ISO 9001:2015 Certified Institution

Balaji Institute of Technology & Science
Laknepally (V), Narsampet (M), Warangal District - 506 331, Telangana State, India
(AUTONOMOUS)
Accredited by NBA (UG - CE, ME, ECE & CSE) **& NAAC A+ Grade**
(Affiliated to JNT University, Hyderabad and Approved by AICTE, New Delhi)
www.bitswgl.ac.in, email: principal@bitswgl.ac.in, Ph:98660 50044, Fax: 08718-230521

Estd.:2001

Course File Contents

<h1 style="text-align:center">I.COVER PAGE</h1>

<div style="text-align:center">

BALAJI INSTITUTE OF TECHNOLOGY & SCIENCE
(AUTONOMOUS)

Department of Artificial Intelligence & Machine Learning

</div>

Name of the Subject : **Data Analytics**

BITS CODE   : **22CS650PC:**   Programme : **UG**

Branch : AI & ML	Version No :
Year : **III**	Document Number : **BITS/AI&ML/**
Semester : **II**	Number of Pages :

Classification status (Unrestricted/Restricted) : Unrestricted

Distribution List: Dept. Library, Dept. Office, Concerned Faculty

Prepared by :	Updated by :
1. Name : 2. Sign : 3. Design: 4. 4) Date:	1. Name : 2. Sign : 3. Design : 4. Date :

Verified by :*For Q.C only

1. Name :	**1.** Name : :
2. Sign :	**2.** Sign
3. Design :	**3.** Design :
4. Date :	**4.** Date :

Approved by (HOD) :

1.   Name :

2.   Sign   :

3.   Date :

CSE (Artificial Intelligence & Machine Learning)

II. VISION AND MISSION OF THE DEPARTMENT

VISION

To be a global leader in Artificial Intelligence and Machine Learning research, innovation, and education, driving transformative advancements that empower industries, enhance human capabilities, and contribute to a smarter, more sustainable world.

MISSION

M1: Innovative Research & Quality Education – To Conduct research on cutting-edge Technologies to address complex real-world problems across diverse domains and provide world-class education and training to equip students with technical expertise, ethical responsibility, and problem-solving skills.

M2: Industry Collaboration & Ethical AI Development –To Foster strong partnerships with industries, academia, and government organizations to develop impactful AI solutions and promote responsible and ethical AI practices that align with societal values and global sustainability.

M3: Entrepreneurship & Innovation – Encourage entrepreneurship and the development of AI-driven start-ups and products that contribute to economic growth.

M4: Community Engagement – Engage with communities to spread AI awareness, inclusivity, and accessibility for societal benefit.

III. PEOs, POs and PSOs
Program Educational Objectives

PEO1: Graduates shall apply the analytical, decision making and prediction skills in AI & ML to formulate and solve complex intelligent computing and multidisciplinary problems.

PEO2: Graduates will be able to take up higher studies, research & development by acquiring in-depth knowledge in Artificial Intelligence & Machine Learning.

PEO3: Graduates will be able to exhibit their employability skills and practice the ethics of their profession with a sense of social responsibility.

Programs Outcomes

PO1: graduate of the Artificial Intelligence & Machine Learning Program will demonstrate:

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO3: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO4: 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.

PO5: 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.

PO6: 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.

PO7: 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.

PO8: 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.

Program Specific Outcomes (PSOs)

PSO1: Apply a set of Artificial Intelligence principles, tools, and techniques to model various real-world business problems, analyze them, and suggest a suitable solution by communicating relevant findings and effectively presenting results using appropriate techniques.

PSO2: Apply the skills of Artificial Intelligence and Machine Learning in the areas of Health Care, Education, Agriculture, e-commerce, financial sector, Smart Systems, and Multi-disciplinary areas of AI.

PSO3: Cultivate the ability to work in teams and learn by participating in Technical Events and Social Welfare Programs and develop the attitude for working productively as an individual and in cross- disciplinary teams to become better citizens in multicultural world.

IV-Syllabus copy and Academic calendar

BALAJI INSTITUTE OF TECHNOLOGY & SCIENCE
(AUTONOMOUS)
22CS650PC: Data Analytics

B.Tech. III Year II Semester

L T P C
3 0 0 3

Prerequisites:

1. A course on "Database Management Systems".
2. Knowledge of probability and statistics.

Course Objectives:

- To explore the fundamental concepts of data analytics.
- To learn the principles and methods of statistical analysis
- Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.
- To understand the various search methods and visualization techniques.

Course Outcomes: After completion of this course students will be able to

- Understand the impact of data analytics for business decisions and strategy
- Carry out data analysis/statistical analysis
- To carry out standard data visualization and formal inference procedures
- Design Data Architecture
- Understand various Data Sources

UNIT - I

Data Management: Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/Signals/GPS etc. Data Management, Data Quality (noise, outliers, missing values, duplicate data) and Data Processing & Processing.

UNIT - II

Data Analytics: Introduction to Analytics, Introduction to Tools and Environment, Application of Modeling in Business, Databases & Types of Data and Variables, Data Modeling Techniques, Missing Imputations etc.
Need for Business Modeling.

UNIT - III

Regression – Concepts, Blue property assumptions, Least Square Estimation, Variable Rationalization, and Model Building etc.
Logistic Regression: Model Theory, Model fit Statistics, Model Construction, Analytics applications to various Business Domains etc.

UNIT - IV

Object Segmentation: Regression Vs Segmentation – Supervised and Unsupervised Learning, Tree Building – Regression, Classification, Overfitting, Pruning and Complexity, Multiple Decision Trees etc.
Time Series Methods: Arima, Measures of Forecast Accuracy, STL approach, Extract features from generated model as Height, Average Energy etc and Analyze for prediction

UNIT - V

Data Visualization: Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.

TEXT BOOKS:

1. Student's Handbook for Associate Analytics – II, III.
2. Data Mining Concepts and Techniques, Han, Kamber, 3rd Edition, Morgan Kaufmann Publishers.

REFERENCE BOOKS:

1. Introduction to Data Mining, Tan, Steinbach and Kumar, Addision Wisley, 2006.
2. Data Mining Analysis and Concepts, M. Zaki and W. Meira
3. Mining of Massive Datasets, Jure Leskovec Stanford Univ. Anand Rajaraman Milliway Labs Jeffrey D Ullman Stanford Univ.

Balaji Institute of Technology & Science
ISO 9001:2015 Certified Institution   Estd. 2001
Laknepally (V), Narsampet (M), Warangal District - 506 331, Telangana State, India
(AUTONOMOUS)
Accredited by NBA (UG - CE, EEE, ME, ECE & CSE) & NAAC A+ Grade
(Affiliated to JNT University, Hyderabad and Approved by AICTE, New Delhi)
www.bitswgl.ac.in, email: principal@bitswgl.ac.in, Ph:98660 50044, Fax: 08718-230521

ACADEMIC CALENDAR FOR B.TECH. III-YEAR FOR THE ACADEMIC YEAR 2024-25

B.Tech- III-YEAR I Semester

S.No	Description	Date		Duration
		From	To	
1	1st Spell of instructions	30-07-2024	25-09-2024	9 Weeks
2	First Mid Term Examinations	26-09-2024	28-09-2024	3 days
3	2nd Spell of Instructions	30-09-2024	05-10-2024	1 week
4	Dussehra Recess	07-10-2024	12-10-2024	1 week
5	2nd Spell of Instructions Continuation	14-10-2024	30-11-2024	7 Weeks
6	Second Mid Term Examinations	02-12-2024	04-12-2024	3 days
7	Preparation Holidays & Practical Examinations	05-12-2024	14-12-2024	9 days
8	End Semester Examinations	16-12-2024	28-12-2024	2 Weeks

