Course Code	PAD21S01J	Course Name	MACHINE LEARNING FOR DATA SCIENCE	Course Category	S	Skill Enhancement Course	L	T	P	C
							4	0	4	6

Pre-requisite Courses	Nil	Co-requisite Courses	Nil	Progressive Courses	Nil
Course Offering Department	Computer Applications	Data Book / Codes/Standards	Nil		

Course Learning Rationale (CLR):	The purpose of learning this course is to:
CLR-1 :	To develop knowledge on Machine Learning fundamentals
CLR-2 :	To understand and analyse various machine learning models – concepts and techniques
CLR-3 :	To understand supervised and unsupervised learning concepts
CLR-4 :	To make decisions and predictions
CLR-5 :	To find ways to use it to help them grow their business
CLR-6 :	To make use of predictive causal analytics, prescriptive analytics

Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:	Learning			Program Learning Outcomes (PLO)														
		1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	Fundamental Knowledge	Application of Concepts	Link with Related Disciplines	Procedural Knowledge	Skills in Specialization	Ability to Utilize Knowledge	Skills in Modeling	Analyze, Interpret Data	Investigative Skills	Problem Solving Skills	Communication Skills	Analytical Skills	ICT Skills	Professional Behavior	Life Long Learning
CLO-1 :	Gain knowledge about basic concepts of Machine Learning	2	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-2 :	Identify machine learning techniques suitable for a given problem	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-3 :	Solve the problems using various machine learning techniques	3	75	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-4 :	Apply Dimensionality reduction techniques	3	85	80	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-5 :	Design application using machine learning techniques	3	85	75	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M
CLO-6 :	Interpret business models and scientific computing paradigms	3	80	70	H	H	H	H	H	M	L	M	H	M	-	H	H	H	M

Duration (hour)		24	24	24	24	24
S-1	SLO-1	*Introduction to Machine Learning*	*Introduction to supervised Learning algorithms –LDA& SVM*	Stochastic Gradient Descent, Naive Bayes, Decision Trees, Ensemblemethods	Multiclass	UnsupervisedLearningAlgorithms
	SLO-2	Definition and types of Machine Learning	Supervised Learning Algorithms	Stochastic Gradient Descent – classification –	Multilabel Algorithms	Clustering

S-2	SLO-1	Machine Learning process	Introduction to Linear Regression- Regularised Regression- Auto Selection of parameters Evaluation of Best Model representation	regression Stochastic gradient descent for sparse data	Feature selection	overview of Clustering methods
	SLO-2	Stages	Introduction to Classification- Regularised Classification	Complexity – stopping criterion – tips on practical use	Multilabel classification format	K-Means
S-3	SLO-1	Machine Learning Development Lifecycle	Auto selection of parameters- Evaluation of best models- Model representation	mathematical formulation – implementation details	One-vs-the-rest	affinity propagation
	SLO-2	Machine Learning Workflow	Lasso- Multi-task Lasso- Least – Angle Regression	Nearest neighbours – classification	One-vs-one	mean shift spectral
S4	SLO-1	Machine Learning Training Process	Bayesian Regression- Robust regression models	Nearest neighbours regression	error-correcting	hierarchical clustering
	SLO-2	Machine Learning Platforms	Polynomial regression- Evaluation of best models- Model representation	nearest neighbour algorithms	output-codes	DBSCAN
S5-8	SLO-1	Collecting and Manipulating data	Write a program to implement and compare SVM, KNN and Logistic regression algorithm to classify the iPhone purchase records data set. Print both correct and wrong predictions. Java/ Python ML library classescan be used for thisproblem	Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test datasets.	Implement multi label in classification	Implement Hierarchical Clustering
S9	SLO-1	Machine Learning in data	Introduction to Random Forest- Auto selection of parameters	nearest centroid classifier	Multioutput regression	OPTICS
	SLO-2	Data Modeling	Bagging and Boosting Models- Model representation	nearest neighbours transformer -	Multioutput classification	**birch – clustering**
S10	SLO-1	Data Processing	Introduction to SVM – Auto selection of parameters –	neighbourhood – components analysis	Classifier chain	**performance evaluation**

