



AMITY UNIVERSITY

UTTAR PRADESH

Annexure 'AAB-CD-01'

Course Title: DATA SCIENCE AND BIG DATA ANALYTICS

Credit Units:

L	T	P/S	SW/F W	TOTAL CREDIT UNITS
3	1	-	-	4

Course Level: PG
Course Code: IT715

Course Objectives:

Approach business problems data-analytically by identifying opportunities to derive business value from data. Know the basics of big data analytics techniques and how they can be applied to extract relevant business intelligence

Pre-requisites: Understanding of basic concepts of Information and Database System, Data Structures and Algorithms and Probabilities and Statistics

Course Contents/Syllabus:

	Weightage
Module I Introduction of Data Science	
Introduction to Data Science, Review of probability and probability distribution, Bayes rules, lifecycle of data science, dimensionality reduction using Principal component analysis and random projection.	20
Module II Data Processing	
Data discovery, data preparation, model planning, model building, communicating results and findings, and operationalizing.	20
Module III Big Data Analytics and Life Cycle methods	
Overview of Big Data Analytics, Big Data Analytics in industry verticals, Key roles for analytics projects, Phases of Big Data Analytics lifecycle.	20
Module IV Big Data Analytics Algorithm	
Naïve Bayesian Classifier, categorization using K-means clustering and association rules, predictive modeling using decision trees, linear and logistic regression, and time-series analysis, and text analysis.	20

Module V Big Data Analytics Applications using Hadoop	20
Introduction to Hadoop, Lifecycle of Hadoop, Hadoop Architecture, case study of Hadoop.	

Student Learning Outcomes:

- Analyze a structured lifecycle approach to data science and big data analytics projects
- Select visualization techniques and tools to analyze big data and create statistical models
- Demonstrate tool HaDoop, and in-database analytics.

Pedagogy for Course Delivery:

Subject will be taught on the basis of lectures, concepts learn in the classroom using various real life situations and discussing case study covering different module with special reference to Data Scientist to analyze the supervised and unsupervised learning with the help of Hadoop tool.

Assessment/ Examination Scheme:

Theory L/T (%)	Lab/Practical/Studio (%)	End Term Examination
100%	NA	100%

Theory Assessment (L&T):

Continuous Assessment/Internal Assessment					End Term Examination
Components (Drop down)	Mid-Term Exam	Assignment	Viva	Attendance	
Weightage (%)	10%	8%	7%	5%	70%

Text Reading:

- Provost & Fawcett, “Data Science for Business: Fundamental principles of data mining and data analytic Thinking”.
- Jeffrey Stanto “Introduction to Data Science”, Syracuse Universit, Tech Data Science2012
- Noreen Burlingame, “The little book on Big Data”, New Street publishers, 2012.
- Norman Matloff, “The Art of R Programming: A Tour of Statistical Software Design”, No Starch Press 1 edition, 2011.
- http://www.johndcook.com/R_language_for_programmers.html.
- <http://bigdatauniversity.com/>.
- <http://home.ubalt.edu/ntsbarsh/stat-data/topics.htm#rintroduction>