



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN – PART I			
Name of the programme and specialization	M.Tech. Computer Science and Engineering		
Course Title	Big Data Analytics and Mining		
Course Code	CS623	No. of Credits	3
Course Code of Pre-requisite subject(s)	NIL		
Session	July 2023	Section (if, applicable)	I Semester
Name of Faculty	Dr. E. Sivasankar	Department	CSE
Email	sivasankar@nitt.edu	Telephone No.	9443744680
Name of Course Coordinator(s) (if, applicable)	NA		
E-mail		Telephone No.	
Course Type	Programme Elective		
Syllabus (approved in Senate)			
<p>UNIT-I Big data overview State of the practice in Analytics-Key roles for new big data ecosystem Data Analytics Lifecycle-Data analytics life cycle overview- Discovery- Data Preparation-Model Planning-Model Building-Communicate Results - operationalize</p> <p>UNIT-II Introduction to R Exploratory Data Analytics-Statistical methods for evaluation Hadoop & Map Reduce framework for R, R with Relational Database Management Systems, R with Non-Relational (NoSQL) DBs</p> <p>UNIT-III Clustering Overview of Clustering-K-means, Association Rules-Overview-Apriori Algorithm-Evaluation of candidate rules-An Example: Transactions in grocery Store-Validation and Testing-Diagnostics, Regression-Linear Regression-Logistic Regression-Reason to choose and Cautions-Additional Regression Models</p> <p>UNIT-IV Classification Decision Trees - Naïve Bayes-Diagnostics of Classifiers-Additional classification methods, Time series Analysis-Overview of Time series analysis-ARIMA Model-Additional methods, Text Analysis-Text analysis steps - A text analysis Example-Collecting raw Text-Representing Text-Term Frequency—Inverse document frequency(TFIDF)-Categorizing documents by Topics-Determining Sentiments-Gaining insights</p> <p>UNIT-V Analytics for Unstructured data The Hadoop ecosystem-NoSQL, In-Database Analytics-SQL Essentials-In-Database Text Analysis-Advanced SQL</p>			



COURSE OBJECTIVES

- To understand big data and data analytics lifecycle
- To learn Basic Data analytic methods using R
- To Get a knowledge on advanced analytical methods, technology and tools

COURSE OUTCOMES (CO)

- Understand the big data concepts
- Utilize and apply the Analytical methods, Technology and tools in the industry.
- Understand hadoop ecosystem and apply to solve real-life problems

Course Outcome (CO)	Aligned programme Outcome
Understand the big data concepts	1, 4
Utilize and apply the Analytical methods, Technology and tools in the industry.	3,7
Understand hadoop ecosystem and apply to solve real-life problems	3,4,6

COURSE PLAN – PART II

COURSE OVERVIEW

This course mainly describes the concepts and techniques to understand big data and data analytics lifecycle. The course introduces the advanced analytical methods, technology and tools.

COURSE TEACHING AND LEARNING ACTIVITIES

S.No.	Week	Topic	Mode of Delivery
1.	I Week	State of the practice in Analytics-Key roles for new big data ecosystem Data Analytics Lifecycle-Data analytics life cycle overview	Chalk and Talk, PPT
2.	II Week	Introduction to R -Exploratory Data Analytics-Statistical methods for evaluation Hadoop & Map Reduce framework for R	Chalk and Talk, PPT
3.	III Week	R with Relational Database Management Systems, R with Non-Relational (NoSQL) DBs	Chalk and Talk, PPT
4.	IV Week	Clustering -Overview of Clustering-K-means, Association Rules-Overview-Apriori Algorithm-Evaluation of candidate rules,	Chalk and Talk, PPT
5.	V Week	An Example: Transactions in grocery Store-Validation and Testing-Diagnostics, Regression-Linear Regression	Chalk and Talk, PPT
6.	VI Week	Logistic Regression-Reason to choose and Cautions-Additional Regression Models	Chalk and Talk, PPT



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

7.	VII Week	Decision Trees - Naïve Bayes-Diagnostics of Classifiers-Additional classification methods	Chalk and Talk, PPT
8.	VIII Week	Time series Analysis-Overview of Time series analysis-ARIMA Model-Additional methods - Text Analysis-Text analysis steps	Chalk and Talk, PPT
9.	IX Week	A text analysis Example-Collecting raw Text-Representing Text. Term frequency—Inverse document frequency(TFIDF)-	Chalk and Talk, PPT
10.	X Week	Categorizing documents by Topics-Determining Sentiments-Gaining insights	Chalk and Talk, PPT
11.	XI Week	Analytics for Unstructured data The Hadoop Ecosystem- NoSQL, In-Database Analytics	Chalk and Talk, PPT
12.	XII Week	SQL Essentials-In-Database Text Analysis-Advanced SQL	Chalk and Talk, PPT

Text Book

1. EMC Education Services, “Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data”, Wiley publishers, 2015.
2. Simon Walkowiak, “Big Data Analytics with R” PackT Publishers, 2016

References Books

1. Bart Baesens, “Analytics in a Big Data World: The Essential Guide to Data Science and its Applications”, Wiley Publishers, 2015.
2. Kim H. Pries and Robert Dunnigan, “Big Data Analytics: A Practical Guide for Managers” CRC Press, 2015.

COURSE ASSESSMENT METHODS-THEORY (shall range from 4 to 6)

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1.	Cycle Test- I	As per Dean (Academic) Schedule	1 hour	20%
2.	Cycle Test- II	As per Dean (Academic) Schedule	1 hour	20%
3.	Mini Project/ Assignment	13 th week	Non-contact Hours	10%
CPA	Compensation Assessment*	14 th week	1 hour	20%
4.	Final Assessment*	As per Dean (Academic) Schedule	2 hours	50%
TOTAL				100%

*mandatory

COURSE EXIT SURVEY (mention the ways in which the feedback about the course shall be assessed)



1. Students' feedback through class committee meetings.
2. Feedback questionnaire from students – from MIS at the end of the semester.

COURSE POLICY (preferred mode of correspondence with students, compensation assessment policy to be specified)

MODE OF CORRESPONDENCE (email/ phone etc)

Mode of Correspondence through Phone, Email, MS Teams.

COMPENSATION ASSESSMENT POLICY

If any student is not able to attend Assessment-1 and/or Assessment-2 due to genuine reasons, student is permitted to attend the compensation assessment (CPA) with 25% weightage.

ATTENDANCE POLICY (A uniform attendance policy as specified below shall be followed)

- At least 75% attendance in each course is mandatory.
- A maximum of 10% shall be allowed under On Duty (OD) category.
- Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded 'V' grade.

ACADEMIC DISHONESTY & PLAGIARISM

- Talking to other students, copying from others during an assessment will be treated as punishable dishonesty.
- Zero mark to be awarded for the offenders. For copying from another student, both students get the same penalty of zero mark.
- The departmental disciplinary committee including the course faculty member, PAC chairperson and the HoD, as members shall verify the facts of the malpractice and award the punishment if the student is found guilty. The report shall be submitted to the Academic office.
- The above policy against academic dishonesty shall be applicable for all the programmers.

ADDITIONAL INFORMATION

The students can get their doubts clarified at any time with their faculty member.

FOR APPROVAL

Course Faculty

CC-Chairperson

HOD