Course Code	PAD21202J	Course Name	BUILDING MACHINE LEARNING PIPELINES	Course Category	C	Professional Core Courses	L	T	P	C
							3	0	4	5

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

Course Learning Rationale (CLR):	The purpose of learning this course is to:
CLR-1 :	Simpler processes to update existing models, Less time spent to reproduce models
CLR-2 :	Help detect potential biases in the datasets or in the trained models
CLR-3 :	Free up development time for data scientists and increase their job satisfaction.
CLR-4 :	Automated machine learning pipelines will free up from maintaining existing models.
CLR-5 :	Publishing the Trained Model as a Web Service for Inference
CLR-6 :	Validating a Recommendation System

Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:	Learning 1 (Bloom)	2 (Expected Proficiency (%))	3 (Expected Attainment (%))	Program Learning Outcomes (PLO) 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	Linkwith 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 :	Recognise a data management plan and understand the purpose of a data management using data designing concept,	3	80	70	L	H	-	H	L	-	-	-	L	L	-	H	-	-	-
CLO-2 :	Understand Basics for Data Science , Linear Algebra, Vector Scalar Multiplication – Vector Norms	3	85	75	M	H	L	M	L	-	-	-	M	L	-	H	-	-	-
CLO-3 :	Understand matrix operations	3	75	70	M	H	M	H	L	-	-	-	M	L	-	H	-	-	-
CLO-4 :	Determinants – Orthogonal matrices Gaussian distribution – Binomial distribution	3	85	80	M	H	M	H	L	-	-	-	M	L	-	H	-	-	-
CLO-5 :	Loading ,Scaling and encoding the data	3	85	75	H	H	M	H	L	-	-	-	M	L	-	H	-	-	-
CLO-6 :	Find principal component using Principal Component Analysis and Normalizing a dataset	3	80	70	L	H	-	H	L	-	-	-	L	L	-	H	-	-	-

Duration (hour)	21	21	21	21	21

S-1	SLO-1	Cloud technologies and Data governance, designing a data governance process	Linear Algebra Basics	Matrix arithmetic , working with matrix, From Scalars and Vectors	Work with Vectors	Loading and exploring a dataset
	SLO-2	managing a Data governance strategy, monitoring a data governance strategy	Linear algebra for Machine learning	Shapes and indexing	Basis and projection of vectors	the binarizer() , the minmaxscaler()
S-2	SLO-1	maintaining a Data governance strategy	Linear Transformations, Intuition	Matrix operations- Addition and Scalar Multiplication	work with – Matrix multiplication	the standard scalar
	SLO-2	Data access governance	Linear Transformations as Vectors and Matrices	Transposition, Matrix Decomposition	Inverse matrix	the normalizer , the maxabsscaler()
S-3	SLO-1	risk and Data safety compliance	Classes of space- scalar	Matrix and PCA-covariant matrix	linear Transformations	label encoding, One-Hot encoding
	SLO-2	governance and its relationship with big data	Vector and its types	Eigen value,		Loading and analyzing a dataset
S 04 - S 07	SLO-1 SLO-2	**LAB :Data collection**	**LAB : Vector addition , vector multiplication**	**LAB : Matrix Transformation in Linear Regression**	**LAB : Creating a Recommendation Engine**	**LAB :Recommending Items Based on Other Items**
S-8	SLO-1	why big data requires governance	Vector space , Subspaces	Eigen vector calculation	Gaussian elimination	building and evaluating a Linear Regression mode
	SLO-2	why is Big Data different?	types of Vector space	sparse matrix	Gaussian elimination-Example problem	
S-9	SLO-1	Cloud technologies and Data governance	Operation on vectors-Addition	Tensor Arithmetic	Determinants	scaling and encoding the data
	SLO-2	designing a Data governance process	Subtraction	Hadamard product and Tensors-		Analyzing the effects of pre-processing
S-10	SLO-1	maintaining a data governance strategy	multiplications	Singular-Value Decomposition	Orthogonal matrices	Standardizing continuous data
	SLO-2	Data access governance , Data access patterns		Probability basics and propositions		Loading a dataset,
S 11 - S 14	SLO-1 SLO-2	**LAB : Manipulating data**	**LAB : Covariant matrix**	**LAB :Using KNN describing Similarity neighborhoods**	**LAB :Recommending Another Item**	**LAB :Evaluating a Recommendation System**
S-15	SLO-1	data breach prevention – least privilege	Scalar and vector multiplications	random variable	Eigenvectors and Linear Transformations	scaling a dataset
	SLO-2	assign and view effective file system permissions		Central limit theorem	Change of Basis	spotting correlations in a dataset
S-16	SLO-1	create an AWS user and group , vulnerability assessments	Linear product Vector	parameter estimation	Linear Transformations in Different Bases	Principal Component Analysis

	SLO-2	Data classification,, data encryption	Theorems related to linear products	Gaussian distribution	Eigen decomposition	
S17	SLO-1	implement security compliance checking	Vector Norms- Definitions, Examples of Norms	Binomial distribution	**Pseudo inverse**	Normalizing a dataset
	SLO-2	Data access monitoring solutions, logging	Norm Representations			
S 18 - S 21	SLO-1	**Lab: matrix addition, Matrix subtraction, Matrix multiplication**	**LAB: Eigen value , Eigen vector**	**LAB :Tensor Hadamard product**	**LAB :Finding Items to Recommend**	**LAB : Validating a Recommendation System**
	SLO-2					

Learning Resources	1. Data Governance: The Definitive Guide, By Evren Eryurek, Uri Gilad, Valliappa Lakshmanan, Anita Kibunguchy and Jessi Ashdown, March 2020 2. Essential Math for Data Science, By Hadrien Jean, November 2020	**1. Feature Engineering for Machine Learning, By Alice Zheng and Amanda Casari, 2018.** **2. Python Feature Engineering Cookbook, By Soledad Galli, January 2020**

Learning Assessment

	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 %		-	

Course Designers

Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
Mr.G.Muruganandam, Group Project Manager, HCL Technologies, Chennai	Dr.Muthu, Professor, Loyola College, Chennai	Ms.A.Rajalakshmi
Mr.M. Hemachandar, Tech Lead, Wipro Limited, Chennai	Dr.Vincent, Associate Professor, VIT	