B.Tech - III-YEAR II Semester

| S.No	Description	Date		Duration
		From	To	
1	Commencement of II Semester class work	30-12-2024		
2	1st Spell of Instructions	30-12-2024	26-02-2025	9 Weeks
3	First Mid Term Examinations	27-02-2025	01-03-2025	3 days
4	2nd Spell of instructions	03-03-2025	26-04-2025	8 Weeks
5	Second Mid Term Examinations	28-04-2025	30-04-2025	3 days
6	Preparation Holidays and Practical Examination	01-05-2025	10-05-2025	9 days
7	Summer Vacation	12-05-2025	24-05-2025	2 Weeks
8	End Semester Examinations	26-05-2025	07-06-2025	2 Weeks

24/7/24

PRINCIPAL

Principal
Balaji Institute of Tech & Science
LAKNEPALLY Narsampet-506 331

Copy to:

1. Dean-Academics
2. All Head of the Departments
3. Examination branch

8

V. BRIEF NOTES ON THE IMPORTANCE OF THE COURSE

Data analytics is an important field that involves the process of collecting, processing, and interpreting data to uncover insights and help in making decisions. Data analytics is the practice of examining raw data to identify trends, draw conclusions, and extract meaningful information. This involves various techniques and tools to process and transform data into valuable insights that can be used for decision-making.
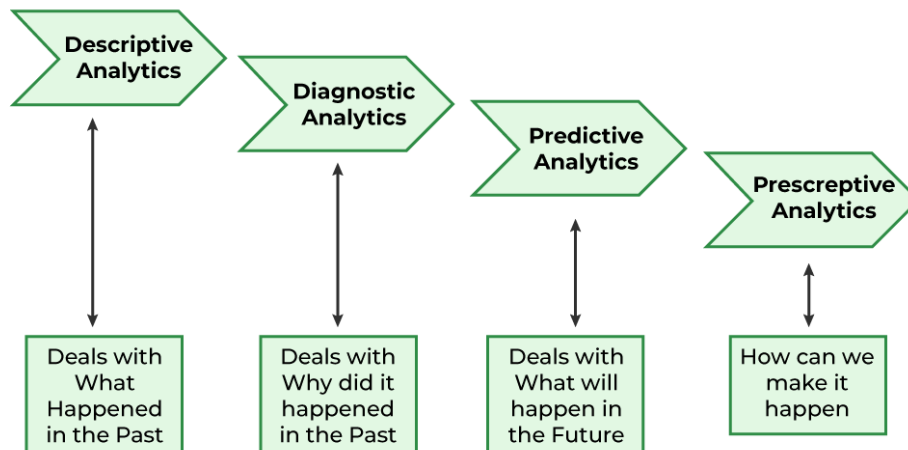
Predictive Analytics

- Time Series Analysis and Forecasting
- Data Mining

Basic Cornerstones of Predictive Analytics

- Predictive modeling
- Decision Analysis and optimization

Future Scope of Data Analytics

- **Retail** : To study sales patterns, consumer behavior, and inventory management, data analytics can be applied in the retail sector. Data analytics can be used by retailers to make data-driven decisions regarding what products to stock, how to price them, and how to best organize their stores.

VI. PREREQUISITES

- **Database Management :** database management is a critical prerequisite for data analytics because it provides the foundation for storing, organizing, and retrieving data.
- **Probability and Statistics :** Probability and Statistics are essential prerequisites for Data Analytics. The backbone of many analytical techniques used in various fields, including machine learning, data science, and business intelligence. Probability and Statistics provide a framework for analyzing and interpreting data, making it possible to draw insights and make informed decisions based on data.

- **Data pre-processing** is an essential step in the data analytics process, where raw data is transformed into a clean, organized, and meaningful dataset that can be analyzed. This step involves several techniques and procedures to transform, clean, and prepare the data for analysis.

- **Data cleaning** involves the identification and removal of errors, inconsistencies, and duplicates from the dataset. This process helps to ensure that the dataset is accurate and complete. Data integration involves the merging of data from multiple sources into a single dataset. This process helps to ensure that the dataset is comprehensive and covers all relevant data.

- **Data transformation** involves the conversion of data into a format that is suitable for analysis. This process includes tasks such as normalization, standardization, and feature scaling. Data reduction involves the reduction of the dataset to a smaller size, while retaining the most important and relevant information.

- Exploratory Data Analysis : Exploratory data analytics is a critical step in the data analytics process, allowing data analysts to gain an initial understanding of the data and identify patterns and trends that may be of interest. Here are some prerequisites for exploratory data analytics

- **Hypothesis testing:** Hypothesis testing involves testing a specific hypothesis or assumption about the data. This may involve performing t-tests or ANOVA tests to compare groups or testing the significance of a correlation coefficient.

- **Data interpretation:** Once the data has been analyzed, data analysts must interpret the results and identify any key findings. This may involve creating a summary report or presentation to communicate the findings to stakeholders.

VII. COURSE OBJECTIVES & OUTCOMES

COURSE OBJECTIVES:

- To explore the fundamental concepts of data analytics.
- To learn the principles and methods of statistical analysis
- Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.
- To understand the various search methods and visualization techniques.

COURSE OUTCOMES (COs):

- Understand the impact of data analytics for business decisions and strategy .
- Carry out data analysis/statistical analysis.
- To carry out standard data visualization and formal inference procedures • Design Data Architecture.
- Understand various Data Sources.

By the end of this course, Students should be able to:

CO Number	Statement
CO1	Understand and analyze various Data types, Concepts and their applications related to Big data
CO2	Apply statistical methods and exploratory data analysis (EDA) to summarize and visualize data effectively.
CO3	Implement machine learning algorithms for classification, clustering, and regression tasks.
CO4	Utilize big data frameworks and tools such as Hadoop and Spark for large-scale data processing.
CO5	Interpret results from data analytics models to make informed business decisions and real-world applications.

VIII. CO-PO, CO-PSO MAPPING & JUSTIFICATION

Name of the Subject: : C311 Data Analytics- 22CS650PC										Year of study: 2022-2023		
CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
C311.1	3	2	2	-	-	-	-	-	-	-	-	-
C311.2	3	3	2	2	-	-	-	-	-	-	-	-
C311.3	2	2	2	2	-	-	-	-	-	-	-	-
C311.4	3	3	3	3	3	-	-	-	-	-	-	-
C311.5	2	2	2	2	2	-	-	-	-	-	-	-
Average	2.6	2.4	2.2	2.25	2.5	-	-	-	-	-	-	-

Course Name: C311 (Data Analytics-22CS650PC) Year of Study: 2024-2025		
CO	PSO1	PSO2
C311.1	3	1
C311.2		2
C311.3	**2**	
C311.4	**3**	
C311.5		3
Average	**2.6**	**2**

JUSTIFICATION FOR COURSE OUTCOMES MAPPING WITH POs AND PSOs

❖ **COURSE OUTCOME 1**
- **PO1 (Engineering Knowledge):** Strong correlation as students apply fundamentals of physics, mathematics, and electrostatics.
- **PO2 (Problem Analysis):** Strong correlation, as students analyze charge distributions and field behavior.
- **PO3 (Design & Development of Solutions):** Strong correlation as students evaluating field distribution calculations.
- **PO4 (Investigations of Complex Problems):** Moderate correlation as students investigating magnetostatics properties using analytical tools.

❖ **COURSE OUTCOME 2**
- **PO1 (Engineering Knowledge):** Strong correlation as students apply fundamentals of physics, mathematics and magnetostatics
- **PO2 (Problem Analysis):** Strong correlation as students analyze current distributions and field behavior in magnetostatics.
- **PO3 (Design & Development of Solutions):** Moderate correlation for applying magnetostatics for magnetic field distribution calculations.
- **PO4 (Investigations of Complex Problems):** Moderate correlation for investigating electromagnetic properties using analytical tools.

