SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES (AUTONOMOUS)

QUESTION BANK

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

SYLLABUS:

UNIT I: INTRODUCTION TO BIG DATA AND ANALYTICS ECOSYSTEM

Definition and Characteristics of Big Data – Volume, Velocity, Variety, Veracity, Value, Types of Analytics: Descriptive, Diagnostic, Predictive, Prescriptive, Big Data Challenges and Opportunities, Hadoop Ecosystem Overview: HDFS, MapReduce, YARN, NoSQL Databases: Key-Value, Columnar, Document, Graph Models, Data Lake vs. Data Warehouse.

UNIT II: BIG DATA TOOLS AND FRAMEWORKS

Apache Spark Architecture and RDDs, Spark SQL, Data Frames, and Datasets, Spark Streaming for Real Time Analytics, Kafka for Data Ingestion and Message Queues, Hive, Pig, and Impala for Big Data Querying, Comparative Analysis of Hadoop vs. Spark

UNIT III: MACHINE LEARNING ON BIG DATA

Introduction to MLlib and Scikit-learn, Data Preprocessing for Big Data ML Pipelines, Supervised Learning: Classification and Regression on Large Datasets, Unsupervised Learning: Clustering and Dimensionality Reduction, Model Evaluation and Validation Techniques, Distributed Training and Optimization Techniques.

UNIT IV: AI APPLICATIONS ON BIG DATA

Predictive Maintenance using Big Data & AI, Fraud Detection in Banking with Machine Learning, AI in Healthcare: Diagnosis, Genomics, Patient Monitoring, Retail and Ecommerce Analytics, AI for Smart Cities and IoT Sensor Data Analysis, Evaluation of Real-Time AI Applications on Streaming Data.

UNIT V: ADVANCED TOPICS AND CASE

Deep Learning on Big Data using Tensor Flow on Spark, Explainable AI (XAI) in Big Data Environments, Ethical Issues and Data Governance in Big Data AI, Edge Computing and AI for Low Latency Applications,

Case Study 1: AI-Powered Big Data in Healthcare, **Case Study 2:** Big Data AI Solution in Smart Manufacturing.

Prepared by M. Deepika   Assistant Professor

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

Max Marks: 10

S.No.	CO	Questions	BT
		Unit I: (INTRODUCTION TO BIG DATA AND ANALYTICS ECOSYSTEM)	
1	1	Define Big Data. Explain the 5V characteristics (Volume, Velocity, Variety, Veracity, Value) with examples.	L2
2	1	Discuss the types of analytics: Descriptive, Diagnostic, Predictive, Prescriptive and their role in Big Data.	L2
3	1	Explain the major challenges and opportunities in Big Data. How does it impact industries?	L2,L4
4	1	Describe the Hadoop ecosystem. Explain HDFS, MapReduce and YARN.	L2
5	1	Explain the architecture of HDFS with Name Node, Data Node and block storage mechanism.	L2
6	1	What is Map Reduce? Explain Map and Reduce phases with an example.	L2
7	1	Discuss the role of YARN in Hadoop. How does it manage cluster resources?	L2 ,L4
8	1	a) Explain NoSQL database models: Key–Value, Columnar, Document, Graph with examples. b) Differentiate between Data Lake and Data Warehouse ?	L2 L4
9	1	a. Define Big Data and explain its characteristics. b. Explain MapReduce working model.	L2
10	1	a) Explain HDFS architecture. b) Explain types of analytics.	L2
11	1	a) Explain Hadoop ecosystem components. b) Differentiate Data Lake vs Data Warehouse.	L2
12	1	a) Explain the 5V's of Big Data in detail.	L2

S.No.	CO	Questions	BT
		b) Mention Big Data opportunities.	
		Unit II: (BIG DATA TOOLS AND FRAMEWORKS)	
1	2	Explain the components of Apache Spark architecture and how RDDs support fault tolerance.	L2
2	2	Explain Apache Kafka and list its core components used in data ingestion.	L2
3	2	Demonstrate how to create and managed Internal and external tables in Hive using suitable examples.	L3
4	2	What is Apache Pig? Explain Pig Latin language and its advantages in Big Data processing.	L2
5	2	Compare Hadoop vs Spark. Discuss differences in architecture, speed, processing model and use cases.	L4
6	2	**a)** Differentiate between RDD, DataFrame, and Dataset. **b)** Explain the working of Spark SQL and its advantages.	L4 L2
7	2	What is Apache Pig? Explain data processing operators in Pig Latin	L3
8	2	Describe Spark SQL, DataFrames, and Datasets in Apache Spark	L2
9	2	Create a Hive table named employees with appropriate data types. Insert sample records and write queries to: a) Display all records b) Retrieve employees with salary greater than 50,000 c) Count number of employees in each department	L6
10	2	Create a DataFrame using SparkSQL for students with columns: roll_no, name, marks, branch. Register it as a SQL view and write queries to: a) Find students scoring above 75 marks b) Calculate average marks c) Display students sorted by marks (descending)	L6

