Course Code	Course Title	L	T	P	C
PMDS505L	Data Mining and Machine Learning	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understand the role of separate database for decision making.
2. Learn the core ideas of data mining techniques in different case studies.
3. Inculcate the concept learning and Machine learning theory.

Course Outcomes

At the end of the course -students will be able to:

1. Gain knowledge over the importance of KDD and Data Mining
2. Recognize the key areas and issues in data mining.
3. Identify data discrepancies and eliminate anomalies and comprehend different types of learning.
4. Predict the outcome based on regression and compute optimal hyperplane and support vectors for data classification.
5. Analyse the data using the machine learning methods to address social, engineering and business problems.

Module:1	Fundamentals of Data Mining	4 hours

Introduction to data mining - Data types -Measures of similarity and dissimilarity - Data mining tools supervised and unsupervised learning - Classification of Data Ming Systems - Data Mining Task Primitives - Major issues in Data mining.

Module:2	Data Warehousing	4 hours

Data Warehousing Components - Multi-Dimensional Data Model - Data Warehouse Architecture - Data Warehouse Implementation - Mapping the Data Warehouse to Multiprocessor Architecture - OLAP - Need - Categorization of OLAP Tools -Uses of data warehouse.

Module:3	The Ingredients of Machine Learning	8 hours

Machine Learning – Types; Data – Getting the data - visualizing the data - preparing the data; Selecting and Training a Model – Fine tuning a Model: Grid Search – Randomized Search - Main Challenges: Data Inadequacy – Non-representativeness – Irrelevant features – -Bias-Variance- Overfitting the Model – Underfitting the Model.

Module:4	Supervised Learning Techniques	8 hours

Binary Classifier – Performance Measures: Cross –Validation – Confusion Matrix – Precision and Recall – Multiclass classification – Mutli-label classification; Linear Regression – Gradient Descent: Batch Gradient – Stochastic Gradient Descent – Mini-batch Gradient Descent; Polynomial Regression –Logistic Regression – Bayes Classification - Estimating Probabilities -Decision Boundaries -Softmax Regression.

Module:5	Ensemble Machine Learning	6 hours

Linear SVM with Soft Margin Classification – Non-linear SVM Classification: Polynomial features –Similarity features – Gaussian Kernel; SVM Regression. Decision Trees and Random Forests: Training and Visualizing a Decision Tree – CART Algorithm – Gini Impurity; Bagging – Pasting – Random Forests – Boosting: Adaboost and Gradient Boosting – Stacking - Explainability.

Module:6	Dimensionality Reduction	6 hours

Main approaches – Projection and Manifold Learning – PCA (Principal Component Analysis): Preserving the Variance – Principal Components – Projecting down to Dimensions – Randomized PCA – Kernel PCA.

Module:7	**Unsupervised Learning Techniques**			**7 hours**
Clustering: K-means Clustering – Limitations – Clustering for Image Segmentation -Preprocessing - Semi supervised learning – DBSCAN – Hierarchical – Partitional - Gaussian Mixtures.				
Module:8	**Contemporary Issues**			**2 hours**
	Total Lecture hours			**45 hours**
Text Book(s)				
1	Alpaydin Ethem, Introduction to Machine Learning, 2019, 3rd Edition, PHI Learning Private Limited.			
2	Mohammed J. Zaki and Wagner Meira, Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2020, 2nd Edition, Cambridge University Press.			
Reference Book(s)				
1	Balas K Natarajan, Machine Learning, 2014, Elsevier Science.			
2	Deisenroth, Marc Peter, A., Aldo Faisal and Cheng Soon Ong., Mathematics for machine learning, 2019, Cambridge University Press.			
Mode of Evaluation: CAT, Assignment, Quiz and FAT				
Recommended by Board of Studies	15-02-2024			
Approved by Academic Council	No. 73	Date	14-03-2024	