

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

SEMESTER - 5

Paper Code: BDS501

Paper Name: *Discrete Mathematics*

Credits: 5

MODULE 1: Introduction to Discrete Mathematics

- - Basic terminologies and concepts in discrete mathematics
- - Sets, set operations, and set theory
- - Functions and relations

MODULE 2: Logic and Proof Techniques

- - Propositional logic: syntax, semantics, and truth tables
- - Logical connectives and truth functions
- - Predicates and quantifiers
- - Proofs: direct proofs, proof by contradiction, and proof by contrapositive

MODULE 3: Methods of Proof

- - Mathematical induction
- - Strong induction
- - Recursive definitions and structural induction

MODULE 4: Combinatorics

- - Basic counting principles: multiplication rule, addition rule, and inclusion-exclusion principle
- - Permutations and combinations
- - Binomial coefficients and Pascal's triangle
- - The pigeonhole principle

MODULE 5: Relations and Functions

- - Relations: properties, equivalence relations, and partial orders
- - Functions: injective, surjective, and bijective functions
- - Composition of functions and inverse functions

MODULE 6: Graph Theory

- - Basic terminologies: vertices, edges, paths, and cycles
- - Graph representations: adjacency matrix and adjacency list
- - Connectivity: connected graphs, components, and bridges
- - Eulerian and Hamiltonian paths and cycles

MODULE 7: Trees

- - Definitions and properties of trees
- - Tree traversal algorithms: depth-first search and breadth-first search
- - Spanning trees and minimum spanning trees
- - Applications of trees in computer science

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

MODULE 8: Boolean Algebra

- - Boolean functions and Boolean expressions
- - Logic gates and Boolean circuits
- - Simplification of Boolean expressions: Karnaugh maps and Quine-McCluskey method
- - Boolean algebra theorems and applications

MODULE 9: Discrete Probability

- - Sample spaces, events, and probability axioms
- - Conditional probability and independence
- - Counting techniques in probability
- - Discrete probability distributions: binomial and geometric distributions

MODULE 10: Recurrence Relations

- - Introduction to recurrence relations
- - Solving linear homogeneous recurrence relations with constant coefficients
- - Solving linear non-homogeneous recurrence relations
- - Generating functions and solving recurrence relations using generating functions

References:

1. *Discrete Mathematical Structures*, Bernard Kolman, Robert C. Busby, Sharon Cutler Ross, Pearson/Prentice Hall, 2009
2. *Discrete Mathematics and Its Applications*, Kenneth Rosen, McGrawhill Education, 2017

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

Paper Code: BDS502& BDS592

Paper Name: *Natural Language Processing*

Credits: 3+2

1. Introduction to Natural Language Processing

- - Definition and scope of NLP
- - Applications of NLP
- - NLP challenges and limitations

2. Text Preprocessing

- - Tokenization
- - Stop word removal
- - Stemming and lemmatization
- - Text normalization

3. Language Modeling

- - N-gram models
- - Hidden Markov Models (HMM)
- - Neural language models

4. Syntactic Analysis

- - Part-of-speech tagging
- - Phrase structure parsing
- - Dependency parsing

5. Semantic Analysis

- - Word sense disambiguation
- - Named entity recognition
- - Semantic role labeling

6. Sentiment Analysis

- - Opinion mining
- - Sentiment classification techniques
- - Aspect-based sentiment analysis

7. Machine Translation

- - Rule-based machine translation
- - Statistical machine translation
- - Neural machine translation

8. Information Extraction

- - Named entity recognition
- - Relation extraction
- - Event extraction

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

9. Question Answering

- - Question classification
- - Passage retrieval
- - Answer extraction

10. Advanced Topics

- - Deep learning for NLP
- - Neural network architectures (e.g., recurrent and convolutional models)
- - Transfer learning in NLP

11. NLP Applications and Research Trends

- - Chatbots and virtual assistants
- - Text summarization
- - Dialogue systems
- - Recent developments and future directions in NLP

12. Project Work and Presentations

- - Students will work on a hands-on NLP project and present their findings to the class.

References:

1. "Natural Language Processing: Basic Techniques and Emerging Trends" by Nitin Indurkha and Fred J. Damerau, CRC Press.
2. "Natural Language Processing: A Paninian Perspective" by Amba Kulkarni, Rajeev Sangal, and Vineet Chaitanya, Prentice-Hall of India.
3. "Foundations of Natural Language Processing" by Pushpak Bhattacharyya, Pearson Education India.

