

Progressive Education Society's
Modern College of Arts, Science and Commerce,
Shivajinagar, Pune - 5
Second Year of B. Sc. Computer Science
(Course under NEP 2020)

Course Code: 24CsStatU4601

Course Name: Data Mining (Machine Learning)

Teaching Scheme: 2 Hours/Week

Credit: 2

Examination Scheme: CIA: 20 Marks

End-Sem: 30 Marks

Prerequisites:

1. Basic knowledge of Probability, Linear Algebra and Statistics
2. Basic knowledge of programming language

Course Objectives:

- To get knowledge by self-study, observation, experience and/or by being taught.
- Data mining deals with the design, analysis and validation of algorithms that enable computers (machines) to learn from data and automatically extract methods to perform intended tasks.
- The main objective of this course is to introduce some standard learning algorithms for tasks such as classification, regression, clustering, outlier detection and association finding.
- To apply data mining techniques for problem solving.
- To use tools and techniques for understanding concept of data mining

Course Outcomes:

On completion of the course, student will be able to:

1. Understand a wide variety of data mining tool.
2. Ability to analyze the data using various data mining tool or various machine learning algorithm
3. Ability to design and implement a learning model
4. Understand how to perform evaluation of learning algorithms and model selection
5. Ability to tackle real world problems of data mining.

Course Contents:

Unit No.	Title	Number of Lectures
Unit 1	Introduction to Data mining	7
	1.1 What is Data mining? 1.2 Applications of data mining. 1.3 Steps in developing learning system, Data preprocessing, Missing data, categorical data, feature scaling, training set vs. test set, Data Modeling phases 1.4 Types of learning - Supervised learning, 1.5 Unsupervised learning, Semi-supervised learning 1.6 Dimensionality reduction- Introduction to dimensionality reduction, Subset selection, Principal Component Analysis(PCA), Factor analysis. 1.7 Case studies	

Unit 2	Supervised Learning	10
	2.1 Supervised learning 2.2 Learning with Classification - Decision Tree, Support Vector Machine, Naive Bayes, K-Nearest Neighbors, Logistic regression 2.3 Learning with Regression - Simple Linear regression, Multiple linear Regression 2.4 Evaluating model performance - False positives, False negatives Confusion matrix, Accuracy, Precision, Recall, Cross Validation and Comparison 2.5 Regression Matrices R- Square, Adjusted R Square 2.6 Classification and Evaluation Matrix 2.7 Case studies	

Unit 3	Unsupervised Learning	8
	3.1 Un-supervised learning 3.2 Clustering algorithms - K-means clustering, Hierarchical clustering 3.3 Rule based learning - Association rule mining, Apriori, Support and Confidence parameters and Comparison 3.4 Case studies	
Unit 4	Learning with Neural Network	5
	4.1 Limitations of Machine Learning 4.2 What is Deep learning? 4.3 Artificial Neural network - Neurons, Activation function, Perceptron learning algorithm, training of perceptron 4.4 Case studies.	
	References: 1. Berson and Smith S.J. (1997): Data warehousing, Data Mining, and OLAP, McGraw-Hill. 2. Breiman J.H Friedman, R.A. Olshen and stone C.J. (1984): Classification and Regression Trees, Wadsworth and Brooks / Cole. 3. Han, J. and Kamber, M. and Pei, J. (2012): Data Mining: Concepts and Techniques. MorganGaufmann.3rd Edition. 4. Mitchell T.M. (1997): Machine Learning, McGraw-Hill. 5. Ripley B.D. (1996): Pattern Recognition and Neural Networks. Cambridge University Press. 6. Vapnik V.N. The nature of Statistical learning theory, Springer. 7. Cristianini N. and Shawe-Taylor J. An Introduction to support vector machines. 8. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram international.	

	9. Hastie T, Tibshirani R, Friedman J, (2009): The elements of statistical Learning, Springer. 10. Chattamvelli, R. (2015). Data mining methods. Alpha Science International	
--	--	--

