes.	4
8	<b>Experiment 8</b> To apply classification algorithms to categorize data instances and evaluate the performance of predictive models.	4
9	<b>Experiment 9</b> To discover hidden patterns and group similar data points into clusters for knowledge discovery and segmentation analysis.	4
10	<b>Experiment 10</b> To identify relationships and correlations among variables using association rule mining techniques such as the Apriori algorithm.	4
11	<b>Experiment 11</b> To reduce dataset dimensionality while preserving important information in order to improve analysis efficiency and visualization.	4
12	<b>Experiment 12</b> To implement an end-to-end data mining pipeline involving data acquisition, preprocessing, visualization, model building, and interpretation of results.	6
<b>Total Hours</b>		<b>52</b>

**Textbook :**

- 1 Data Mining: Concepts and Techniques, Han, J., Kamber, M., & Pei, J., Morgan Kaufmann (Elsevier), 2011
- 2 Introduction to Data Mining, Tan, P. N., Steinbach, M., & Kumar, V., 2nd Edition, Pearson Education, Boston, MA, USA, 2018. ISBN: 978-0133128901, 2018

**References:**

- 1 Data Mining: Practical Machine Learning Tools and Techniques, Data Mining: Practical Machine Learning Tools and Techniques, I. H., Frank, E., Hall, M. A., & Pal, 4th Edition, Morgan Kaufmann (Elsevier), 2016
- 2 Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Géron, A., 2nd Edition, O'Reilly Media, Sebastopol, 2019
- 3 Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, Bruce, Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, Bruce, P., Bruce, A., & Gedeck, P., 2nd Edition, O'Reilly Media, Sebastopol, CA, USA, 2020
- 4 Storytelling with Data: A Data Visualization Guide for Business Professionals, Storytelling with Data: A Data Visualization Guide for Business Professionals, Knaflic, C. N., 1st Edition, John Wiley & Sons, , Hoboken, NJ, USA, 2015
- 5 SQL for Data Analysis: Advanced Techniques for Transforming Data into Insights, SQL for Data Analysis: Advanced Techniques for Transforming Data into Insights, Tanimura, C., O'Reilly Media, 2021

**Suggested Theory Distribution:**

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation					
Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking / Creative
0.00	0.00	50.00	25.00	25.00	0.00

**Instructional Method:**

- 1 PPT
- 2 Computational Facility and programming software

**Supplementary Resources:**

- 1 <https://datasetsearch.research.google.com/>
- 2 <https://scikit-learn.org/>
- 3 <https://public.tableau.com/>
- 4 <https://learn.microsoft.com/power-bi>