



Estd. 2001

# Sri Indu

College of Engineering & Technology

UGC Autonomous Institution

Recognized under 2(f) & 12(B) of UGC Act 1956,

NAAC, Approved by AICTE &

Permanently Affiliated to JNTUH



## NAAC

NATIONAL ASSESSMENT AND  
ACCREDITATION COUNCIL



# HANDOUT

III Year (AIML) – Semester II

DEPARTMENT OF AIML

ACADEMIC YEAR 2025-26

## DEPARTMENT OF AIML

# HANDOUT-INDEX

| S.No | Contents                                          |
|------|---------------------------------------------------|
| 1    | Vision, Mission, PEOs, POs, PSOs & COs            |
| 2    | Institution Academic Calendar                     |
| 3    | Department Academic Calendar                      |
| 4    | Subject wise                                      |
| i)   | Syllabus Copy                                     |
| ii)  | Lesson Plan                                       |
| iii) | Question Bank                                     |
| iv)  | End Examination Questions (Previous3AcademicYear) |
| v)   | Mid-1 & Mid-2 Questions (Previous3AcademicYear)   |



# **RIINDUCOLLEGE OF ENGINEERING & TECHNOLOGY**

## **B. TECH – CSE (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

### **INSTITUTION VISION**

To be a premier Institution in Engineering & Technology and Management with competency, values and social consciousness.

### **INSTITUTION MISSION**

- IM1** Provide high quality academic programs, training activities and research facilities.
- IM2** Promote Continuous Industry-Institute interaction for employability, Entrepreneurship, leadership and research aptitude among stakeholders.
- IM3** Contribute to the economic and technological development of the region, state and nation.

### **DEPARTMENT VISION**

To develop competent engineers in the domain of Artificial Intelligence & Machine Learning for noteworthy contributions to the society.

### **DEPARTMENT MISSION**

The Department has following Missions:

- DM1** To strengthen academic collaborations for better exposure.
- DM2** Promote professional environment to imbibe ethical values and entrepreneurial skills.
- DM3** Encourage research development by emphasizing on innovation.
- DM4** Facilitate Industry-Institute collaboration for mutual benefit

## **PROGRAM EDUCATIONAL OBJECTIVES (PEOs)**

**PEO1:** Generate contributions towards advancements in Artificial Intelligence and Machine Learning.

**PEO2:** Promote design, research and implementation of products through strong communication skills, leadership and entrepreneurial skills.

**PEO3:** Apply basic principles and practices of AIML to successfully complete software related projects to meet customer business objectives and/or productively engage in research.

| PO                               | Description                                                                                                                                                                                                                                                                                                     |
|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PO 1                             | Engineering Knowledge: Apply knowledge of mathematics, natural science, computing, engineering fundamentals and an engineering specialization as specified in WK1 to WK4 respectively to develop to the solution of complex engineering problems.                                                               |
| PO 2                             | Problem Analysis: Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development. (WK1 to WK4)                                                                                                      |
| PO 3                             | Design/Development of Solutions: Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required. (WK5) |
| PO 4                             | Conduct Investigations of Complex Problems: Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions. (WK8).                                                            |
| PO 5                             | Engineering Tool Usage: Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems. (WK2 and WK6)                                                                     |
| PO 6                             | The Engineer and The World: Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment. (WK1, WK5, and WK7).                                       |
| PO 7                             | Ethics: Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws. (WK9)                                                                                                                                                       |
| PO 8                             | Individual and Collaborative Team work: Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.                                                                                                                                                                   |
| PO 9                             | Communication: Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.                     |
| PO 10                            | Project Management and Finance: Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.                                          |
| PO 11                            | Life-Long Learning: Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change. (WK8)                                                  |
| <b>Program Specific Outcomes</b> |                                                                                                                                                                                                                                                                                                                 |
| PSO 1                            | Register mathematical methodology to crack problems using suitable data structures                                                                                                                                                                                                                              |

|              |                                                                                                              |
|--------------|--------------------------------------------------------------------------------------------------------------|
| <b>PSO 2</b> | Competence to design and develop software for web based and mobiles androids under real world environment    |
| <b>PSO 3</b> | Skill to design the algorithms for machine learning, data compression can be used in different applications. |

### **DATA ANALYTICS(R22CSE3141)**

At the end of the course student will be able to:

| <b>Course Outcomes</b> | <b>Statements</b>                                                            |
|------------------------|------------------------------------------------------------------------------|
| C322.1                 | Understand the impact of data analytics for business decisions and strategy. |
| C322.2                 | Carry out data analysis /statistical analysis.                               |
| C322.3                 | To carry out standard data visualization and formal inference procedures.    |
| C322.4                 | Design data architecture.                                                    |
| C322.5                 | Understand various data sources.                                             |
| C322.6                 | Evaluate segmentation models using appropriate performance metrics.          |

### Course Articulation Matrix:

| Course Outcome | PO1  | PO2  | PO3  | PO4 | PO5  | PO6 | PO7 | PO8 | PO9  | PO10 | PO11 | PSO1 | PSO2 | PSO3 |
|----------------|------|------|------|-----|------|-----|-----|-----|------|------|------|------|------|------|
| C322.1         | 2    | 2    | -    | -   | 3    | -   | -   | 1   | -    | -    | -    | 2    | 1    | 2    |
| C322.2         | 1    | 1    | 3    | -   | 3    | -   | -   | 1   | -    | -    | 1    | 2    | 2    | 2    |
| C322.3         | 2    | 3    | 3    | 1   | -    | -   | -   | 2   | -    | -    | 1    | 2    | -    | 2    |
| C322.4         | 3    | 2    | 2    | 2   | 3    | -   | -   | 1   | -    | -    | -    | 3    | 3    | 2    |
| C322.5         | 2    | 3    | 3    | -   | 1    | -   | -   | 1   | 1    | -    | 2    | 1    | -    | 2    |
| C322.6         | 1    | 3    | -    | 3   | -    | -   | -   | -   | 1    | -    | 2    | 2    | -    | 2    |
| C322           | 1.83 | 2.33 | 1.83 | 1   | 1.66 | -   | -   | 1   | 0.33 | -    | 1    | 2    | 1    | 2    |



**SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY**  
(An Autonomous Institution under UGC, New Delhi)  
Recognized under 2(f) and 12(B) of UGC Act 1956  
NBA & NAAC Accredited, Approved by AICTE and Permanently affiliated to JNTUH  
Sheriguda (V), Ibrahimpatnam, R.R.Dist, Hyderabad - 501 510

D4

**ANNEXURE - III**

**BR-22**

**Lr.No.SICET/AUTO/DAE/II B.Tech Academic Calendar/30/2025**

**B.TECH II-YEAR I-SEM & II-SEM ACADEMIC CALENDAR**  
**(FOR ACADEMIC YEAR : 2025-26)**

**Academic Calendar for B.Tech – II<sup>nd</sup> Year Students (2024 - 25 Batch), BR-22 Regulation.**

**I – Semester**

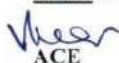
| S.No.                                                                    | EVENT                                                                          | PERIOD                     | DURATION |
|--------------------------------------------------------------------------|--------------------------------------------------------------------------------|----------------------------|----------|
| 1.                                                                       | Commencement of class work.                                                    | <b>01.08.2025 (Monday)</b> |          |
| 2.                                                                       | 1 <sup>st</sup> Spell of Instructions for covering First Two and a half Units. | 01.08.2025 – 07.10.2025    | 9 Weeks  |
| 3.                                                                       | I Mid Term Examinations.                                                       | 08.10.2025 – 10.10.2025    | 3 Days   |
| 4.                                                                       | Submission of I Mid Term Examination Marks.                                    | 16.10.2025                 |          |
| 5.                                                                       | 2 <sup>nd</sup> Spell of Instructions for Remaining Two and a half Units.      | 13.10.2025 – 13.12.2025    | 9 Weeks  |
| 6.                                                                       | II Mid Term Examinations.                                                      | 15.12.2025 – 17.12.2025    | 3 Days   |
| 7.                                                                       | Preparation Holidays and Practical Examinations                                | 18.12.2025 – 27.12.2025    | 1 Week   |
| 8.                                                                       | Submission of II Mid Term Examination Marks.                                   | 23.12.2025                 |          |
| 9.                                                                       | I Semester End Examinations.                                                   | 29.12.2025 – 10.01.2026    | 2 Weeks  |
| <b>Commencement of Class-Work for II B.Tech - II Semester 19.01.2026</b> |                                                                                |                            |          |

**Note:** No of Working/instructional days : 92

**II – Semester**

| S.No. | EVENT                                                                          | PERIOD                     | DURATION       |
|-------|--------------------------------------------------------------------------------|----------------------------|----------------|
| 1.    | Commencement of class work.                                                    | <b>19.01.2026 (Monday)</b> |                |
| 2.    | 1 <sup>st</sup> Spell of Instructions for covering First Two and a half Units. | 19.01.2026 – 14.03.2026    | 8 Weeks        |
| 3.    | I Mid Term Examinations.                                                       | 16.03.2026 – 18.03.2026    | 3 Days         |
| 4.    | Submission of I Mid Term Examination Marks.                                    | 24.03.2026                 |                |
| 5.    | 2 <sup>nd</sup> Spell of Instructions for Remaining Two and a half Units.      | 20.03.2026 – 09.05.2026    | 7 Weeks        |
| 6.    | Continuation of remaining 2 <sup>nd</sup> Spell of Instructions                | 25.05.2026 – 10.06.2026    | <b>2 Weeks</b> |
| 7.    | II Mid Term Examinations.                                                      | 11.06.2026 – 13.06.2026    | 3 Days         |
| 8.    | Preparation Holidays and Project Evaluation.                                   | 15.06.2026 – 20.06.2026    | 1 Week         |
| 9.    | Submission of II Mid Term Examination Marks.                                   | 18.06.2026                 |                |
| 10.   | II Semester End Examinations                                                   | 22.06.2026 – 04.07.2026    | 2 Weeks        |

**Note:** No of Working/instructional days : 92

  
ACE

  
CE

  
DIRECTOR

  
PRINCIPAL

**Controller of Examination**  
Sri Indu College of Engineering & Technology  
(An Autonomous Institution Under UGC)  
Sheriguda (V), Ibrahimpatnam, R.R. Dist-501510

**DIRECTOR**  
Sri Indu College of Engineering & Technology  
(An Autonomous Institution Under UGC)  
Sheriguda (V), Ibrahimpatnam, R.R. Dist-501510

**PRINCIPAL**  
Sri Indu College of Engineering & Technology  
(An Autonomous Institution Under UGC)  
Sheriguda (V), Ibrahimpatnam, R.R. Dist-501510.

Sign:   
**Dr. M.V.S.S. Giridhar**  
Prof. of CEA; JNTUH Nominee

Sign:   
**Dr. T. Venu Gopal**  
Prof. of CSE; JNTUH Nominee

Sign:   
**Dr. D. Ramesh**  
Prof. of CSE; JNTUH Nominee

(Copy to DAP&E and Copy to all the Dept. Heads)

# Department Calender

**SRI INDU COLLEGE OF ENGINEERING AND TECHNOLOGY(AUTONOMOUS)**  
**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING(AI&ML) DEPARTMENT CALENDAR – 2025-2026(SEMESTER-II)**

|           |    |                         |    |                              | Feb-26 |                                                | Mar-26 |                                        |    |                             |    |                            |                |                                                 |
|-----------|----|-------------------------|----|------------------------------|--------|------------------------------------------------|--------|----------------------------------------|----|-----------------------------|----|----------------------------|----------------|-------------------------------------------------|
| SUNDAY    |    | Dec-25                  |    |                              | 1      |                                                | 1      |                                        |    |                             |    | Jun-26                     |                |                                                 |
| MONDAY    | 1  |                         |    |                              | 2      |                                                | 2      |                                        |    |                             |    | 1                          |                |                                                 |
| TUESDAY   | 2  |                         |    |                              | 3      | II YR CRT TRAINING                             | 3      | II YR CRT TRAINING                     |    | Apr-26                      |    | 2                          | II YR TRAINING |                                                 |
| WEDNESDAY | 3  |                         |    | Jan-26                       | 4      |                                                | 4      | HOLI HOLIDAY                           | 1  |                             |    | 3                          |                |                                                 |
| THURSDAY  | 4  |                         | 1  | NEW YEAR HOLIDAY             | 5      |                                                | 5      |                                        | 2  | II YR TRAINING              |    | May-26                     | 4              |                                                 |
| FRIDAY    | 5  | AI CLUB MEETING         | 2  |                              | 6      | II YR CRT TRAINING                             | 6      | II YR CRT TRAINING                     | 3  | GOOD FRIDAY HOLIDAY         | 1  | IV-II MID II EXAMS         | 5              | II YR TRAINING                                  |
| SATURDAY  | 6  | DEPARTMENT MEETING      | 3  |                              | 7      | AI CLUB MEETING                                | 7      |                                        | 4  | AI CLUB MEETING             | 2  | IV-II MID II EXAMS         | 6              |                                                 |
| SUNDAY    | 7  | HOLIDAY                 | 4  | HOLIDAY                      | 8      | HOLIDAY                                        | 8      | HOLIDAY                                | 5  | HOLIDAY                     | 3  | HOLIDAY                    | 7              | HOLIDAY                                         |
| MONDAY    | 8  |                         | 5  |                              | 9      |                                                | 9      |                                        | 6  |                             | 4  | IV-II PREPARATION HOLIDAYS | 8              |                                                 |
| TUESDAY   | 9  |                         | 6  |                              | 10     | II YR CRT TRAINING                             | 10     | II YR CRT TRAINING                     | 7  | II YR TRAINING              | 5  | IV-II PREPARATION HOLIDAYS | 9              |                                                 |
| WEDNESDAY | 10 |                         | 7  |                              | 11     |                                                | 11     |                                        | 8  |                             | 6  | IV-II PREPARATION HOLIDAYS | 10             |                                                 |
| THURSDAY  | 11 |                         | 8  |                              | 12     |                                                | 12     |                                        | 9  |                             | 7  | SUBMISSION OF MID-II MARKS | 11             | II-II&III-II MID II EXAM                        |
| FRIDAY    | 12 |                         | 9  |                              | 13     | II YR CRT TRAINING                             | 13     | II YR CRT TRAINING                     | 10 | II YR TRAINING              | 8  | IV-II PREPARATION HOLIDAYS | 12             | II-II&III-II MID II EXAM                        |
| SATURDAY  | 13 |                         | 10 | DEPARTMENT MEETING           | 14     |                                                | 14     |                                        | 11 | DEPARTMENT MEETING          | 9  | IV-II PREPARATION HOLIDAYS | 13             | II-II&III-II MID II EXAM                        |
| SUNDAY    | 14 | HOLIDAY                 | 11 | HOLIDAY                      | 15     | HOLIDAY                                        | 15     | HOLIDAY                                | 12 | HOLIDAY                     | 10 | HOLIDAY                    | 14             | HOLIDAY                                         |
| MONDAY    | 15 | IV-II COMMENCEMENT WORK | 12 |                              | 16     | IV-II MID I EXAM                               | 16     | II & III-II MID I EXAM                 | 13 |                             | 11 | IV-II SEM EXAMS            | 15             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| TUESDAY   | 16 |                         | 13 | PONGAL HOLIDAYS              | 17     | IV-II MID I EXAM& II YR TRAINING               | 17     | II & III-II MID I EXAM,II YR TRAINING  | 14 | DR.AMBEDKAR JAYANTI HOLIDAY | 12 | IV-II SEM EXAMS            | 16             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| WEDNESDAY | 17 |                         | 14 | PONGAL HOLIDAYS              | 18     | IV-II MID I EXAM                               | 18     | II & III-II MID I EXAM                 | 15 |                             | 13 | IV-II SEM EXAMS            | 17             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| THURSDAY  | 18 |                         | 15 | PONGAL HOLIDAYS              | 19     | IV-II 2ND SPELL                                | 19     |                                        | 16 |                             | 14 | IV-II SEM EXAMS            | 18             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| FRIDAY    | 19 |                         | 16 | PONGAL HOLIDAYS              | 20     | II YR CRT TRAINING                             | 20     | II -II & III-II/ II SPELL INSTRUCTIONS | 17 | II YR TRAINING              | 15 | IV-II SEM EXAMS,           | 19             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| SATURDAY  | 20 |                         | 17 | PONGAL HOLIDAYS              | 21     |                                                | 21     |                                        | 18 |                             | 16 | IV-II SEM EXAMS            | 20             | II-II&III-II PREP HOILDAYS & PROJECT EVALUATION |
| SUNDAY    | 21 | HOLIDAY                 | 18 | HOLIDAY                      | 22     | HOLIDAY                                        | 22     | HOLIDAY                                | 19 | HOLIDAY                     | 17 | HOLIDAY                    | 21             | HOLIDAY                                         |
| MONDAY    | 22 |                         | 19 | II &III-II COMMENCEMENT WORK | 23     |                                                | 23     |                                        | 20 |                             | 18 |                            | 22             | II-II & III-II SEM END EXAMS                    |
| TUESDAY   | 23 |                         | 20 |                              | 24     | SUBMISSION OF IV-II MID I MARKS,II YR TRAINING | 24     | SUBMISSION OF III-II MID I MARKS       | 21 | II YR TRAINING              | 19 | II YR TRAINING             | 23             | II-II & III-II SEM END EXAMS                    |

