DATA ANALYTICS

Semester	Subject Code	Category	Lecture Hrs		Theory Hrs		Practical		Credits
			Per week	Per Sem	Per week	Per Sem	Per week	Per Sem	
VI		Elective – IV	5	75	5	75	0	0	3

COURSE OBJECTIVE

- This paper provides an exposure to Big data, to learn the different ways of Data Analysis, to be familiar with data streams, to learn the mining and clustering and able to be familiar with the visualization concept.

COURSE OUTCOME

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level (K1 – K4)
CO1	Learn the concept behind Big Data and apply the statistical analysis methods.	**K1, K2 & K3**
CO2	Learn and Apply different Data Analysis Techniques	**K1, K2 & K3**
CO3	Learn Data Streams	**K1, K2**
CO4	Learn and Apply the Mining and Clustering Techniques	**K1, K2 & K3**
CO5	Learn and apply the different data Visualization Concepts	**K2&k3**

Knowledge Level – K1 – Remember, K2 – Understand, K3 – Apply, K4 – Analyze

MAPPING WITH PROGRAMME OUTCOME

COS	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6
CO1	S	S	S	M	S	S
CO2	S	S	S	M	S	M
CO3	S	S	M	M	S	M
CO4	S	S	S	S	S	M
CO5	S	S	S	M	S	M

S – Strong ***M – Medium*** ***L – Low***

SYLLABUS

UNIT I – Introduction to Big Data **14 hrs**

Introduction to Big Data Platform – Challenges of conventional systems - Web data – Evolution of Analytic scalability, analytic processes and tools, Analysis VS Reporting - Modern data analytic tools, Statistical concepts: Sampling distributions, re-sampling, statistical inference, prediction error.

UNIT II – Data Analysis **16 hrs**

Regression modeling, Multivariate analysis, Bayesian modeling, inference and Bayesian networks, Support vector and kernel methods, Analysis of time series: linear systems analysis, nonlinear dynamics - Rule induction - Neural networks: learning and generalization, competitive learning, principal component analysis and neural networks.

UNIT III – Mining Data Streams **14 hrs**

Introduction to Streams Concepts – Stream data model and architecture - Stream Computing, Sampling data in a stream – Filtering streams – Counting distinct elements in a stream –Estimating moments – Counting oneness in a window – Decaying window – Real Time Analytics Platform(RTAP) applications.

UNIT IV – Frequent Itemsets and Clustering **16 hrs**

Mining Frequent Itemsets - Market based model – 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 – CLIQUE and PROCLUS – Frequent pattern based clustering methods – Clustering in non-Euclidean space – Clustering for streams and Parallelism.

UNIT V – Frameworks and Visualization **15 hrs**

MapReduce – Hadoop, Hive, Map R – Sharding – NoSQL Databases - S3 - Hadoop Distributed file systems – Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications

Distribution of Marks: Theory: 80% and Application: 20%

TEXT BOOKS

S.No	Authors	Title	Publishers	Year of Publication
1.	Michael Berthold, David J. Hand	Intelligent Data Analysis	Springer	2007
2	Anand Rajaraman, Jeffrey David Ullman	Mining of Massive Datasets	Cambridge University Press	2012

REFERENCE BOOKS

S.No	Authors	Title	Publishers	Year of Publication
1.	Bill Franks	Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with advanced analytics	John Wiley & sons	2012
2.	Glenn J. Myatt	Making Sense of Data	John Wiley & Sons	2007

3.	Pete Warden	Big Data Glossary	O"Reilly	2011
4.	Jiawei Han, Micheline Kamber	Data Mining Concepts and Techniques	Elsevier	Reprinted 2008
5.	Viktor Mayer Schonberger, Kenneth Cukier	Big Data: A Revolution That Will Transform How We Live, Work and Think	John Murray Publishers	2013
6	Gareth james, Daniela Witten, Trevor Hastie, Robert Tibshirani	An Introduction to Statistical Learning with Applications in R	Springer	2013
7	EMC Education Service	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data	EMC Education Service - Wiley	2015
8	Daniel Covington	Analytics: Data Science, Data Analysis and Predictive Analytics for Business	Audible	5th Edition

WEB RESOURCES

1. https://data-flair.training/blogs/big-data-tutorials-home/

TEACHING METHODOLOGY

- Class room teaching.
- Group discussions
- Seminars and Smart Class room
- Demo using systems
- Chart/Assignment
- Simulation Model

SYLLABUS DESIGNERS

Mrs.B.ARULMOZHI , Assistant Professor and Head, Dept of Computer Science

Mrs.S.SHANTHI., Assistant Professor, Dept of Computer science