DATA ANALYTICS (4–0–0–4--4)

Subject Code: UE18CS312 **No. of Hours: 56**

Class #	Chapter Title/Reference Literature	Topics to be Covered	% of Portions Covered	
			Reference Chapter	Cumulative
1	**Unit: 1** **Exploratory Data Analysis and Visualization** T1:1, 2 R1: 1.3,2, 3	Introduction to data analytics, Types of Data Analytics(T1:Chapter 1)	18	18
2		Data sources and representations (R1:Chapter 1:sec 1.3)		
3		The R programming environment, Exploring data - basic statistics (T1:Chapter 2)		
4		Data visualization – motivation, general concepts(T1:Chapter 2)		
5		Data visualization and R Graphics (ref. in ppt)		
6		ggplot(ref. in ppt)		
7		Data preprocessing - Data Cleaning(R1:Chapter 3)		
8		Data Integration, Data Reduction(R1:Chapter 3)		
9		Data Transformation and Discretization(R1:Chapter 3)		
10		Case Study(ref. in ppt)		
11	**Unit : 2** **Regression Analysis**	Distance and similarity measures(R1:Chapter 2:Sec:2.4) , Correlation(T1:Chapter 8)	21	39

12	T1: 8, 9,10, 11 R1:2.4	Simple regression, Linear regression(T1:Chapter 9)		
13		Linear Regression contd., (T1:Chapter 9)		
14		Multiple regression (T1:Chapter 10)		
15		Multiple regression Contd., (T1:Chapter 10)		
16		Multivariate regression(ref. in ppt)		
17		Linear regression approaches (OLS Vs Gradient descent) (ref. in ppt)		
18		Ridge Regression, Lasso Regression(ref. in ppt)		
19		Non-linear regression (ref. in ppt)		
20		Logistic regression (concept of odds, odds ratio) (T1:Chapter 11)		
21		Concept of training, validation and testing Confusion matrices and Metrics, RoC and AUC(T1:Chapter 11:Sec:11.6)		
22		Case study		
23	**Unit :3** **Time Series** T1: 13	Introduction to Time series data and components, Types of time series and Decomposition.	21	60
24		Forecasting techniques and accuracy(Exponential smoothing Holt's and Holt Winter's model		

25	**13.1, 13.2** **13.4 to 13.7** **13.7.1, 1.8** **13.3** **13.9, 13.9.1** **13.14.1** **13.14.3**	'Decomposing' a time series signal – into seasonal, trend and irregular components		
26		Trend analysis – simple and exponential smoothing		
27		Regression Model for forcasting.		
28		Concept of ACF and PACF and Correlogram.		
29		Forecasting using AR, MA ARMA and ARIMA		
30		Concept of stationarity, DF and ADF test and transforming non stationary process to a stationary one.		
31		ARIMA and SARIMA		
32		Ljung Box and Theil's coefficient		
33		ARIMAX and Box-Jenkins		
34		Spectral Analysis of time series data. Time series feature extraction using Fourier and wavelets, Using filters, ML for prediction		
35	**Unit : 4** **Recommendation Systems** T1: 12, 14	Introduction to recommendation systems	20	80
36		Collaborative filtering		
37		Knowledge based filtering using knn		
38		Decision trees – CART,		
39		Ensemble methods and Random Forest		

40	R1: 6, 8, 9	Brief review of other classifiers: SVM, ANN and data driven approaches		
41		Brief review of unsupervised learning – clustering algorithms – DBSCAN		
42		Content based analysis – dealing with textual data		
43		Text classification and clustering		
44		Market basket analysis (Apriori algorithm)		
45		Generation and evaluation of association rules from frequent item sets		
46		Case Study		
47	**Unit : 5** **Advanced techniques** T1: 16 + Additional Reference material	Sparse data processing,	20	100
48		LSA		
49		Sparse PCA		
50		Concept of hidden variables		
51		Concept of confounding variables		
52		Introduction to stochastic models		
53		Introduction to Markov processes (first order)		
54		Introduction to discrete Markov Chains		
55		Interpreting business values		
56		Case study		

Literature

Book Type	Code	Title & Author	Publication Information		
			Edition	Publisher	Year
Text Book	T1	Business Analytics, The Science of Data-Driven Decision Making, U. Dinesh Kumar		Wiley	2017
Reference Book	R1	Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber and Jian Pei	3rd	The Morgan Kaufmann Series in Data Management Systems	
	R2	The Elements of Statistical Learning, Trevor Friedman, Robert Tibshirani and Jerome Hastie		Data Mining, Inference and Prediction, Springer	2001
	R3	Practical Data Science with R, Nina Zumel and John Mount		Manning Publications	2014

Programming language:

1. R