|           |    |                       |    |                            |    |                       |    |                |    |                    |    |                |    |                                 |
|-----------|----|-----------------------|----|----------------------------|----|-----------------------|----|----------------|----|--------------------|----|----------------|----|---------------------------------|
| WEDNESDAY | 24 |                       | 21 |                            | 25 |                       | 25 |                | 22 |                    | 20 |                | 24 | II-II & III-II SEM<br>END EXAMS |
| THURSDAY  | 25 | CHRISTMAS<br>HOLIDAY  | 22 |                            | 26 |                       | 26 |                | 23 |                    | 21 |                | 25 | II-II & III-II SEM<br>END EXAMS |
| FRIDAY    | 26 | BOXING DAY<br>HOLIDAY | 23 | AI CLUB MEETING            | 27 | II YR CRT<br>TRAINING | 27 | II YR TRAINING | 24 | II YR TRAINING     | 22 | II YR TRAINING | 26 | MUHARRAM<br>HOLIDAY             |
| SATURDAY  | 27 | DEPARTMENT<br>MEETING | 24 |                            | 28 | DEPARTMENT<br>MEETING | 28 |                | 25 |                    | 23 |                | 27 | II-II & III-II SEM<br>END EXAMS |
| SUNDAY    | 28 | HOLIDAY               | 25 | HOLIDAY                    |    | HOLIDAY               | 29 | HOLIDAY        | 26 | HOLIDAY            | 24 | HOLIDAY        | 28 | HOLIDAY                         |
| MONDAY    | 29 |                       | 26 | REPUBLIC DAY               |    |                       | 30 |                | 27 |                    | 25 |                | 29 | II-II & III-II SEM<br>END EXAMS |
| TUESDAY   | 30 |                       | 27 | II& III YR CRT<br>TRAINING |    |                       | 31 | II YR TRAINING | 28 | II YR TRAINING     | 26 | II YR TRAINING | 30 | II-II & III-II SEM<br>END EXAMS |
| WEDNESDAY | 31 | AI CLUB MEETING       | 28 | II& III YR CRT<br>TRAINING |    |                       |    |                | 29 | AI CLUB MEETING    | 27 | BAKRID HOILDAY |    |                                 |
| THURSDAY  |    |                       | 29 | II& III YR CRT<br>TRAINING |    |                       |    |                | 30 | IV-II MID II EXAMS | 28 |                |    |                                 |
| FRIDAY    |    |                       | 30 | II& III YR CRT<br>TRAINING |    |                       |    |                |    |                    | 29 | II YR TRAINING |    |                                 |
| SATURDAY  |    |                       | 31 | II& III YR CRT<br>TRAINING |    |                       |    |                |    |                    | 30 |                |    |                                 |
| SUNDAY    |    | HOLIDAY               |    | HOLIDAY                    |    | HOLIDAY               |    | HOLIDAY        |    | HOLIDAY            | 31 | HOLIDAY        |    | HOLIDAY                         |

HOD

PRINCIPAL

**SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY**  
**(An Autonomous Institution under UGC, New Delhi)**

**B.Tech. - III Year – II Semester**

| L | T | P | C |
|---|---|---|---|
| 3 | 0 | 0 | 3 |

**(R22CSE3141) DATA ANALYTICS**

**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. Data Mining Concepts and Techniques, Han, Kamber, 3<sup>rd</sup> Edition, Morgan Kaufmann Publishers.
2. Big Data Analytics, Seema Acharya, Subhasini Chellappan, Wiley 2015.



**SRIINDUCOLLEGE OF ENG & TECH**  
**LESSON PLAN**  
**(Regulation: R22)**  
**Department of AIML**

Prepared on:  
11.12.2024

|                            |                                                                        |        |  |
|----------------------------|------------------------------------------------------------------------|--------|--|
| Sub.Code & Title           | R22CSE3141 Data Analytics                                              |        |  |
| AcademicYear:2025-2026     | Year/Sem./Section                                                      | III-II |  |
| Faculty Name & Designation | MR.DR.M. NARESH ASSOCIATE PROFESSOR, MR.M. AVINASH ASSISTANT PROFESSOR |        |  |

| Unit/<br>Item<br>No. | Topic(s)                                               | Book<br>Reference | Page (s) |    | Teaching<br>Methodology    | Proposed<br>No. of<br>Periods | Actual Date<br>ofHandled | CO/RBT  |
|----------------------|--------------------------------------------------------|-------------------|----------|----|----------------------------|-------------------------------|--------------------------|---------|
|                      |                                                        |                   | From     | To |                            |                               |                          |         |
|                      | UNIT-I                                                 |                   |          |    |                            |                               |                          |         |
| I                    | Data Management                                        |                   |          |    |                            | 17                            |                          |         |
| 1.1                  | Introduction to Data Management                        | T1                |          |    | Blackboard                 | 01                            |                          | CO1,L1  |
| 1.2                  | Design Data Architecture                               | T1                |          |    | Blackboard                 | 01                            |                          | CO1,L1  |
| 1.3                  | Manage the data for analysis                           | T1                |          |    | PowerPoint<br>Presentation | 02                            |                          | CO1,,L2 |
| 1.4                  | Understand various sources of Data<br>like Sensors     | T1                |          |    | Blackboard                 | 01                            |                          | CO1,L2  |
| 1.5                  | Understand various sources of Data<br>like Signals     | T1                |          |    | MODEL                      | 01                            |                          | CO1,L2  |
| 1.6                  | Understand various sources of Data<br>like GPS         | T1                |          |    | Chart                      | 01                            |                          | CO1,L1  |
| 1.7                  | Understand various sources of Data<br>- Primary Data   | T1                |          |    | Blackboard                 | 01                            |                          | CO1,L2  |
| 1.8                  | Understand various sources of Data<br>- Secondary Data | T1                |          |    | Blackboard                 | 01                            |                          | C01,L2  |
| 1.9                  | Comparison of sources of data                          | T1                |          |    | PowerPoint<br>Presentation | 01                            |                          | CO1, L1 |
| 1.10                 | Data Management                                        | T1                |          |    | Blackboard                 | 01                            |                          | CO1, L1 |
| 1.11                 | Data Quality                                           | T1                |          |    | Blackboard                 | 02                            |                          | CO1, L3 |

|         |                                                               |                                  |        |  |                          |    |  |         |
|---------|---------------------------------------------------------------|----------------------------------|--------|--|--------------------------|----|--|---------|
| 1.12    | Data processing                                               | T1                               | 84-110 |  | Blackboard               | 02 |  | CO1, L1 |
| 1.13    | Data Preprocessing, Steps involved in data pre preprocessing. | W1                               |        |  | Seminars                 | 02 |  | CO1, L3 |
|         | Review                                                        | Signature of the HOD/Coordinator |        |  |                          |    |  |         |
| UNIT-II |                                                               |                                  |        |  |                          |    |  |         |
| II      | Data Analytics                                                |                                  |        |  |                          | 10 |  |         |
| 2.1     | Introduction to Analytics                                     | W2                               |        |  | Blackboard               | 01 |  | CO2,L1  |
| 2.2     | Introduction to Tools and Environment                         | W2                               |        |  | Blackboard               | 01 |  | CO2,L1  |
| 2.3     | Application of modeling a business                            | T1                               |        |  | Power point Presentation | 01 |  | CO2,L2  |
| 2.4     | Need for business Modeling                                    | T1                               |        |  | Power point Presentation | 01 |  | CO2,L2  |
| 2.5     | Types of Data and variables                                   | T1                               |        |  | Blackboard               | 02 |  | CO2,L2  |
| 2.6     | Data Modeling Techniques                                      | T1                               |        |  | Blackboard               | 02 |  | CO2,L3  |
| 2.7     | Types of Data Models                                          | T1                               |        |  | Blackboard               | 01 |  | CO2,L3  |
| 2.8     | Missing Imputations                                           | R1                               |        |  | Blackboard               | 01 |  | CO2,L2  |
|         | Review                                                        | Signature of the HOD/Coordinator |        |  |                          |    |  |         |

| Unit/<br>Item<br>No. | Topic(s)                                       | Book<br>Reference | Page (s) |  | Teaching<br>Methodology  | Proposed<br>No. of<br>Periods | Actual<br>Date of<br>Handled | CO/RBT |
|----------------------|------------------------------------------------|-------------------|----------|--|--------------------------|-------------------------------|------------------------------|--------|
| UNIT-III             |                                                |                   |          |  |                          |                               |                              |        |
| III                  | Regression                                     |                   |          |  |                          | 13                            |                              |        |
| 3.1                  | Basic Concepts of Regression                   | T1                |          |  | Seminar                  | 02                            |                              | CO3,L3 |
| 3.2                  | Blue property assumptions                      | T1                |          |  | Blackboard               | 01                            |                              | CO3,L4 |
| 3.3                  | The Gauss-Markov Assumptions in Algebra        | T1                |          |  | Blackboard               | 02                            |                              | CO3,L3 |
| 3.4                  | Variable Rationalization, and Model Building   | T1                |          |  | Blackboard               | 02                            |                              | CO3,L4 |
| 3.5                  | Least Square Estimation or Least Square Method | T1                |          |  | Blackboard               | 02                            |                              | CO3,L4 |
| 3.6                  | Logistic Regression                            | T1                |          |  | Power point Presentation | 01                            |                              | CO4,L4 |

|      |                                                                     |                                  |  |  |                          |    |  |         |
|------|---------------------------------------------------------------------|----------------------------------|--|--|--------------------------|----|--|---------|
| 3.7  | Model fit Statistics                                                | T1                               |  |  | Blackboard               | 01 |  | CO4,L4  |
| 3.8  | Model Construction                                                  | T1                               |  |  | Blackboard               | 01 |  | CO4,L3  |
| 3.9  | Analytics applications to various Business Domains                  | W3                               |  |  | Blackboard               | 01 |  | CO4,L3  |
|      | Review                                                              | Signature of the HOD/Coordinator |  |  |                          |    |  |         |
|      | UNIT-IV                                                             |                                  |  |  |                          |    |  |         |
| IV   | Object Segmentation                                                 |                                  |  |  |                          | 13 |  |         |
| 4.1  | Regression Vs Segmentation                                          | T1                               |  |  | Blackboard               | 01 |  | CO5,L2  |
| 4.2  | Logistic regression                                                 | T1                               |  |  | Blackboard               | 02 |  | CO5,L2  |
| 4.3  | Supervised and Unsupervised Learning                                | T1                               |  |  | Blackboard               | 01 |  | CO5, L4 |
| 4.4  | Tree Building                                                       | T1                               |  |  | Blackboard               | 02 |  | CO5,L4  |
| 4.5  | CHAID                                                               | T1                               |  |  | Blackboard               | 01 |  | CO5,L4  |
| 4.6  | CART (Classification And Regression Tree)                           | W4                               |  |  | Power point Presentation | 02 |  | CO5,L2  |
| 4.7  | Model overfitting, ARIMA (Autoregressive Integrated Moving Average) | T1                               |  |  | Blackboard               | 01 |  | CO5,L4  |
| 4.8  | Measures of Forecast Accuracy, STL Approach                         | T1                               |  |  | Blackboard               | 01 |  | CO5,L2  |
| 4.9  | Extract features from generated model as Height.                    | T1                               |  |  | Blackboard               | 01 |  | CO5,L2  |
| 4.10 | Average Energy, Analyze for prediction.                             | T1                               |  |  | Seminars                 | 01 |  | CO5,L4  |
|      | Review                                                              | Signature of the HOD/Coordinator |  |  |                          |    |  |         |
| V    | Data Visualization                                                  |                                  |  |  |                          | 09 |  |         |
| 5.1  | Introduction to Data Visualization                                  | W5                               |  |  | Blackboard               | 01 |  | CO6,L3  |
| 5.2  | Pixel-Oriented Visualization Techniques                             | T1                               |  |  | Power point Presentation | 01 |  | CO6,L4  |
| 5.3  | Geometric Projection Visualization Techniques                       | T1                               |  |  | Power point Presentation | 02 |  | CO6,L4  |
| 5.4  | Icon-Based Visualization Techniques                                 | T1                               |  |  | Blackboard               | 01 |  | CO6,L6  |
| 5.5  | Hierarchical Visualization Techniques                               | T1                               |  |  | Blackboard               | 01 |  | CO6,L6  |
| 5.6  | Visualizing Complex Data and Relations.                             | T1                               |  |  | Blackboard               | 01 |  | CO6,L6  |
| 5.7  | Real Time Project.                                                  |                                  |  |  | Power Point Presentation | 02 |  | CO6,L6  |
|      | Review                                                              | Signature of the HOD/Coordinator |  |  |                          |    |  |         |

## **LIST OF TEXTBOOKS AND REFERENCES**

### **TEXTBOOKS:**

1. Data Mining Concepts and Techniques, Han, Kamber, 3<sup>rd</sup> Edition, Morgan Kaufmann Publishers.
2. Big Data Analytics, Seema Acharya, Subhasini Chellappan, Wiley 2015.