❖ **COURSE OUTCOME 3**
- **PO1 (Engineering Knowledge):** Strong correlation as students apply fundamentals of physics, mathematics and electromagnetics
- **PO2 (Problem Analysis):** Strong correlation as students analyze time varying charge distributions and time varying current distributions field behavior.
- **PO3 (Design & Development of Solutions):** Strong correlation, as students calculate electric and magnetic field for time varying charge and time varying current distributions
- **PO4 (Investigations of Complex Problems):** Strong correlation as students investigates boundary conditions in different media.

❖ **COURSE OUTCOME 4**
- **PO1 (Engineering Knowledge):** Strong correlations as students apply fundamentals of wave propagation in different media and related concepts like reflection and refraction.
- **PO2 (Problem Analysis):** Strong correlations as students Identifies and evaluates wave behavior in conducting and dielectric media.
- **PO3 (Design & Development of Solutions):** Strong correlations as students Solve practical engineering problems involving electromagnetic wave propagation.
- **PO4 (Investigations of Complex Problems):** Moderate correlations as students investigate wave propagation parameters for different media.

❖ **COURSE OUTCOME 5**

- **PO1 (Engineering Knowledge):** Strong correlations as students apply fundamentals like line equations, impedance matching, smith chart etc.
- **PO2 (Problem Analysis):** Strong correlation as students analyzes transmission line parameters, input impedance, VSWR, and reflection coefficient in transmission lines
- **PO3 (Design & Development of Solutions):** Strong correlation as student applies advanced mathematical models to evaluate transmission line performance and designs circuits involving transmission lines using the Smith chart
- **PO4 (Investigations of Complex Problems):** Strong correlation as student investigates transmission line properties using theoretical approaches

.

IX. CLASS TIME TABLE & INDIVIDUAL TIME TABLE

Day	P1	P2	P3	P4	P5	P6	P7
MON	**DA**						
TUE			**DA**	**DA-LAB**			
WED							**DA**
THU							**DA**
FRI		**DA**					
SAT							

X. METHOD OF TEACHING

Teaching methods for a course on Data analytics' should be diverse and interactive to cater to the complexity and depth of the subject matter. Here are some effective methods:

1. LECTURES

TRADITIONAL LECTURES: Using clear explanations, real-world examples, and visual aids like slides and diagrams for explaining key concepts such as Maxwell's equations, wave propagation, and transmission line theory.

GUEST LECTURES: Invited industry experts/researchers for providing insights into current trends and applications of electromagnetic fields and transmission lines.

2. INTERACTIVE LEARNING

PROBLEM-SOLVING SESSIONS: Conducting sessions where students solve problems in real-time, encouraging participation and collaborative learning.

Q&A SESSIONS: Regularly holding sessions where students can ask questions and engage in discussions to clarify doubts and deepen their understanding.

3. LABORATORY EXERCISES

Hands-On EXPERIMENTS: Setting up laboratory experiments where students can observe and measure electromagnetic phenomena, such as wave propagation and impedance matching.

4. PROJECTS AND CASE STUDIES

DESIGN PROJECTS: Assign projects where students design components such as antennas, transmission lines, or RF circuits, applying theoretical knowledge to practical problems.

CASE STUDIES: Analyzing real-world case studies of electromagnetic field applications in various industries, encouraging students to think critically about practical challenges and solutions.

5. FLIPPED CLASSROOM

PRE-CLASS ASSIGNMENTS: Providing reading materials, videos, and online resources for students to study before class. This prepares them for more in-depth discussions and activities during class time.

INTERACTIVE CLASS ACTIVITIES: Using class time for interactive activities such as group discussions, problem-solving, and hands-on experiments, reinforcing the pre-class material.

6. ASSESSMENT AND FEEDBACK

QUIZZES AND TESTS: Regular quizzes and tests to assess understanding and providing feedback on areas needing improvement.

PEER REVIEW: Incorporating peer review sessions for project presentations and reports, fostering collaborative learning and constructive criticism.

7. VISUAL AND MULTIMEDIA AIDS

VIDEOS AND ANIMATIONS: Using videos and animations to illustrate complex electromagnetic phenomena, making abstract concepts more tangible.

8. COLLABORATIVE LEARNING

GROUP PROJECTS: Encouraging teamwork through group projects where students can collaborate on designing and testing electromagnetic systems.

STUDY GROUPS: Forming study groups to promote peer-to-peer learning and discussion outside of formal class hours.

9. SUPPLEMENTARY RESOURCES

ONLINE FORUMS: Creating online forums or discussion boards for students to ask questions, share resources, and discuss course material.

READING MATERIALS: Provide a list of recommended textbooks, research papers, and articles for further reading and exploration of advanced topics.

S.No.	Question	Response (No of students) (%)	No of students (%) Response: b
1.	Preferred the conventional lecture by "Talk and chalk".	(95%)	20 (5%)
2.	Diagrams should be shown by drawing on board.	(85%)	60 (15%)
3.	Concepts become clearer by "Talk and Chalk".	(90%)	40 (10%)
4.	Teachers take more time to explain the concept rather than changing the slides fast.	(92%)	32 (8%)
5.	Easier to take down notes when taught by "Talk and chalk" method because power point slides are changed very fast.	(90%)	40 (10%)
6.	Diagrams are easier to follow when drawn on board step by step.	(70%)	120 (30%)
7.	They connect better with the teacher during Talk and chalk lecture.	(80%)	80 (20%)

XI: Lecture Schedule

Department of AI & ML

Lesson Plan & Delivery Report

Subject: *Data Analytics (*22CS650PC)   **Class: B.Tech III year - II SEM, CSM**

Faculty Name: **Regulation: R22**

Academic Year: 2024-25   **Commencement of Class Work: 30-12-24**

Topics (as per syllabus)	Sub Topics	Lect. No.	Scheduled Date	Topic Delivered Date
UNIT –I Data Management & Data Processing & Processing.
Data Analytics - Introduction	• Data Management:	L1	30-12-2024	
	• Design Data Architecture	L2	31-12-2024	
	• Manage the data for analysis,	L3	03-01-2025	
	• Understand various sources of Data	L4	04-01-2025	
	• Sensors	L5	06-01-2025	
	• Signals	L6	07-01-2025	
Data Management	• GPS	L7	10-01-2025	
	• Data Management,	L8	11-01-2025	
	• Data cleaning	L9	17-01-2025	
	• Data Integration	L10	18-01-2025	
	• Data Quality	L11	20-01-2025	

	• Noise, Data Processing & Processing	L12	21-01-2025	
	• Outliers, missing values, duplicate data	L13	21-01-2025	
	Overview of unit-1	L14	24-01-2025	
	UNIT -1 SLIP TEST		20-01-2025	

UNIT-II Data Analytics: (No. of Lectures -13)

Topics (as per syllabus	Sub Topics	Lect . No.	Scheduled Date	Topic Delivered Date
Introduction to Analytics with Tools	• Introduction to Analytics	L15	25-01-2025	
	• Applications	*L16*	27-01-2025	
	• Tools	L17	28-01-2025	
	• Spark	L18	31-01-2025	
	• Apache	L19	01-02-2025	
	• Excel	L20	03-02-2025	
	• MS Acess	L21	07-2-2025	
	• Application of Modeling in Business	L22	8-02-2025	
Databases & Types of Data and Variables	• Data Modeling Techniques in Business, Databases	L23	10-02-2025	
		L24	11-02-2025	
	• Types of Data and Variables, Data Modeling Techniques, Need for Business Modeling.	L25	12-02-2025	
		L26	14-02-2025	
	• Overview of unit 2	L27	15-02-2025	
	UNIT -2 SLIP TEST		18-02-2028	

