

**B.Tech(Hons) in Computer Science & Engineering
with specialization in Big Data Analytics
Semester wise Break Up of Courses**

SEMESTER V														
S. No.	Cours e Code	Course Name	Cour se Type	Periods			CCA				ES E	Tota l		Credi t
				L	T	P	C T	A T	Tot al	P S	TE	P E		
1	BBME-501	Introduction to Big Data and Data Analytics	PECM-1	3	0	0	20	10	30	-	70	-	100	3
2	BBME-551	Data Mining & Predictive Modelling	PECM-2	0	0	4	-	-	-	15	-	35	50	2
TOTAL												150	5	

BBME-501	Introduction to Big Data and Data Analytics	L: 3	T: 0	P: 0	C: 3
----------	---	------	------	------	------

Introduction to Big Data and Data Analytics

Objective

The principle objective of this course is to enable understudies to learn, comprehend, and practice enormous information investigation and Big Data approaches, which incorporate the investigation of present day registering enormous information innovations and scaling up data analytics strategies concentrating on industry applications. For the most part the course goals are: conceptualization and synopsis of enormous information and Big Data, inconsequential information versus enormous information, huge information registering advances,

COURSE OUTCOME

1. Capacity to take care of issues related with bunch learning and internet learning, and the enormous information qualities, for example, high dimensionality, progressively developing information and specifically adaptability issues.
2. Ability to comprehend and apply scaling up AI methods and related registering procedures and innovations.
3. Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.
4. Ability to perceive and actualize different methods for choosing appropriate model parameters for various AI systems.
5. Ability to coordinate AI libraries and scientific and factual instruments with present day advances like hadoop and map reduce.

UNIT I: INTRODUCTION TO BIG DATA AND HADOOP Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, analyzing Data with Unix tools, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Info sphere Big Insights and Big Sheets.

UNIT II: HDFS(Hadoop Distributed File System) The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

UNIT III: Map Reduce Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Unit IV: Hadoop Eco System Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive: Hive Shell, Hive Services, Hive Meta store, Comparison with Traditional Databases, Hive QL, Tables, Querying Data and User Defined Functions. H base: H Basics, Concepts, Clients, Example, H base versus RDBMS. Big SQL: Introduction

UNIT V: Data Analytics with R Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with Big R.

Text Books

- Tom White “Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012.
- Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

References

- Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
- Jay Liebowitz, “Big Data and Business Analytics” Auerbach Publications, CRC press (2013)
- Tom Plunkett, Mark Hornick, “Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop”, McGraw-Hill/Osborne Media (2013), Oracle press.
- Anand Rajaraman and Jeffrey David Ulman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
- Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
- Glen J. Myat, “Making Sense of Data”, John Wiley & Sons, 2007
- Pete Warden, “Big Data Glossary”, O’Reily, 2011.

Data Mining & Predictive Modelling

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-551	Data Mining & Predictive Modelling	0	0	4	2

OBJECTIVES:

- To analyze the data using statistical methods.
- To understand and demonstrate data mining.

List of Experiments

1. Data Analysis- Getting to know the Data (Using ORANGE,WEKA)
 - Parametric - Means, T-Test, Correlation
 - Prediction for numerical outcomes - Linear regression
 - Correlation analysis
 - Preparing data for analysis
 - Pre-processing techniques
2. Data Mining (Using ORANGE,WEKA or any open source data mining tool)
 - Implement clustering algorithm
 - Implement classification using
 - Decision tree
 - Back propagation
 - Visualization methods.

COURSE OUTCOMES:

Students who complete this course will be able to

- use statistical techniques to carry out the analysis of data.
- gain hands-on skills and experience on data mining tools.

SEMESTER VI														
S. No.	Course Code	Course Name	Course Type	Periods			CCA				ES	Total		Credit
				L	T	P	CT	AT	Total	PS	TE	PE		
1	BBME-601	Algorithms for advanced Analysis	PECM-3	3	0	0	20	10	30	-	70	-	100	3
2	BBME-651	Programming in R for Data Analysis	PECM-4	0	0	4	-	-	-	15	-	35	50	2
TOTAL												150	5	

Algorithms for advanced Analysis

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-601	Algorithms for advanced Analysis	3	0	0	3

UNIT-I Algorithm Analysis: Asymptotic Notation, Amortization.

