

**B.Sc. (Data Science)**  
**Discipline Specific Course (DSC)**  
**Semester VI**  
**BSDB33602T: Applications of Data Science**

**Total Marks: 100**  
**External Marks: 70**  
**Internal Marks: 30**  
**Credits: 4**  
**Pass Percentage: 40%**

**Objective**

To identify the characteristics of datasets and compare the trivial data and big data for various applications. Students will be able to implement machine learning techniques and computing environment that will suitable for the applications under consideration.

**INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER**

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

**INSTRUCTIONS FOR THE CANDIDATES**

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

**Section A**

**Unit I: Social Network Analysis for Big Data-** Introduction of Social Network Analysis, Graph Essentials, Graph Basics-Graph, Data Collection and Preprocessing.

**Unit II: Big Data Analysis and Predictive Analytics** -Evolution of analytic scalability, analytic tools- Predictive Analytics using Supervised – Unsupervised learning, Neural networks, Clustering Methods.

**Unit III: Financial Data Analytics:** Framework, techniques, and metrics, News events impact market sentiment, relating news analytics to stock returns

**Unit IV Financial Time Series Analytics** and Their Characteristics, Common Financial Time Series models, Autoregressive models, Markov chain models, and Time series models with leading indicators, Long term forecasting

**Section B**

**Unit V Recommendation In Social Media And Behavior:** Analytics, Challenges-Classical Recommendation Algorithms, Recommendation Using Social Context, Behavior Analytics: Individual Behavior, Collective Behavior.

**Unit VI Introduction to Healthcare Data Analytics**, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems.

**Unit VII Natural Language Processing** and Data Mining for Clinical Text: Core NLP Components, Information Extraction and Named Entity Recognition, Social Media Analytics for Healthcare: Tracking of Infectious Disease Outbreaks, Readmission risk Prediction.

**Unit VIII Retail Analytics:** Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.

**Suggested Readings**

1. Mike Frampton, Mastering Apache Spark, Packt Publishing, 2015
2. Mohammed Guller, Big Data Analytics with Spark, Apress, 2016