UNIT-III Regression & Logistic Regression: (No. of Lectures -14)

Topics (as per syllabus)	Sub Topics	Lect . No.	Scheduled Date	Topic Delivered Date

Regression	• Concepts, Regression	L28	21-02-2025	
	• Regression Examples	L29	22-02-2025	
	• Implementations	L30		
	• Blue property assumptions,	L31	24-02-2025	
		L32	24-02-2025	
	• Least Square Estimation	L33	24-02-2025	
	• Variable Rationalization,	L34	24-02-2025	
Logistic Regression	• Model Building etc.	L35	03-03-2025	
	• Model Theory	L36	04-03-2025	
	• Model fit Statistics,	L37	07-03-2025	
	• Model Construction • Analytics applications to various Business Domains etc.	L38	08-03-2025	
	• Overview of unit-3	L39	15-03-2025	
Review about Mid I Exam	• Review of theory Questions • Review of Objective questions • Plan for mid I exam • Tips to get good marks	L40	17-03-2025	
Mid I Schedule: 04-09-2024 to 06-09-2024 days)		**Mid I Exam (OS)**		
Mid I Marks Distribution	• Marks Distribution • Discussion about Paper • Counsel the students (AB/got poor marks)	L41	18-03-2025	
	UNIT -3 SLIP TEST		21-03-2025	

UNIT-IV Object Segmentation: Regression Vs Segmentation (No. of Lectures -13)				
Topics (as per syllabus)	**Sub Topics**	**Lect . No.**	**Scheduled Date**	**Topic Delivered Date**
Object Segmentation: Regression Vs Segmentation	• Supervised and	L42	22-03-2025	
	• Unsupervised Learning	L43	22-03-2025	
		L44	24-03-2025	
	• Tree Building – Regression,	L45	24-03-2025	
	• Classification,	L46	25-03-2025	
	• Overfitting,	L47	29-03-2025	
	• Pruning and Complexity,	L48 L49	29-03-2025	
	• Multiple Decision Trees etc	L50	29-03-2025	
Time Series Methods	• Multiple Decision Trees etc	L51	29-03-2025	
	• Multiple Decision Trees etc	L52	29-03-2025	
	• Multiple Decision Trees etc	L53	29-03-2025	
	• Multiple Decision Trees etc	L54	29-03-2025	
UNIT -4 SLIP TEST			01-04-2025	
UNIT-V Data Visualization: (No. of Lectures -9)				
Topics (as per syllabus)	**Sub Topics**	**Lect . No.**	**Scheduled Date**	**Topic Delivered Date**

Data Visualization:	• Pixel-Oriented Visualization Techniques,	L55	11-04-2025	
	• Geometric Projection Visualization Techniques	L56	12-04-2025	
	• Geometric Projection Visualization Techniques	L57	17-04-2025	
	• Icon-Based Visualization Techniques	L58	19-04-2025	
	• Icon-Based Visualization Techniques	L59	21-04-2025	
	• Hierarchical Visualization Techniques • Visualizing Complex Data and • Relations	L60	24-04-2025	
	• overview of unit 5	L61	25-04-2025	
Review about Mid II Exam	• Review of theory Questions • Review of Objective questions • Plan for mid I exam • Tips to get good marks	L62	26-04-2025	
Mid II Schedule:				
Mid II Marks Distribution	• Marks Distribution • Discussion about Paper • Counsel the students (AB/got poor marks)	L63	28-04-2025	

XIII. ADDITIONAL TOPICS

- **Data Manipulation and Data Analysis**

  Have you ever given thought to why data manipulation is one of the crucial prerequisites to becoming data analytics?

  Well, it is important for several reasons and hence when you see the job description you must see that data analysis and data manipulation are the core skillset. Data manipulation is important for data preparation, ensuring the quality of data, and data integration (to integrate and merge data from different resources and create a dataset for analysis), feature engineering, exploring in-depth data (data exploration), testing hypotheses, and custom analysis. It is the foundation of data analysis.

- **Learn Data Visualization**
  Another important prerequisite to becoming a data analyst is learning data visualization. If you want to excel in your career as a data analyst, you must know how to present data clearly and concisely. It is important that your data is visually appealing and easily communicable in front of stakeholders. So, one should be able to create compelling charts, graphs, and dashboards to enhance the impact of your analysis. Data Analysts should be skilled in visualization tools like Tableau, Power BI, Matplotlib, and Seaborn in Python.

- Predictive Analytics and Machine Learning

  As data analysis evolves, knowledge and understanding of machine learning and predictive analysis have become one of the main aspects of prerequisites to becoming a data analyst. It is important to understand the basic of machine learning algorithms, namely linear tools for data analysts, decision trees, and clustering. This will enable you to build predictive models and identify patterns in data.

- **Manufacturing** : Data analytics can be used to examine production data, spot trends in production methods, and boost production efficiency in the manufacturing sector. Data analytics can be used by manufacturers to cut costs and enhance product quality.
- **Transportation** : To evaluate logistics data, spot trends in transportation routes, and improve transportation routes, the transportation sector can employ data analytics. Data analytics can help transportation businesses cut expenses and speed up delivery times.
- All processing happens on a dedicated central server that hosts all the data.
- Distributed processing
- Data is distributed and stored on different servers.
- Batch processing
- Pieces of data accumulate over time and are processed in batches.
- Real-time processing
- Data is processed continually, with computational tasks finishing in seconds.

Answer any four questions.   4X5=20M

1. Consider the time series data given below: Use the **least square method** to determine the equation of line of best fit for the data. Then plot the line.

x_i	8	3	2	10	11	3	6	5	6	8
y_i	4	12	1	12	9	4	9	6	1	14

2. What are the Data Modelling Techniques in Data Analytics?

3. a) Find the Probability of Pass for the Student who studied for 33 Hours for the given Table

   b) At least How many hours student should study that makes he will pass the course with Probability of more than 95% , **Given log(odds) =-64+2*hours**

Hour Study	Pass(1) & Fail (0)
29	0
15	0
33	1
28	1
39	1

4..What is Data Preprocessing &Data Cleaning Various Techniques with Example?

5.What is the difference between SQL and NoSQL?

6.What are the Tools &Environment in Data Analytics?

MCQ

1. Data Analytics uses ___ to get insights from data.

   A. Statistical figures
   B. Numerical aspects
   C. Statistical methods
   D. None of the mentioned above

2.With reference to data, dependent and independent variables should be quantitative.

   A. True
   B. False

3.Alternative Hypothesis is also called as?

   A. Null Hypothesis
   B. Research Hypothesis
   C. Simple Hypothesis
   D. None of the mentioned above

4.Customer analytics refers -

   A. Customer Relationship Management: churn analysis and prevention
   B. Marketing: cross-sell, up-sell
   C. Pricing: leakage monitoring, promotional effects tracking, competitive price responses
   D. All of the mentioned above

5 __ are the basic building blocks of qualitative data.

   A. Categories
   B. Data chunk
   C. Numeric figures
   D. None of the mentioned above

6.Customer behavior analytics is about understanding how your customers act -

   A. True
   B. False

7..Amongst which of the following is / are the types of Linear Regression,

    A. Simple Linear Regression
    B. Multiple Linear Regression
    C. Both A and B
    D. None of the mentioned above

8.Data Analysis is a process of,

    A. Inspecting data
    B. Data Cleaning
    C. Transforming of data
    D. All of the mentioned above

9.Customer analytics refers -

    A. Customer Relationship Management: churn analysis and prevention
    B. Marketing: cross-sell, up-sell
    C. Pricing: leakage monitoring, promotional effects tracking, competitive price responses
    D. All of the mentioned above

10.Customer behavior analytics is about understanding how your customers act -

    A. True
    B. False

Fill in the Blanks

1. Tableau is a ___ tool.
2. …………clustering technique needs the merging approach
3. The self-organizing maps can also be considered as the instance of _____ type of learning.
4. Clustering belongs to ___ data analysis.
5. What is the primary goal of data analysis…………..
6. What is the difference between mean and median………………..
7. .Expand **MAR………….**

8.For each value of the ____, the distribution of the dependent variable must be normal.

