21AID10	DATA ANALYTICS	L	T	P	C
		3	0	0	3

Course Objectives

- To understand the basic principles of Data Analytics
- To learn the various Data Analytic methods
- To understand the various clustering algorithms and its application on data
- To work with stream data model and computing

UNIT I	INTRODUCTION TO DATA ANALYTICS	9 Hours

Introduction to Data Analytics - Types of Data Analytics - Predictive Analytics - Simple linear regression - Multiple linear regression - Auto regression - Moving Average - Autoregressive Integrated Moving Average - Data Preprocessing - Data Cleaning - Data Integration and Transformation - Data Reduction - Descriptive data analytics - measures of central tendency - measures of location of dispersions.

UNIT II	ASSOCIATION RULE MINING	9 Hours

Association Rule Mining: Efficient and Scalable Frequent Item set Mining Methods - Mining Various Kinds of Association Rules - Association Mining to Correlation Analysis - Constraint Based Association Mining - Cluster Analysis: Types of Data in Cluster Analysis - A Categorization of Major Clustering Methods - Partitioning Methods - Hierarchical methods.

UNIT III	STREAM CONCEPTS	9 Hours

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 - case studies - real time sentiment analysis - stock market predictions.

UNIT IV	GRAPH ANALYTICS	9 Hours

Using Graph Analytics for Big Data: Graph Analytics - The Graph Model - Representation as Triples - Graphs and Network Organization - Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs - Features of a Graph Analytics Platform - Considerations: Dedicated Appliances for Graph - Graph QL

UNIT V	NoSQL DATABASES	9 Hours

NoSQL Databases - Schema-less Models - Increasing Flexibility for Data Manipulation - Key Value Stores - Document Stores - Tabular Stores - Object Data Stores - Graph Databases Hive-Sharding-Hbase - Analyzing big data with twitter - Big data for E-Commerce - Big data for blogs - Review of Basic Data Analytic Methods using R.

UNIT VI	LATEST TRENDS	

Latest Trends

TOTAL PERIODS: 45

Course Outcomes:

- Evaluate the use of data from acquisition through cleaning, warehousing, analytics, and visualization to the ultimate business decision

- Mine data and carry out predictive modeling and analytics to support business decision-making
- Suggest prescriptive modeling techniques for real-world problems
- Execute real-time analytical methods on streaming datasets to react quickly to customer needs

Text books:

1. Jiawei Han, MichelineKamber, Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2011.
2. A. Rajaraman, J. Ullman, "Mining Massive Data Sets", Cambridge University Press, 2012.
3. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, No SQL, and Graph", 2013.

Reference Books:

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, "Probability & Statistics for Engineers & Scientists", Ninth Edition, Prentice Hall Inc.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning, Data Mining, Inference, and Prediction", Second Edition, Springer, 2014.
3. G James, D. Witten, T Hastie, R. Tibshirani, "An Introduction to Statistical Learning: With Applications in R", Springer, 2013.
4. Mohammed J. Zaki, Wagner Meira, "Data Mining and Analysis", Cambridge, 2012.
5. E. Alpaydin, "Introduction to Machine Learning", MIT Press, 2014.