Basic Data Structures: Stacks and Queues, Vectors, Lists and Sequences, Trees, Priority Queues, Heaps, Dictionaries and Hash Tables.

Search Trees: Ordered Dictionaries and Binary Search Trees, AVL Trees, Bounded-Depth Search Trees, and Splay Trees.

UNIT-II Fundamental Techniques: The Greedy Method, Divide-and-Conquer, and Dynamic Programming.

Graphs: The Graph Abstract Data Type, Data Structures for Graphs, Graph Traversal, Directed Graphs.

UNIT-III Weighted Graphs: Single-Source Shortest Paths, All-Pairs Shortest Paths, Minimum Spanning Trees.

Network Flow and Matching: Flows and Cuts, Maximum Bipartite Matching, Minimum-Cost Flow.

UNIT-IV Text Processing: Strings and Pattern Matching Algorithms, Tries, Text Compression, Text Similarity Testing. Number Theory and Cryptography: Fundamental Algorithms involving numbers, Cryptographic Computations, Information Security Algorithms and Protocols.

UNIT-V Computational Geometry: Range Trees, Priority Search Trees, Quad trees and k-D Trees, Convex Hulls.

REFERENCES

- M T Goodrich, R Tomassia. “Algorithm Design –Foundations, Analysis, and Internet Algorithms”, John Wiley,2006.
- E Horowitz S Sahani, S Rajasekaran, “Computer Algorithms”,Second Edition,Silicon Press,2007.
- Aho, A V Hopcraft, Ullman J D, “The Design and Analysis of Computer Algorithms”, Pearson Education, 2007.
- Hari Mohan Pandey, “Design Analysis and Algorithms”, Firewall Media, 2008.
- Cormen, Lieserson, Rivest “Introduction to Algorithms”, 2ndEdition, MIT Press, 2009.

Programming in R for Data Analysis

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-651	Programming in R for Data Analysis	0	0	4	2

Course Outcomes:

Students will be able to:

1. Recognize and make appropriate use of different types of data structures
2. Use R to create sophisticated figures and graphs
3. Identify and implement appropriate control structures to solve a particular programming problem
4. Design and write functions in R and implement simple iterative algorithms.

Lab:

- Introduction to R Overview of R programming, Evolution of R, Applications of R programming, Basic syntax.
- Basic Concepts of R Reserved Words, Variables & Constants.
- Data structures in R Vectors, Matrix
- Control flow If...else, If else() Function
- Functions R Functions, Function Return Value
- String construction rules
- Packages Study of different packages in R
- R Data Reshaping Joining Columns and Rows in a Data Frame
- Working with files Read and writing into different types of files
- R object and Class Object and Class, R S3 Class, R S4 Class
- Data visualization in R and Data Management Bar Chart, Dot Plot
- Statistical modelling and Databases in R Mean, mode, median

SEMESTER VII														
S. No.	Course Code	Course Name	Course Type	Periods			CCA				ES	Total		Credit
				L	T	P	C	A	Total	PS	TE	PE		
1	BBME-701	E-Commerce and Social Media Analytics	PECM-5	3	0	0	20	10	30	-	70	-	100	3
TOTAL												100	3	

E-Commerce and Social Media Analytics

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-701	E-Commerce and Social Media Analytics	3	0	0	3

COURSE OUTCOME:-

1. Students can learn how to leverage new models in business and e-commerce to increase profitability
2. Students are able to generate Successful Social media and digital marketing techniques
3. Students can interact with real-world application of digital marketing and e-commerce methods

Unit 1: Social Media Defined Social media design framework –social media examples –The Network perspective –types of networks

Unit 2: Web Analytics 2.0 paradigm Clickstream Analysis –Eight critical web metrics –Bounce rate –Exit rate –Conversion rate –Engagement –Attributes of great metrics –A Web Analytics Primer Understanding Visitor Acquisition Strengths –Click Density Analysis –Measuring Visits to Purchase –Search Engine Optimization (SEO) Analysis –Direct Traffic Analysis

Unit 3: Measuring outcome Key Performance Indicators (KPIs) –Moving beyond Conversion Rates –Measuring Macro and Micro Conversions –Measuring Success for a Non-ecommerce Website –Lab Usability tests –Surveys –Types of Surveys.

