MANIPAL UNIVERSITY JAIPUR

School of Computing and IT

Department of Computer Science and Engineering
Course Hand-out

Big Data Analytics | CS 1701 | 4 Credits | 3 1 0 4

Session: Aug 20– Dec 20 | Faculty: Dr. Santosh Kumar Vishwakarma /Mr. Priyank Singh Hada / Ms. Bali Devi | Class: VII Semester

A. **Introduction:** This course provides practical foundation level in big data Analytics. The course provides grounding in basic and advanced methods to big data technology and tools, including Hadoop and its ecosystem.

B. **Course Outcomes:** At the end of the course, students will be able to

[1701.1] Apply the concept of Data Analytics for Industry-specific Big Data Challenges.

[1701.2] Design structural lifecycle approach to data science and big data analytics for project development.

[1701.3] Analyse big data applications and create statistical models to solve the industry challenges that impact on increase employability skills.

[1701.4] Interpret various theoretical models as experiment of big data analytics using the R tool/R studio.

[1701.5] Apply analytical platforms such as Hadoop, NoSQL DBMS software for performing advanced data analytics.

C. **PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES**

[PO.1] Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems

[PO.2] Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences

[PO.3] Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations

[PO.4] Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions

[PO.5] Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations

[PO.6] The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice

[PO.7] Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development

[PO.8] Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices

[PO.9] Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings

[PO.10] Communication: Communicate effectively_on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions

[PO.11] Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments

[PO.12] Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

[PSO.1] Will be able to design, develop and implement efficient software for a given real life problem.

[PSO.2] Will be able to apply knowledge of AI, Machine Learning and Data Mining in analyzing big data for extracting useful information from it and for performing predictive analysis.

[PSO.3] Will be able to design, manage and secure wired/ wireless computer networks for transfer and sharing of information.

D. Assessment Plan:

Criteria	Description	Maximum Marks
Internal Assessment (Summative)	Sessional Exam I (Closed Book)	15
	Sessional Exam II (Closed Book)	15
	Quizzes and Assignments (Accumulated and Averaged)	30
End Term Exam (Summative)	End Term Exam (Closed Book)	40
	Total	100
Attendance (Formative)	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking up the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves.	
Make up Assignments (Formative)	Students who misses a class will have to report to the teacher about the absence. A makeup assignment on the topic taught on the day of absence will be given which has to be submitted within a week from the date of absence. No extensions will be given on this. The attendance for that day of absence will be marked blank, so that the student is not accounted for absence. These assignments are limited to a maximum of 5 throughout the entire semester.	
Homework/ Home Assignment/ Activity Assignment (Formative)	There are situations where a student may have to work in home, especially before a flipped classroom. Although these works are not graded with marks. However, a student is expected to participate and perform these assignments with full zeal since the activity/ flipped classroom participation by a student will be assessed and marks will be awarded.	

E. SYLLABUS

Introduction to big data: definition, need and evolution of BDA, applications of Big Data, **Big Data Analytics:** Analysing big data, sources of big data, characteristics of big data (4 V's), Drivers of BDA, types of data, structured vs. unstructured data, data marts, Case study based tutorial, Differences between traditional DWDM and BDA, Limitations of traditional RDBMSs to store and analyse Big Data.

Data science, definition and concepts, data scientists: key competencies and characteristics of data scientists, more discussions on data science: data wrangling, data munging, data jujitsu, Tutorial based on data science applications.

Big Data Analytics Ecosystem, State of the Practice in Analytics: Data Analytics Lifecycle and discussions, Roles for a Successful Analytics Project; Case Study to apply the data analytics lifecycle, Analytical databases and DW appliances; Hadoop distributions – Comparing various BDA tools

Analyzing and Exploring the Data: Challenges when managing and analysing big data, the role of Data Virtualization in a Big Data environment; Why to visualize data.

Statistics for Model Building and Evaluation: Statistics in the Analytic Lifecycle, Hypothesis Testing, Difference of means **Advanced Analytics – Theory and Methods Overview:** K-means clustering, Association Rules, Linear Regression, Logistic Regression, Naïve Bayesian Classifiers, Tutorial based on advanced analytics, Decision Trees, Time Series Analysis, Text Analytics; Tutorial based on analytics

Big Data Platforms and Storage Options: The new multi-platform Analytical Ecosystem; Beyond the Data Warehouse

- Analytical databases, Hadoop and NoSQL DBMSs

TEXT BOOKS

T. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services.

