Year: B. Tech III (Semester VI)

Subject Name: Machine Learning **Subject Code:** BTCO14603
Type of course: Professional Elective II and III
Prerequisite (if any): Python Programming, Probability and Statistics

List of Courses where this course will be prerequisite : --

Rationale: In today's world, we see various applications of machine learning like face recognition, recommendation on e-commerce portals, weather forecasting, prediction in various domains like medicine, finance and cyber crimes. This course aims to make machines emulate the human learning process. The course gives a detailed insight into various types of machine learning techniques which will assist in data analysis and analytical model building for various applications.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments/projects/open book tests/closed book tests CA2: Sincerity in attending classes/class tests/ timely submissions of assignments/self-learning attitude/solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work/Quality of work submitted/Active participation in lab sessions/viva on practical skills learned in course

Content:

Sr. No.	Content	Total Hrs
1	**Fundamentals of Machine Learning:** Introduction to Machine Learning,, Overview of Supervised Learning, Unsupervised Learning and Reinforcement Learning, Elements of Computational Learning Theory, Loss Functions, Evaluation Metrics	6
2	**Dimensionality Reduction:** Feature Selection, Feature Extraction, Variance and Covariance, Eigen values and Eigen vectors, Principal Component Analysis,	6
3	**Regression Methods:** Introduction to Regression, Linear Regression, Multiple Linear Regression, Polynomial Regression, Logistic Regression	8
4	**Classification Methods:** Decision Trees, Naive Bayes Classification, k-Nearest Neighbour, Ensemble Learning: Bagging, Random Forest, Boosting; Kernel Methods for non-linear data, Support Vector Machines, Probability and Bayes learning	11
5	**Clustering Methods:** Clustering: K-means Clustering, Hierarchical Clustering, Cluster Quality Metrics, Outlier Detection	08

6	**Artificial Neural Networks:** Overview of Artificial Neural Networks, Multilayer Perceptrons, Activation Functions and Hyper parameter Tuning, Introduction to Deep Neural Network	06

Suggested Specification table with Marks (Theory): (For B. Tech only)

Distribution of Theory Marks					
R Level	**U Level**	**A Level**	**N Level**	**E Level**	**C Level**
6	6	8	15	15	10

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

Sr No	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication / Publication Edition
1.	Machine Learning	Tom Mitchell	McGraw- Hill	Latest Edition
2.	Introduction to Machine Learning	Ethem Alpaydin	PHI Learning	
3.	Pattern Recognition. Techniques and Applications	Rajjan Shinghal	OXFORD	

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Compare and contrast supervised and unsupervised machine learning techniques	15%
CO-2	Examine the important Dimensionality Reduction methods for machine learning	15%
CO-3	Implement regression models for numerical and categorical data	15%
CO-4	Construct and evaluate classification models using supervised learning methods	25%
CO-5	Construct and evaluate clustering models using unsupervised learning methods	20%
CO-6	Analyze various advanced concepts of machine learning	10%

List of Open learning website:

- https://nptel.ac.in/courses/106/106/106106139/
- https://nptel.ac.in/courses/106/105/106105152/
- https://www.coursera.org/learn/machine-learning

List of Open Source Software:

- Python
- WEKA- Waikato Environment for Knowledge Analysis
- Orange
- Knime- KoNstanz Information MinEr

List of Experiments:

Sr. No.	Practical
1.	Implement Logistic Regression in Python using machine learning for different applications (Sample dataset : Social_Networks_Ads.csv)
2.	Implement and compare PCA and LDA on some sample dataset.
3.	Implement filter based method for feature selection on the Titanic dataset.
4.	Implement k- Nearest Neighbour algorithm.
5.	i) Study the Python library sklearn.ensemble for a bagging and boosting algorithm. ii) Implement the following Ensemble algorithms on the Iris dataset or the labelled dataset of your choice. Prepare a comparison table of accuracy measures for different numbers of estimators for the given algorithms. 1. Random Forest 2. AdaBoost 3. Gradient Boost
6.	Calculate the Cost metrics and/or the Cost function for the following datasets after preprocessing, if necessary. Further, use cross validation and record the results. Use Linear Regression on Boston Housing dataset
7.	Calculate the Cost metrics and/or the Cost function for the following datasets after preprocessing, if necessary. Further, use cross validation and record the results. Use Classification or Logistic Regression on following datasets: a). Titanic dataset b). Coronavirus patients dataset
8.	Implement a User – User Collaborative Filtering Recommendation System on the Movie dataset.
9.	Implement k-means clustering on a sample unlabelled dataset.
10.	Design a basic feed forward neural network and evaluate the error.