### **REFERENCEBOOK:**

1. Data Mining Analysis and Concepts, M.Zaki and W.Meira
2. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business, Michael Minelli, Michehe Chambers, 1<sup>st</sup> Edition, Ambiga Dhiraj, WileyCIOSeries, 2013.
3. Introduction to Data Mining, Tan, Stein bach and Kumar, Addison Wisley, 2006.
4. Mining of Massive Datasets, Jure Leskovec Stanford Univ. Anand Rajaraman Millway Labs Jeffrey D Ullman Stanford Univ.

### **WEBLINKS**

W1: <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/>

W2 : <https://nptel.ac.in/courses/106107220>

W3: <https://www.simplilearn.com/business-analytics-applications-and-use-cases-article>

W4:

[http://ndl.iitkgp.ac.in/he\\_document/nptel/nptel/courses\\_106\\_107\\_106107220\\_video\\_lec60?e=0|CART%20\(Classification%20And%20Regression%20Tree\)](http://ndl.iitkgp.ac.in/he_document/nptel/nptel/courses_106_107_106107220_video_lec60?e=0|CART%20(Classification%20And%20Regression%20Tree))

W5: <https://www.geeksforgeeks.org/data-visualization-and-its-importance/>

**CONTENT BEYOND THE SYLLABUS**

| <b>S.No.</b> | <b>Topics</b>      | <b>Books&amp; Journals</b>                                                                                                                                                                       | <b>CourseOutcomes</b> |
|--------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| 02           | Machine Learning   | Mathematics for Machine Learning<br>Marc Peter Deisenroth, 2020<br><a href="#">Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies</a> | CO5                   |
| 01           | Data Visualization | Story Telling with Data, Wiley                                                                                                                                                                   | CO6                   |

**SELF STUDY TOPICS**

| <b>S.No.</b> | <b>Topics</b>                              | <b>Books&amp; Journals</b>                                   | <b>CourseOutcomes</b> |
|--------------|--------------------------------------------|--------------------------------------------------------------|-----------------------|
| 01           | Statistical Foundations for Data Analytics | Statistics for Business and Economics by Paul Newbold et al. | CO1                   |
| 02           | Tools & Technologies in Data Analytics     | Learning SQL by Alan Beaulieu                                | CO6                   |



## ASSIGNMENT – I

SRIINDUCOLLEGE OF ENGINEERING & TECHNOLOGY, Sheriguda(V), R.R. Dist.  
(An Autonomous Institution under UGC)

DEPARTMENT OF AIML

YEAR/SEMESTER: III YEAR/I SEM A.Y:2025-26

SUBJECT NAME: (R22CSE3141) & DATA ANALYTICS

ASSIGNMENT-I (AIML-C)

| Batch. No | Assignment Questions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Register Number                | Course outcome                  | Level                      |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|---------------------------------|----------------------------|
| 1.        | <ol style="list-style-type: none"><li>What is data? How to handle large collections of data? Discuss the various sources of data.</li><li>Explain the key steps involved in data preprocessing and discuss their importance in ensuring high-quality data for analysis.</li><li>Define Data Analytics and explain its importance in business decision-making.</li><li>Apply data visualization techniques to summarize a dataset and provide business insights.</li><li>Explain the key concepts of regression and its importance in predictive modeling.</li></ol> | 23D41A66C9<br>To<br>23D41A66D3 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L1<br>L2<br>L3<br>L5<br>L2 |
| 2.        | <ol style="list-style-type: none"><li>List the common sources of data such as sensors, signals, and GPS, and briefly describe their applications.</li><li>List out the characteristics of data quality. Explain them in brief.</li><li>List and describe different types of data used in analytics.</li><li>Describe the process of missing data imputation and explain why it is necessary.</li><li>What are the BLUE (Best Linear Unbiased Estimator) properties in regression? Explain their significance.</li></ol>                                             | 24D41A66D4<br>To<br>22D41A66D8 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L3<br>L1<br>L1<br>L2<br>L1 |
| 3.        | <ol style="list-style-type: none"><li>Define the concept of data architecture and its importance in managing data for analysis.</li><li>How does data normalization help in the analysis of sensor data?</li><li>Explain the difference between structured, semi-structured, and unstructured data.</li><li>Compare and contrast three popular data analytics tools (e.g., Python, R, and Excel) in terms of usability, features, and application areas.</li><li>What is variable rationalization? How does it impact model building?</li></ol>                     | 22D41A66D9<br>To<br>22D41A66E3 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L3<br>L1<br>L2<br>L4<br>L1 |
| 4.        | <ol style="list-style-type: none"><li>Consider a dataset with missing values, demonstrate how to handle these missing values using different imputation techniques.</li><li>List out the characteristics of data quality. Explain them in brief.</li><li>Differentiate between descriptive, predictive, and prescriptive analytics with examples.</li><li>Illustrate how data analytics can be applied to a retail business to improve sales and customer satisfaction.</li><li>Explain the step-by-step process of building a regression model.</li></ol>          | 22D41A66E4<br>To<br>22D41A66E8 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L2<br>L1<br>L3<br>L1<br>L1 |
| 5.        | <ol style="list-style-type: none"><li>Compare the challenges posed by noise and outliers in a dataset, and suggest methods to address each issue.</li><li>What is data? How to handle large collections of data? Discuss the various sources of data.</li><li>Explain the role of databases in data analytics and discuss different types of databases.</li><li>Examine the strengths and weaknesses of various data imputation</li></ol>                                                                                                                           | 23D41A66E9<br>To<br>23D41A66F3 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L4<br>L1<br>L2<br>L2       |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                               |                                                        |                                                   |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|--------------------------------------------------------|---------------------------------------------------|
|     | <p>methods in handling missing data.</p> <p>5. What are the key assumptions of linear regression? How do violations impact model accuracy?</p>                                                                                                                                                                                                                                                                                                                                                                                  |                                               |                                                        | L1                                                |
| 6.  | <p>1. Distinguish between primary data and secondary data in brief.</p> <p>2. How does data normalization help in the analysis of sensor data?</p> <p>3. Discuss the importance of data modelling techniques in business applications.</p> <p>4. Define Data Analytics and explain its importance in business decision-making.</p> <p>5. Explain multicollinearity in regression models. How can it be detected and addressed?</p>                                                                                              | <p>23D41A66F4</p> <p>To</p> <p>23D41A66F8</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L2</p> <p>L1</p> <p>L1</p> <p>L3</p> <p>L1</p> |
| 7.  | <p>1. List out the characteristics of data quality. Explain them in brief.</p> <p>2. What is data? How to handle large collections of data? Discuss the various sources of data.</p> <p>3. Apply data visualization techniques to summarize a dataset and provide business insights.</p> <p>4. List and describe different types of data used in analytics.</p> <p>5. Explain the key concepts of regression and its importance in predictive modeling.</p>                                                                     | <p>23D41A66F9</p> <p>To</p> <p>23D41A66G3</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L1</p> <p>L1</p> <p>L5</p> <p>L1</p> <p>L2</p> |
| 8.  | <p>1. What is data preprocessing. List out the steps in data preprocessing.</p> <p>2. Distinguish between primary data and secondary data in brief.</p> <p>3. Analyze the impact of poor data quality on business decision-making and performance.</p> <p>4. Discuss the importance of business modeling in strategic decision-making. Support your answer with relevant examples.</p> <p>5. What are the BLUE (Best Linear Unbiased Estimator) properties in regression? Explain their significance.</p>                       | <p>23D41A66G4</p> <p>To</p> <p>23D41A66G8</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L1</p> <p>L2</p> <p>L5</p> <p>L1</p> <p>L1</p> |
| 9.  | <p>1. Describe the process of data preprocessing and its significance in preparing data for machine learning and analytical models.</p> <p>2. Compare the challenges posed by noise and outliers in a dataset, and suggest methods to address each issue.</p> <p>3. Describe the process of missing data imputation and explain why it is necessary.</p> <p>4. Define Data Analytics and explain its importance in business decision-making.</p> <p>5. What is variable rationalization? How does it impact model building?</p> | <p>23D41A66G9</p> <p>To</p> <p>23D41A66H3</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L2</p> <p>L4</p> <p>L2</p> <p>L3</p> <p>L2</p> |
| 10. | <p>1. How does data normalization help in the analysis of sensor data?</p> <p>2. List the common sources of data such as sensors, signals, and GPS, and briefly describe their applications.</p> <p>3. Examine the strengths and weaknesses of various data imputation methods in handling missing data.</p> <p>4. Illustrate how data analytics can be applied to a retail business to improve sales and customer satisfaction.</p> <p>5. What is variable rationalization? How does it impact model building?</p>             | <p>23D41A66H4</p> <p>To</p> <p>23D41A66H9</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L1</p> <p>L1</p> <p>L2</p> <p>L1</p> <p>L1</p> |
| 11. | <p>1. How can you identify and remove duplicate data entries in a dataset?</p> <p>2. Distinguish between primary data and secondary data in brief.</p> <p>3. Analyze the role of primary keys and foreign keys in database design. How do they ensure data integrity and relationships?</p> <p>4. Illustrate how data analytics can be applied to a retail business to improve sales and customer satisfaction.</p> <p>5. Explain the step-by-step process of building a regression model.</p>                                  | <p>23D41A66J0</p> <p>To</p> <p>23D41A66J4</p> | <p>CO1</p> <p>CO1</p> <p>CO2</p> <p>CO2</p> <p>CO3</p> | <p>L1</p> <p>L2</p> <p>L5</p> <p>L1</p> <p>L2</p> |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                    |                                 |                                  |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|---------------------------------|----------------------------------|
|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                    |                                 |                                  |
| 12. | <ol style="list-style-type: none"> <li>1. Explain the key steps involved in data preprocessing and discuss their importance in ensuring high-quality data for analysis.</li> <li>2. Describe the process of data preprocessing and its significance in preparing data for machine learning and analytical models.</li> <li>3. Discuss the importance of business modeling in strategic decision-making. Support your answer with relevant examples.</li> <li>4. Explain the role of databases in data analytics and discuss different types of databases.</li> <li>5. Explain multicollinearity in regression models. How can it be detected and addressed?</li> </ol> | 23D41A66J5<br><br>To<br><br>23D41A66J9                             | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L2<br>L2<br>L1<br>L2<br>L2       |
| 13. | <ol style="list-style-type: none"> <li>1. What is data preprocessing. List out the steps in data preprocessing.</li> <li>2. Define the concept of data architecture and its importance in managing data for analysis.</li> <li>3. Illustrate how data analytics can be applied to a retail business to improve sales and customer satisfaction.</li> <li>4. Examine the strengths and weaknesses of various data imputation methods in handling missing data.</li> <li>5. What are the key assumptions of linear regression? How do violations impact model accuracy?</li> </ol>                                                                                       | 23D41A66K0<br>23D41A66K1<br>23D41A66K2<br>22D41A66B2<br>22D41A6608 | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L1<br>L3<br>L1<br>L2<br>L1       |
| 14. | <ol style="list-style-type: none"> <li>1. List out the characteristics of data quality. Explain them in brief.</li> <li>2. Distinguish between primary data and secondary data in brief.</li> <li>3. Compare and contrast three popular data analytics tools (e.g., Python, R, and Excel) in terms of usability, features, and application areas.</li> <li>4. List and describe different types of data used in analytics.</li> <li>5. Explain multicollinearity in regression models. How can it be detected and addressed?</li> </ol>                                                                                                                                | 24D45A6608<br>24D45A6609<br>24D45A6610                             | CO1<br>CO1<br>CO2<br>CO2<br>CO3 | L2<br>L2<br>L2<br>L4<br>L1<br>L2 |

