

# 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 Engg		
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)	B
Name of Faculty	Dr. S. Jaya Nirmala	Department	CSE
Official Email	<a href="mailto:sjaya@nitt.edu">sjaya@nitt.edu</a>	Telephone No.	0431-2503200
Name of Course Coordinator(s) (if, applicable)	---		
Official E-mail	--	Telephone No.	
Course Type (please tick appropriately)	<input type="checkbox"/> Core course	<input type="checkbox"/> Elective course	

### Syllabus (approved in BoS)

#### 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

**MAPPING OF COs with POs**

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Evaluate the use of data from acquisition through cleaning, warehousing, analytics, and visualization to the ultimate business decision	1,3,8,11
2. Mine data and carry out predictive modeling and analytics to support business decision-making	1,3,6,8,11
3. Suggest prescriptive modeling techniques for real-world problems	2,3,6,11
4. 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**

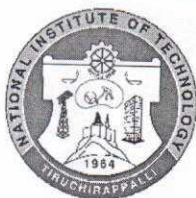
The course first introduces the different types of Data Analytics followed by the various stages in it. The various association rule mining and clustering techniques are discussed followed by stream analytics. The course also discusses about Graph Analytics and graph Databases and concludes with the use of R for Data Mining.

**COURSE TEACHING AND LEARNING ACTIVITIES**

S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	1/3	Introduction to Data Analytics - Types of Data Analytics – Predictive Analytics - Simple linear regression - Multiple linear regression - Examples	Chalk and Talk
2	2/3	Auto regression - Moving Average - Autoregressive Integrated Moving Average - Data Pre- processing - Data Cleaning - Data Integration and Transformation	Chalk and Talk



3	3/3	Data Reduction - Descriptive data analytics - measures of central tendency - measures of location of dispersions.	Chalk and Talk
4	4/3	UNIT-2: Association Rule Mining: Efficient and Scalable Frequent Item set Mining Methods - Mining Various Kinds of Association Rules	PPT, Chalk and Talk
5	5/3	Association Mining to Correlation Analysis - Constraint Based Association Mining - Cluster Analysis: Types of Data in Cluster Analysis	PPT, Chalk and Talk
6	6/1	Tutorial	PPT, Chalk and Talk
7	7/3	UNIT-3: Introduction to Streams Concepts - Stream data model and architecture - Stream Computing -	PPT, Chalk and Talk
8	8/3	Sampling data in a stream - Filtering streams - Counting distinct elements in a stream - Estimating moments- Counting oneness in a window - Decaying window	PPT, Chalk and Talk
9	9/3	Real Time Analytics Platform (RTAP) applications - case studies - real time sentiment analysis - stock market prediction- Tutorials	PPT, Chalk and Talk
10	10/3	UNIT- 4: Using Graph Analytics for Big Data: Graph Analytics - The Graph Model - Representation as Triples - Graphs and Network Organization -	PPT, Chalk and Talk
11	11/3	Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs	PPT, Chalk and Talk
12	12/2	Features of a Graph Analytics Platform - Considerations: Dedicated Appliances for Graph - Graph QL	PPT, Chalk and Talk
13	13/2	Tutorial and UNIT 5: NoSQL Databases - Schema-less Models - Increasing Flexibility for Data Manipulation - Key Value Stores - Document Stores	PPT, Chalk and Talk



14	14/1	Tabular Stores - Object Data Stores	PPT, Chalk and Talk
15	15/3	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	PPT, Chalk and Talk

**COURSE ASSESSMENT METHODS** (shall range from 4 to 6)

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test - 1	As per academic schedule	1 hour	15
2	Cycle Test - 2	As per academic schedule	1 hour	15
3	Assignment I and II	4 <sup>th</sup> and 10 <sup>th</sup> week	--	20
CPA	Compensation Assessment*	As per academic schedule	1 hour	15
4	Final Assessment *	As per academic schedule	3 hours	50

**COURSE EXIT SURVEY**

1. Feedbacks are collected after the end semester exam in the feedback forms.
2. Suggestions from the students are incorporated for making the course more interesting.
3. Students may give their feedback at any time, through their class representatives to the concerned faculty and also in the class committee meetings.

**COURSE POLICY**

**MODE OF CORRESPONDENCE (email/ phone etc)**

- Through email

**COMPENSATION ASSESSMENT**

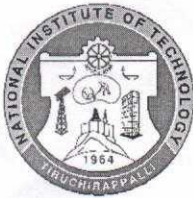
Compensation assessment (Retest) will be conducted for absentees in the cycle tests only after the submission of medical or On-Duty certificates signed by competent authority.

**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**

- Possessing a mobile phone, carrying bits of paper, talking to other students, copying



## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

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 programmes.

### ADDITIONAL INFORMATION, IF ANY

The students can clarify their doubts at any time during working hours from the faculty with prior appointment.

### FOR APPROVAL

Course Faculty *S. Jaya Nirmala*  
(Dr. S. Jaya Nirmala)

CC- Chairperson *R. Mohan*  
(Dr. R. Mohan)

HOD *S. Mary Saira Bhanu*  
(Dr. S. Mary Saira Bhanu)