

MAT5010	Foundations of Data science	L	T	P	J	C
		3	0	0	0	3
Pre-requisite	NIL	Syllabus version				
		1.0				
Course Objectives :						
The course is aimed at						
1. Building the fundamentals of data science.						
2. Imparting design thinking capability to build big-data						
3. Developing design skills of models for big data problems						
4. Gaining practical experience in programming tools for data sciences						
5. Empowering students with tools and techniques used in data science						
Expected Course Outcome :						
At the end of the course the student should be able to:						
1. Apply data visualisation in big-data analytics						
2. Utilise EDA, inference and regression techniques						
6. Utilize Matrix decomposition techniques to perform data analysis						
7. Apply data preprocessing techniques						
8. Apply Basic Machine Learning Algorithms						
Student Learning Outcomes (SLO):		1,7,9,17,18				
Module:1	Introduction				4 hours	
Big Data and Data Science - Big Data Analytics, Business intelligence vs Big data, big data frameworks, Current landscape of analytics, data visualisation techniques, visualisation software						
Module:2	EDA				6 hours	
Exploratory Data Analysis (EDA), statistical measures, Basic tools (plots, graphs and summary statistics) of EDA, Data Analytics Lifecycle, Discovery						
Module:3	Basic Statistical Inference				6 hours	
Developing Initial Hypotheses, Identifying Potential Data Sources, EDA case study, testing hypotheses on means, proportions and variances						
Module:4	Regression models				6 hours	
Regression models: Simple linear regression, least squares principle, MLR, logistic regression, Multiple correlation, Partial correlation						
Module:5	Linear Algebra Basics				6 hours	
Matrices to represent relations between data, Linear algebraic operations on matrices – Matrix decomposition: Singular Value Decomposition (SVD) and Principal Component Analysis (PCA).						
Module:6	Data Preprocessing and Feature Selection				7 hours	
Data cleaning - Data integration - Data Reduction - Data Transformation and Data Discretization, Feature Generation and Feature Selection, Feature Selection algorithms: Filters-Wrappers - Decision Trees - Random Forests						
Module:7	Basic Machine Learning Algorithms				8 hours	
Classifiers - Decision tree - Naive Bayes - k-Nearest Neighbors (k-NN), k-means – SVM Association Rule mining – Ensemble methods						
Module:8	Expert Lecture				2 hours	
Skillsets required for a Data Scientist						
	Total Lecture hours:				45 hours	
Text Book(s)						
1.	Mining of Massive Datasets. v2.1, Jure Leskovek, Anand Rajaraman and Jeffrey Ullman., Cambridge University Press. (2019). (free online)					
2.	Big Data Analytics, paperback 2 nd ed., Seema Acharya, Subhasini Chellappan, Wiley (2019).					
Reference Books						
1.	Doing Data Science, Straight Talk From The Frontline, Cathy O'Neil and Rachel Schutt,,					

	O'Reilly (2014).		
2	Data Mining: Concepts and Techniques”, Third Edition, Jiawei Han, Micheline Kamber and Jian Pei, ISBN 0123814790,(2011).		
3	Big Data and Business Analytics, Jay Liebowitz, CRC press (2013)		
Mode of Evaluation: CAT / Assignment / Quiz / FAT / Project / Seminar			
Recommended by Board of Studies		24-06-2020	
Approved by Academic Council		No. 59	Date 24-09-2020