9. Alternative Hypothesis is also called as………..

10. Velocity is the speed at which the data is processed -…………..

XV University Question papers of previous years

D

Code No: 117JU

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, November/December – 2017
BIG DATA ANALYTICS
(Common to CSE, IT)

Time: 3 Hours   Max. Marks: 75

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART- A
(25 Marks)

1. a) Is it desirable to build a system that helps users perform appropriate analysis? Explain. [2]
   b) Whether use of a centralized server for the MDW database is the best Practice and Caveat for Management Data Warehouse? Justify. [3]
   c) What is Apache Spark? Explain key features of Spark. [2]
   d) Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief. [3]
   e) Explain some of the applications of Big data. [2]
   f) What is impedance mismatch? What are the major difficulties faced by big data application developers? [3]
   g) What do you mean by hypothesis? Explain. [2]
   h) Discuss the criteria for evaluating case study. [3]
   i) What do you mean by data visualization? Explain. [2]
   j) Write names of some data visualization tools. Also, discuss properties of different tools. [3]

PART-B
(50 Marks)

2. What role should data quality and governance play in any organization? Also how are data quality and governance related? Explain. [10]

OR

3. What software requirements does user end data analysis impose upon a data management project? Explain with a suitable example. [10]

4. Write Map Reduce code for counting occurrences of specific words in the input text file(s). Also write the commands to compile and run the code. [10]

OR

5. What is Hbase? Discuss in detail the data model and Implementation aspect of Hbase. [10]

6. What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities. [10]

OR

7. What is RDD? Explain about transformations and actions in the context of RDDs. State and explain RDD operations in brief. [10]

XVI   Unit-wise quiz questions

UNIT-1

MCQ

2. Data Analytics uses ___ to get insights from data.

   E. Statistical figures
   F. Numerical aspects
   G. Statistical methods
   H. None of the mentioned above

2.Amongst which of the following is / are the branch of statistics which deals with the development of statistical methods is classified as ___.

   A. Industry statistics
   B. Economic statistics
   C. Applied statistics
   D. None of the mentioned above

3.Data Analysis is a process of,

   A. Inspecting data
   B. Data Cleaning
   C. Transforming of data
   D. All of the mentioned above

4.What is a hypothesis?

   A. A statement that the researcher wants to test through the data collected in a study
   B. A research question the results will answer
   C. A theory that underpins the study
   D. A statistical method for calculating the extent to which the results could have happened by chance

5.With reference to data, dependent and independent variables should be quantitative.

   C. True
   D. False

6.For each value of the ___, the distribution of the dependent variable must be normal.

   A. Independent variable

B. Depended variable
C. Intermediate variable
D. None of the mentioned above

7. Amongst which of the following is / are not a major data analysis approach?

A. Predictive Intelligence
B. Business Intelligence
C. Text Analytics
D. Data Mining

8.By 2025, the volume of data will increase to,

A. TB
B. YB
C. ZB
D. EB

9.Alternative Hypothesis is also called as?

E. Null Hypothesis
F. Research Hypothesis
G. Simple Hypothesis
H. None of the mentioned above

10. Amongst which of the following is / are not an example of social media?

2. Twitter
3. Instagram
4. Both A and B
5. None of the mentioned above

Fill in the blanks

1. ___ refers to the ability to turn your data useful for business.
2. Correlation is the relationship between two variables -………
3. **Expand MCR……………………**
4. Expand **MAR………….**
5. Expand **MNAR)…………………..**
6. Expand Nann…………
7. **.isnull() is used for………………….**
8. **.isna() is used for…………………**
9. **fillna() used for…………**
10. In smoothing by bin means, each value in the bin is replaced by _____

UNIT-2

1.Logistic regression is used to find the probability of event = Success and event = _____.

    A. Failure
    B. Success
    C. Both A and B
    D. None of the mentioned above

2.A good data analytics solution includes a viable self-service ___.

    E. Data mining
    F. Data wrangling
    G. Data warehouse
    H. None of the mentioned above

3. To glean insights from the data, many analysts and data scientists rely on ___.

    A. Data mining
    B. Data visualization
    C. Data warehouse
    D. All of the mentioned above

4.With reference to Predictive analytics, it allows organizations to predict customer behavior -

    A. True
    B. False

5. Customer analytics refers -

    E. Customer Relationship Management: churn analysis and prevention
    F. Marketing: cross-sell, up-sell
    G. Pricing: leakage monitoring, promotional effects tracking, competitive price responses
    H. All of the mentioned above

6. Data Modeling is the process of analyzing the data objects -

    A. True
    B. False

7 __ are the basic building blocks of qualitative data.

    E. Categories
    F. Data chunk
    G. Numeric figures
    H. None of the mentioned above

8. Metadata and data modeling tools support the creation and documentation of models -

A. True
B. False

9. The Process of describing the data that is huge and complex to store and process is known as ____.

A. Analytics mining
B. Data cleaning
C. Big data
D. None of the mentioned above

10. Customer behavior analytics is about understanding how your customers act -

C. True
D. False

Fill in the Blanks

1. Tableau is a ____ tool.
2. …………clustering technique needs the merging approach
3. The self-organizing maps can also be considered as the instance of _____ type of learning.
4. ____ is a category, also called supervised machine learning methods in which the data is split on two parts.
5. Clustering belongs to ____ data analysis.
6. What is the primary goal of data analysis…………..
7. What is the difference between mean and median………………..
8. What does the term "outlier" mean in data analysis………….
9. How is the range of a dataset calculated………………
10. What is the significance of the term "standard deviation…………………

UNIT-3
MCQs on Machine Learning for Data Analytics (Unit 3)

1. **Which of the following is a type of supervised learning?**
   a) Clustering
   b) Linear Regression
   c) Principal Component Analysis (PCA)
   d) Association Rule Mining

2. **Which machine learning algorithm is used for classification problems?**
   a) K-Means
   b) Decision Tree

c) Apriori Algorithm

d) DBSCAN

3. **Which of the following is an example of an unsupervised learning algorithm?**
   a) Logistic Regression
   b) Support Vector Machine
   c) K-Means Clustering
   d) Random Forest

4. **Overfitting in a machine learning model occurs when:**
   a) The model generalizes well to new data
   b) The model performs well on training data but poorly on test data
   c) The model has low variance and high bias
   d) The dataset has too many missing values

5. **Which metric is commonly used to evaluate the performance of a classification model?**
   a) Mean Absolute Error (MAE)
   b) R-Squared
   c) Confusion Matrix
   d) Sum of Squared Errors

6. **Which of the following is NOT a commonly used machine learning library?**
   a) TensorFlow
   b) PyTorch
   c) NumPy
   d) PostgreSQL

7. **Which algorithm is most suitable for predicting continuous numerical values?**
   a) Naïve Bayes
   b) Linear Regression
   c) K-Means Clustering
   d) Apriori Algorithm

8. **Which of the following is a key advantage of using Neural Networks?**
   a) Requires less training data
   b) Can capture complex patterns and relationships in data
   c) Is always faster than traditional algorithms
   d) Does not require any parameter tuning

9. **In machine learning, feature selection is important because:**
   a) It reduces the number of irrelevant or redundant features
   b) It increases the number of features to improve accuracy

c) It makes the model more complex

d) It eliminates the need for training the model

10. **Which of the following is a deep learning framework?**
    a) Hadoop
    b) Keras
    c) Tableau
    d) SQL

Fill in the Blanks Questions (Unit 3: Machine Learning for Data Analytics)

1. _____ learning is a type of machine learning where the model learns from labeled data.