FACULTY

HOD



DEPARTMENT OF AIML  
YEAR / SEMESTER: III YEAR / II SEM     A.Y:2025-26  
SUBJECT NAME: (R22CSE3141) & DATA ANALYTICS  
ASSIGNMENT- II (AIML-C)

| Batch.<br>No | Assignment Questions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Register<br>Number                 | Course<br>outcome                   | Level                          |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|-------------------------------------|--------------------------------|
| 1.           | 6. Discuss about the following Model Fit Statistics:i) Hosmer Lemeshow Test ii) Error Matrix.<br>7. Explain the differences between object segmentation and regression in image processing.<br>8. Describe how extracted features from a model can be analyzed for better predictions.<br>9. Explain pixel-oriented visualization techniques and their advantages in handling large datasets.<br>5. Write down steps involved in Data Visualization in Tableau.                                                                                | 23D41A66C9<br><br>To<br>23D41A66D3 | CO4<br><br>CO5<br>CO5<br>CO6<br>CO6 | L6<br><br>L2<br>L2<br>L2<br>L2 |
| 2.           | 1. Discuss model fit statistics in logistic regression, including AIC, BIC, and likelihood ratio tests.<br>2. Compare and contrast supervised and unsupervised learning with real-world applications.<br>3. How does complexity impact machine learning models? Discuss with reference to bias-variance tradeoff.<br>4. Describe different geometric projection visualization techniques with suitable examples.<br>5. Evaluate an in-depth analysis of how pixel-oriented visualization is applied in time-series data.                       | 24D41A66D4<br><br>To<br>22D41A66D8 | ^1                                  | L5<br><br>L5<br>L1<br>L4<br>L5 |
| 3.           | 1. Explain the process of constructing a logistic regression model with real-world applications.<br>2. Describe the tree-building process in decision trees and explain how it helps in regression and classification tasks.<br>3. Describe how extracted features from a model can be analyzed for better predictions.<br>4. What are icon-based visualization techniques? Provide use cases where they are beneficial.<br>5. How can complex data and relationships be visualized effectively? Give real-world examples.                     | 22D41A66D9<br><br>To<br>22D41A66E3 | CO4<br>CO5<br>CO5<br>CO6<br>CO6     | L1<br>L2<br>L2<br>L1<br>L1     |
| 4.           | 6. How logistic regression is used in business analytics applications such as fraud detection and customer segmentation?<br>7. What is overfitting in machine learning models? How does pruning help in reducing overfitting?<br>8. Compare and contrast supervised and unsupervised learning with real-world applications.<br>9. Explain hierarchical visualization techniques and discuss their significance in data representation.<br>5. Describe how hierarchical visualization can be used for representing large-scale structured data. | 22D41A66E4<br><br>To<br>22D41A66E8 | CO4<br>CO5<br>CO5<br>CO6<br>CO6     | L1<br><br>L1<br>L5<br>L2<br>L4 |

|    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                |                                 |                            |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|---------------------------------|----------------------------|
| 5. | <ol style="list-style-type: none"> <li>1. What are the different types of link functions in logistic regression? Explain their use cases.</li> <li>2. Explain complexity in decision trees and discuss methods to control it.</li> <li>3. Compare decision trees with other predictive modeling techniques in terms of accuracy and interpretability.</li> <li>4. How can complex data and relationships be visualized effectively? Give real-world examples.</li> <li>5. Describe different geometric projection visualization techniques with suitable examples.</li> </ol>                                                | 23D41A66E9<br>To<br>23D41A66F3 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L5<br>L1<br>L4 |
| 6. | <ol style="list-style-type: none"> <li>6. Discuss the evaluation metrics for logistic regression models, such as accuracy, precision, recall, and AUC-ROC.</li> <li>7. Describe multiple decision trees and how they contribute to ensemble learning techniques like Random Forest.</li> <li>8. Compare decision trees with other predictive modeling techniques in terms of accuracy and interpretability.</li> <li>9. Differentiate between pixel-oriented and icon-based visualization techniques.</li> <li>10. Compare and contrast different visualization techniques for large, multi-dimensional datasets.</li> </ol> | 23D41A66F4<br>To<br>23D41A66F8 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L5<br>L4<br>L5 |
| 7. | <ol style="list-style-type: none"> <li>6. Explain how feature selection techniques, such as stepwise regression and LASSO, improve model performance.</li> <li>7. Explain the ARIMA model in time series forecasting. What are its key components?</li> <li>8. Explain the differences between object segmentation and regression in image processing.</li> <li>9. Discuss the strengths and limitations of geometric projection techniques in data visualization.</li> <li>10. Explain hierarchical visualization techniques and discuss their significance in data representation.</li> </ol>                              | 23D41A66F9<br>To<br>23D41A66G3 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L2<br>L6<br>L2 |
| 8. | <ol style="list-style-type: none"> <li>6. Discuss the evaluation metrics for logistic regression models, such as accuracy, precision, recall, and AUC-ROC.</li> <li>7. What are the different measures of forecast accuracy? Explain with examples.</li> <li>8. What is Overfitting? How to Prevent Overfitting? Compare and Contrast between ARMA and ARIMA.</li> <li>9. Describe how hierarchical visualization can be used for representing large-scale structured data.</li> <li>10. How can complex data and relationships be visualized effectively? Give real-world examples.</li> </ol>                              | 23D41A66G4<br>To<br>23D41A66G8 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L1<br>L1<br>L4<br>L1 |
| 9. | <ol style="list-style-type: none"> <li>6. How logistic regression is used in business analytics applications such as fraud detection and customer segmentation?</li> <li>7. Explain the STL (Seasonal-Trend Decomposition) approach in time series analysis and its significance.</li> <li>8. Explain the differences between object segmentation and regression in image processing.</li> <li>9. Demonstrate Geometric projection visualization techniques. List out various applications of data visualization.</li> </ol>                                                                                                 | 23D41A66G9<br>To<br>23D41A66H3 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L2<br>L4<br>L4 |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                    |                                 |                            |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|---------------------------------|----------------------------|
|     | 10. Describe how hierarchical visualization can be used for representing large-scale structured data.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                    |                                 |                            |
| 10. | <ol style="list-style-type: none"> <li>1. Discuss model fit statistics in logistic regression, including AIC, BIC, and likelihood ratio tests.</li> <li>2. How can features such as height and average energy be extracted from a generated model?</li> <li>3. What is over fitting in machine learning models? How does pruning help in reducing over fitting?</li> <li>4. Compare and contrast different visualization techniques for large, multi-dimensional datasets.</li> <li>5. Discuss the strengths and limitations of geometric projection techniques in data visualization.</li> </ol>                               | 23D41A66H4<br>To<br>23D41A66H9                                     | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L5<br>L1<br>L1<br>L5<br>L6 |
| 11. | <ol style="list-style-type: none"> <li>1. What are the different types of link functions in logistic regression? Explain their use cases.</li> <li>2. Describe how extracted features from a model can be analyzed for better predictions.</li> <li>3. Compare and contrast supervised and unsupervised learning with real-world applications.</li> <li>4. How does interactive visualization improve the interpretation of complex data relationships?</li> <li>5. Discuss how data visualization principles align with cognitive psychology to improve understanding.</li> </ol>                                              | 23D41A66J0<br>To<br>23D41A66J4                                     | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L5<br>L1<br>L6 |
| 12. | <ol style="list-style-type: none"> <li>1. Explain the process of constructing a logistic regression model with real-world applications.</li> <li>2. Compare decision trees with other predictive modeling techniques in terms of accuracy and interpretability.</li> <li>3. Describe the tree-building process in decision trees and explain how it helps in regression and classification tasks.</li> <li>4. Write down steps involved in Data Visualization In Tableau.</li> <li>5. How can complex data and relationships be visualized effectively? Give real-world examples.</li> </ol>                                    | 23D41A66J5<br>To<br>23D41A66J9                                     | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L5<br>L2<br>L2<br>L1 |
| 13. | <ol style="list-style-type: none"> <li>6. Explain how feature selection techniques, such as stepwise regression and LASSO, improve model performance.</li> <li>7. Explain the role of pruning in decision trees and differentiate between pre-pruning and post-pruning techniques.</li> <li>8. Compare and contrast supervised and unsupervised learning with real-world applications.</li> <li>9. Discuss how data visualization principles align with cognitive psychology to improve understanding.</li> <li>10. Explain pixel-oriented visualization techniques and their advantages in handling large datasets.</li> </ol> | 23D41A66K0<br>23D41A66K1<br>23D41A66K2<br>22D41A66B2<br>22D41A66O8 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L1<br>L2<br>L2<br>L6<br>L2 |

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                        |                                 |                            |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|---------------------------------|----------------------------|
| 14. | 1. Discuss about the following Model Fit Statistics:i) Hosmer Lemeshow Test ii) Error Matrix.<br>2. How does complexity impact machine learning models? Discuss with reference to bias-variance tradeoff.<br>3. What is overfitting in machine learning models? How does pruning help in reducing overfitting?<br>4. Evaluate an in-depth analysis of how pixel-oriented visualization is applied in time-series data.<br>5. How to perform visualization of the data using a hierarchical partitioning into subspaces? Explain with examples . | 24D45A6608<br>24D45A6609<br>24D45A6610 | CO4<br>CO5<br>CO5<br>CO6<br>CO6 | L6<br>L1<br>L1<br>L5<br>L1 |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|---------------------------------|----------------------------|

FACULTY

HOD

|                                                                                    |                                                                                                                                                                     |                             |                                                                        |
|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|------------------------------------------------------------------------|
|  | <b>SRI INDU COLLEGE OF ENGG &amp; TECH</b><br><b>QUESTION BANK</b><br><b>(Regulation: R22)</b><br><b>Department of Artificial Intelligence and Machine Learning</b> |                             | <b>(Regulation: R22)</b><br><b>Prepared on : 30.12.25</b>              |
|                                                                                    | Sub. Code & Title                                                                                                                                                   | R22CSE3141 & DATA ANALYTICS |                                                                        |
|                                                                                    | Academic Year: 2025 - 26                                                                                                                                            | Year/Sem                    | III/II - AIML                                                          |
|                                                                                    | Faculty Name & Designation                                                                                                                                          |                             | MR.DR.M. NARESH ASSOCIATE PROFESSOR, MR.M. AVINASH ASSISTANT PROFESSOR |

QUESTION BANK

**Data Analytics**  
**QUESTION BANK WITH BLOOMS TAXONOMY LEVEL (BTL)**  
**(1. Remembering 2. Understanding 3. Applying 4. Analyzing 5. Evaluating 6. Creating)**

| <b>UNIT – I: DATA MANAGEMENT</b> |                                                                                                                                                                                               |                 |                       |
|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|-----------------------|
| <b>SNO</b>                       | <b>Multiple Choice Questions</b>                                                                                                                                                              | <b>BT LEVEL</b> | <b>Course Outcome</b> |
| 1C-1                             | What is the primary goal of data architecture?<br>A. Visualizing data trends<br>B. Structuring and managing data efficiently<br>C. Cleaning noisy data<br>D. Building machine learning models | L1              | CO1                   |
| 1C-2                             | Which of the following is NOT a type of data source?<br>A. Sensors<br>B. Signals<br>C. GPS<br>D. Hyperlinks                                                                                   | L1              | CO1                   |
| 1C-3                             | GPS data primarily includes:<br>A. Temperature and pressure readings<br>B. Latitude, longitude, and altitude<br>C. Textual descriptions<br>D. Signal strength                                 | L3              | CO1                   |
| 1C-4                             | What type of data do sensors typically produce?<br>A. Text<br>B. Real-time numerical data<br>C. Video streams<br>D. Images                                                                    | L1              | CO1                   |
| 1C-5                             | Which layer of a data architecture is responsible for raw data collection?<br>A. Presentation layer<br>B. Data ingestion layer<br>C. Transformation layer<br>D. Analytical layer              | L1              | CO1                   |
| 1C-6                             | The process of integrating data from multiple sources is called:<br>A. Data migration<br>B. Data integration<br>C. Data visualization<br>D. Data mining                                       | L4              | CO1                   |
| 1C-7                             | Which of the following is a method for handling missing data?<br>A. Clustering<br>B. Imputation<br>C. Aggregation<br>D. Replication                                                           | L1              | CO1                   |
| 1C-8                             | What is the first step in a typical data pipeline?<br>A. Data storage<br>B. Data analysis<br>C. Data collection<br>D. Data visualization                                                      | L1              | CO1                   |
| 1C-9                             | Duplicate data in a dataset is also referred to as:<br>A. Noisy data<br>B. Redundant data                                                                                                     | L3              | CO1                   |

|                    |                                                                                                                                                                                                 |    |     |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
|                    | C. Erroneous data<br>D. Structured data                                                                                                                                                         |    |     |
| 1C-10              | A dataset containing irrelevant or random variations is said to have:<br>A. Outliers<br>B. Noise<br>C. Missing values<br>D. Bias                                                                | L4 | CO1 |
| 1C-11              | Outliers in data can be detected using:<br>A. Clustering algorithms<br>B. Normalization techniques<br>C. Statistical methods like Z-scores<br>D. Text analysis                                  | L4 | CO1 |
| 1C-12              | What is a key characteristic of high-quality data?<br>A. It is always large in size<br>B. It is consistent, complete, and accurate<br>C. It has a lot of noise<br>D. It is unstructured and raw | L1 | CO1 |
| 1C-13              | The interquartile range (IQR) method is used for:<br>A. Detecting missing values<br>B. Identifying duplicate data<br>C. Detecting outliers<br>D. Normalizing data                               | L2 | CO1 |
| 1C-14              | Which of the following is a common cause of noise in data?<br>A. User errors<br>B. Data compression<br>C. Inaccurate sensor readings<br>D. Lack of storage                                      | L1 | CO1 |
| 1C-15              | Normalization in data preprocessing ensures that:<br>A. All features have the same scale<br>B. Outliers are removed<br>C. Data is stored efficiently<br>D. Missing values are handled           | L5 | CO1 |
| Fill in the Blanks |                                                                                                                                                                                                 |    |     |
| 1F-1               | _____ provides a framework for managing and organizing data to ensure its accessibility and usability.                                                                                          | L1 | CO1 |
| 1F-2               | A data pipeline involves the collection, transformation, and _____ of data for analysis                                                                                                         | L1 | CO1 |
| 1F-3               | Data integrity ensures that data remains _____ and unaltered during its lifecycle.                                                                                                              | L1 | CO1 |
| 1F-4               | Removing noise from data often involves techniques like smoothing or _____.                                                                                                                     | L1 | CO1 |
| 1F-5               | Data quality dimensions include accuracy, completeness, and _____                                                                                                                               | L1 | CO1 |
| 1F-6               | Outliers are data points that deviate significantly from the _____ pattern.                                                                                                                     | L1 | CO1 |
| 1F-7               | _____ refers to random variations in data that can distort analysis results.                                                                                                                    | L1 | CO1 |
| 1F-8               | IoT devices like sensors generate _____ data in real-time.                                                                                                                                      | L1 | CO1 |
| 1F-9               | Duplicate data can lead to _____ in analysis results if not handled properly.                                                                                                                   | L1 | CO1 |

