

Data Analytics (KIT 601)		
Course Outcome (CO)		Bloom's Knowledge Level (KL)
At the end of course , the student will be able to		
CO 1	Discuss various concepts of data analytics pipeline	K ₁ , K ₂
CO 2	Apply classification and regression techniques	K ₃
CO 3	Explain and apply mining techniques on streaming data	K ₂ , K ₃
CO 4	Compare different clustering and frequent pattern mining algorithms	K ₄
CO 5	Describe the concept of R programming and implement analytics on Big data using R.	K ₂ , K ₃
DETAILED SYLLABUS		3-0-0
Unit	Topic	Proposed Lecture
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.	08
II	Data Analysis: 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 generalisation, competitive learning, principal component analysis and neural networks, fuzzy logic: extracting fuzzy models from data, fuzzy decision trees, stochastic search methods.	08
III	Mining Data Streams: 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.	08
IV	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, CLIQUE and ProCLUS, frequent pattern based clustering methods, clustering in non-euclidean space, clustering for streams and parallelism.	08
V	Frame Works and Visualization: MapReduce, Hadoop, Pig, Hive, HBase, MapR, Sharding, NoSQL Databases, S3, Hadoop Distributed File Systems, Visualization: visual data analysis techniques, interaction techniques, systems and applications. Introduction to R - R graphical user interfaces, data import and export, attribute and data types, descriptive statistics, exploratory data analysis, visualization before analysis, analytics for unstructured data.	08
Text books and References:		
1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer 2. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press. 3. John Garrett, Data Analytics for IT Networks : Developing Innovative Use Cases, Pearson Education		