COURSE CONTENT

DATA MINING AND MACHINE LEARNING							
V Semester: CSE | IT | CSE (CS)							
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
ACSD19	Core	**L**	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
		3	0	0	3	40	60	100
Contact Classes:48	**Total Tutorials: Nil**	**Total Practical Classes: Nil**				**Total Classes: 48**		
Prerequisites: Database Management Systems								

I. COURSE OVERVIEW:

The Data Mining and Machine Learning course provides a deep dive into the theories and practical applications of data analysis, mining, and machine learning techniques. Students will explore how to extract meaningful insights from large datasets, recognize patterns, and use predictive modeling techniques. The course also delves into machine learning algorithms and their implementation in solving real-world problems across various domains, including healthcare, finance, and technology.

II. COURSE OBJECTIVES:

The students will try to learn:

I. The core concepts and techniques of data mining and machine learning.
II. The data preprocessing and feature engineering methods to real-world datasets.
III. The various supervised and unsupervised machine learning algorithms.
IV. The advanced topics in data mining, such as clustering, anomaly detection, and association rule mining.

III. COURSE OUTCOMES:

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

CO1 Understand the fundamental concepts of data mining and machine learning algorithms.
CO2 Apply data preprocessing and transformation techniques to prepare data for mining and learning.
CO3 Analyze classification, clustering, and association algorithms for different data-driven problems.
CO4 Evaluate the performance of machine learning models using suitable metrics and validation techniques.
CO5 Implement various supervised and unsupervised learning algorithms using tools like Python or R.
CO6 Design data-driven applications using appropriate mining and learning techniques to solve real-world problems.

IV. COURSE CONTENT:

MODULE 1: INTRODUCTION (9)

Introduction to Data Mi9ning: Data Mining, Knowledge Discovery from Data (KDD), Data Mining integrating with Datawarehouse, Data to be mined, Patterns to be mined, Data objects and Attributes, statistical description of data, Measures of Similarity and Dissimilarity-Basics, Introduction to Machine Learning: Classic and Adaptive Machines, Learning Problems and Scenarios, Type of Machine Learning, Important Elements in Machine Learning

MODULE 2: DATA PREPROCESSING AND MINING ASSOCIATION RULES (10)

Data Preprocessing: Overview, Major Tasks in Data Pre-processing: Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. Association Rules: Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods Frequent Itemsets, Closed Itemsets, and Association Rules. Apriori Algorithm: Finding Frequent Itemsets by Confined Candidate Generation, Generating Association Rules from Frequent Itemsets, Improving the Efficiency of Apriori, FP-Growth Algorithm, Association Analysis to Correlation Analysis.

MODULE 3: SUPERVISED LEARNING (10)

Supervised Learning: Basic Concepts, Attribute selection measures, Decision Tree Classification, Bayes Classification Methods, Metrics for Evaluating Classifier Performance, Classification by Backpropagation.

Support Vector Machines (SVM), k-Nearest-Neighbor Classifiers (KNN), Linear Regression, Logistic Regression

MODULE 4: UNSUPERVISED LEARNING (10)

Cluster Analysis: Basic Concepts of clustering Analysis, Partitioning Methods: k-Means, k-Medoids, Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering, ds, DBSCAN: Density-Based Clustering, Grid-Based Methods: STING, CLIQUE, Evaluation of Clustering, Probabilistic model-based clustering, Graph Clustering Method.

MODULE 5: ENSEMBLE LEARNING AND REINFORCEMENT LEARNING (9)

Ensemble Learning Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, boosting: Adaboost, Gradient Boosting, Xg boost, Stacking,
Reinforcement learning overview, Elements of Reinforcement learning, Model based learning, Temporal difference learning

V. TEXT BOOKS:

1. Data Mining: Concepts and Techniques Jiawei Han, Micheline Kamber, Jian Pei (2012), 3rd Edition, Elsevier, United States of America.
2. Giuseppe Bonaccorso Machine Learning Algorithms 2nd Edition, Packt, 2018.
3. Introduction to Machine Learning, Ethem Alpaydın, The MIT Press, Cambridge, Massachusetts, London, 2nd Edition 2019. (M5)

VI. REFERENCE BOOKS:

1. Kevin P. Murphy, "*Machine Learning: A Probabilistic Perspective*", MIT Press, 1st Edition, 2012
2. Ethem Alpaydin, *Introduction to Machine Learning*, MIT Press, 4th Edition, 2020.
3. Andreas C. Müller, Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly Media, 1st Edition, 2016.

VII. ELECTRONICS RESOURCES:

1. https://www.coursera.org/learn/machine-learning /
2. https://www.kaggle.com /
3. https://cloud.google.com/docs

VIII. MATERIALS ONLINE

1. Course template
2. Tutorial question bank
3. Tech talk topics
4. Open-ended experiments
5. Definitions and terminology

6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. PowerPoint presentation
11. E-Learning Readiness Videos (ELRV)