21CSH-286	MACHINE LEARNING	L	T	P	S	C	CH	Course type
Version 1.00		2	0	4	0	4	6	Elective
					21CSH-286			
Pre-requisites/ Exposure	Python Programming Language, Basics of Statistics							
Co-requisites								
Anti-requisites								

PC: Program Core; PE: Program Elective; OE: Open Elective;
PR: Project; AE: Ability Enhancement; SE: Skill Enhancement

A pre-requisite means a course or other requirement that a student must have taken prior to enrolling in a specific course or program.
A co-requisite means a course or other requirement that a student must take at the same time as another course or requirement.
Anti-requisites are courses that overlap sufficiently in course content that both cannot be taken for credit.

A. COURSE DESCRIPTION

This course introduces several fundamental concepts and methods for machine learning. Topics include: supervised learning (Regression, KNN, Decision Trees, Random Forest, Support Vector Machine, and Naïve Bayes); unsupervised learning (K- means Clustering, Principle Component Analysis, Performance metrics for clustering), and reinforcement learning. Several software libraries and data sets publicly available will be used to illustrate the application of these algorithms. The emphasis will be thus on machine learning algorithms and applications, with some broad explanation of the underlying principles.

B. COURSE OBJECTIVES

The course aims to:

1. Understand and apply various data handling and visualization techniques.
2. Understand about some basic learning algorithms and techniques and their applications, as well as general questions related to analysing and handling large data sets.
3. To develop skills of supervised and unsupervised learning techniques and implementation of these to solve real life problems.
4. To develop basic knowledge on the machine techniques to build an intellectual machine for making decisions behalf of humans.
5. To develop skills for selecting suitable model parameters and apply them for designing optimized machine learning applications.

C. COURSE OUTCOMES

On completion of this course, the students shall be able to

CO1	Understand machine learning techniques and computing environment that are suitable for the applications under consideration.
CO2	Understand data pre-processing techniques and apply these for data cleaning.
CO3	Identify and implement simple learning strategies using data science and statistics principles.
CO4	Evaluate machine learning model's performance and apply learning strategy to improve the performance of supervised and unsupervised learning model.
CO5	Develop a suitable model for supervised and unsupervised learning algorithm and optimize the model on the expected accuracy.

D. SYLLABUS

Unit-1	**Introduction to Machine Learning**	**Contact Hours: 10 Hours**
Introduction to Machine Learning	Definition and Classification of Machine Learning algorithms, Supervised Learning, Unsupervised Learning, Reinforcement Learning, Semi-Supervised Learning, Applications of Machine Learning.	
Data Pre-Processing and Feature Extraction	Handling Missing data, Encoding Categorical data, Feature Scaling, Handling Time Series data; Feature Selection techniques, Data Transformation, Normalization, Dimensionality reduction	
Data Visualization	Data Frame Basics, Different types of plots, Plotting fundamentals using Matplotlib, Plotting fundamentals using Seaborn.	
Unit-2	**Supervised Learning**	**Contact Hours: 10 Hours**
Regression	Linear regression, Ridge Regression, Lasso Regression, and Bayesian Linear Regression.	
Classification	Types of Classification Algorithm: Binary Classification and Multi-Class Classification, Logistic Regression, K-Nearest Neighbours, Decision Trees, Random Forest, Support Vector Machine.	
Performance Metrics	Performance metrics for Regression: Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, R-Squared; Performance metrics for classification: Confusion Matrix, Accuracy, Precision, Recall, F1 score.	
Unit-3	**Unsupervised Learning**	**Contact Hours: 10 Hours**
Clustering	Types of Clustering: Centroid-based clustering, Density-based clustering, Distribution-based Clustering and Hierarchical clustering; K- Means Clustering, KNN (K-Nearest Neighbours), DBSCAN clustering algorithm; Performance metrics for clustering: Silhouette Score	
Association Rule Learning	Apriori algorithm, F-P Growth Algorithm, Applications of Association Rule Learning, Market Basket Analysis.	
Reinforcement Learning	Types of Reinforcement learning, Key Features of Reinforcement Learning, Elements of Reinforcement Learning, Applications of Reinforcement Learning.	

TEXT BOOKS:

There is no single textbook covering the material presented in this course. Here is a list of books recommended for further reading in connection with the material presented:

T1: Tom.M.Mitchell, "Machine Learning, McGraw Hill International Edition".
T2: Ethern Alpaydin," Introduction to Machine Learning. Eastern Economy Edition, Prentice Hall of India, 2005".
T3: Andreas C. Miller, Sarah Guido, Introduction to Machine Learning with Python, O'REILLY (2001).

REFERENCE BOOKS:

R1 Sebastian Raschka, Vahid Mirjalili, Python Machine Learning, (2014)
R2 Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification, Wiley, 2nd Edition".
R3 Christopher Bishop, "Pattern Recognition and Machine Learning, illustrated Edition, Springer, 2006".

E. MODE OF EVALUATION: The performance of students is evaluated as follows:

	Theory	
Components	**Continuous Internal Assessment (CAE)**	**Semester End Examination (SEE)**
Marks	40	60
Total Marks	100	

Internal Evaluation Component:

S. No.	Type of Assessment	Weightage of actual conduct	Frequency of Task	Final Weightage in Internal Assessment	Remarks (Graded/Non-Graded)
1	Assignment	10 marks for each assignment	One per unit	10	Graded
2	Exam	20 marks for one MST	2 per semester	20	Graded
3	Quiz/Test	4 marks for each quiz	2 per unit	4	Graded
4.	Surprise test	12 marks for each test	One per unit	4	Graded
5	Homework	NA	One per lecture topic (of 2 questions)	NA	Non-Graded
6	Case study	NA	NA	NA	Non-Graded
7	Discussion Forum	NA	One per unit	NA	Non-Graded
8	Presentation	NA	NA	NA	Non-Graded
9	Attendance	NA	NA	2	Graded

F. CO-PO MAPPING:

Course Outcome	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO 11	PO 12	PSO1	PSO2	PSO3	PSO4
CO1	2	3	2	2	2	1	-	1	2	1	1	1	3	2	3	3
CO2	2	3	2	2	2	1	-	1	2	1	1	1	3	2	3	3
CO3	2	3	3	3	2	1	-	1	2	1	2	2	3	2	2	3
CO4	3	3	3	2	2	1	-	1	3	1	3	2	3	3	3	3
CO5	3	3	3	2	2	1	-	1	3	1	3	2	3	3	3	3

CO PO correlation matrix of each subject to be mapped with
High correlation (3); Medium correlation (2); Low correlation (1)