2. _____ regression is used to predict continuous numerical values.

3. The _____ algorithm is commonly used for classification tasks and works by splitting data based on feature values.

4. _____ is an unsupervised learning technique used for grouping similar data points together.

5. In machine learning, _____ occurs when a model performs well on training data but poorly on new, unseen data.

6. A _____ matrix is used to evaluate the performance of a classification model.

7. _____ learning is a type of machine learning where the model learns from unlabeled data.

8. _____ is a deep learning framework developed by Google and widely used for building neural networks.

9. The process of selecting the most relevant features for a machine learning model is called _____.

10. The _____ algorithm is commonly used in recommendation systems and market basket analysis.

MCQs on (Unit 4)

1. **What is the primary characteristic that distinguishes Big Data from traditional data?**
   a) Small size
   b) Limited variety
   c) High volume, velocity, and variety
   d) Manual processing
2. **Which of the following is NOT considered one of the 5 Vs of Big Data?**
   a) Volume
   b) Variety
   c) Visualization
   d) Veracity
3. **What is the primary function of Hadoop Distributed File System (HDFS)?**
   a) To store and manage structured data only
   b) To provide a distributed storage system for large datasets
   c) To execute real-time queries on relational databases
   d) To process images and multimedia data
4. **Which component of the Hadoop ecosystem is responsible for parallel processing?**
   a) HDFS
   b) MapReduce
   c) Pig
   d) Sqoop
5. **Apache Spark is preferred over MapReduce because:**
   a) It supports batch processing only
   b) It provides in-memory processing for faster computation
   c) It can run only on Hadoop clusters
   d) It does not support machine learning
6. **Which tool is used for real-time Big Data processing?**
   a) Apache Storm
   b) Apache Pig
   c) Apache Hive
   d) Apache Sqoop
7. **Which of the following is NOT a Big Data framework?**
   a) Hadoop
   b) Spark
   c) SQL Server
   d) Flink
8. **What is the role of Apache Hive in Big Data Analytics?**
   a) It is a NoSQL database for storing Big Data
   b) It is a distributed computing framework
   c) It provides SQL-like querying for Big Data stored in Hadoop
   d) It is a machine learning library for Big Data

9. **Which of the following best describes "Data Lake"?**
   a) A centralized repository that stores structured, semi-structured, and unstructured data
   b) A relational database that processes Big Data
   c) A small storage unit for real-time data processing
   d) A cloud-based data warehouse
10. **Which cloud service is widely used for Big Data Analytics?**
    a) Amazon Redshift
    b) Microsoft Excel
    c) Google Drive
    d) Tableau

Fill in the Blanks Questions (Unit 4:)

1. Big Data is characterized by the five Vs: _____, _____, _____, _____, and _____.
2. _____ is the distributed storage system used in Hadoop to store large datasets across multiple nodes.
3. _____ is a programming model used for processing large datasets in a parallel and distributed manner in Hadoop.
4. Apache _____ is a Big Data processing framework that provides in-memory computation for faster data processing.
5. _____ is a cloud-based service by Amazon for data warehousing and analytics.
6. _____ is a SQL-like query language used in Hadoop for querying large datasets stored in HDFS.
7. _____ is an open-source real-time data processing system used for streaming data analytics.
8. The process of extracting data from relational databases and transferring it to Hadoop is done using _____.
9. A _____ is a centralized repository that allows you to store structured, semi-structured, and unstructured data at any scale.
10. _____ is an open-source framework used for large-scale data analytics and machine learning.

MCQs (Unit 5)

1. **Which of the following is NOT a common application of Data Analytics?**
   a) Fraud detection
   b) Predictive maintenance
   c) Video editing
   d) Customer segmentation
2. **Which industry uses data analytics for credit risk assessment and fraud detection?**
   a) Healthcare
   b) Retail

c) Banking & Finance
d) Education
3. **In healthcare, data analytics is primarily used for:**
a) Enhancing video quality
b) Diagnosing diseases and predicting outbreaks
c) Managing classroom schedules
4. **Which type of analytics is used for making future predictions based on historical data?**
a) Descriptive Analytics
b) Diagnostic Analytics
c) Predictive Analytics
d) Prescriptive Analytics
5. **Retail companies use data analytics mainly for:**
a) Product recommendations and demand forecasting
b) Weather forecasting
c) Agricultural yield predictions
d) Managing vehicle traffic
6. **Which of the following is a popular data visualization tool used in business analytics?**
a) Hadoop
b) Spark
c) Tableau
d) MongoDB
7. **Sentiment analysis in social media analytics is an application of:**
a) Text Mining
b) Image Processing
c) Network Security
d) Video Compression
8. **Which data analytics technique is widely used for detecting fraudulent transactions?**
a) Classification
b) Clustering
c) Association Rule Mining
d) Regression
9. **Prescriptive analytics helps businesses by:**
a) Describing past events
b) Diagnosing why something happened
c) Providing recommendations for decision-making
d) Predicting future trends
10. **Which of the following is NOT a challenge in data analytics implementation?**
a) Data privacy concerns
b) Lack of data quality
c) Unlimited computing resources
d) High cost of storage and processing

Fill in the Blanks Questions (Unit 5:)

1. _____ analytics is used to analyze past data to understand what happened.

2. Predictive analytics helps businesses _____ future trends based on historical data.

3. _____ is a widely used data visualization tool in business intelligence.

4. In healthcare, data analytics is used for _____ prediction and treatment planning.

5. _____ is an analytics technique used for detecting fraudulent transactions in banking.

6. The process of analyzing customer behavior to offer personalized recommendations is known as _____.

7. In social media analytics, _____ is used to analyze people's opinions and emotions.

8. _____ analytics provides actionable recommendations to improve decision-making.

9. _____ is an example of a Big Data application used in e-commerce for personalized recommendations.

10. A major challenge in data analytics implementation is ensuring _____ and security of sensitive data.

XVII TUTORIAL PROBLEMS WITH BLOOMS MAPPING

Unit	Bloom's Level	Question
UNIT-I **Data Analytics - Introduction**	**Applying**	Define clear and specific learning outcomes for each data analytics course or task. Once objectives are clarified, Bloom's levels can be applied more consistently.
	Analyzing	Interpret trends and patterns in structured data.
	Creating	Develop simple data visualizations to represent key insights.
UNIT-II **Introduction to Analytics with Tools**	**Applying**	Implement data cleaning techniques (handling missing values, removing duplicates).
	Analyzing	Assess data quality and identify inconsistencies.
	Creating	Design a workflow for automated data preprocessing.
UIT-III **Regression**	**Applying**	Use Python/R libraries (Pandas, NumPy, Matplotlib) for EDA.
	Evaluating	Identify correlations and distributions within datasets.
	Creating	Generate reports and interactive dashboards for exploratory insights.
UNIT-IV Object Segmentation: Regression Vs Segmentation	**Understanding**	Train basic predictive models using regression/classification.
	Applying	Evaluate model performance using accuracy, precision, recall.
	Evaluating	Optimize machine learning models for real-world applications.
UNIT-V	**Applying**	Use storytelling techniques to present data insights.

Data Visualization	Analyzing	Interpret business impact from analytical results.
	Evaluating	Develop a case study with data-driven recommendations.

