Course Code	UDS21404J	Course Name	DATA SCIENCE FOR ENTERPRISE	Course Category	C	Professional Core Course	L	T	P	C
							4	0	2	5

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,	Learning	Program Learning Outcomes (PLO)

		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
		1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-1 :	To provide the participants with the comprehensive knowledge of different advanced data Science concepts with some the mathematical functions that are the foundational building blocks of the machine learning models																		
CLR-2 :	To educate the participants with concepts of tree based models in machine learning, it working and applications to represent how different input variables can be used to predict a target value.																		
CLR-3 :	To educate the participants on Advanced regression techniques used in building machine learning applications.																		
CLR-4 :	To help the participants understand, diagnose, and refine a machine learning model with the help of interactive visualization techniques, for solving real-world artificial intelligence and data mining problems																		
CLR-5 :	To educate the participants on the concepts of Sqoop, it working, architecture.																		
CLR-6 :	To educate the participants on how to Visualize data using charts, graphs, and maps is one of the most impactful ways to communicate complex data.																		
Course Learning Outcomes (CLO):	At the end of this course, learners will be able to:																		
CLO-1 :	Have excellent hands-on skills, knowledge and expertise in creating tree based models for classification and regression tasks in machine learning	2	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H
CLO-2 :	Have excellent hands-on skills, knowledge and expertise in understanding, diagnosing, and refining a machine learning models for solving real-world artificial intelligence and data mining problems	3	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H
CLO-3 :	Be able to collect and transport huge amounts of data such as events, log files, etc. from several sources to one central data store.	3	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H
CLO-4 :	Have excellent hands-on skills, knowledge and expertise in Amazon Redshift to create a data warehouse with launch a set of nodes, called an Amazon Redshift cluster	3	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H
CLO-5 :	Have excellent hands-on skills, knowledge and expertise in the basic concepts, principles, and major algorithms in text analytics and their potential applications	3	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H
CLO-6 :	Effectively use best visualization for your dataset, and interpret common plot types like histograms, scatter plots, line plots and bar plots	3	85	80	H	H	H	M	H	H	H	H	H	H	M	H	H	H	H

Note: All our curriculum, study materials, assignments, quizzes, lab works, and learning resources are personalized and dynamically generated using machine learning models based on the learner's learning ability. Users can review our learning curriculum only through our intelligent learning management platform (iLMSP), and our learning resources and lab infrastructures are available only in the digital form on our cloud infrastructures.

Duration (hour)		18	18	18	18	18
S-1	SLO-1	**Unit 1: Data Science for Enterprise - Deep Dive**	Generalized Discriminant Analysis (GDA)	Importance Machine Learning Model Analysis	Characteristics of Apache Flume, Business Benefits of Apache Flume	Regular expression, Word tokenization, Named Entity Recognition, Stemming and lemmatization
	SLO-2	Data Science for Enterprise overview	Multi-Dimension Scaling (MDS)	Business Benefits and Challenges of Machine Learning Model Analysis	Applications of Apache Flume, SQOOP vs Flume, SQOOP vs HDFS	Word cloud, Bag-of-words, Term Frequency Inverse Document Frequency
S-2	SLO-1	The core of data mining process	IsoMap	How to perform machine learning model analysis	**Unit 9: Amazon RedShift for Querying Data**	Text Regression (Automated Machine Learning and Deep Learning), Text Classification (Automated Deep Learning)
	SLO-2	Numerical optimization	Autoencoders	**Unit 6: Introduction to structured frameworks like 5W, 5WHYs, and SPIN**	Overview of Amazon Web Services	**Unit 13: Time Series Analysis - Analysis and Forecast of Series of Data that varies with time**
S-3	SLO-1	Streaming algorithms	How Dimensionality Reduction Works with PCA?	5W overview	Overview of Amazon RedShift	Time Series Analysis Overview
	SLO-2	High-dimensional regression and variable selection	Eigenvalue Decomposition	Root cause Analysis overview	Amazon Management Console	Business Benefits of Time Series Analysis Overview
S-4	SLO-1	Compression and error detection	Matrix Decomposition	Business Benefits and Challenges of 5W?	Creating Amazon RedShift Cluster ✓ Cluster Configuration ✓ Database Configuration ✓ Additional Configuration	Business Challenges of Time Series Analysis Overview
	SLO-2	Lossless coding	Eigenvectors	5WHY's overview	Querying AWS Redshift Cluster	When to use Time Series Analysis
S-5 & S-6	SLO-1 SLO-2	Lab 1 :	Lab 4 :	Lab 7:	Lab 10 :	Lab 13:
S-7	SLO-1	Entropy	Eigenvalues	Business Benefits and Challenges of 5WHY's?	**Unit 10: Building Automated Pipelines with Apache Airflow**	Components of Time Series
	SLO-2	Shannon's theorem	Singular value decomposition	SPIN overview	Apache AirFlow overview	Stationary and Non Stationary Time Series
S-8	SLO-1	**Unit 2: Machine Learning Tree Models - Deep Dive**	**Unit 4: Advanced Regression Techniques**	Business Benefits and Challenges of SPIN	When to use Airflow ✓ Reasons to choose Airflow ✓ Reasons not to choose Airflow	Auto Regressive model for Time series Implementation

