CVM UNIVERSITY
Aegis: Charutar Vidya Mandal (Estd.1945)

FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2020-21

Programme: **Bachelor of Engineering (Information Technology)**

Semester: **VI**

Course Code: **102045609**

Course Title: **Machine Learning**

Course Group: **Professional Core Course**

Course Objectives: This course is centered around Machine learning techniques on how to imitate learning mechanism which can improve automatically through experience by the use of data. For that several aspects of Machine learning like model preparation, how it can be evaluated, what are all different techniques and algorithms to construct a learning model, relevant applications and advance concepts are covered.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
				Theory		J/V/P*		Total
Lecture	Tutorial	Practical		Internal	External	Internal	External	
3	0	2	4	40 / 14	60 / 21	20 / 7	30 / 10	150 / 52

* **J**: Jury; **V**: Viva; **P**: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	**Introduction to Machine Learning:** Types of human learning, What is machine learning?, Types of machine learning– supervised, unsupervised, semi-supervised and reinforcement learning, machine learning activities, applications of machine learning.	03
2	**Model Preparation, Evaluation and feature engineering:** Types of data in machine learning, Exploring structure of data, Data pre-processing, Model selection and training (for supervised learning), Model representation and interpretability, Evaluating machine learning algorithms and performance enhancement of models. What is feature engineering?, Feature transformation, Feature subset selection. Principal component analysis.	09
3	**Supervised Learning - Regression:** Introduction of regression, Regression algorithms: Simple linear regression, Multiple linear regression, Polynomial regression model, Logistic regression, Maximum likelihood estimation.	05

4	**Supervised Learning - Classification:** Introduction of supervised learning, Classification model and learning steps, Classification algorithms: Naïve Bayes classifier, k-Nearest Neighbour (kNN), Decision tree, Support vector machines, Random forest.	**08**
5	**Unsupervised Learning:** Introduction of unsupervised learning, Unsupervised vs supervised learning, Application of unsupervised learning, Clustering and its types, Partitioning method: k-Means and K-Medoids, Hierarchical clustering, Density-based methods – DBSCAN.	**05**
6	**Artificial Neural Network:** Biological neuron, Artificial neuron, Activation functions, Architectures of neural network, , Perceptron, Learning process in ANN, Back propagation.	**05**
7	**Advance topics in Machine Learning:** Introduction to deep learning, overview of reinforcement learning, Case-study of ML applications: Image recognition, speech recognition, Email spam filtering, Online fraud detection and other.	**05**
	Total	**40**

List of Practicals / Tutorials:

1	Study Python ecosystem for Machine learning: Python, SciPy, Scikit-learn.
2	Study of preprocessing methods. Write a program to find following statistics from a given dataset. Mean, mode, median, variance, standard deviation, quartiles, interquartile range.
3	Study and implement PCA in python.
4	Study and implement simple linear regression.
5	Write a program to demonstrate the working of the decision tree-based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
6	Write a program to implement the Naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
7	Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Compute the accuracy of the classifier, considering few test data sets.
8	Write a program to implement SVM algorithm to classify the iris data set. Compute the accuracy of the classifier, considering few test data sets.
9	Write a program to implement and check Sklearn's K-Fold, Shuffled K-fold, Repeated K-Fold and Leave-One-Out validation technique for appropriate classification algorithm and dataset.
10	Write a program to implement k-Means clustering algorithm for a sample training data set stored as a .CSV file.
11	Write a program to recognize handwritten digit using Artificial Neural Network.
12	Study Weka toolkit for demonstration of regression, classification and clustering models on it.

Reference Books:

1	Machine Learning, Saikat Dutt, S. Chjandramouli, Das, Pearson
2	Ethem Alpaydin, Machine Learning: The New AI, MIT Press-2016
3	Machine Learning, Anuradha Srinivasaraghavan, Vincy Joseph, Wiley
4	Mitchell T, Machine Learning, McGraw-Hill
5	Neural Networks, Fuzzy Systems and Evolutionary Algorithms: Synthesis and Applications, S. Rajasekaran, G. A. Vijayalakshmi Pai, PHI.

Supplementary learning material:	
1	NPTEL - Swayam Course: Introduction to machine learning by Prof. Balaraman Ravindran, IIT Madras - https://onlinecourses.nptel.ac.in/noc20_cs29/preview
2	Coursera Courses: Machine Learning with Python by SAEED AGHABOZORGI
3	GitHub, UCI, Kaggle

Pedagogy:

- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks in %						**R**: Remembering; **U**: Understanding; **A**: Applying; **N**: Analyzing; **E**: Evaluating; **C**: Creating
R	**U**	**A**	**N**	**E**	**C**	
20%	25%	25%	15%	10%	5%	

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.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Understand the concepts of learning models and how to evaluate the models generated from data set.	30%
CO-2	Synthesize concept of supervised learning classification techniques.	30%
CO-3	Demonstrate the numerical prediction using regression techniques.	20%
CO-4	Examine and apply unsupervised techniques on unlabelled dataset.	20%

Curriculum Revision:	
Version:	1.0
Drafted on (Month-Year):	June-2020
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2025