Unit 4: A/B Testing Multivariate Testing –Competitive Intelligence Data Sources, Types, and Secrets –Website Traffic Analysis –Search and Keyword Analysis –Audience Identification and Segmentation Analysis.

References:-

- Derek L. Hansen, Ben Sheiderman, Marc A. Smith, .AnalyzingSocialMediaNetworkswith NodeXL, Morgan Kaufmann, 2011
- Avinash Kaushik. 2009. Web Analytics 2.0, Wiley Publishing, Inc, 2010.LTPC3003

SEMESTER VIII														
S. No.	Course Code	Course Name	Course Type	Periods			CCA				ES	Total		Credit
				L	T	P	CT	AT	Total	PS	TE	PE		
1	BBME-801	Big Data Analytics for IOT	PECM-6	3	0	0	20	10	30	-	70	-	100	3
2	BBME-802	Web Social and Mobile Analytics	PECM-7	3	0	0	20	10	30	-	70	-	100	3
3	BBME-851	Big Data Technology	PECM-8	0	0	4	-	-	-	15	-	35	50	2
TOTAL												250	8	

Big Data Analytics for IOT

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-801	Big Data Analytics for IOT	3	0	0	3

Mode of Teaching: Regular Mode

OBJECTIVES

1. To learn the concepts of big data analytics
2. To learn the concepts about Internet of things
3. To understand and implement smart systems

COURSE OUTCOME

1. Ability to know about Internet of Things, also called as, IOT. Get to know the eco-system of IOT, Historical perspective, Industrial automation technology fused into information technology.
2. Ability to describe that in what way the big data is creating great opportunity in IOT domain for aspiring big data analytics enthusiasts. Learn the steps to define the cognitive problem statement and then turn to solutions.
3. Ability to Start thinking rationally on IOT application with intelligence built into it.

UNIT I -BIG DATA PLATFORMS FOR THE INTERNET OF THINGS

Big Data Platforms for the Internet of Things: network protocol-data dissemination –current state of art-Improving Data and Service Interoperability with Structure, Compliance, Conformance and Context Awareness: interoperability problem in the IoT context-Big Data Management Systems for the Exploitation of Pervasive Environments -Big Data challenges and requirements coming from different Smart City applications.

UNIT II -RFID FALSE AUTHENTICATIONS

On RFID False Authentications: YA TRAP –Necessary and sufficient condition for false authentication prevention -Adaptive Pipelined Neural Network Structure in Self-aware Internet of Things: self-healing systems-Role of adaptive neural network-Spatial Dimensions of Big Data: Application of Geographical Concepts and Spatial Technology to the Internet of Things-Applying spatial relationships, functions, and models.

UNIT III-FOG COMPUTING

Fog Computing: A Platform for Internet of Things and Analytics: a massively distributed number of sources -Big Data Metadata Management in Smart Grids: semantic inconsistencies – role of metadata.

UNIT IV -WEB ENHANCED BUILDING

Toward Web Enhanced Building Automation Systems: heterogeneity between existing installations and native IP devices -loosely-coupled Web protocol stack –energy saving in smart building-Intelligent Transportation Systems and Wireless Access in Vehicular Environment Technology for Developing Smart Cities: advantages and achievements-Emerging Technologies in Health Information Systems: Genomics Driven Wellness Tracking and Management System (GO-WELL) –predictive care –personalized medicine.

UNIT V -SUSTAINABILITY DATA AND ANALYTICS

Sustainability Data and Analytics in Cloud-Based M2M Systems -potential stakeholders and their complex relationships to data and analytics applications -Social Networking Analysis - Building a useful understanding of a social network -Leveraging Social Media and IoT to Bootstrap Smart Environments : lightweight Cyber Physical Social Systems -citizen actuation.