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

S.No.	CO	Questions	BT
— Unit III: (**UNIT III: MACHINE LEARNING ON BIG DATA**)			
1	3	Explain the role of MLlib in Apache Spark. How is it useful for Machine Learning on Big Data?	L4
2	3	Discuss Scikit-learn framework. Compare MLlib and Scikit-learn in Big Data applications.	L3
3	3	Explain Data Preprocessing steps in Big Data ML pipelines such as cleaning, normalization, feature extraction and scaling.	L4
4	3	Describe supervised learning on large datasets. Explain classification techniques with examples.	L3
5	3	Explain regression techniques used in Big Data Machine Learning. Discuss linear regression on distributed datasets.	L5
6	3	Explain PCA (Principal Component Analysis) and its importance in handling high-dimensional Big Data.	L4
7	3	Discuss various model evaluation metrics used for classification and regression in Big Data environments.	L3
8	3	Explain Accuracy, Precision, Recall, F1-Score, RMSE, and R^2 with examples.	L5
9	3	Explain Gradient Descent, Stochastic Gradient Descent (SGD), and their role in large-scale model training.	L4
10	3	Explain model Evaluation and validation Techniques in ML	L3

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

S.No.	CO	Questions	BT
Unit IV: (AI APPLICATIONS ON BIG DATA)			
1	4	Explain Predictive Maintenance using Big Data and Artificial Intelligence.	L4
2	4	Describe how Machine Learning is used for fraud detection in banking systems.	L3
3	4	Discuss the role of Artificial Intelligence in medical diagnosis, Genomics, Patient Monitoring	L4
4	4	Explain how Big Data and AI are used in retail and e-commerce analytics.	L3
5	4	Explain the role of AI in Smart City applications. Discuss traffic management, energy optimization, public safety, and waste management using Big Data analytics.	L5
6	4	Describe how AI techniques are applied to IoT sensor data. Explain data collection, stream processing, anomaly detection, and predictive modeling.	L4
7	4	Explain the evaluation metrics and performance criteria for real-time AI applications on streaming data.	L3
8	4	Discuss data ingestion, stream processing, model deployment, and monitoring components.	L5
9	4	a) Explain Smart City AI applications with examples. b) Mention benefits of Smart City analytics.	L4
10	4	a) Explain IoT Sensor Data analysis using AI. b) List challenges in IoT streaming data.	L3
11	4	a) Explain evaluation of real-time AI on streaming data. b) Mention two performance metrics.	L3

Prepared by M. Deepika   Assistant Professor

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

S.No.	CO	Questions	BT
Unit V: (ADVANCED TOPICS AND CASE)			
1	5	Explain Deep Learning in the context of Big Data using TensorFlow on Apache Spark.	L2
2	5	Describe the importance of data governance in Big Data AI systems.	L2
3	5	Apply XAI techniques to interpret predictions made by a Big Data deep learning model.	L3
4	5	Demonstrate how TensorFlow can be integrated with Spark for training a deep learning model on large datasets	L3
5	5	Compare cloud-based AI processing with Edge Computing for low-latency applications.	L4
6	5	Evaluate the effectiveness of Explainable AI techniques in improving transparency and trust in Big Data AI systems.	L5
7	5	Design a Big Data AI architecture that integrates Deep Learning, XAI, ethical governance, and Edge Computing for a real-time application.	L6
8	5	Describe the importance of AI-Powered Big Data in Healthcare	L2
9	5	Demonstrate How the Big Data AI Solution in Smart Manufacturing	L3
10	5	Illustrate the workflow of implementing a distributed deep learning pipeline on Spark using TensorFlow for a real-time Big Data application.	L3
11	5	Analyze the importance of Explainable AI (XAI) techniques in interpreting deep learning models trained on Big Data.	L4

Note: L1-Remembering, L2-Understanding, L3-Applying, L4-Analyzing, L5-Evaluating, and L6-Creating

Prepared by M. Deepika   Assistant Professor

Year / Semester: III B.Tech VI Semester   **Regulation: R23**

Subject and Code: Bigdata for AI Applications & 23CSM371T

Instruction to Faculty Members:

The Six Levels of Bloom's Taxonomy:

1. **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from long-term memory (e.g., list, define, name, locate).

2. **Understanding:** Constructing meaning, explaining ideas, or concepts (e.g., summarize, interpret, classify, compare).

3. **Applying:** Using information in new situations or implementing procedures to solve problems (e.g., solve, use, demonstrate, implement).

4. **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure (e.g., contrast, categorize, distinguish, diagram).

5. **Evaluating:** Making judgments based on criteria and standards through checking and critiquing (e.g., judge, critique, justify, defend, argue).

6. **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure (e.g., design, construct, develop, formulate).