XVIII Assignment Questions with Blooms

Unit 1:

Assignment Questions:

1. Define Data Analytics and explain its importance in decision-making. (Remembering - Level 1)

2. Compare and contrast structured, unstructured, and semi-structured data with examples. (Understanding - Level 2)

3. Demonstrate different data preprocessing techniques using a real-world dataset. (Applying - Level 3)

4. Analyze the challenges faced in data cleaning and transformation processes. (Analyzing - Level 4)

5. Justify the role of data analytics in business intelligence and industry applications. (Evaluating - Level 5)

Unit 2

Assignment Questions:

1. Explain the difference between descriptive and inferential statistics with examples. (Understanding - Level 2)

2. Compute and interpret mean, median, mode, and standard deviation for a given dataset. (Applying - Level 3)

3. Examine the use of data visualization techniques (histograms, scatter plots) for data exploration. (Analyzing - Level 4)

4. Critically evaluate the impact of probability distributions on data analytics models. (Evaluating - Level 5)

5. Design a case study demonstrating hypothesis testing in real-world applications. (Creating - Level 6)

Unit 3:

Assignment Questions:

1. Describe the key differences between supervised and unsupervised learning techniques. (Understanding - Level 2)

2. Implement a linear regression model using Python and interpret the results. (Applying - Level 3)

3. Compare the effectiveness of different classification algorithms (e.g., Decision Trees, SVM, Naïve Bayes). (Analyzing - Level 4)

4. Evaluate the performance of a clustering algorithm on a dataset using suitable metrics. (Evaluating - Level 5)

5. Develop a predictive analytics model for a business problem of your choice. (Creating - Level 6)

Unit 4:

Assignment Questions:

1. Define Big Data and explain the 5 Vs (Volume, Velocity, Variety, Veracity, and Value). (Remembering - Level 1)

2. Illustrate the working of the Hadoop ecosystem, including HDFS and MapReduce. (Understanding - Level 2)

3. Use Apache Spark to perform a basic ETL (Extract, Transform, Load) operation. (Applying - Level 3)

4. Analyze the challenges of real-time big data processing and storage. (Analyzing - Level 4)

5. Recommend suitable big data analytics tools for different industry use cases. (Evaluating - Level 5)

Assignment Questions:

1. Identify different industries where data analytics is applied and discuss its impact. (Remembering - Level 1)

2. Discuss ethical considerations in data privacy and security in analytics. (Understanding - Level 2)

3. Apply data analytics techniques to a healthcare or finance dataset to derive insights. (Applying - Level 3)

4. Examine the role of data analytics in fraud detection and risk management. (Analyzing - Level 4)

5. Propose a new application of data analytics for a real-world problem. (Creating - Level 6)

XIX -List of students

Course - B.Tech. Branch - CSM (AI&ML), Year & Sem: III/ II
Subject: Data Analytics

Sl. No.	Roll Number	Name of the Candidate
1	22C31A6601	ADEPU SUDHEER KUMAR
2	22C31A6602	AJMEERA KUSUMA SRI
3	22C31A6603	ALIKANTI SURAJ KUMAR
4	22C31A6604	AMAROJU RAVALI
5	22C31A6605	ANANDHAPU SRAVANI
6	22C31A6606	ANUMULA SAHODAR REDDY
7	22C31A6607	BABBERA RAVALI
8	22C31A6608	BANDAM KEERTHI REDDY
9	22C31A6609	BANDI NAMRATHA
10	22C31A6610	BISUPAKA NITHIN
11	22C31A6611	BYRI TEJASWINI
12	22C31A6612	CHITTIMALLA SRI RAM
13	22C31A6613	DASARI CHETHANA
14	22C31A6614	DEVULAPELLI MURALIKRISHNA
15	22C31A6615	DHARAVATH BHUMIKA
16	22C31A6616	DOMMATI DIVYA
17	22C31A6617	DUBBAKULA SAI KALYANI
18	22C31A6618	DUPAKI SIJJU VARMA
19	22C31A6619	GANGARAPU AKSHAYA
20	22C31A6620	GANGULA VAGDEVI
21	22C31A6621	GUGULOTHU SARITHA
22	22C31A6622	GUNTI SRIKANTH
23	22C31A6623	GURRAM BHARGAV
24	**22C31A6624**	**JERIPOTHULA AJAY KUMAR**
25	22C31A6625	KAITHOJU SHIVAVARAPRASAD
26	22C31A6626	KANNE NARESH
27	**22C31A6627**	**KURAKULA MAHENDER**
28	22C31A6628	KURAPATI ASHRITHA
29	**22C31A6629**	**LOKATI PAVAN**
30	22C31A6630	MADISHETTY VIJAYKUMAR
31	22C31A6631	MALAKUMMARI SHIVA
32	22C31A6632	MANDA VIGNAN
33	22C31A6633	MANDALA BHAVITHA

34	22C31A6634	MANEM MAHENDAR
35	22C31A6635	MD SANIYA
36	22C31A6636	MEKALA SOUJANYA
37	22C31A6637	MERUGU KALYAN
38	22C31A6638	MERUGU SAI SHIVA
39	22C31A6639	MOHAMMAD ALTAF HUSSAIN
40	22C31A6640	MOHAMMAD ARSHINAAZ
41	22C31A6641	MOHAMMAD GHOUSE KHAN
42	22C31A6642	MOHAMMAD MUSKAN
43	22C31A6643	MOHAMMED SOHAIL
44	22C31A6644	MORAPAKA MANIKANTA
45	22C31A6645	MUDUTHANAPELLY SRI SAI MADHURVIND
46	22C31A6646	MUTHUNURI SUNIL
47	22C31A6647	NAGAPURI RISHIVARDHAN
48	22C31A6648	NALAMASA NAVEEN
49	22C31A6649	NALLELLA DEVIKA
50	22C31A6650	PALLE JAYARAM
51	22C31A6651	PATTABI SHRUTHI
52	22C31A6652	PENDRA ASHOK
53	22C31A6653	PITTA ANJI
54	22C31A6654	PITTA VINAY
55	22C31A6655	POLU INDU
56	22C31A6656	RATHNA RAKESH
57	22C31A6657	RAVULA PRABHAS
58	22C31A6658	SADA ASHOK
59	22C31A6659	SAMBAR VINAY
60	22C31A6660	SAYED SANIYA
61	22C31A6661	SHENKESHI SRIJA
62	22C31A6662	SRIRAM VARA LAXMI
63	22C31A6663	THODUPUNURI SHARVANI
64	22C31A6664	YARA GANESH
65	23C35A6601	BATHIKA DILEEP
66	23C35A6602	GUGGILLA SRIKAR
67	23C35A6603	KOKKONDA SRINITHA
68	23C35A6604	POLABOINA PAVAN SAI
69	23C35A6605	RAINI MARUTHI
70	23C35A6606	TALLA SAGAR

XX. SCHEME AND SOLUTION OF INTERNAL TESTS.

In CIE, for theory subjects, during a semester, there shall be two mid-term examinations. Each Mid-Term examination consists of two parts i) **Part – A** for 10 marks, ii) **Part – B** for 20 marks with a total duration of 2 hours as follows:

1. **Mid Term Examination for 30 marks:**
    a. Part - A : Objective/quiz paper for 10 marks.
    b. Part - B : Descriptive paper for 20 marks.

The objective/quiz paper is set with multiple choice, fill-in the blanks and match the following type of questions for a total of 10 marks. The descriptive paper shall contain 6 full questions out of which, the student has to answer 4 questions, each carrying 5 marks. The **average of the two Mid Term Examinations** shall be taken as the final marks for Mid Term Examination (for 30 marks).

