Reg. No. | | | | | | | | |

G. VENKATASWAMY NAIDU COLLEGE (AUTONOMOUS), KOVILPATTI – 628 502.

UG DEGREE END SEMESTER EXAMINATIONS - NOVEMBER 2025.
(For those admitted in June 2023 and later)

PROGRAMME AND BRANCH: B.Sc., COMPUTER SCIENCE

SEM	CATEGORY	COMPONENT	COURSE CODE	COURSE TITLE
V	PART - III	CORE ELECTIVE-2	U23CS5E2A	INTRODUCTION TO DATA SCIENCE

Date & Session: 11.11.2025/FN **Time : 3 hours** **Maximum: 75 Marks**

Course Outcome	Bloom's K-level	Q. No.	SECTION – A (10 X 1 = 10 Marks) Answer ALL Questions.
CO1	K1	1.	Which of the following is the most important language for Data Science? a) R b) Java c) Ruby d) hadoop
CO1	K2	2.	Which of the following is the top most important thing in data science? a) Data b) Question c) programcode d) algorithms
CO2	K1	3.	Which of the following is **not** typically involved in the data retrieval phase? a) quering database b) Scraping websites c) Building predictive models d) Accessing APIs
CO2	K2	4.	What is the main goal of Exploratory Data Analysis (EDA)? a) To create final visual reports b) perform statistical tests for model validation c) To explore data patterns and trends d) To transform data for machine learning
CO3	K1	5.	Which of these is an example of an unsupervised learning task? a) Predicting house prices b) Classifying emails as spam or not spam c) Grouping customers by purchasing behaviour d) Detecting fraudulent transactions
CO3	K2	6.	What is the key characteristic of semi-supervised learning? a) Uses only unlabeled data b) Requires no training data c) Combines a small amount of labeled data with a large amount of unlabeled data d) Trains multiple models and combines results
CO4	K1	7.	Which of the following best describes NoSQL databases? a) Only support relational schemas b) Are optimized for transactional consistency only c) Support flexible data models and scale horizontally d) Only run on Hadoop clusters

CO4	K2	8.	What is the primary function of the Hadoop Distributed File System (HDFS)**?** a) Encrypting data b) Distributing data across multiple servers c) Generating machine learning models d) Storing data on local machines only
CO5	K1	9.	What is the best source for retrieving health-related data? a) Amazon Web Services b) WHO or UCI Machine Learning Repository c) Stack Overflow d) GitHub repositories
CO5	K2	10.	Which algorithm is most commonly used for binary disease classification? a) K-means b) Linear Regression c) Logistic Regression d) Apriori Algorithm
Course Outcome	**Bloom's K-level**	**Q. No.**	**SECTION – B (5 X 5 = 25 Marks)** **Answer ALL Questions choosing either (a) or (b)**
CO1	K3	11a.	Explain any five benefits and real-world uses of Data Science with examples. **(OR)**
CO1	K3	11b.	Differentiate between traditional data science and big data ecosystems.
CO2	K3	12a.	Illustrate the steps involved in transforming raw data before analysis in a data science workflow. **(OR)**
CO2	K3	12b.	Write short note on the role of Exploratory Data Analysis (EDA) in uncovering patterns and guiding model selection.
CO3	K4	13a.	Conclude any five commonly used python tools in machine learning and how it is used in Data science. **(OR)**
CO3	K4	13b.	Examine the Application of machine learning and how it is used in DataScience process.
CO4	K4	14a.	Outline the concept of Spark and its application. **(OR)**
CO4	K4	14b.	Analyze how Apache Spark addresses the limitations of the MapReduce model in terms of performance and data processing.
CO5	K5	15a.	Critically evaluate the role of data exploration and disease profiling in building a predictive model. **(OR)**
CO5	K5	15b.	Evaluate the importance of setting clear research goals in a disease prediction case study.

Course Outcome	**Bloom's K-level**	**Q. No.**	**SECTION – C (5 X 8 = 40 Marks)** **Answer ALL Questions choosing either (a) or (b)**
CO1	K3	16a.	Describe the Data Science process in detail, from setting research goals to model building. **(OR)**
CO1	K3	16b.	Evaluate the importance of different facets of Data Science in solving real-world problems.

CO2	K4	17a.	Examine the different steps involved in Retrieving Data ,cleansing, integrating and transforming data in data science process . **(OR)**
CO2	K4	17b.	Comment on the process of building the models in data science.
CO3	K4	18a.	Distinguish between different types of machine learning with suitable examples **(OR)**
CO3	K4	18b.	Inspect the four steps involved in modelling phase of machine learning
CO4	K5	19a.	Assess the architecture of the Hadoop framework and explain how Apache Spark overcomes the limitations of MapReduce. **(OR)**
CO4	K5	19b.	Conclude the fundamental concepts of NoSQL databases and analyze how the CAP theorem and BASE properties influence their design.
CO5	K5	20a.	Evaluate how data exploration, profiling, and automated reporting contribute to better healthcare decision-making. **(OR)**
CO5	K5	20b.	Prioritize the steps of disease prediction in a data science process and justify your reasoning