Course Code	Course Title						
116h54C401	Applied Data Science						
	TH		P	TUT			Total
Teaching Scheme(Hrs.)	03		--	--			03
Credits Assigned	03		--	--			03
Examination Scheme	Marks						
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites (if any):

Students are expected to have basic knowledge of algorithms and programming experience.

Course Objectives

- To develop understanding of the applied data science in the real world problems.
- To get the understanding of R programming language with respect to data analysis.
- To understand the application of Machine Learning Algorithms for data modeling.
- To apply various data visualization techniques using real-world data sets and analyze the graphs and charts.
- To understand various analytics metrics, processing unstructured text/data, and the ability to investigate hidden patterns.

Course Outcomes

At the end of successful completion of the course the student will be able to

CO1	Develop an understanding of data science and business analytics.
CO2	Application of Exploratory data analysis (EDA) on Real world problems.
CO3	Understand the basic concept and techniques of Machine Learning regression and classification.
CO4	Understand the basic concept and techniques of Machine Learning clustering.

Module No.	Unit No.	Details	Hrs.	CO
1	**Introduction to Applied Data Science and Data Scraping Process**		**10**	**CO1**
	1.1	Introduction to Applied Data Science: What is Data Science? - Big Data and Data Science, Datafication - Current landscape of perspectives - Skill sets needed and various application areas. Challenges and skill Sets needed and various applications areas.		
	1.2	Impact of applying Data Science in business scenario, Introduction to need of estimation and validation for added value due to data science		
	1.3	**Introduction to the mathematical foundations required for data science.** Statistical Inference: Populations and samples, Statistical modeling, Probability distribution, Fitting a model Normal Distribution, Skewness and Kurtosis, Heteroskedasticity, Descriptive Statistics, Higher-Order Moments, Matrices, Maximum-likelihood, Introduction to Brownian Motions, Monte Carlo		
	1.4	Data Scrapping: Introduction, Need, Sources, Web Scrapping, Scrapping of Images, Data Wrangling, ETL Process, Data Munging		
2	**Exploratory Data Analysis**		**08**	**CO2**
	2.1	Exploratory Data Analysis and the Data Science Process, Basic tools (plots, graphs and summary statistics) of EDA, Measuring similarity and dissimilarity		
	2.2	Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; Data Transformation & Data Discretization: Normalization, Binning, Histogram Analysis and Concept hierarchy generation.		
	2.3	The Data Science Process : Case Study on Online E-Commerce Dataset		
3	**Machine Learning Algorithms – Regression**		**05**	**CO3**
	3.1	Introduction to Machine learning, Linear Regression, Building model, Model Diagnostic		
	3.2	Multiple Linear regression, Logistic Regression,		
4	**Machine Learning Algorithms: Classification**		**12**	**CO3**

	4.1	Decision Trees , Naive Bayes theorem, Bayes Classifier, Rule-Based Classification, Bayesian Belief networks, CART, Random Forests, k-Nearest Neighbors (k-NN), Hidden Markov Models, Support Vector Machines		
	4.2	Dimensionality Reduction Techniques: Principal Component Analysis, Independent Component Analysis, Singular value decomposition		
	4.3	Model Evaluation and Selection: Metrics, Confusion Matrix, Precision and Recall, Accuracy, False Positives, Techniques to Improve Classification Accuracy		
5	**Machine Learning Algorithms: Clustering**		**10**	**CO4**
	5.1	Cluster Analysis: Basic Concepts, Partitioning Methods: KMeans, KMediods and hierarchical methods: Agglomerative. Expectation Maximization Algorithm, Radial Basis functions		
	5.2	Cost Function, how to Minimize cost function, coefficients of determination.		
	5.3	Introduction Text processing, plain and simple, Text Classification, Text Summarization.		
		# Self Learning –Mini Project		
		Total	**45**	

 # Students should prepare all Self Learning topics on their own. Self-learning topics will enable students to gain extended knowledge of the topic. Assessment of these topics may be included in IA and Laboratory Experiments.

Recommended Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Han, J., Kamber, M., Pei, J.	*Data mining concepts and techniques*	Morgan Kaufmann	2011
2.	James, G., Witten, D., Hastie, T., Tibshirani, R.	*An introduction to statistical learning with applications in R*	Springer	2013
3.	Cathy O'Neil and Rachel Schutt	*Doing Data Science, Straight Talk From The Frontline*	O'Reilly	2014
4.	Kevin P. Murphy	*Machine Learning: A Probabilistic Perspective*	ISBN 0262018020	2013
5.	Mohammed J. Zaki and Wagner Miera Jr.	*Data Mining and Analysis: Fundamental Concepts and Algorithms*	Cambridge University Press	2014
6.	Avrim Blum, John Hopcroft, and RavindranKannan	*Foundations of Data Science*	ONLINE	2014
7	C R Kothari	Research Mythology	New Age International Publishers	4th edition 2019

Course Code	Course Title							
116h54L401	Applied Data Science							
	TH			P		TUT		Total
Teaching Scheme(Hrs.)	-			02		--		02
Credits Assigned	-			01		--		01
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	-	-	-	25	25	--	--	50

Term-Work:

Term work will consist of experiments/ tutorials covering entire syllabus of the course 'Applied Data Science'. Students will be graded based on continuous assessment of their term work.