



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN – PART I			
Name of the programme and specialization	B.Tech. Computer Science and Engineering		
Course Title	Data Analytics		
Course Code	CSPE64	No. of Credits	3
Course Code of Pre-requisite subject(s)	CSPC52		
Session	January 2023.	Section (if, applicable)	III Year A Section
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)			
UNIT I			
Introduction to Data Analytics - Types of Data Analytics - Predictive Analytics - Simple linear regression - Multiple linear regression - Auto regression - Moving Average - Autoregressive Integrated Moving Average -Data Pre-processing - Data Cleaning - Data Integration and Transformation - Data Reduction - Descriptive data analytics - measures of central tendency - measures of location of dispersions.			
UNIT II			
Association Rule Mining: Efficient and Scalable Frequent Item set Mining Methods - Mining Various Kinds of Association Rules - Association Mining to Correlation Analysis - Constraint Based Association Mining – Cluster Analysis: Types of Data in Cluster Analysis - A Categorization of Major Clustering Methods – Partitioning Methods - Hierarchical methods.			
UNIT III			
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.			
UNIT IV			
Using Graph Analytics for Big Data: Graph Analytics - The Graph Model - Representation as Triples – Graphs and Network Organization - Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs - Features of a Graph Analytics Platform - Considerations: Dedicated Appliances for Graph - Graph QL			
UNIT V			
NoSQL Databases - Schema-less Models - Increasing Flexibility for Data Manipulation - Key Value Stores - Document Stores - Tabular Stores - Object Data Stores - Graph Databases Hive-Sharding- Hbase -			



Analyzing big data with twitter - Big data for E-Commerce - Big data for blogs - Review of Basic Data Analytic Methods using R.

COURSE OBJECTIVES

- To understand the basic principles of Data Analytics
- To learn the various Data Analytic methods
- To understand the various clustering algorithms and its application on data
- To work with stream data model and computing

COURSE OUTCOMES (CO)

Upon completion of this course, the students will be able to:

- Evaluate the use of data from acquisition through cleaning, warehousing, analytics, and visualization to the ultimate business decision
- Mine data and carry out predictive modeling and analytics to support business decision-making
- Suggest prescriptive modeling techniques for real-world problems
- Execute real-time analytical methods on streaming datasets to react quickly to customer needs

Course Outcome (CO)	Aligned programme Outcome
Evaluate the use of data from acquisition through cleaning, warehousing, analytics, and visualization to the ultimate business decision.	1,3,8,11
Mine data and carry out predictive modeling and analytics to support business decision-making.	1,3,6,8,11
Suggest prescriptive modeling techniques for real-world problems.	2,3,6,11
Execute real-time analytical methods on streaming datasets to react quickly to customer needs.	1,3,5,9,12

COURSE PLAN – PART II

COURSE OVERVIEW

This course mainly describes the concepts for knowledge discovery in an information repository using analytical tools and techniques. The course introduces the basic functionalities provided by data analytics for time series analysis, stream computing and social network analysis

COURSE TEACHING AND LEARNING ACTIVITIES

S.No.	Week	Topic	Mode of Delivery
1.	I Week	Introduction to Data Analytics - Types of Data Analytics - Predictive Analytics - Simple linear regression -Multiple linear regression	Chalk and Talk, PPT
2.	II Week	Auto regression - Moving Average - Autoregressive Integrated Moving Averages	Chalk and Talk, PPT



3.	III Week	Data Pre-processing - Data Cleaning - Data Integration and Transformation - Data Reduction - Descriptive data analytics - measures of central tendency - measures of location of dispersions	Chalk and Talk, PPT
4.	IV Week	Association Rule Mining: Efficient and Scalable Frequent Item set Mining Methods - Mining Various Kinds of Association Rules - Association Mining to Correlation Analysis - Constraint Based Association Mining	Chalk and Talk, PPT
5.	V Week	Cluster Analysis: Types of Data in Cluster Analysis - A Categorization of Major Clustering Methods - Partitioning Methods - Hierarchical methods.	Chalk and Talk, PPT
6.	VI Week	Introduction to Streams Concepts - Stream data model and architecture - Stream Computing - Sampling data in a stream - Filtering streams - Counting distinct elements in a stream	Chalk and Talk, PPT
7.	VII Week	Estimating moments - Counting oneness in a window - Decaying window - Real Time Analytics Platform (RTAP) applications -	Chalk and Talk, PPT
8.	VIII Week	case studies - real time sentiment analysis - stock market predictions. Using Graph Analytics for Big Data: Graph Analytics - The Graph Model - Representation as Triples-Graphs and Network Organization	Chalk and Talk, PPT
9.	IX Week	Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs	Chalk and Talk, PPT
10.	X Week	Features of a Graph Analytics Platform - Considerations: Dedicated Appliances for Graph - Graph QL	Chalk and Talk, PPT
11.	XI Week	NoSQL Databases - Schema-less Models - Increasing Flexibility for Data Manipulation - Key Value Stores - Document Stores - Tabular Stores	Chalk and Talk, PPT
12.	XII Week	Object Data Stores - Graph Databases Hive-Sharding-Hbase - Analyzing big data with twitter	Chalk and Talk, PPT
13.	XIII Week	Big data for E-Commerce - Big data for blogs - Review of Basic Data Analytic Methods using R.	Chalk and Talk, PPT

Text Book

1. Jiawei Han, MichelineKamber, Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Elsevier,2011.
2. A. Rajaraman, J. Ullman, "Mining Massive Data Sets", Cambridge University Press, 2012.
3. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, No SQL, and Graph", 2013.



References Books

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, "Probability & Statistics for Engineers & Scientists", Ninth Edition, Prentice Hall Inc.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, "The Elements of Statistical Learning, Data Mining, Inference, and Prediction", Second Edition, Springer, 2014.
3. G James, D. Witten, T Hastie, R. Tibshirani, "An Introduction to Statistical Learning: With Applications in R", Springer, 2013.
4. Mohammed J. Zaki, Wagner Meira, "Data Mining and Analysis", Cambridge, 2012.
5. E. Alpaydin, "Introduction to Machine Learning", MIT Press, 2014.

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

S. No.	Mode of Assessment	Week/Date	Duration	% Weightage
1.	Cycle Test- 1	As per Dean (Academic) Schedule	1 hour	20%
2.	Cycle Test- 11	As per Dean (Academic) Schedule	1 hour	20%
3.	Assignment	13 th week	Non-contact Hours	10%
4.	CPA (Compensation Assessment*)	14 th week	1 hour	20%
5.	Final Assessment*	As per Dean (Academic) Schedule	3 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.



NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

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

S. S. Sathya

CC-Chairperson

R. Mohan

HOD

M. S. Sathya