REFERENCE BOOKS

R. Michael Minelli, Michele Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends", John wiley,2013

F. Lecture Plan:

Lecture No.	Topics	Session outcome	Mode of Delivery	Corresponding CO	Modes of Accessing the CO
1.	Definition, Need and evolution of BDA	Will get to know the need of the BDA	Lecture	1701.1	Mid Term I, Quiz & End Term
2.	Applications of Big Data	Analyse on various real time applications of BDA	Discussion	1701.1	Mid Term I, Quiz & End Term
3.	Analyzing big data	Identify the Need of Big Data	Lecture/PPT	1701.1	Mid Term I, Quiz & End Term
4.	sources of big data, Drivers of BDA	Conclude on various Data sources	Lecture/PPT	1701.1	Mid Term I, Quiz & End Term
5.	characteristics of big data (4 V's)	List out the key characteristics of the BDA	Lecture/PPT	1701.1	Mid Term I, Quiz & End Term
6.	Types of data, structured vs. unstructured data, data marts,	Deal with variety of the data	Flipped Class	1701.1	Mid Term I, Quiz & End Term
7.	Case study-based tutorial	Real time application – Discussion	Activity	1701.1	Mid Term I, Quiz & End Term
8.	Differences between traditional DWDM and BDA, Limitations of traditional RDBMSs to store and analyse Big Data,	Identify the differences between Data Mining and Data Analytics	Lecture/PPT	1701.1	Mid Term I, Quiz & End Term
9.	Data science, definition and concepts,	Would define the definition of Data Science	Flipped Classes	1701.1	Mid Term I, Quiz & End Term
10.	key competencies and characteristics of data scientists	List out the key competencies and characteristics of data scientists	Lecture	1701.2	Mid Term I, Quiz & End Term
11.	More discussions on data science: data wrangling, data munging	Define Data Wrangling and munging	Flipped Classes	1701.2	Mid Term I, Quiz & End Term
12.	data jujitsu, Tutorial based on data science applications	Work on data jujitsu	Lecture	1701.2	Mid Term I, Quiz & End Term
13.	Big Data Analytics Ecosystem,	Analyse on BDA Ecosystem	Lecture	1701.2	Mid Term I, Quiz & End Term
14.	Data Analytics Lifecycle and discussions	Explain on BDA lifecycle	Discussion	1701.2	Mid Term II, Quiz & End Term
15.	Roles for a Successful Analytics Project	Identify the key indicators to make analytics project	Lecture	1701.2	Mid Term II, Quiz & End Term
16.	Case Study to apply the data analytics lifecycle	Implement some real time issues	Activity	1701.2	Mid Term II, Quiz & End Term
17.	Analytical databases and DW appliances	Compare the DW and Analytical DB's	Flipped Class	1701.2	Mid Term II, Quiz & End Term

18.	Hadoop distributions – Comparing various BDA tools	Access the Hadoop platforms	Lecture	1701.2	Mid Term II, Quiz & End Term
19.	Challenges when managing and analysing big data	List the challenges in BDA process.	Discussion	1701.3	Mid Term II, Quiz & End Term
20.	The role of Data Virtualization in a Big Data environment	Create Data Virtualization	Lecture/PPT	1701.3	Mid Term II, Quiz & End Term
21.	Why to visualize data	Mention the advantage of visualization	Lecture/PPT	1701.3	Mid Term II, Quiz & End Term
22.	Statistics in the Analytic Lifecycle	Will do Statistics in the Analytic Lifecycle	Lecture/PPT	1701.3	Mid Term II, Quiz & End Term
23.	Hypothesis Testing	Will perform Hypothesis Testing	Activity	1701.3	Mid Term II, Quiz & End Term
24.	Difference of means	Perform Difference of means	Activity	1701.3	Mid Term II, Quiz & End Term
25.	K-means clustering	Perform K-means clustering	Activity	1701.4	Mid Term II, Quiz & End Term
26	Association Rules	Perform Association Rules	Activity	1701.4	Mid Term II, Quiz & End Term
27	Linear Regression, Logistic Regression	Perform Linear Regression, Logistic Regression	Activity	1701.4	Mid Term II, Quiz & End Term
28	Naïve Bayesian Classifiers	Perform Naïve Bayesian Classifiers	Activity	1701.4	Mid Term II, Quiz & End Term
29	Tutorial based on advanced analytics	Will implement real time issues	Activity	1701.4	Mid Term II, Quiz & End Term
30	Decision Trees	Perform Decision Trees	Activity	1701.4	Mid Term II, Quiz & End Term
31	Time Series Analysis	Perform Time Series Analysis	Activity	1701.4	Mid Term II, Quiz & End Term
32	Text Analytics	Perform Text Analytics	Activity	1701.4	Mid Term II, Quiz & End Term
33	Tutorial based on analytics	Will implement real time issues	Activity	1701.4	Mid Term II, Quiz & End Term
34	The new multi-platform Analytical Ecosystem	Understand Analytical ecosystem	Lecture	1701.5	Quiz & End Term
35	Beyond the Data Warehouse - Analytical databases	Understand Analytical DB's	Lecture/PPT	1701.5	Quiz & End Term
36	Introduction to Hadoop and Hadoop ecosystem	Will learn basics of Hadoop	Lecture/PPT	1701.5	Quiz & End Term
37,38	Advantages of Hadoop & HDFS	Will learn basics of Hadoop file systems	Lecture/PPT	1701.5	Quiz & End Term
39,40	MAPREDUCE, Hadoop Clustering	Will learn basics of Hadoop clustering	Lecture/PPT	1701.5	Quiz & End Term
41,42,43	Ecosystem component: Oozie, Spark, Scala, ML lib, Pig, Hive, kafka, flume	Will run applications on different modern tools	Flipped Class	1701.5	Quiz & End Term
44	Hadoop Usage for Applications	Implement projects on Hadoop platform	Activity	1701.5	Quiz & End Term
45	Introduction to NoSQL Data base, Advantages of NoSQL	Will learn basics of NoSQL	Lecture	1701.5	Quiz & End Term
46, 47	Mongo dB	Incur knowledge on advantages of No SQL	Lecture/PPT	1701.5	Quiz & End Term

47,48	Cassandra	Incur knowledge on No SQL	Lecture/PPT	1701.5	Quiz & End Term
50	NoSQL DBMSs projects & case studies	Incur knowledge on advantages of No SQL	Activity	1701.5	Quiz & End Term