The remaining 10 marks of Continuous Internal Evaluation are distributed as:

2. **Assignment for 5 marks. (Average of 2 Assignments each for 5 marks)**
3. **Subject Viva-Voce/PPT/Poster Presentation/ Case Study on a topic in the concerned subject for 5 marks.**

While the first mid-term examination shall be conducted on 50% of the syllabus, the second mid-term examination shall be conducted on the remaining 50% of the syllabus. Five (5) marks are allocated for assignments (as specified by the subject teacher concerned). The first assignment should be submitted before the conduct of the first mid-term examination, and the second assignment should be submitted before the conduct of the second mid-term examination. The average of the two assignments shall be taken as the final marks for assignment (for 5 marks). Subject Viva-Voce/PPT/Poster Presentation/ Case Study on a topic in the subject concerned for 5 marks before II Mid-Term Examination.

XX-II Marks sheet

ISO 9001:2015 Certified Institution — Balaji Institute of Technology & Science
Laknepally (V), Narsampet (M), Warangal District - 506 331, Telangana State, India
(AUTONOMOUS)
Accredited by NBA (UG - CE, ME, ECE & CSE) & NAAC A+ Grade
(Affiliated to JNT University, Hyderabad and Approved by AICTE, New Delhi)
www.bitswgl.ac.in, email: principal@bitswgl.ac.in, Ph:98660 50044, Fax: 08718-230521
Estd.:2001

Department of Artificial Intelligence & Machine Learning

Sl. No.	Roll Number	Name of the Candidate	Part - A (Quiz -10)	A+B=30	Assessment(5)	Grand Total(35)
1	22C31A6601	ADEPU SUDHEER KUMAR	10	30	5	35
2	22C31A6602	AJMEERA KUSUMA SRI	10	29	5	34
3	22C31A6603	ALIKANTI SURAJ KUMAR	10	30	5	35
4	22C31A6604	AMAROJU RAVALI	10	29	5	34
5	22C31A6605	ANANDHAPU SRAVANI	10	29	5	34
6	22C31A6606	ANUMULA SAHODAR REDDY	10	18	4	s
7	22C31A6607	BABBERA RAVALI	10	25	4	29
8	22C31A6608	BANDAM KEERTHI REDDY	10	23	4	27
9	22C31A6609	BANDI NAMRATHA	10	30	5	35
10	22C31A6610	BISUPAKA NITHIN				
11	22C31A6611	BYRI TEJASWINI	9	24	4	28
12	22C31A6612	CHITTIMALLA SRI RAM	10	23	3	26
13	22C31A6613	DASARI CHETHANA	10	29	5	34
14	22C31A6614	DEVULAPELLI MURALIKRISHNA	10	18	4	22
15	22C31A6615	DHARAVATH BHUMIKA	9	27	4	31
16	22C31A6616	DOMMATI DIVYA	9	22	4	26
17	22C31A6617	DUBBAKULA SAI KALYANI	9	29	5	34
18	22C31A6618	DUPAKI SIJJU VARMA	9	21	5	26
19	22C31A6619	GANGARAPU AKSHAYA	10	30	5	35
20	22C31A6620	GANGULA VAGDEVI	7	23	4	27

21	22C31A6621	GUGULOTHU SARITHA	7	25	5	30
22	22C31A6622	GUNTI SRIKANTH	10	17	4	21
23	22C31A6623	GURRAM BHARGAV	10	30	4	34
24	**22C31A6624**	**JERIPOTHULA AJAY KUMAR**	**Absent**			
25	22C31A6625	KAITHOJU SHIVAVARAPRASAD	10	29	4	33
26	22C31A6626	KANNE NARESH	10	28	4	32
27	**22C31A6627**	**KURAKULA MAHENDER**	**Absent**			
28	22C31A6628	KURAPATI ASHRITHA	10	30	5	35
29	**22C31A6629**	**LOKATI PAVAN**				
30	22C31A6630	MADISHETTY VIJAYKUMAR	10	18	4	22
31	22C31A6631	MALAKUMMARI SHIVA	9	20	5	25
32	22C31A6632	MANDA VIGNAN	7	15	4	19
33	22C31A6633	MANDALA BHAVITHA	9	27	5	32
34	22C31A6634	MANEM MAHENDAR	9	29	5	34
35	22C31A6635	MD SANIYA	6	19	5	24
36	22C31A6636	MEKALA SOUJANYA	**Absent**			
37	22C31A6637	MERUGU KALYAN	8	15	5	20
38	22C31A6638	MERUGU SAI SHIVA	10	30	5	35
39	22C31A6639	MOHAMMAD ALTAF HUSSAIN	8	24	4	28
40	22C31A6640	MOHAMMAD ARSHINAAZ	10	27	5	32
41	22C31A6641	MOHAMMAD GHOUSE KHAN	10	26	5	31
42	22C31A6642	MOHAMMAD MUSKAN	10	25	4	29
43	22C31A6643	MOHAMMED SOHAIL	10	28	5	33
44	22C31A6644	MORAPAKA MANIKANTA	10	15	4	19
45	22C31A6645	MUDUTHANAPELLY SRI SAI MADHURVIND	7	24	5	29
46	22C31A6646	MUTHUNURI SUNIL	9	24	4	28
47	22C31A6647	NAGAPURI RISHIVARDHAN	9	25	5	30
48	22C31A6648	NALAMASA NAVEEN	9	19	4	23
49	22C31A6649	NALLELLA DEVIKA	9	29	5	34

50	22C31A6650	PALLE JAYARAM	10	30	5	35
51	22C31A6651	PATTABI SHRUTHI	7	26	5	31
52	22C31A6652	PENDRA ASHOK	9	24	4	28
53	22C31A6653	PITTA ANJI	5	22	4	26
54	22C31A6654	PITTA VINAY	5	21	4	25
55	22C31A6655	POLU INDU	9	29	5	34
56	22C31A6656	RATHNA RAKESH	10	27	4	31
57	22C31A6657	RAVULA PRABHAS	10	28	3	31
58	22C31A6658	SADA ASHOK	4	17	4	21
59	22C31A6659	SAMBAR VINAY	10	30	4	34
60	22C31A6660	SAYED SANIYA	9	28	5	33
61	22C31A6661	SHENKESHI SRIJA	7	24	4	28
62	22C31A6662	SRIRAM VARA LAXMI	7	21	5	26
63	22C31A6663	THODUPUNURI SHARVANI	7	20	4	24
64	22C31A6664	YARA GANESH	9	28	5	33
65	23C35A6601	BATHIKA DILEEP	4	22	3	25
66	23C35A6602	GUGGILLA SRIKAR	2	17	4	21
67	23C35A6603	KOKKONDA SRINITHA	4	17	4	21
68	23C35A6604	POLABOINA PAVAN SAI	7	19	3	22
69	23C35A6605	RAINI MARUTHI	6	18	4	22
70	23C35A6606	TALLA SAGAR	6	19	4	23

XX-III Result analysis for internal Exams (tests) with respect to COs-POs

XX-IV . Result analysis for external exams (university)

XX-V . CO and PO attainment sheet

XX-VI. REFERENCES, JOURNALS, WEBSITES AND E-LINKS IF ANY

TEXT BOOKS:

1. Student's Handbook for Associate Analytics – II, III.

2.Data Mining Concepts and Techniques, Han, Kamber, 3rd Edition, Morgan Kaufmann Publishers.

REFERENCE BOOKS:

1. Introduction to Data Mining, Tan, Steinbach and Kumar, Addision Wisley, 2006.

2. Data Mining Analysis and Concepts, M. Zaki and W. Meira 3. Mining of Massive Datasets, Jure Leskovec Stanford Univ. Anand Rajaraman Milliway Labs Jeffrey D Ullman Stanford Univ.