Paper Code: ARI 305	L	T/P	Credits
Subject : Data Analytics	4	-	4

Marking Scheme:
Teachers Continuous Evaluation: As per university examination norms from time to time.
End Term Theory Examination: As per university examination norms from time to time.

INSTRUCTIONS TO PAPER SETTERS: Maximum Marks: As per University Norms

➢ There should be 9 questions in the end term examination question paper.

➢ Question No. 1 should be compulsory and cover the entire syllabus. This question should have objective or short answer type questions. It should be of 15 marks.

➢ Apart from Question No. 1, the rest of the paper shall consist of four units as per the syllabus. Every unit should have two questions. However, students may be asked to attempt only 1 question from each unit. Each question should be 15 marks.

➢ The questions are to be framed keeping in view the learning outcomes of course/paper. The standard/ level of the questions to be asked should be at the level of the prescribed textbooks.

➢ The requirement of (scientific) calculators/ log-tables/ data-tables may be specified if required.

Course Outcomes[Bloom's Knowledge Level (KL)]:

CO1	Ability of students to understand the basics concepts of Data Analytics **[K1, K2]**
CO2	Ability of students to apply and analyze various classification and regression techniques **[K3,K4]**
CO3	Ability of students to understand mining frequent itemsets and apply clustering techniques [K1, **K2, K3]**
CO4	Ability of students to understand Big data frameworks **[K1,K2]**

CO/ PO	**PO01**	**PO02**	**PO03**	**PO04**	**PO05**	**PO06**	**PO07**	**PO08**	**PO09**	**PO10**	**PO11**	**PO12**
CO1	3	2	1	1	3	-	-	-	2	2	1	2
CO2	3	3	3	3	3	-	-	-	3	3	4	2
CO3	3	2	1	1	3	-	-	-	2	2	1	3
CO4	3	3	3	3	3	-	-	-	3	3	3	3

Course Content	**No of lectures**

Unit I **Introduction to Data Analytics**: Sources and nature of data, classification of data (structured, semi-structured, unstructured), characteristics of data, introduction to Big Data platform, need of data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools, applications of data analytics. **Data Analytics Lifecycle**: Need, key roles for successful analytic projects, various phases of data analytics lifecycle – discovery, data preparation, model planning, model building, communicating results, operationalization.	[10]
Unit II **Data Analysis:** Data preprocessing, feature engineering, dimension reduction, Regression modelling: linear regression, non linear regression, regularized regression, Neural Networks: learning and generalisation, perceptron, logistic regression, Bayesian modeling, support vector and kernel methods, K- Nearest Neighbour Classifiers, analysis of time series: linear systems analysis & nonlinear dynamics.	[12]
Unit III **Frequent Itemsets and Clustering:** Mining frequent itemsets, market based modelling, Apriori algorithm, handling large data sets in main memory, limited pass algorithm, counting frequent itemsets in a stream, clustering techniques: hierarchical, K-means, clustering high dimensional data, DBSCAN, CLIQUE and ProCLUS.	[10]
Unit IV **Frame Works:** MapReduce, Hadoop, Pig, Hive, HBase, MapR, Sharding, NoSQL Databases, S3, Hadoop Distributed File Systems **Visualization:** visual data analysis techniques, interaction techniques, systems and applications. **Case studies –** Real time sentiment analysis, stock market predictions.	[10]
Text Books: [T1] David Dietrich, Barry Heller, Beibei Yang, (2015). Data Science and Big Data Analytics, EMC Education Series, John Wiley	
Reference Books: [R1] Sebastian Raschka,Vahid Mirjalili, (2019), Python Machine Learning - Third Edition, Pact Publisher. [R2] Tom M. Mitchell, (1997). Machine Learning, McGraw-Hill [R3] Duda, R. O. & Hart, P. E. (2006). Pattern Classification. John Wiley & Sons.	