Semester: 2	
Programme: M.Sc. Data Science	
Course: Machine Learning 1	
Paper code: M2C4DS26022CT / M2C4DS26022CP	**Credits: 6**
Hours/week: 8 (4Th + 4Pr)	
Category: Core/MDC/SEC/VAC: Core	
Theory / Practical / Composite: Composite	
No of Modules: 1	

Course Outcome:

1. **Remember** the fundamental concepts and definitions of machine learning systems (supervised, unsupervised, semi-supervised, reinforcement), alongside the properties of multivariate data, random vectors, mean vectors, dispersion matrices, and multivariate distributions
2. **Understand** the theoretical foundations of machine learning learning processes, the bias-variance trade-off, the curse of dimensionality in high-dimensional data, and the principles underlying regression models, Principal Component Analysis (PCA), and clustering algorithms.
3. **Apply** supervised regression techniques (such as simple and multiple linear regression, K-nearest neighbor regression, and qualitative predictors with dummy variables) as well as unsupervised techniques (PCA, K-Means clustering, Hierarchical clustering, and DBSCAN) to process and analyze multivariate datasets.
4. **Analyze** regression diagnostics and model health by examining outlier detection, leverage points, Cook's distance, multicollinearity, and partial or multiple correlation coefficients.
5. **Evaluate** model selection and predictive performance using criteria such as AIC, BIC, adjusted R-squared, and stepwise subset selection methods (forward and backward selection).
6. **Create** comprehensive machine learning workflows by synthesizing multivariate data preprocessing, feature dimension reduction via PCA, supervised regression modeling, and unsupervised cluster analysis to solve complex analytical problems

Prerequisites: None

SYLLABUS

UNIT/Module	CONTENT	HOURS or NUMBER OF CLASSES	CO Mapping	COGNITIVE LEVEL
I.	**Introduction:** Why use Machine Learning? Types of Machine Learning systems–supervised, unsupervised, semi-supervised and reinforcement. Challenges of Machine Learning. Applications of Machine Learning. Bias-variance trade off. How does a Machine Learning model learn?	**10**	**CO 1, CO 2**	**K1, K2**
II.	**Multivariate data & distributions:** Concept of random vector, mean vector and dispersion matrix, probability mass and density functions, marginal and conditional distributions, multiple and	**10**	**CO 1, CO 2, CO 4**	**K1, K2, K4**

	partial correlation coefficient. Multinomial distribution, multivariate normal distribution.			
III.	**Supervised learning techniques:** Least square method, simple and multiple linear regression and related tests. Use of multiple and partial correlations. Qualitative predictors and dummy variables. K-nearest neighbour regression. Regression diagnostics- outlier detection, leverage, influential point, Cook's distance, model selection via AIC, BIC, adjusted R-squared. Subset selection method (forward and backward stepwise selection). Multicollinearity.	**20**	**CO 2, CO 3, CO 4, CO 5, CO 6**	**K2, K3, K4, K5, K6**
IV.	**Unsupervised learning techniques:** High dimensional data and curse of dimensionality. Principal Component Analysis. Cluster Analysis – K-Means clustering and its variants; Hierarchical clustering; DBSCAN.	**12**	**CO 2, CO 3, CO 6**	**K2, K3, K6**

Text Books

1. **An Introduction to Statistical Learning with Applications in Python by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor. Springer.**
2. **Machine Learning by Tom M. Mitchell. McGraw Hill.**

Suggested readings

1. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly.**
2. **Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar. Wiley.**

Web Resources

1. **Introduction to Machine Learning by Prof. Balaraman Ravindran, IIT Madras, NPTEL Course**
2. **Machine Learning Specialization by Andrew Ng, Coursera**

Evaluation

Marks	CIA: Class Performance: 2 Attendance: 3 Mid-term Exam: 10 Total: 15 End Semester Exam: 45	Continuous Assessment: 40
Paper Structure for Theory Semester Exam	Short Questions: 5 marks each	Long Questions: 10 marks each
	3 out of 5	3 out of 5

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	**Remember** the fundamental concepts and definitions of machine learning systems (supervised, unsupervised, semi-supervised,	K1

	reinforcement), alongside the properties of multivariate data, random vectors, mean vectors, dispersion matrices, and multivariate distributions	
CO2	**Understand** the theoretical foundations of machine learning learning processes, the bias-variance trade-off, the curse of dimensionality in high-dimensional data, and the principles underlying regression models, Principal Component Analysis (PCA), and clustering algorithms.	K2
CO3	**Apply** supervised regression techniques (such as simple and multiple linear regression, K-nearest neighbor regression, and qualitative predictors with dummy variables) as well as unsupervised techniques (PCA, K-Means clustering, Hierarchical clustering, and DBSCAN) to process and analyze multivariate datasets.	K3
CO4	**Analyze** regression diagnostics and model health by examining outlier detection, leverage points, Cook's distance, multicollinearity, and partial or multiple correlation coefficients.	K4
CO5	**Evaluate** model selection and predictive performance using criteria such as AIC, BIC, adjusted R-squared, and stepwise subset selection methods (forward and backward selection).	K5
CO6	**Create** comprehensive machine learning workflows by synthesizing multivariate data preprocessing, feature dimension reduction via PCA, supervised regression modeling, and unsupervised cluster analysis to solve complex analytical problems	K6