|                     |                                                                                                             |                                                                                                                                                                                                                         |                       |
|---------------------|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| 1F-10               | Data preprocessing involves cleaning, transforming, and _____ data for analysis.                            | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-11               | _____ is the term used for combining data from multiple sources into a unified format.                      | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-12               | A data processing pipeline often ends with storing data in a _____ or presenting it for analysis.           | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-13               | Sensors, signals, and GPS are examples of _____ data sources.                                               | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-14               | Data outliers can be detected using statistical methods like the _____ range rule.                          | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-15               | The process of identifying and fixing errors in a dataset is known as data _____.                           | L2                                                                                                                                                                                                                      | CO1                   |
| 1F-16               | Missing values in a dataset can be addressed using _____ or deletion techniques.                            | L2                                                                                                                                                                                                                      | CO1                   |
| Match the Following |                                                                                                             |                                                                                                                                                                                                                         |                       |
| 1M-1                | 1. Data Architecture<br>2. GPS<br>3. Sensors<br>4. Signals                                                  | a. Provides location data using satellites<br>b. Represents data transmitted through waves<br>c. Framework for managing and organizing data<br>d. Devices that collect real-time numerical data                         | L1<br><br><br><br>CO1 |
| 1M-2                | 1. Normalization<br>2. Imputation<br>3. Smoothing<br>4. One-hot encoding                                    | a. Converting categorical data into binary features<br>b. Reducing noise in data<br>c. Filling in missing values<br>d. Scaling data to a common range                                                                   | L1<br><br><br><br>CO1 |
| 1M-3                | 1. Noise<br>2. Outliers<br>3. Missing values<br>4. Duplicate data                                           | a. Absence of data points in a dataset<br>b. Random variations in data affecting accuracy<br>c. Extreme values that deviate from the expected pattern<br>d. Absence of data points in a dataset                         | L1<br><br><br><br>CO1 |
| 1M-4                | 1. Data preprocessing<br>2. Data storage<br>3. Data transformation<br>4. Data analysis                      | a. Extracting insights from processed data<br>b. Converting raw data into an analysis-ready format<br>c. Saving data for future retrieval and use<br>d. Cleaning, transforming, and integrating data                    | L2<br><br><br><br>CO1 |
| 1M-5                | 1. Data integration<br>2. Data pipeline<br>3. Data quality<br>4. Data cleaning                              | a. End-to-end process for data collection and analysis<br>b. Measures like accuracy, consistency, and completeness<br>c. End-to-end process for data collection and analysis<br>d. Combining data from multiple sources | L2<br><br><br><br>CO1 |
| Questions           |                                                                                                             |                                                                                                                                                                                                                         |                       |
| 1D-1                | What is data? How to handle large collections of data? Discuss the various sources of data.                 | L1                                                                                                                                                                                                                      | CO1                   |
| 1D-2                | List the common sources of data such as sensors, signals, and GPS, and briefly describe their applications. | L1                                                                                                                                                                                                                      | CO1                   |
| 1D-3                | Define the concept of data architecture and its importance in managing data for analysis.                   | L1                                                                                                                                                                                                                      | CO1                   |

|        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |        |     |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----|-----|-----|-----|-----|----|--------|-----|-----|-----|-----|-----|-----|--------|-----|-----|-----|----|-----|-----|-----|-----|-----|----|-----|
| 1D-4   | Consider a dataset with missing values, explain how to handle these missing values using different imputation techniques.                                                                                                                                                                                                                                                                                                                                                                                                                    | L6     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-5   | Compare the challenges posed by noise and outliers in a dataset, and suggest methods to address each issue.                                                                                                                                                                                                                                                                                                                                                                                                                                  | L4     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-6   | Distinguish between primary data and secondary data in brief.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | L4     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-7   | List out the characteristics of data quality. Explain them in brief.                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | L1     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-8   | What is data preprocessing. List out the steps in data preprocessing.                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | L1     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-9   | Describe the process of data preprocessing and its significance in preparing data for machine learning and analytical models.                                                                                                                                                                                                                                                                                                                                                                                                                | L2     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-10  | How does data normalization help in the analysis of sensor data?                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | L1     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-11  | How can you identify and remove duplicate data entries in a dataset?                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | L1     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-12  | Design a data processing pipeline for a smart city project that involves collecting and analyzing data from multiple sources like traffic sensors, weather stations, and GPS systems. Include steps for data cleaning, integration, and preparation for analysis.                                                                                                                                                                                                                                                                            | L6     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-13  | <div>Three different kinds of food are tested on three groups of rats for 5 weeks. The objective is to check the difference in mean weight(in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to the following data:</div> <table><tr><td>Food 1</td><td>8</td><td>12</td><td>19</td><td>8</td><td>6</td><td>11</td></tr><tr><td>Food 2</td><td>4</td><td>5</td><td>4</td><td>6</td><td>9</td><td>7</td></tr><tr><td>Food 3</td><td>11</td><td>8</td><td>7</td><td>13</td><td>7</td><td>9</td></tr></table> | Food 1 | 8   | 12  | 19  | 8   | 6   | 11 | Food 2 | 4   | 5   | 4   | 6   | 9   | 7   | Food 3 | 11  | 8   | 7   | 13 | 7   | 9   | L3  | CO1 |     |    |     |
| Food 1 | 8                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 12     | 19  | 8   | 6   | 11  |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| Food 2 | 4                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 5      | 4   | 6   | 9   | 7   |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| Food 3 | 11                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 8      | 7   | 13  | 7   | 9   |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-14  | <div>The following data show the number of worms quarantined from the GI areas of four groups of muskrats in a carbon tetrachloride anthelmintic study. Conduct a one-way ANOVA test.</div> <table><tr><td>I</td><td>338</td><td>324</td><td>268</td><td>147</td><td>309</td></tr><tr><td>II</td><td>412</td><td>387</td><td>400</td><td>233</td><td>212</td></tr><tr><td>III</td><td>124</td><td>353</td><td>469</td><td>222</td><td>111</td></tr><tr><td>IV</td><td>389</td><td>432</td><td>255</td><td>133</td><td>265</td></tr></table>  | I      | 338 | 324 | 268 | 147 | 309 | II | 412    | 387 | 400 | 233 | 212 | III | 124 | 353    | 469 | 222 | 111 | IV | 389 | 432 | 255 | 133 | 265 | L3 | CO1 |
| I      | 338                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 324    | 268 | 147 | 309 |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| II     | 412                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 387    | 400 | 233 | 212 |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| III    | 124                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 353    | 469 | 222 | 111 |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| IV     | 389                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 432    | 255 | 133 | 265 |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |
| 1D-15  | Explain the key steps involved in data preprocessing and discuss their importance in ensuring high-quality data for analysis.                                                                                                                                                                                                                                                                                                                                                                                                                | L1     | CO1 |     |     |     |     |    |        |     |     |     |     |     |     |        |     |     |     |    |     |     |     |     |     |    |     |

#### UNIT – II : Data Analytics

| SNO  | Multiple Choice Questions                                                                                                                                                                                       | BT LEVEL | Course Outcome |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------|
| 2C-1 | <p>Which type of analytics focuses on analyzing historical data to understand trends?</p> <p>A. Predictive Analytics<br/>B. Descriptive Analytics<br/>C. Prescriptive Analytics<br/>D. Diagnostic Analytics</p> | L1       | CO2            |
| 2C-2 | <p>Predictive analytics is primarily used to:</p> <p>A. Summarize past performance<br/>B. Suggest the best course of action<br/>C. Predict future outcomes<br/>D. Clean large datasets</p>                      | L1       | CO2            |
| 2C-3 | <p>The primary goal of analytics is to:</p> <p>A. Store large amounts of data</p>                                                                                                                               | L1       | CO2            |

|       |                                                                                                                                                                                                                   |    |     |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
|       | B. Extract meaningful insights<br>C. Build complex databases<br>D. Automate reporting                                                                                                                             |    |     |
| 2C-4  | Which of the following is NOT a data analytics tool?<br>A. Python<br>B. Tableau<br>C. Hadoop<br>D. Adobe Photoshop                                                                                                | L1 | CO2 |
| 2C-5  | A Business Intelligence tool commonly used for creating dashboards is:<br>A. R<br>B. Excel<br>C. Power BI<br>D. SQL                                                                                               | L2 | CO2 |
| 2C-6  | Which technique is best suited for segmenting customers based on buying behavior?<br>A. Regression<br>B. Clustering<br>C. Classification<br>D. Time-series analysis                                               | L1 | CO2 |
| 2C-7  | Optimization models in business are designed to:<br>A. Maximize or minimize an objective function<br>B. Summarize historical data<br>C. Clean and preprocess data<br>D. Visualize relationships between variables | L1 | CO2 |
| 2C-8  | What type of data is represented by categories without a natural order?<br>A. Ordinal<br>B. Nominal<br>C. Interval<br>D. Ratio                                                                                    | L1 | CO2 |
| 2C-9  | Which key ensures a unique identifier for each row in a database table?<br><br>A. Foreign key<br>B. Composite key<br>C. Primary key<br>D. Unique key                                                              | L1 | CO2 |
| 2C-10 | Decision trees are primarily used for:<br>A. Analyzing unstructured data<br>B. Predicting outcomes based on rules<br>C. Reducing dimensionality<br>D. Segmenting data clusters                                    | L2 | CO2 |
| 2C-11 | Which technique replaces missing values with the mean or median of the variable?<br>A. List wise deletion<br>B. Imputation<br>C. Interpolation<br>D. Normalization                                                | L1 | CO2 |
| 2C-12 | Discrete data is:<br>A. Measured on a continuous scale<br>B. Countable and has specific values<br>C. Always numeric and ranked<br>D. Unstructured and text-based                                                  | L1 | CO2 |

|                     |                                                                                                                                                                                                                                                                                                                              |    |     |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 2C-13               | Which type of environment is used for large-scale data storage and processing?<br>A. Cloud-based<br>B. Local computing<br>C. Desktop applications<br>D. Spreadsheet software                                                                                                                                                 | L1 | CO2 |
| 2C-14               | In predictive modeling, the variable we aim to predict is called the:<br>A. Independent variable<br>B. Dependent variable<br>C. Categorical variable<br>D. Feature variable                                                                                                                                                  | L1 | CO2 |
| 2C-15               | Why is business modeling important for an organization?<br><br>A) It helps the company reduce its operational costs.<br>B) It helps in identifying business objectives and aligning strategies.<br>C) It provides a clear roadmap for product development.<br>D) It focuses solely on the financial aspects of the business. | L1 | CO2 |
| Fill in the Blanks  |                                                                                                                                                                                                                                                                                                                              |    |     |
| 2F-1                | Data analytics is the process of examining _____ to extract meaningful insights.                                                                                                                                                                                                                                             | L1 | CO2 |
| 2F-2                | _____ analytics focuses on using past data to predict future outcomes.                                                                                                                                                                                                                                                       | L1 | CO2 |
| 2F-3                | _____ is a popular programming language for data analysis and machine learning.                                                                                                                                                                                                                                              | L2 | CO2 |
| 2F-4                | The three types of analytics are descriptive, _____, and prescriptive analytics.                                                                                                                                                                                                                                             | L2 | CO2 |
| 2F-5                | In a data analytics environment, _____ tools like Tableau and Power BI are used for data visualization.                                                                                                                                                                                                                      | L1 | CO2 |
| 2F-6                | A cloud-based environment, such as _____, allows users to analyze data without needing local infrastructure.                                                                                                                                                                                                                 | L1 | CO2 |
| 2F-7                | In business modeling, _____ models are used to classify data into predefined categories.                                                                                                                                                                                                                                     | L1 | CO2 |
| 2F-8                | _____ modeling is widely used to optimize resource allocation in supply chain management                                                                                                                                                                                                                                     | L2 | CO2 |
| 2F-9                | A _____ is a collection of organized data that allows easy access, management, and updating                                                                                                                                                                                                                                  | L1 | CO2 |
| 2F-10               | Data that can be counted, such as the number of employees, is classified as _____ data.                                                                                                                                                                                                                                      | L1 | CO2 |
| 2F-11               | In databases, a _____ key uniquely identifies each record in a table                                                                                                                                                                                                                                                         | L2 | CO2 |
| 2F-12               | _____ trees are used to make decisions and predictions based on data attributes.                                                                                                                                                                                                                                             | L2 | CO2 |
| 2F-13               | Regression models estimate the relationship between a dependent variable and one or more _____ variables.                                                                                                                                                                                                                    | L2 | CO2 |
| 2F-14               | Missing data can be handled by _____, which fills gaps using averages, medians, or other methods.                                                                                                                                                                                                                            | L1 | CO2 |
| 2F-15               | A key benefit of business modeling is its ability to _____ uncertainties and assess risks.                                                                                                                                                                                                                                   | L1 | CO2 |
| Match the Following |                                                                                                                                                                                                                                                                                                                              |    |     |