	SLO-2	Tree Based Models Overview	Advanced Regression overview	**Unit 7: Introduction to structured frameworks like 5W, 5WHYs, and SPIN**	Data Pipelines as graphs	ARMA and ARIMA
S-9	SLO-1	Decision Tree Models	What is Regression Analysis?	Linear Optimization overview	Executing a Pipeline graph	Steps of time series implementation
	SLO-2	Classification and Regression Trees	Why do we use Regression Analysis?	MPsolver Interface	Pipeline graph vs sequential scripts	
S-10	SLO-1	How to Create decision tree models	How to select the right Regression Model?	Solving an MP problem	Running pipeline using workflow managers	
	SLO-2	Bias-Variance Trade-off	Polynomial Regression.	Advanced MP problem	Scheduling and executing pipelines	
S-11 & S-12	SLO-1	Lab 2 :	Lab 5 :	Lab 8:	Lab 11:	Lab 14:
	SLO-2					
S-13	SLO-1	Ensemble methods	Stepwise Regression.	Integer Optimization overview ✓ Integer variables ✓ Boolean variables	**Unit 11: Text Analytic Processing**	**Unit 14: Introduction to Data Visualization - Visual Representation of Data**
	SLO-2	Bagging and Random Forests	Ridge Regression.	Solving MIP problem	Text Analytics Overview	Data Visualization Overview
S-14	SLO-1	Boosting and Gradient Boosting	Lasso Regression.	Using Arrays to define a model	Text Analytics Business Benefits	Business Benefits of Data Visualization
	SLO-2	**Unit 3: Machine Learning Tree Models - Deep Dive**	ElasticNet Regression	**Unit 8: Introduction to Apache Sqoop And Apache Flume**	Text Analytics Business Challenges	Business Challenges of Data Visualization
S-15	SLO-1	Dimensionality Reduction Overview	Poisson's Regression	Apache Sqoop Overview, Why do we need Apache Sqoop, Apache Sqoop Architecture	Examples of Text Analytics	Data Visualization Tools
	SLO-2	The Curse of Dimensionality	Non Linear Regression	How to data transfer using sqoop	Analyse your data Import the data Define your tags Model Selection Train Model Test Model	Data Visualization Techniques
S-16	SLO-1	Principal Component Analysis (PCA)	**Unit 5: Advanced Machine Learning Model Analysis and Recursion Techniques**	Importing Data, Exporting Data, Sqoop Connectors	**Unit 12: Text based Predictive Modelliung**	Data Visualization examples
	SLO-2	Linear Discriminant Analysis (LDA)	Machine Learning Model Analysis Overview	Apache Flume Overview, Why do we need Apache Flume	Text Based Predictive Modelling Overview, Steps in Text Based predictive modelling	Bar chart, Column Chart, Pie Chart, Scatter Plot, Data Visualization in Tableau
S-17 &	SLO-1	Lab 3:	Lab 6:	Lab 9:	Lab 12:	Lab 15:
	SLO-2					

S-18						

Learning Resources	1. Vijay Kotu, Bala Deshpande, " Data Science Concepts and Practice", Second Edition, Morgan Kaufmann Publishers, 2019 2. Clinton Sheppard, " Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting" 2017 3. Olga Korosteleva, " Advanced Regression Models with SAS and R", First Edition,Published by Chapman and Hall/CRC, 2020 4. John Hearty, " Advanced Machine Learning with Python", Packt Publisher, 2016	5. https://data-flair.training/blogs/flume-books/ 6. Shruti Worlikar, Thiyagarajan Arumugam, Harshida Patel, " Amazon Redshift, Cookbook", Packt Publisher, 2021 7. Bas P. Harenslak and Julian Rutger de Ruiter, " Data Pipelines with Apache Airflow", Manning Publications, 2021 8. Chengqing Zong, Rui Xia, Jiajun Zhang, "Text Data Mining", First Edition, 2021 9. Claus O. Wilke, " Fundamentals of Data Visualization — A Primer on Making Informative and Compelling Figures", 2019

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%	15%	20%	15%	20%	15%	20%	15%	20%	15%
	Understand										
Level 2	Apply	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
	Analyze										
Level 3	Evaluate	10%	15%	10%	15%	10%	15%	10%	15%	10%	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

Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
Mr.Jothi, Periyasamy , Chief AI Architect DeepSphere.AI, CA, USA	Dr.S.Gopinathan, Associate Professor, University of Madras, Chennai	Mrs.Kanmani,SRMIST