GANPAT UNIVERSITY

FACULTY OF ENGINEERING & TECHNOLOGY

Programme	Master of Technology	Branch/Spec.	Computer Engineering (Artificial Intelligence)
Semester	I	Version	1.0.0.0
Effective from Academic Year	2025-26	Effective for the Batch admitted in	July 2025
Course Code	3CEAIPE105	Course Name	Data Science with Python

Teaching Scheme						Examination Scheme (Marks)			
(Per week)	Lecture (DT)		Practical (Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	3	-	1	-	4	Theory	50	50	100
Hours	3	-	2	-	5	Practical	30	20	50

Pre-requisites

Introduction to programming, Probability

Course Outcomes

On successful completion of the course, the students will be able to:

CO1	Understand the fundamental concepts and techniques of machine learning, including types of learning, data preprocessing, mathematical foundations, and performance evaluation.
CO2	Apply regression and classification techniques to build predictive models, assess their effectiveness using appropriate evaluation metrics, and improve accuracy through hyperparameter tuning.
CO3	Analyze data characteristics, select optimal machine learning algorithms and tuning techniques, and evaluate model performance using cross-validation and ensemble methods.
CO4	Design end-to-end machine learning solutions for real-world problems, encompassing data wrangling, model development, performance improvement, and deployment.

Theory Syllabus

Unit	Content	Hrs.
1	**Introduction to Machine Learning:** The origins of machine learning-How machines learn - Machine learning in practice - Exploring and understanding state-of-the-art methods.	02
2	**Data pre-processing/Data Wrangling Techniques:** Aggregation, Feature Cleaning - Missing values, Scaling, Outliers, Special values, Obvious inconsistencies, sampling, discretization, binarization, attribute transformation, Feature Imputation - Hot-Deck, Cold-Deck, Mean-substitution, Feature Engineering - Decompose, Discretization - Continuous Features and Categorical Features, Reframe Numerical Quantities, Crossing, Feature Encoding - Label Encoding, One Hot Encoding, Feature Normalization or Scaling - Re-scaling, Standardization.	05
3	**Dimensionality Reduction:** Introduction, Subset Selection, Factor Analysis, Singular Value Decomposition, Multidimensional Scaling, Canonical Correlation Analysis, Discriminant analysis, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA).	02
4	**Parametric methods & Nonparametric Methods:** Introduction, Maximum Likelihood Estimation: Bernoulli, binomial, Poisson distributions, Gaussian Density, Nonparametric Density Estimation: Histogram Estimator, Kernel Estimator-K Nearest Neighbor Estimator, Generalization to Multivariate Data, Nonparametric classification, Distance Based Classification, Outlier Detection.	03
5	**Regression:** Linear vs non-linear regression model, simple linear regression, multi linear regression, polynomial regression, SVR, Random Forest, Decision Tree, performance measurement Matrix.	04

6	**Classification:** Logistic regression, Naive Bayes, K-Nearest Neighbors, Perceptron Learning, SVM-Formulation, SVM-Interpretation & Analysis, SVMs for Linearly Non-Separable data, SVM Kernels, Subset Selection, Shrinkage Methods, Decision Tree, Regression trees, Stopping Criteria & Pruning, Loss Function for Classification, Categorical Attributes, Multiway Splits, Instability & Smoothness & Repeated Subtrees.	10
7	**Clustering:** Common distance measures, Hierarchical algorithms – agglomerative and divisive, Partitioning algorithms – k-means & k-Medoids, CURE Algorithm, Density-based Clustering.	06
8	**Ensemble Methods:** Bagging & Committee Machines and Stacking, Boosting, Gradient Boosting, Random Forest.	05
9	**Machine Learning Hardware:** Machine Learning Hardware TensorFlow TPU, machine learning algorithm implementation framework (open-source software libraries - Caffe, Torch, Theano), machine learning algorithms on hardware like GPU, CPU and FPGA.	04
10	**Case Study**	04
Practical Content		
Practicals, assignments and tutorials are based on the syllabus.		
Text Books		
1	Pattern Recognition and Machine Learning, by C.M. Bishop, Springer.	
2	Machine Learning: A Probabilistic Perspective by K. P. Murphy, MIT Press.	
Reference Books		
1	Modern Database Management, by Hoffer, Prescott and Mcfadden, Prentice Hall.	
2	Data Sciences, by Jain V.K., Khanna Publishing House, Delhi.	
3	Big Data and Hadoop, by Jain V.K., Khanna Publishing House, Delhi.	
4	Machine Learning, by Jeeva Jose, Khanna Publishing House, Delhi.	
5	Machine Learning, by Chopra Rajiv, Khanna Publishing House, Delhi.	
6	Deep Learning, by Ian Goodfellow, Yoshua Bengio and Aaron Courville, MIT Press.	
7	Data Mining Concepts and Techniques, by Jiawei Han and Jian Pei, Morgan Kaufmann Publishers	
8	Foundations of Machine Learning, by Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar , MIT Press.	
9	Data Mining and Predictive Analytics, Wiley Series on Methods and Applications in Data Mining.	
10	Introduction to Machine Learning, by Ethem Alpaydin, MIT Press, Prentice Hall of India.	
11	Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis, by Bruce Ratner	
ICT/MOOCs Reference		
1	https://nptel.ac.in/courses/106106139	
2	https://www.tensorflow.org/resources/learn-ml/basics-of-machine-learning	
3	https://developers.google.com/machine-learning/crash-course	
4	https://www.coursera.org/specializations/mathematics-machine-learning	

Mapping of CO with PO and PSO:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	1	2	2	0	0	0	0	1	1	3	2	3
CO2	2	3	2	2	3	0	0	0	0	1	2	2	3	3
CO3	2	2	3	3	3	0	0	0	1	1	2	2	3	3
CO4	2	2	3	3	3	1	0	0	1	2	3	1	3	3