|           |                                                                                                                                              |                                                                                                                                            |    |     |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 2M-1      | 1. Descriptive Analytics<br>2. Predictive Analysis<br>3. Prescriptive Analysis<br>4. Diagnostic Analysis                                     | a. Recommends actions based on data<br>b. Explains reasons for past events<br>c. Forecasts future outcomes<br>d. Summarizes past data      | L1 | CO2 |
| 2M-2      | 1. Tableau<br>2. Python<br>3. Hadoop<br>4. Google Cloud Platform                                                                             | a. Cloud-based analytics environment<br>b. Data visualization<br>c. Programming for analytics<br>d. Big data processing framework          | L2 | CO2 |
| 2M-3      | 1. Clustering<br>2. Regression<br>3. Optimization<br>4. Classification                                                                       | a. Resource allocation<br>b. Categorizing data into groups<br>c. Predicting continuous outcomes<br>d. Customer segmentation                | L1 | CO2 |
| 2M-4      | 1. Primary Key<br>2. Foreign Key<br>3. Structured Data<br>4. Unstructured Data                                                               | a. Text, images, videos<br>b. Organized in rows and columns<br>c. Links tables in a database<br>d. Uniquely identifies a row               | L1 | CO2 |
| 2M-5      | 1. Decision Trees<br>2. Mean Imputation<br>3. Time-Series Analysis<br>4. Neural Networks                                                     | a. Rule-based predictions<br>b. Replacing missing values with mean<br>c. Analyzing trends over time<br>d. Complex data pattern recognition | L1 | CO2 |
| Questions |                                                                                                                                              |                                                                                                                                            |    |     |
| 2D-1      | Define Data Analytics and explain its importance in business decision-making.                                                                |                                                                                                                                            | L1 | CO2 |
| 2D-2      | List and describe different types of data used in analytics.                                                                                 |                                                                                                                                            | L1 | CO2 |
| 2D-3      | Explain the difference between structured, semi-structured, and unstructured data.                                                           |                                                                                                                                            | L2 | CO2 |
| 2D-4      | Differentiate between descriptive, predictive, and prescriptive analytics with examples.                                                     |                                                                                                                                            | L4 | CO2 |
| 2D-5      | Explain the role of databases in data analytics and discuss different types of databases.                                                    |                                                                                                                                            | L2 | CO2 |
| 2D-6      | Discuss the importance of data modelling techniques in business applications.                                                                |                                                                                                                                            | L1 | CO2 |
| 2D-7      | Consider a dataset with missing values, demonstrate how you would handle missing data using different imputation techniques.                 |                                                                                                                                            | L2 | CO2 |
| 2D-8      | Apply data visualization techniques to summarize a dataset and provide business insights.                                                    |                                                                                                                                            | L3 | CO2 |
| 2D-9      | Analyze the impact of poor data quality on business decision-making and performance.                                                         |                                                                                                                                            | L4 | CO2 |
| 2D-10     | Describe the process of missing data imputation and explain why it is necessary.                                                             |                                                                                                                                            | L3 | CO2 |
| 2D-11     | Examine the strengths and weaknesses of various data imputation methods in handling missing data.                                            |                                                                                                                                            | L4 | CO2 |
| 2D-12     | Analyze the role of primary keys and foreign keys in database design. How do they ensure data integrity and relationships?                   |                                                                                                                                            | L4 | CO2 |
| 2D-13     | Discuss the importance of business modeling in strategic decision-making. Support your answer with relevant examples.                        |                                                                                                                                            | L1 | CO2 |
| 2D-14     | Illustrate how data analytics can be applied to a retail business to improve sales and customer satisfaction.                                |                                                                                                                                            | L2 | CO2 |
| 2D-15     | Compare and contrast three popular data analytics tools (e.g., Python, R, and Excel) in terms of usability, features, and application areas. |                                                                                                                                            | L5 | CO2 |

| UNIT – III : Regression |                                                                                                                                                                                            |          |                |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------|
| SNO                     | Multiple Choice Questions                                                                                                                                                                  | BT LEVEL | Course Outcome |
| 3C-1                    | Which of the following is NOT an assumption of linear regression?<br>A. Linearity<br>B. Homoscedasticity<br>C. Independence of errors<br>D. Multicollinearity                              | L1       | CO3            |
| 3C-2                    | Which property ensures that the least square estimators are unbiased and have minimum variance?<br>A. Normality<br>B. BLUE<br>C. Heteroscedasticity<br>D. Residual Analysis                | L1       | CO3            |
| 3C-3                    | What does the Least Squares method minimize in linear regression?<br>A. Sum of squared residuals<br>B. Absolute error<br>C. Mean squared error<br>D. Log-likelihood                        | L1       | CO3            |
| 3C-4                    | Which of the following is an example of a dependent variable in a regression model predicting house prices?<br>A. Square footage<br>B. Number of bedrooms<br>C. House price<br>D. Location | L1       | CO3            |
| 3C-5                    | Which technique is used to address multicollinearity in regression models?<br>A. Data normalization<br>B. Variance Inflation Factor (VIF)<br>C. Residual Analysis<br>D. Cross-validation   | L1       | CO3            |
| 3C-6                    | What type of problem does logistic regression solve?<br>A. Regression<br>B. Classification<br>C. Clustering<br>D. Dimensionality Reduction                                                 | L1       | CO3            |
| 3C-7                    | Which function is used in logistic regression to transform linear outputs into probabilities?<br>A. ReLU function<br>B. Softmax function<br>C. Sigmoid function<br>D. Linear function      | L1       | CO4            |
| 3C-8                    | Which metric is commonly used to evaluate the performance of logistic regression?<br>A. R-squared<br>B. Mean Absolute Error<br>C. AUC-ROC<br>D. Sum of Squares Error                       | L1       | CO4            |
| 3C-9                    | What is the range of predicted values in logistic regression?<br>A. $-\infty$ to $+\infty$                                                                                                 | L1       | CO4            |

|                    |                                                                                                                                                                                                                       |    |     |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
|                    | B. 0 to 1<br>C. -1 to 1<br>D. Any real number                                                                                                                                                                         |    |     |
| 3C-10              | What does the term “Odds Ratio” represent in logistic regression?<br>A. Ratio of two probabilities<br>B. Difference between two coefficients<br>C. Probability of an event occurring<br>D. Likelihood of residuals    | L1 | CO4 |
| 3C-11              | Which of the following is NOT a model fit statistic for logistic regression?<br>A. AIC<br>B. BIC<br>C. RMSE<br>D. Log-likelihood                                                                                      | L1 | CO4 |
| 3C-12              | Which regularization technique is used to prevent overfitting in logistic regression?<br>A. Ridge regression<br>B. LASSO regression<br>C. Both a and b<br>D. None of the above                                        | L1 | CO4 |
| 3C-13              | Which optimization algorithm is commonly used for training logistic regression models?<br>A. Newton-Raphson<br>B. Gradient Descent<br>C. Backpropagation<br>D. Expectation-Maximization                               | L1 | CO4 |
| 3C-14              | What does the confusion matrix help measure in logistic regression?<br>A. Bias in the model<br>B. Model interpretability<br>C. Classification performance<br>D. Overfitting                                           | L1 | CO4 |
| 3C-15              | In logistic regression, what happens when multicollinearity is high?<br>A. The model remains unaffected<br>B. Coefficient estimates become unstable<br>C. The model fit improves<br>D. The dependent variable changes | L1 | CO4 |
| Fill in the Blanks |                                                                                                                                                                                                                       |    |     |
| 3F-1               | The main goal of regression analysis is to establish a _____ between the dependent and independent variables.                                                                                                         | L1 | CO3 |
| 3F-2               | The _____ method is used to estimate the coefficients in linear regression by minimizing the sum of squared residuals.                                                                                                | L1 | CO3 |
| 3F-3               | The _____ property states that the Ordinary Least Squares (OLS) estimators are the Best Linear Unbiased Estimators.                                                                                                   | L1 | CO3 |
| 3F-4               | The presence of _____ in data violates the assumption of independence of errors and can lead to incorrect inferences.                                                                                                 | L1 | CO3 |
| 3F-5               | In multiple regression, adding unnecessary independent variables can lead to a problem called _____.                                                                                                                  | L1 | CO3 |
| 3F-6               | Logistic regression is mainly used when the dependent variable is _____.                                                                                                                                              | L2 | CO3 |
| 3F-7               | The acronym BLUE in regression stands for _____                                                                                                                                                                       | L1 | CO3 |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                            |    |     |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 3F-8                | The _____ function in logistic regression converts any real-valued number into a probability value between 0 and 1.                                                                                                                                                                                                                                                                                                        | L1 | CO4 |
| 3F-9                | The _____ statistic is used to assess the goodness of fit for a logistic regression model.                                                                                                                                                                                                                                                                                                                                 | L1 | CO4 |
| 3F-10               | The odds ratio in logistic regression is calculated as the exponentiation of the _____.                                                                                                                                                                                                                                                                                                                                    | L1 | CO4 |
| 3F-11               | When applying regression in business analytics, customer churn prediction is often modelled using _____.                                                                                                                                                                                                                                                                                                                   | L1 | CO4 |
| 3F-12               | A key difference between linear and logistic regression is that logistic regression predicts the _____ of an event occurring.                                                                                                                                                                                                                                                                                              | L1 | CO4 |
| 3F-13               | In logistic regression, the _____ test is used to evaluate the significance of each independent variable.                                                                                                                                                                                                                                                                                                                  | L2 | CO4 |
| 3F-14               | _____ helps in reducing the risk of overfitting in logistic regression models.                                                                                                                                                                                                                                                                                                                                             | L2 | CO4 |
| 3F-15               | In logistic regression, the dependent variable is usually _____ or categorical.                                                                                                                                                                                                                                                                                                                                            | L1 | CO4 |
| Match the Following |                                                                                                                                                                                                                                                                                                                                                                                                                            |    |     |
| 3M-1                | <div> <div>1. Least Squares Estimation</div> <div>2. Multicollinearity</div> <div>3. Homoscedasticity</div> <div>4. R – Squared</div> </div> <div> <div>a. Measures proportion of variance explained</div> <div>b. Constant variance of error terms</div> <div>c. High correlation among independent variables</div> <div>d. Minimizes sum of squares</div> </div>                                                         | L1 | CO3 |
| 3M-2                | <div> <div>1. Sigmoid Function</div> <div>2. Odds Ratio</div> <div>3. Maximum Likelihood Estimation</div> <div>4. Binary Outcome</div> </div> <div> <div>a. Dependent variable has two categories</div> <div>b. Method to estimate coefficients in logistic regression .</div> <div>c. Exponentiation of regression coefficient</div> <div>d. Converts output into probability</div> </div>                                | L2 | CO3 |
| 3M-3                | <div> <div>1. Customer churn Prediction</div> <div>2. Credit Risk Assessment</div> <div>3. Marketing campaign Response</div> <div>4. Fraud detection</div> </div> <div> <div>a. Identifies suspicious financial transactions</div> <div>b. Predicts Customer Engagement with advertisements</div> <div>c. Identifies customers likely to leave a service</div> <div>d. Determines probability of loan default</div> </div> | L1 | CO3 |
| 3M-4                | <div> <div>1. BLUE Property</div> <div>2. No Autocorrelation</div> <div>3. Normality Assumption</div> <div>4. Variable Rationalization</div> </div> <div> <div>a. Selecting relevant predictors for the model</div> <div>b. Residuals follow a normal distribution</div> <div>c. errors are independent of each other</div> <div>d. best Linear Unbiased Estimator</div> </div>                                            | L1 | CO4 |
| 3M-5                | <div> <div>1. Likelihood Ratio Test</div> <div>2. Pseudo R-squared</div> <div>3. Wald Test</div> </div> <div> <div>a. Evaluates classification accuracy of model</div> <div>b. Checks statistical significance of coefficient</div> <div>c. Measures goodness of fit in logistic regression</div> </div>                                                                                                                   | L1 | CO4 |

|           |                                                                                                                       |                                              |    |     |
|-----------|-----------------------------------------------------------------------------------------------------------------------|----------------------------------------------|----|-----|
|           | 4. Confusion Matrix additional                                                                                        | d. Compares model improvement with variables |    |     |
| Questions |                                                                                                                       |                                              |    |     |
| 3D-1      | Explain the key concepts of regression and its importance in predictive modeling.                                     |                                              | L2 | CO3 |
| 3D-2      | What are the BLUE (Best Linear Unbiased Estimator) properties in regression? Explain their significance.              |                                              | L1 | CO3 |
| 3D-3      | Describe the Least Squares Estimation method and its role in linear regression.                                       |                                              | L2 | CO3 |
| 3D-4      | What is variable rationalization? How does it impact model building?                                                  |                                              | L1 | CO3 |
| 3D-5      | Explain the step-by-step process of building a regression model.                                                      |                                              | L2 | CO3 |
| 3D-6      | Compare and contrast simple and multiple linear regression models.                                                    |                                              | L5 | CO3 |
| 3D-7      | What are the key assumptions of linear regression? How do violations impact model accuracy?                           |                                              | L1 | CO3 |
| 3D-8      | Explain multicollinearity in regression models. How can it be detected and addressed?                                 |                                              | L2 | CO3 |
| 3D-9      | Discuss about the following Model Fit Statistics:i) Hosmer Lemeshow Test<br>ii) Error Matrix.                         |                                              | L6 | CO4 |
| 3D-10     | Discuss model fit statistics in logistic regression, including AIC, BIC, and likelihood ratio tests.                  |                                              | L5 | CO4 |
| 3D-11     | Explain the process of constructing a logistic regression model with real-world applications.                         |                                              | L1 | CO4 |
| 3D-12     | How is logistic regression used in business analytics applications such as fraud detection and customer segmentation? |                                              | L1 | CO4 |
| 3D-13     | What are the different types of link functions in logistic regression? Explain their use cases.                       |                                              | L1 | CO4 |
| 3D-14     | Discuss the evaluation metrics for logistic regression models, such as accuracy, precision, recall, and AUC-ROC.      |                                              | L1 | CO4 |
| 3D-15     | Explain how feature selection techniques, such as stepwise regression and LASSO, improve model performance.           |                                              | L1 | CO4 |

| UNIT – IV : OBJECT SEGMENTATION |                                                                                                                                                                                                 |          |                |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------|
| SNO                             | Multiple Choice Questions                                                                                                                                                                       | BT LEVEL | Course Outcome |
| 4C-1                            | Which of the following is NOT a type of supervised learning?<br>A. Classification<br>B. Clustering<br>C. Regression<br>D. Decision Tree                                                         | L1       | CO5            |
| 4C-2                            | In object segmentation, which technique is used to separate foreground objects from the background?<br>A. Regression<br>B. Supervised Learning<br>C. Image Thresholding<br>D. Data Augmentation | L2       | CO5            |
| 4C-3                            | Which method is commonly used to prevent overfitting in decision trees?<br>A. Increasing the depth of the tree<br>B. Pruning<br>C. Adding more features<br>D. Removing training data            | L1       | CO5            |

