Machine Learning

COURSE CODE: 18B1WCI634

COURSE CREDITS: 2

CORE/ELECTIVE: ELECTIVE

L-T-P: 2-0-0

Pre-requisite: None

Course Objectives:

1. Have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
2. Have an understanding of the strengths and weaknesses of many popular machine learning approaches.
3. Appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
4. Be able to design and implement various machine learning algorithms in a range of real-world applications.

Course Outcomes:

S. No.	Course Outcomes	Level of Attainment
CO-1	To learn the basic concepts and terminology in machine learning	Familiarity
CO-2	To learn about the definition of learning systems, their goals and applications in machine learning	Familiarity
CO-3	To understand concepts associated with classification and experimental evaluation of classification algorithms	Assessment
CO-4	To learn concepts associated with decision trees and experimental evaluation of classification algorithms	Assessment
CO-5	To learn about instance-based learning, clustering and unsupervised learning	Usage

Course Contents:

Unit	Contents	Lectures required
1	**Machine Learning Theoretical Foundations**: What is machine learning? What kind of problems can be tackled using machine learning? Some standard learning tasks, Learning stages, Learning scenarios, Generalization. Types of Machine Learning. Three Prerequisites to apply Machine Learning. **PAC Learning Framework**: The PAC learning mode, Guarantees for finite hypothesis sets — consistent case, Guarantees for finite hypothesis sets — inconsistent case, Generalities, Deterministic versus stochastic scenarios, Bayes error and noise	5
2	**Rademacher Complexity and VC-Dimension**: Rademacher Complexity, Growth function, VC-dimension, Lower bounds. **Model Selection**: Estimation and approximation errors, Empirical risk minimization (ERM), Structural risk minimization (SRM), Cross-validation, n-Fold cross-validation, Regularization-based algorithms, Convex surrogate losses	5

3	**Bayesian Decision Theory**: Introduction, Classification, Losses and Risks, Discriminant Functions, Utility Theory, Association Rules. **Metrics of Classification:** Precision, Recall, F1-Score etc**.** **Parametric Methods:** Introduction, Maximum Likelihood Estimation, Bernoulli Density, Multinomial Density, Gaussian (Normal) Density, evaluating an Estimator: Bias and Variance, The Bayes' Estimator, Parametric Classification, Regression, Tuning Model Complexity: Bias/Variance Dilemma, Model Selection Procedures	**5**
4	**Decision Tree Learning**: Representing concepts as decision trees. Recursive induction of decision trees. Picking the best splitting attribute: entropy and information gain. Overfitting, noisy data, and pruning, Linear regression	**5**
5	**Support Vector Machine:** Kernel function and Kernel SVM Instance-Based Learning: m Constructing explicit generalizations versus comparing to past specific examples. k-Nearest Neighbor algorithm. Case-based learning. **Dimensionality Reduction:** Introduction, Subset Selection, Principal Components Analysis, Factor Analysis, Multidimensional Scaling, Linear Discriminant Analysis	**6**
6	**Clustering and Unsupervised Learning**: Learning from unclassified data. Hierarchical Agglomerative Clustering. k-means partitioned clustering. Expectation maximization (EM) for soft clustering. Semi-supervised learning with EM using labeled and unlabeled data	**6**
Total lecture		**32**

Suggested Text Book(s):

1. Tom Mitchell, "Machine Learning", McGraw Hill, 1997, ISBN 0070428077
2. Introduction to Machine Learning Edition 2, by Ethem Alpaydin.
3. Foundations of machine learning, Edition 2, by Mehryar Mohari et al.

Suggested Reference Book(s):

1. Richard o. Duda, Peter E. Hart, and David G. Stork, "Pattern Classification", John Wiley Asia, 2006
2. T. Hastie, R. Tibshirani, & J. H. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer Verlag, 2001.
3. Ian H. Witten & Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann, 1999.
4. S. M. Weiss & C. A. Kulikowski, "Computer Systems that Learn", Morgan Kaufman Publishers, San Fancisco, CA, 1991

Other useful resource(s):

1. Link to NPTEL course contents: https://onlinecourses.nptel.ac.in/noc18_cs40/preview
2. Link to topics related to course:
   i. https://in.udacity.com/course/intro-to-machine-learning--ud120-india
   ii. https://www.edx.org/learn/machine-learning
   iii. https://www.datacamp.com/courses/introduction-to-machine-learning-with-r
   iv. https://www.simplilearn.com/big-data-and-analytics/machine-learning-certification-training-course

Evaluation Scheme:

S. No	Exam	Mark	Duratio	Coverage / Scope of Examination
1	T 1	15	1 Hour.	Syllabus covered upto T 1
2	T 2	25	1.5 Hour	Syllabus covered upto T 2

3.	T 3	35	2 Hours	Entire Syllabus
4.	Teaching Assessment		Entire Semester	Assignment (2) - 10 Quizzes (2) - 10 Attendance - 5

Course Outcomes (COs) contribution to the Programme Outcomes(POs)

Course outcomes (Machine Learning)	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	PO-10	PO-11	PO-12	Average
CO-1	3	3	3	2	2	3	2	2	2	3	1	3	**2.4**
CO-2	3	3	3	2	3	2	3	2	2	3	1	3	**2.5**
CO-3	3	3	3	2	3	3	2	2	3	3	1	3	**2.6**
CO-4	3	3	3	2	3	3	2	2	2	3	1	3	**2.5**
CO-5	3	3	3	2	3	3	2	2	2	3	1	3	**2.5**
Average	**3**	**3**	**3**	**2**	**2.8**	**2.8**	**2.2**	**2**	**2.2**	**3**	**1**	**3**	