BIG DATA ANALYTICS

Course Code: BECSE3C007
Course Title: BIG DATA ANALYTICS
Semester: VI
Credits: 03

Rationale

The study of Big Data Analytics is essential for equipping learners with the skills needed to analyse, process, and derive insights from large and complex datasets. The role of Big Data Analytics extends to optimizing business processes, enhancing customer experiences, and fostering innovation through data-driven strategies.

Course Outline

Contents	No. of Lectures
Unit - I **Introduction to Big Data:** Big Data, Characteristics of Big Data and Dimensions of Scalability, Getting Value out of Big Data, Foundations of Big Data Systems and Programming, Getting Started with Hadoop.	10
Unit - II **Introduction to Data Analytics:** Data Analytics, Obtaining and Scrubbing Data, Exploring and Modelling Data, Mining and Visualizing Data and Communicating Results, Open AI in Data Analytics.	10
Unit - III **Python for Data Analytics:** Introduction to Python for Big Data, Data Manipulation, Data Cleaning, Data Analysis, Data Visualization, Exploratory Data Analysis, Big Data Structures and Algorithms	10
Unit - IV **NoSQL Big Data Management:** Introduction, NoSQL Data Store, Data Architecture Patterns, NoSQL to Manage Big Data, Architecture for Big Data Tasks, MongoDB Database and Cassandra Database	10
Unit - V	10

Machine Learning Algorithms for Big Data Analytics: Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering, Big Data Analytics with python.	

COURSE OUTCOMES

Upon successful completion of the course the student will be able to:

1. Recall the fundamental concepts of Big Data and explain its role in addressing business challenges.
2. Describe the components of Hadoop and evaluate the functionality of tools in the Hadoop ecosystem.
3. Demonstrate the ability to access, process, and manage data using distributed file systems like HDFS.
4. Design and develop efficient Big Data solutions by leveraging Hadoop and its associated tools.
5. Assess data analysis challenges and implement machine learning models

Text Books:

1. Nathan Marz and James Warren, *Big Data: Principles and Best Practices of Scalable Real-Time Data Systems*, Manning Publications (2015).
2. Tom White, *Hadoop: The Definitive Guide*, O'Reilly Media (2015) 4th Edition.

Reference Books:

1. Jure Leskovec, Anand Rajaraman, and Jeffrey Ullman, *Mining of Massive Datasets*, Cambridge University Press (2020) 3rd Edition.
2. Frank J. Ohlhorst, *Big Data Analytics: Turning Big Data into Big Money*, Wiley (2012).
3. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Mariner Books (2013).
4. Bill Chambers and Matei Zaharia, *Spark: The Definitive Guide*, O'Reilly Media (2018).