|       |                                                                                                                                                                                                         |    |     |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 4C-4  | Which of the following is NOT a component of ARIMA?<br>A. Auto-regression<br>B. Moving average<br>C. Seasonality<br>D. Differencing                                                                     | L1 | CO5 |
| 4C-5  | Which of the following measures is used to evaluate forecast accuracy?<br>A. Mean Absolute Error (MAE)<br>B. Cross-entropy<br>C. Confusion Matrix<br>D. ROC Curve                                       | L1 | CO5 |
| 4C-6  | What is the purpose of decision tree pruning?<br>A. To increase model complexity<br>B. To reduce overfitting<br>C. To add more training data<br>D. To increase tree depth                               | L1 | CO5 |
| 4C-7  | Which of the following is NOT a part of the STL decomposition?<br>A. Seasonal component<br>B. Trend component<br>C. Noise component<br>D. Regression component                                          | L1 | CO5 |
| 4C-8  | Which of the following techniques combines multiple decision trees to improve performance?<br>A. Gradient Boosting<br>B. Linear Regression<br>C. Principal Component Analysis<br>D. K-Means Clustering  | L1 | CO5 |
| 4C-9  | Which evaluation metric is suitable for comparing time series forecasts?<br>A. Accuracy Score<br>B. Precision-Recall Curve<br>C. Root Mean Square Error (RMSE)<br>D. F1 Score                           | L1 | CO5 |
| 4C-10 | Which method helps in extracting key features from time series data?<br>A. Fourier Transform<br>B. Regression Analysis<br>C. Logistic Regression<br>D. Decision Tree                                    | L1 | CO5 |
| 4C-11 | Which of the following is an example of unsupervised learning?<br>A. Support Vector Machines<br>B. K-Means Clustering<br>C. Random Forest<br>D. Logistic Regression                                     | L1 | CO5 |
| 4C-12 | What is the main advantage of multiple decision trees in ensemble learning?<br>A. It always increases accuracy<br>B. It avoids under fitting<br>C. It reduces variance<br>D. It minimizes training time | L1 | CO5 |
| 4C-13 | Which of the following is NOT a forecasting method?<br>A. ARIMA<br>B. STL<br>C. Decision Trees<br>D. K-Means Clustering                                                                                 | L1 | CO5 |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                             |    |     |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 4C-14               | What does the term 'height' in a decision tree refer to?<br>A. Number of internal nodes<br>B. Number of leaf nodes<br>C. Longest path from root to a leaf node<br>D. Total number of nodes in the tree                                                                                                                                                                                                      | L1 | CO5 |
| 4C-15               | Which of the following is an application of time series forecasting?<br>A. Image Classification<br>B. Spam Detection<br>C. Stock Price Prediction<br>D. Sentiment Analysis                                                                                                                                                                                                                                  | L1 | CO5 |
| Fill in the Blanks  |                                                                                                                                                                                                                                                                                                                                                                                                             |    |     |
| 4F-1                | In supervised learning, the model is trained using labeled _____.                                                                                                                                                                                                                                                                                                                                           | L1 | CO5 |
| 4F-2                | Unsupervised learning does not use _____ labels for training.                                                                                                                                                                                                                                                                                                                                               | L1 | CO5 |
| 4F-3                | Decision tree pruning helps prevent _____.                                                                                                                                                                                                                                                                                                                                                                  | L1 | CO5 |
| 4F-4                | A decision tree with high depth is more likely to suffer from _____.                                                                                                                                                                                                                                                                                                                                        | L2 | CO5 |
| 4F-5                | In a decision tree, the root node represents the _____ feature in the dataset.                                                                                                                                                                                                                                                                                                                              | L1 | CO5 |
| 4F-6                | The ARIMA model is commonly used for _____ forecasting.                                                                                                                                                                                                                                                                                                                                                     | L1 | CO5 |
| 4F-7                | STL stands for Seasonal-Trend _____.                                                                                                                                                                                                                                                                                                                                                                        | L2 | CO5 |
| 4F-8                | The height of a decision tree is the longest path from the _____ to a leaf node.                                                                                                                                                                                                                                                                                                                            | L1 | CO5 |
| 4F-9                | Mean Absolute Error (MAE) is used to measure _____ accuracy.                                                                                                                                                                                                                                                                                                                                                | L1 | CO5 |
| 4F-10               | Feature extraction from a model helps in improving _____ quality.                                                                                                                                                                                                                                                                                                                                           | L2 | CO5 |
| 4F-11               | Overfitting occurs when a model learns too much from _____ noise.                                                                                                                                                                                                                                                                                                                                           | L2 | CO5 |
| 4F-12               | A Random Forest model consists of multiple _____.                                                                                                                                                                                                                                                                                                                                                           | L2 | CO5 |
| 4F-13               | The moving average component in ARIMA helps smooth out _____.                                                                                                                                                                                                                                                                                                                                               | L1 | CO5 |
| 4F-14               | A complex model with high variance is more prone to _____ error.                                                                                                                                                                                                                                                                                                                                            | L1 | CO5 |
| 4F-15               | Stock market predictions are a real-world application of _____ methods                                                                                                                                                                                                                                                                                                                                      | L1 | CO5 |
| Match the Following |                                                                                                                                                                                                                                                                                                                                                                                                             |    |     |
| 4M-1                | <div> <div>1. Regression</div> <div>2. Segmentation</div> <div>3. Instance Segmentation</div> <div>4. Semantic Segmentation</div> </div> <div> <div>a. Identifies individual objects separately in an image</div> <div>b. Predicts continuous numerical values</div> <div>c. Assigns a class label to each pixel without distinguishing</div> <div>d. Divides an image into meaningful regions</div> </div> | L1 | CO5 |
| 4M-2                | <div> <div>Supervised Learning</div> <div>Unsupervised Learning</div> <div>Classification</div> <div>Clustering</div> </div> <div> <div>a. Groups similar data points together</div> <div>b. Requires labeled data for training</div> <div>c. Clusters data based on patterns without labels</div> <div>d. Categorizes data into predefined labels</div> </div>                                             | L2 | CO5 |
| 4M-3                | <div> <div>1. Decision Tree</div> <div>2. Classification Tree</div> <div>3. Overfitting.</div> <div>4. Pruning</div> </div> <div> <div>a. Regression Predicts continuous values</div> <div>b. Reduces complexity by removing unnecessary branches</div> <div>c. Predicts categorical labels based on input data</div> <div>d. Model learns noise, reducing generalization ability</div> </div>              | L1 | CO5 |

|                               |                                                                                                                                     |                                                                                                                                                                                      |          |                |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------|
| 4M-4                          | 1. Bagging<br>2. Boosting predictions<br>3. Random Forest better accuracy<br>4. Complexity Control reduce variance                  | a. Prevents overfitting by limiting tree depth<br>b. Uses multiple decision trees to enhance<br>c. Sequentially improves weak learners for<br>d. Combines multiple decision trees to | L1       | CO5            |
| 4M-5                          | 1. Height of Model Output. model fitting<br>2. Average Energy predictions<br>3. Seasonality Strength series<br>4. Residual Analysis | a. Evaluates remaining variance after<br>b. Determines complexity or depth of<br>c. Measures overall trend strength in time<br>d. Captures repeating patterns in data                | L1       | CO5            |
| Questions                     |                                                                                                                                     |                                                                                                                                                                                      |          |                |
| 4D-1                          | Explain the differences between object segmentation and regression in image processing.                                             |                                                                                                                                                                                      | L2       | CO5            |
| 4D-2                          | Compare and contrast supervised and unsupervised learning with real-world applications.                                             |                                                                                                                                                                                      | L5       | CO5            |
| 4D-3                          | Describe the tree-building process in decision trees and explain how it helps in regression and classification tasks.               |                                                                                                                                                                                      | L2       | CO5            |
| 4D-4                          | What is overfitting in machine learning models? How does pruning help in reducing overfitting?                                      |                                                                                                                                                                                      | L1       | CO5            |
| 4D-5                          | Explain complexity in decision trees and discuss methods to control it.                                                             |                                                                                                                                                                                      | L2       | CO5            |
| 4D-6                          | Describe multiple decision trees and how they contribute to ensemble learning techniques like Random Forest.                        |                                                                                                                                                                                      | L2       | CO5            |
| 4D-7                          | Explain the ARIMA model in time series forecasting. What are its key components?                                                    |                                                                                                                                                                                      | L2       | CO5            |
| 4D-8                          | What are the different measures of forecast accuracy? Explain with examples.                                                        |                                                                                                                                                                                      | L1       | CO5            |
| 4D-9                          | Explain the STL (Seasonal-Trend Decomposition) approach in time series analysis and its significance.                               |                                                                                                                                                                                      | L2       | CO5            |
| 4D-10                         | How can features such as height and average energy be extracted from a generated model?                                             |                                                                                                                                                                                      | L1       | CO5            |
| 4D-11                         | Describe how extracted features from a model can be analyzed for better predictions.                                                |                                                                                                                                                                                      | L2       | CO5            |
| 4D-12                         | Compare decision trees with other predictive modeling techniques in terms of accuracy and interpretability.                         |                                                                                                                                                                                      | L5       | CO5            |
| 4D-13                         | Explain the role of pruning in decision trees and differentiate between pre-pruning and post-pruning techniques.                    |                                                                                                                                                                                      | L2       | CO5            |
| 4D-14                         | How does complexity impact machine learning models? Discuss with reference to bias-variance tradeoff.                               |                                                                                                                                                                                      | L1       | CO5            |
| 4D-15                         | What is Overfitting? How to Prevent Overfitting?Compare and Contrast between ARMA and ARIMA.                                        |                                                                                                                                                                                      | L1       | CO5            |
| UNIT – V : DATA VISUALIZATION |                                                                                                                                     |                                                                                                                                                                                      |          |                |
| SNO                           | Multiple Choice Questions                                                                                                           |                                                                                                                                                                                      | BT LEVEL | Course Outcome |

|      |                                                                                                                                                                                                            |    |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 5C-1 | Which of the following visualization techniques is best suited for displaying large datasets in limited space?<br>A. Geometric projection<br>B. Pixel-oriented<br>C. Icon-based<br>D. Hierarchical         | L1 | CO6 |
| 5C-2 | Which of the following is an example of a geometric projection technique?<br>A. Scatter Plot<br>B. Parallel Coordinates<br>C. TreeMap<br>D. Heatmap                                                        | L1 | CO6 |
| 5C-3 | Which visualization technique uses small graphical representations like glyphs or icons to encode data?<br>A. Pixel-oriented<br>B. Icon-based<br>C. Hierarchical<br>D. Geometric                           | L1 | CO6 |
| 5C-4 | Which technique is best suited for representing hierarchical structures like file systems?<br>A. Tree visualization<br>B. Parallel coordinates<br>C. Scatter plot<br>D. Histogram                          | L1 | CO6 |
| 5C-5 | Which of the following is NOT an advantage of pixel-oriented techniques?<br>A. High data density<br>B. Efficient use of screen space<br>C. Good for large datasets<br>D. Easy interpretation for beginners | L1 | CO6 |
| 5C-6 | What is a major disadvantage of geometric projection techniques?<br>A. High computational cost<br>B. Limited data representation<br>C. Poor scalability<br>D. Lack of color mapping                        | L1 | CO6 |
| 5C-7 | Which technique is most effective for visualizing hierarchical data relationships?<br>A. Dendrogram<br>B. Scatterplot<br>C. Line graph<br>D. Boxplot                                                       | L1 | CO6 |
| 5C-8 | What type of data is best visualized using pixel-oriented techniques?<br>A. Unstructured text data<br>B. High-dimensional numerical data<br>C. Categorical data<br>D. Low-dimensional datasets             | L1 | CO6 |
| 5C-9 | Which hierarchical visualization method is widely used for hierarchical clustering?<br>A. Bar chart<br>B. Dendrogram<br>C. Histogram<br>D. Pie chart                                                       | L1 | CO6 |

|                    |                                                                                                                                                                                                                  |    |     |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 5C-10              | Which of the following is a challenge in visualizing complex data relationships?<br>A. Lack of computing power<br>B. Overlapping data points<br>C. Limited data storage<br>D. Network issues                     | L1 | CO6 |
| 5C-11              | Which visualization technique is most effective for analyzing social network relationships?<br>A. Pixel-based visualization<br>B. Force-directed graphs<br>C. Icon-based representation<br>D. Pie chart          | L1 | CO6 |
| 5C-12              | Which factor is crucial for effective visualization of multi-dimensional data?<br>A. Data redundancy<br>B. Proper selection of visual encoding<br>C. Increased storage capacity<br>D. Use of primary colors only | L1 | CO6 |
| 5C-13              | Which of the following techniques is commonly used for representing high-dimensional data?<br>A. PCA (Principal Component Analysis)<br>B. Line charts<br>C. Histograms<br>D. Bar charts                          | L1 | CO6 |
| 5C-14              | Which visualization method can handle millions of data points effectively?<br>A. Scatterplot matrix<br>B. Pixel-oriented visualization<br>C. Pie chart<br>D. Boxplot                                             | L1 | CO6 |
| 5C-15              | Which of the following techniques best visualizes relationships between entities in a dataset?<br>A. Force-directed graphs<br>B. Treemap<br>C. Pie chart<br>D. Heatmap                                           | L1 | CO6 |
| Fill in the Blanks |                                                                                                                                                                                                                  |    |     |
| 5F-1               | _____ visualization techniques involve the use of individual pixels to represent data points or values.                                                                                                          | L1 | CO6 |
| 5F-2               | In pixel-oriented visualization, each pixel typically represents a single _____.                                                                                                                                 | L2 | CO6 |
| 5F-3               | _____ projection is a method of visualizing data by mapping high-dimensional data onto lower-dimensional spaces, like a 2D plane.                                                                                | L1 | CO6 |
| 5F-4               | _____ visualization uses geometric shapes like points, lines, and polygons to represent complex data structures.                                                                                                 | L1 | CO6 |
| 5F-5               | _____ visualization is used to depict data hierarchies, where elements are grouped and represented in tree structures or nested layouts.                                                                         | L1 | CO6 |
| 5F-6               | Geometric projection techniques often include methods such as _____ for reducing dimensions of the data for visualization.                                                                                       | L1 | CO6 |
| 5F-7               | Pixel-oriented visualizations are particularly effective for visualizing large datasets with numerous _____.                                                                                                     | L2 | CO6 |
| 5F-8               | In icon-based visualizations, the size or color of the icon often represents the _____ of the data.                                                                                                              | L1 | CO6 |

