



Savitribai Phule Pune University Fourth Year of Computer Engineering (2019 Course) 410242: Machine Learning		
Teaching Scheme:	Credit	Examination Scheme:
TH: 03 Hours/Week	03	In-Sem (Paper): 30 Marks End-Sem (Paper): 70 Marks
Prerequisite Courses: Data Science and Big Data Analytics(310251)		
Companion Course: Laboratory Practice III(410246)		
Course Objectives: <ul style="list-style-type: none"> • To understand the need for Machine learning • To explore various data pre-processing methods. • To study and understand classification methods • To understand the need for multi-class classifiers. • To learn the working of clustering algorithms • To learn fundamental neural network algorithms. 		
Course Outcomes: On completion of the course, student will be able to– CO1: Identify the needs and challenges of machine learning for real time applications. CO2: Apply various data pre-processing techniques to simplify and speed up machine learning algorithms. CO3: Select and apply appropriately supervised machine learning algorithms for real time applications. CO4: Implement variants of multi-class classifier and measure its performance. CO5 :Compare and contrast different clustering algorithms. CO6: Design a neural network for solving engineering problems.		
Course Contents		
Unit I	Introduction To Machine Learning	07 Hours
Introduction to Machine Learning, Comparison of Machine learning with traditional programming, ML vs AI vs Data Science. Types of learning: Supervised, Unsupervised, and semi-supervised, reinforcement learning techniques, Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and non-parametric models. Important Elements of Machine Learning- Data formats, Learnability, Statistical learning approaches		
#Exemplar/Case Studies	Suppose you are working for Uber where a task to increase sales is given. Understand the requirements of the client	
*Mapping of Course Outcomes for Unit	CO1	
Unit II	Feature Engineering	07 Hours

Concept of Feature, Preprocessing of data: Normalization and Scaling, Standardization, Managing missing values, Introduction to Dimensionality Reduction, Principal Component Analysis (PCA), Feature Extraction: Kernel PCA, Local Binary Pattern.

Introduction to various Feature Selection Techniques, Sequential Forward Selection, Sequential Backward Selection.

Statistical feature engineering: count-based, Length, Mean, Median, Mode etc. based feature vector creation.

Multidimensional Scaling, Matrix Factorization Techniques.

#Exemplar/Case Studies	<p>You are a Data Scientist, and a client comes to you with their data. Client is running a few campaigns from the past few months, but no campaign seems effective. Client provides you the data of customers, product sales and past campaign success. They want to increase their sales and figure out which marketing strategy is working the best for them?</p> <p>Questions for data scientists:</p> <ol style="list-style-type: none"> 1. What data analysis approach will you follow? 2. What statistical approach do you need to follow? <p>How will you select important features?</p>
-------------------------------	--

***Mapping of Course Outcomes for Unit II**

CO2

Unit III	Supervised Learning : Regression	06 Hours
-----------------	---	-----------------

Bias, Variance, Generalization, Underfitting, Overfitting, Linear regression, Regression: Lasso regression, Ridge regression, Gradient descent algorithm.

Evaluation Metrics: MAE, RMSE, R2

#Exemplar/Case Studies	Stock market price prediction
*Mapping of Course Outcomes for Unit III	CO3

Unit IV	Supervised Learning : Classification	08 Hours
----------------	---	-----------------

Classification: K-nearest neighbour, Support vector machine.

Ensemble Learning: Bagging, Boosting, Random Forest, Adaboost.

Binary-vs-Multiclass Classification, Balanced and Imbalanced Multiclass Classification

Problems, Variants of Multiclass Classification: One-vs-One and One-vs-All

Evaluation Metrics and Score: Accuracy, Precision, Recall, Fscore, Cross-validation, Micro-Average Precision and Recall, Micro-Average F-score, Macro-Average Precision and Recall, Macro-Average F-score.

#Exemplar/Case Studies	Prediction of Thyroid disorders such as Hyperthyroid, Hypothyroid, Euthyroid-sick, and Euthyroid using multiclass classifier.
*Mapping of Course Outcomes for Unit IV	CO4

Unit V	Unsupervised Learning	07 Hours
---------------	------------------------------	-----------------

K-Means, K-medoids, Hierarchical, and Density-based Clustering, Spectral Clustering. Outlier analysis: introduction of isolation factor, local outlier factor.

Evaluation metrics and score: elbow method, extrinsic and intrinsic methods

#Exemplar/Case Studies	Market basket analysis/ Customer Segmentation	
*Mapping of Course Outcomes for Unit V	CO5	
Unit VI	Introduction To Neural Networks	07 Hours
Artificial Neural Networks: Single Layer Neural Network, Multilayer Perceptron, Back Propagation Learning, Functional Link Artificial Neural Network, and Radial Basis Function Network, Activation functions, Introduction to Recurrent Neural Networks and Convolutional Neural Networks		
#Exemplar/Case Studies	Movie Recommendation System	
*Mapping of Course Outcomes for Unit VI	CO6	
Learning Resources		
Text Books:		
<ol style="list-style-type: none">1. Bishop, Christopher M., and Nasser M. Nasrabadi, "Pattern recognition and machine learning", Vol. 4.No. 4. New York: springer, 2006.2. Ethem Alpaydin, "Introduction to Machine Learning", PHI 2nd Edition-2013		
Reference Books:		
<ol style="list-style-type: none">1. Tom Mitchell, "Machine learning", McGraw-Hill series in Computer Science, 19972. Shalev-Shwartz, Shai, and Shai Ben-David, "Understanding machine learning: From theory to algorithms", Cambridge university press, 2014.3. Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining: Concepts and Techniques", Elsevier Publishers Third Edition, ISBN: 9780123814791, 97801238148074. Hastie, Trevor, et al., "The elements of statistical learning: data mining, inference, and prediction", Vol. 2. New York: springer, 2009.5. McKinney, "Python for Data Analysis ", O'Reilly media, ISBN : 978-1-449-31979-36. Trent hauk, "Scikit-learn", Cookbook , Packt Publishing, ISBN: 97817872863827. Goodfellow I.,Bengio Y. and Courville, " A Deep Learning", MIT Press, 2016		
e-Books :		
<ol style="list-style-type: none">1. Python Machine Learning : http://www.ru.ac.bd/wp-content/uploads/sites/25/2019/03/207_05_01_Rajchka_Using-Python-for-machine-learning-2015.pdf2. Foundation of Machine Learning: https://cs.nyu.edu/~mohri/mlbook/3. Dive into Deep Learning: http://d2l.ai/4. A brief introduction to machine learning for Engineers: https://arxiv.org/pdf/1709.02840.pdf5. Feature selection: https://dl.acm.org/doi/pdf/10.5555/944919.9449686. Introductory Machine Learning Nodes : http://lcs1.mit.edu/courses/ml/1718/MLNotes.pdf		
MOOC Courses Links:		
<ul style="list-style-type: none">• Introduction to Machine Learning : https://nptel.ac.in/courses/106105152• Introduction to Machine Learning (IIT Madras): https://onlinecourses.nptel.ac.in/noc22_cs29/preview• Deep learning: https://nptel.ac.in/courses/106106184		