			Evaluation of best models – Model representation			
	SLO-2	Architecture for ML in Enterprises	Linear and Quadratic Discriminant Analysis	naive bayes – gaussian naive bayes – multinomial naive bayes – complement naive bayes	regressor chain	introduction to – Unsupervised Learning
S11	SLO-1	Software	Dimensionality reduction using Linear Discriminant Analysis	Bernoulli naive bayes – categorical naive bayes	**feature selection**	Auto selection
	SLO-2	Architecture to Model ML Apps in Production	Mathematical formulation of the LDA and QDA classifiers	out-of-core naive bayes model fitting	removing features with low variance	evaluation of best model representation
S12	SLO-1	Model Machine Learning apps	Mathematical formulation of LDA dimensionality reduction	decision trees – classification – regression	**univariate feature selection**	introduction to dimensional reduction
	SLO-1	ML Reference Architecture	Shrinkage – Estimation algorithms	multi-output problems- – complexity – tips on practical use	Recursive feature elimination	auto selection of parameters
S13 -16	SLO-1	Implementing Data Preprocessing	Write a program to implement Logistic Regression algorithm to classify the housing price data set. Print both correct and wrong predictions. Java/Python ML library classescanbe used for this problem.	Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task Calculate the accuracy, precision, and recall for your data set.	Implement the Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs	Compare the results of two algorithms and comment on the quality of clustering
S17	SLO-1	Building Blocks	Kernel ridge regression	tree algorithms – cart – mathematical formulation – minimal cost- complexity pruning	Feature selection using select from model	evaluation of best model
	SLO-2	Evolvable Architectures	Support Vector Machines	ensemble methods – bagging meta- estimator	Univariate Selection	model representation
S18	SLO-1	Migration	Classification	forests of randomized trees	Feature Importance	introduction to nearest neighbours
	SLO-2	Pitfalls of Evolutionary Architecture	Regression	AdaBoost – gradient tree boosting	Correlation Matrix with Heatmap	auto selection of parameters
S19	SLO-1	Anti patterns	Density estimation- novelty detection - Complexity	histogram-based gradient boosting	Feature selection as part of a pipeline	evaluation of best models
	SLO-2	Setting Up ML Solutions	Tips on Practical Use: Kernel functions	voting – classifier	**Select K Best**	Compute a distance value between the item to be classified

						and every item in the training data-set
S20	SLO-1	Fitness Function and Categories	Mathematical formulation	voting regressor	**Select From Model**	**Euclidean distance**
	SLO-2	.Architecture for Refinement and Production Readiness	Implementation details	stacked generalization	**Cross-Validation on Pipelines**	Model representation
S21 -24	SLO-1	Describing Similarity neighbourhoods	Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.	Construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set.	Demonstrate Feature Selection	To implement k-Nearest Neighbour algorithm to classify the iris data set.

Learning Resources	TEXTBOOKS: • Introduction to Machine Learning with Python, By Andreas C. Müller and Sarah Guido, October 2016 • Essential Machine Learning and Pragmatic AI, By Noah Gift, December 2018 :	REFERENCE BOOKS/VIDEOS • Stanford Lectures of Andrew Ng. • Machine Learning Yearning by Andrew Ng, deeplearning.ai, 2018 • Hands-On Unsupervised Learning Using Python, By Ankur A. Patel, March 2019 • Clustering and Unsupervised Learning, By Angie Ma, Gary Willis and Alessandra Stagliano, August 2017

Learning Assessment

Level	Bloom's Level of Thinking	Continuous Learning Assessment (50% weightage)								Final Examination (50% weightage)	
		CLA – 1 (10%)		CLA – 2 (10%)		CLA – 3 (20%)		CLA – 4 (10%)#			
		Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember	20%	20%	15%	15%	15%	15%	15%	15%	15%	15%
	Understand										
Level 2	Apply	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Analyze										
Level 3	Evaluate	10%	10%	15%	15%	15%	15%	15%	15%	15%	15%
	Create										
	Total	100 %		100 %		100 %		100 %		100 %	

CLA – 4 can be from any combination of these: Assignments, Seminars, Tech Talks, Mini-Projects, Case-Studies, Self-Study, MOOCs, Certifications, Conf. Paper etc.,

Course Designers