REFERENCES

1. Stackowiak, R.,Licht, A.,Mantha, V.,Nagode, L.,” Big Data and The Internet of Things Enterprise Information Architecture for A New Age”, Apress, 2015.
2. Dr. John Bates, “Thingalytics -Smart Big Data Analytics for the Internet of Things”, john Bates, 2015
3. Stackowiak, R.,Licht, A.,Mantha, V.,Nagode, L.,” Big Data and The Internet of Things Enterprise Information Architecture for A New Age”, Apress, 2015.
- 4.Dr. John Bates, “Thingalytics -Smart Big Data Analytics for the Internet of Things”, john Bates, 2015.

Big Data Technology

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-851	Big Data Technology	0	0	4	2

Course outcome

1. Appreciate the software needs of an IoT project
2. Understand how data is managed in an IoT network
3. Apply software solutions for different systems and Big Data to your IoT concept designs
4. Create Python scripts to manage large data files collected from sensor data and interact with the real world via actuators and other output devices.

Lab

- NoSQL Lab using (MongoDB/Redis/Cassandra/CouchDB/Hbase using HDFs etc): Introduction to Nosql, Difference between RDBMS to NOSQL,JSON and BSON documents.
- Introduction to Mongo DB/..and its Features, Database, Collection and Documents, Various Data Types in Mongo DB/.
- Introduction to mongo/..shell, CRUD Operations, Database Operations, Read and Write Operations, Aggregation, Data Modeling Introduction, Data Modeling Concept, Storage Engine, Indexing, Replication Concept, Failover & Recovery
- Multidimensional Data Modeling using OLAP: Introduction of Data warehousing and OLAP, example of a Data Warehouse and Data mart, Data Cleaning and integration, Data analysis techniques, Transformation algorithms, Integrations.

Web Social and Mobile Analytics

SL. NO	Engineering Minor Elective		L	T	P	C
1	BBME-802	Web Social and Mobile Analytics	3	0	0	3

COURSE OUTCOME

1. Ability to describe SMAC and the scope of each of its core components
2. Ability to describe social media marketing and its impact on business and impact using STATA module.
3. Ability to explain mobile platforms and their uses in various industries
4. Ability to explain the impact of big data and analytics in businesses

Unit-I

History of Social media, Basics of Social media and business models, NY Times Paywall Case, Trends in social and digital marketing. Paid/Earned/Owned media and Inbound/Outbound, Ford Fiesta Case, Data-driven decision making, Metrics in social media and digital marketing, Web search engine ranking , Search engine marketing, Air France Internet Marketing Case Analyses in class via Excel.

Unit-II

STATA module on basics of econometrics and in-class data analytics, BBVA Budget Allocation Digital Case, Hub spot Case, STATA module on paid search advertising and in-class data analytics, Field experiments in digital marketing.

Unit-III

Mobile path to purchase, mobile couponing, mobile show rooming and location based advertising, STATA module on mobile services adoption and in-class data analytics, Mobile advertising, cross-device synergies, mobile commerce, and mobile apps, STATA module on demand estimation of mobile apps and in-class data analytics.

Unit-IV

Bank of America Mobile Banking Case, Big Data analytics and sentiment analysis , Text Mining of user-generated content(UGC), Tripadvisor case, STATA module on relationship between UGC and product prices, Social communities, Minnesota Wild Case and Facebook Insights Data Analytics via Excel.

Unit-V

Open Innovation and harnessing the wisdom of the crowds, Crowd funding, Mobile ecosystem, Digital attribution analyses via Excel, Socialization of TV, Bluefin Twitter Labs case, Amazon, FB, Google and Apple case, STATA module on crowd funding and in-class data analytics.

REFERENCES:

1. Gustavo Alonso, Fabio Casati, Harumi Kuno, Vijay Machiraju, “Web Services: Concepts, Architectures and Applications”,Springer,ISBN: 3662108763,2013.
2. Olaf Zimmermann, Mark Tomlinson, Stefan Peuser, “Perspectives on Web Services: Applying SOAP, WSDL and UDDI to Real-World”, Springer, ISBN:978-3-642-62468-1,2012.
3. Alex Belotserkovskiy, Stephen Kaufman, Nikhil Sachdeva, “Building Web Services with Microsoft Azure”, first edition, Packt publishing, ISBN: 978-1-78439-8,2015.