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |    |     |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----|
| 5F-9                | A common type of geometric projection used in data visualization is the _____ plot, which visualizes multidimensional data in 2D space.                                                                                                                                                                                                                                                                                                                                                                             | L1 | CO6 |
| 5F-10               | Complex data relations can be visualized effectively through _____ techniques that show both hierarchical and relational aspects of the data.                                                                                                                                                                                                                                                                                                                                                                       | L2 | CO6 |
| 5F-11               | In data visualization, one of the main goals of icon-based techniques is to make the data more _____ and relatable to human cognition.                                                                                                                                                                                                                                                                                                                                                                              | L1 | CO6 |
| 5F-12               | _____ visualizations are especially valuable for illustrating processes or flows, where the data elements move through a series of stages.                                                                                                                                                                                                                                                                                                                                                                          | L2 | CO6 |
| 5F-13               | A _____ diagram is an example of an icon-based visualization technique, where each icon represents a data point.                                                                                                                                                                                                                                                                                                                                                                                                    | L1 | CO6 |
| 5F-14               | _____ visualizations provide a method for showing large datasets in an intuitive manner by breaking them into simpler components.                                                                                                                                                                                                                                                                                                                                                                                   | L1 | CO6 |
| 5F-15               | _____ visualization techniques use icons or pictograms as graphical elements to represent abstract data.                                                                                                                                                                                                                                                                                                                                                                                                            | L1 | CO6 |
| Match the Following |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |    |     |
| 5M-1                | <div> <div>1. Recursive Pattern queries</div> <div>2. Pixel Bar Chart</div> <div>3. Query-Independent Layout for efficient visualization</div> <div>4. Query-Dependent Layout data values in bars</div> </div> <div> <div>a. Pixel arrangement changes based on specific queries</div> <div>b. Fixed pixel positioning regardless of data queries</div> <div>c. Arranges pixels in a repeated manner for efficient visualization</div> <div>d. Uses small dots or pixels to encode data values in bars</div> </div> | L2 | CO6 |
| 5M-2                | <div> <div>1. Parallel Coordinates</div> <div>2. Principal Component Analysis (PCA)</div> <div>3. Multidimensional Scaling (MDS)</div> <div>4. Scatterplot Matrix</div> </div> <div> <div>a. Shows pairwise relationships among multiple variables</div> <div>b. Maps high-dimensional data into a lower-dimensional space</div> <div>c. Reduces dimensionality while preserving variance</div> <div>d. Maps high-dimensional data into a lower-dimensional space.</div> </div>                                     | L1 | CO6 |
| 5M-3                | <div> <div>1. Chernoff Faces</div> <div>2. Star Glyphs</div> <div>3. Stick Figures</div> <div>4. Shape Coding data</div> </div> <div> <div>a. Represents data using human-like stick figures</div> <div>b. Uses different shapes to encode categorical data</div> <div>c. Displays multiple variables as radiating lines</div> <div>d. Uses facial features to represent multivariate data</div> </div>                                                                                                             | L1 | CO6 |
| 5M-4                | <div> <div>1. Tree map</div> <div>2. Sunburst Chart</div> <div>3. Dendrogram</div> <div>4. Icicle Plot</div> </div> <div> <div>a. Uses nested rectangles to represent hierarchical structures</div> <div>b. Represents hierarchical relationships in a top-down fashion</div> <div>c. Commonly used in hierarchical clustering</div> <div>d. Displays hierarchical data in a radial layout</div> </div>                                                                                                             | L1 | CO6 |
| 5M-5                | <div> <div>1. Sankey Diagram</div> <div>2. Network Graph</div> <div>3. Heat map</div> <div>4. Scatterplot Matrix</div> </div> <div> <div>a. Shows relationships between multiple numerical variables</div> <div>b. Uses color intensity to represent data values  </div> <div>c. Displays connections between nodes and edges</div> <div>d. Represents flow relationships between entities</div> </div>                                                                                                             | L1 | CO6 |

| Questions |                                                                                                                    |    |     |
|-----------|--------------------------------------------------------------------------------------------------------------------|----|-----|
| 5D-1      | Explain pixel-oriented visualization techniques and their advantages in handling large datasets.                   | L2 | CO6 |
| 5D-2      | Describe different geometric projection visualization techniques with suitable examples.                           | L4 | CO6 |
| 5D-3      | What are icon-based visualization techniques? Provide use cases where they are beneficial.                         | L1 | CO6 |
| 5D-4      | Explain hierarchical visualization techniques and discuss their significance in data representation.               | L2 | CO6 |
| 5D-5      | How can complex data and relationships be visualized effectively? Give real-world examples.                        | L1 | CO6 |
| 5D-6      | Differentiate between pixel-oriented and icon-based visualization techniques.                                      | L4 | CO6 |
| 5D-7      | Discuss the strengths and limitations of geometric projection techniques in data visualization.                    | L6 | CO6 |
| 5D-8      | Describe how hierarchical visualization can be used for representing large-scale structured data.                  | L4 | CO6 |
| 5D-9      | Demonstrate Geometric projection visualization techniques. List out various applications of data visualization.    | L4 | CO6 |
| 5D-10     | Compare and contrast different visualization techniques for large, multi-dimensional datasets.                     | L5 | CO6 |
| 5D-11     | How does interactive visualization improve the interpretation of complex data relationships?                       | L1 | CO6 |
| 5D-12     | Discuss how data visualization principles align with cognitive psychology to improve understanding.                | L6 | CO6 |
| 5D-13     | Evaluate an in-depth analysis of how pixel-oriented visualization is applied in time-series data.                  | L5 | CO6 |
| 5D-14     | Write down steps involved in Data Visualization In Tableau.                                                        | L2 | CO6 |
| 5D-15     | How to perform visualization of the data using a hierarchical partitioning into subspaces? Explain with examples . | L1 | CO6 |

## MODEL PAPERS

Subject code: R22CSE3141



### SRI INDU COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC, New Delhi)

Recognized under 2(f) and 12(B) of UGC Act 1956

III B. Tech- II Semester – MID Examinations (Model paper)

DATA ANALYTICS

CSE(AIML)

Duration : 2 Hrs.

Max Marks: 30M

#### SECTION A

I. Answer all the multiple-choice questions.

10 x ½ = 5M

|                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Which of the following is NOT a type of data source?<br>A. Sensors<br>B. Signals<br>C. GPS<br>D. Hyperlinks                                                                    |
| 2. GPS data primarily includes:<br>A. Temperature and pressure readings<br>B. Latitude, longitude, and altitude<br>C. Textual descriptions<br>D. Signal strength                  |
| 3. Decision trees are primarily used for:<br>A. Analyzing unstructured data<br>B. Predicting outcomes based on rules<br>C. Reducing dimensionality<br>D. Segmenting data clusters |
| 4. Which technique replaces missing values with the mean or median of the variable?<br>A. List wise deletion                                                                      |

|                                                                                                                                                                                                                              |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>B. Imputation</p> <p>C. Interpolation</p> <p>D. Normalization</p>                                                                                                                                                         |
| <p>5. Which technique is used to address multicollinearity in regression models?</p> <p>A. Data normalization</p> <p>B. Variance Inflation Factor (VIF)</p> <p>C. Residual Analysis</p> <p>D. Cross-validation</p>           |
| <p>6. What type of problem does logistic regression solve?</p> <p>A. Regression</p> <p>B. Classification</p> <p>C. Clustering</p> <p>D. Dimensionality Reduction</p>                                                         |
| <p>7. Which evaluation metric is suitable for comparing time series forecasts?</p> <p>A. Accuracy Score</p> <p>B. Precision-Recall Curve</p> <p>C. Root Mean Square Error (RMSE)</p> <p>D. F1 Score</p>                      |
| <p>8. Which method helps in extracting key features from time series data?</p> <p>A. Fourier Transform</p> <p>B. Regression Analysis</p> <p>C. Logistic Regression</p> <p>D. Decision Tree</p>                               |
| <p>9. Which of the following visualization techniques is best suited for displaying large datasets in limited space?</p> <p>A. Geometric projection</p> <p>B. Pixel-oriented</p> <p>C. Icon-based</p> <p>D. Hierarchical</p> |
| <p>10. Which of the following is an example of a geometric projection technique?</p> <p>A. Scatter Plot</p> <p>B. Parallel Coordinates</p> <p>C. Tree Map</p> <p>D. Heatmap</p>                                              |

**Section A2: Answer all the fill in the blanks**

**6Q x ½ = 3M**

11. Data preprocessing involves cleaning, transforming, and \_\_\_\_\_ data for analysis.
12. A \_\_\_\_\_ is a collection of organized data that allows easy access, management, and updating.
13. The odds ratio in logistic regression is calculated as the exponentiation of the \_\_\_\_\_.
14. Pixel-oriented visualizations are particularly effective for visualizing large datasets with numerous \_\_\_\_\_

15. In icon-based visualizations, the size or colour of the icon often represents the \_\_\_\_\_ of the data.

16. The moving average component in ARIMA helps smooth out \_\_\_\_\_

**SECTION A3: Answer all the Match the following Questions.**

**2Q x 1M =2M**

**17.**

- |                      |                                                  |
|----------------------|--------------------------------------------------|
| 1. Data Architecture | a. Provides location data using satellites       |
| 2. GPS               | b. Represents data transmitted through waves     |
| 3. Sensors           | c. Framework for managing and organizing data    |
| 4. Signals           | d. Devices that collect real-time numerical data |

**18.**

- |                              |                                                |
|------------------------------|------------------------------------------------|
| 1. BLUE Property             | a. Selecting relevant predictors for the model |
| 2. No Autocorrelation        | b. Residuals follow a normal distribution      |
| 3. Normality Assumption.     | c.Errors are independent of each other         |
| 4. Variable Rationalization. | d.best Linear Unbiased Estimator               |

**PART – B (20 MARKS)**

**Answer any FOUR questions.**

**Marks: 4Q x 5M = 20M**

19. Discuss the importance of data modelling techniques in business applications.

20. Examine the strengths and weaknesses of various data imputation methods in handling missing data.

21. What is data? How to handle large collections of data? Discuss the various sources of data.

22. Explain the ARIMA model in time series forecasting. What are its key components?

23. What are the different measures of forecast accuracy? Explain with examples.

24. Demonstrate Geometric projection visualization techniques. List out various applications of data visualization.

**\*\*\*\*\***

# EXTERNAL MODEL PAPER

Subject code: R22CSE3141



**SRI INDUCOLLEGE OF ENGINEERING& TECHNOLOGY**

(An Autonomous Institution under UGC, New Delhi)

Recognized under 2(f) and 12(B) of UGC Act 1956

**III B. Tech- II Semester – End Examinations (Model paper)**

**DATA ANALYTICS**

**CSE(AIML)**

**Duration: 3 Hrs**

**Max Marks: 60**

## Section – A

| Answer <u>All</u> the following questions |                                                                    | Marks: 10Qx1M = 10M |                |          |
|-------------------------------------------|--------------------------------------------------------------------|---------------------|----------------|----------|
| 1                                         |                                                                    | Marks               | Course Outcome | BT Level |
| a                                         | What is Data? What are the sources of data?                        | 1                   | CO1            | L1       |
| b                                         | What is primary and secondary data?                                | 1                   | CO1            | L1       |
| c                                         | Explain the importance of data in business decision making.        | 1                   | CO2            | L2       |
| d                                         | Define Data imputation.                                            | 1                   | CO2            | L1       |
| e                                         | What is BLUE Property?                                             | 1                   | CO3            | L1       |
| f                                         | What is Logistic regression?                                       | 1                   | CO4            | L1       |
| g                                         | Define Object segmentation.                                        | 1                   | CO5            | L1       |
| h                                         | List the properties of decision tree.                              | 1                   | CO5            | L1       |
| i                                         | List out 2 differences between pixel and icon-based visualization. | 1                   | CO6            | L4       |
| j                                         | What is Data Visualization?                                        | 1                   | CO6            | L1       |

## Section – B

|                                                                                                                |                                                           |                   |                |                                  |
|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|-------------------|----------------|----------------------------------|
| Answer any <u>FIVE</u> questions choosing at least one from each Unit Marks:                                   |                                                           |                   | 5Qx10M = 50M   |                                  |
|                                                                                                                |                                                           | Marks             | Course Outcome | BT Level                         |
| UNIT-I                                                                                                         |                                                           |                   |                |                                  |
| 2.                                                                                                             |                                                           |                   |                |                                  |
| a. List the common sources of data such as sensors, signals, and GPS, and briefly describe their applications. | 10                                                        | CO1               | 3,1            |                                  |
| (OR)                                                                                                           |                                                           |                   |                |                                  |
| b. How does data normalization help in the analysis of sensor data?                                            |                                                           |                   |                |                                  |
| UNIT-II                                                                                                        |                                                           |                   |                |                                  |
| 3.                                                                                                             |                                                           |                   |                |                                  |
| a. Differentiate between descriptive, predictive, and prescriptive analytics with examples.                    | 10                                                        | CO2               | 4,5            |                                  |
| (OR)                                                                                                           |                                                           |                   |                |                                  |
| b. Examine the strengths and weaknesses of various data imputation methods in handling missing data.           |                                                           |                   |                |                                  |
| UNIT – III                                                                                                     |                                                           |                   |                |                                  |
| 4.                                                                                                             |                                                           |                   |                |                                  |
| a. What are the BLUE (Best Linear Unbiased Estimator) properties in regression? Explain their significance.    | 10                                                        | CO3               | 1,1            |                                  |
| (OR)                                                                                                           |                                                           |                   |                |                                  |
| b. What are the different types of link functions in logistic regression? Explain their use cases.             |                                                           |                   |                |                                  |
| UNIT – IV                                                                                                      |                                                           |                   |                |                                  |
| 5.                                                                                                             |                                                           |                   |                |                                  |
| a. What is over fitting in machine learning models? How does pruning help in reducing over fitting?            | 10                                                        | CO4               | 1,4            |                                  |
| (OR)                                                                                                           |                                                           |                   |                |                                  |
| b. Compare decision trees with other predictive modeling techniques in terms of accuracy and interpretability. |                                                           |                   |                |                                  |
| UNIT – V                                                                                                       |                                                           |                   |                |                                  |
| 6.                                                                                                             |                                                           |                   |                |                                  |
| a. Explain hierarchical visualization techniques and discuss their Significance in data representation.        | 10                                                        | CO5               | 2,5            |                                  |
| (OR)                                                                                                           |                                                           |                   |                |                                  |
| b. Evaluate an in-depth analysis of how pixel-oriented visualization is applied in time-series data.           |                                                           |                   |                |                                  |
| S.NO                                                                                                           | Bloom Taxonomy Level                                      | No. Of. Questions | %              | Name & Signature of Paper Setter |
| 1                                                                                                              | Fundamental Knowledge from level I & II                   | 14                | 70             |                                  |
| 2                                                                                                              | Knowledge on application & Analysis from level III & IV   | 4                 | 20             |                                  |
| 3                                                                                                              | Critical thinking and ability to design from level V & VI | 2                 | 10             |                                  |

## **PREVIOUS YEAR QUESTION PAPERS**