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

SEMESTER - 6

Paper Code: BDS601 & BDS691

Paper Name: *Big Data Analysis*

Credits: 3+2

Module 1: Introduction to Big Data Analysis

- - Overview of big data and its characteristics
- - Introduction to Hadoop and MapReduce
- - Introduction to Apache Spark

Module 2: Data Preprocessing and Cleaning

- - Data quality assessment and data cleaning techniques
- - Handling missing data and outliers
- - Feature selection and dimensionality reduction

Module 3: Data Transformation and Integration

- - Data normalization and standardization
- - Encoding categorical variables
- - Data integration techniques

Module 4: Exploratory Data Analysis

- - Descriptive statistics and data visualization
- - Univariate and multivariate analysis
- - Outlier detection and handling

Module 5: Big Data Processing Frameworks

- - Introduction to Apache Hadoop ecosystem (HDFS, Hive, Pig)
- - Introduction to Apache Spark (RDDs, DataFrames, Spark SQL)

Module 6: Big Data Processing with Spark

- - Data processing with Spark RDDs
- - Data processing with Spark DataFrames and Spark SQL
- - Introduction to Spark MLlib for machine learning

Module 7: Machine Learning with Big Data

- - Supervised learning techniques for big data
- - Unsupervised learning techniques for big data
- - Model evaluation and validation

Module 8: Advanced Topics in Big Data Analysis

- - Time series analysis with big data
- - Text mining and natural language processing
- - Graph analysis and social network analysis

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)

Syllabus of *B.Sc. in Data Science*

Effective from academic session 2023-2024

Module 9: Ethical Considerations in Big Data Analysis

- - Privacy issues and data anonymization
- - Bias and fairness in big data analysis
- - Ethical frameworks and guidelines

Module 10: Case Studies and Project

- - Analysis of real-world big data datasets
- - Application of big data analysis techniques
- - Presentation of project findings

References:

1. Pyne, S., & Agrawal, P. (2016). Big Data Analytics: Methods and Applications. CRC Press.
2. White, T. (2015). Hadoop: The Definitive Guide. O'Reilly Media.
3. Zaharia, M., Karau, H., Konwinski, A., & Wendell, P. (2021). Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly Media.

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

Paper Code: BDS602 & BDS692

Paper Name: *Time Series Analysis*

Credits: 3+2

Module 1: Introduction to Time Series Analysis

- - Definition of time series
- - Components of a time series
- - Time series data visualization
- - Applications of time series analysis

Module 2: Descriptive Analysis of Time Series

- - Measures of central tendency and dispersion
- - Time series plots and patterns
- - Seasonality and trend analysis

Module 3: Autocorrelation and Partial Autocorrelation

- - Autocorrelation function (ACF)
- - Partial autocorrelation function (PACF)
- - Identification of autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) models

Module 4: Stationarity and Non-Stationarity

- - Definition and characteristics of stationarity
- - Unit root tests (e.g., Augmented Dickey-Fuller test)
- - Transformations and differencing for achieving stationarity
- - Trend and seasonality decomposition

Module 5: Time Series Forecasting

- - Simple moving average and weighted moving average
- - Exponential smoothing methods (e.g., Holt-Winters method)
- - ARIMA models for forecasting
- - Evaluation and accuracy measures for forecasting models

Module 6: Time Series Regression

- - Simple and multiple linear regression models
- - Incorporating lagged variables in regression models
- - Testing and interpreting the regression coefficients
- - Model diagnostics and residuals analysis

Module 7: Advanced Topics in Time Series Analysis (choose a subset)

- - ARCH and GARCH models for volatility modeling
- - Vector Autoregressive (VAR) models
- - Seasonal ARIMA models (SARIMA)
- - State-space models and Kalman filtering
- - Spectral analysis and periodogram

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

Module 8: Applications of Time Series Analysis

- - Economic and financial time series analysis
- - Environmental and climate data analysis
- - Engineering and industrial applications
- - Social and health-related time series analysis

Module 9: Software Tools for Time Series Analysis

- - Introduction to R or Python programming for time series analysis
- - Hands-on exercises using statistical software packages
- - Implementing forecasting models and analyzing real-world data

Module 10: Case Studies and Project

- - Review of case studies from various domains
- - Student projects involving data analysis and modeling
- - Presentation and discussion of project results

References:

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis: Forecasting and Control. Wiley-Interscience. ISBN: 978-0470272848.
2. Chatfield, C. (2019). Introduction to Time Series Analysis and Forecasting. Chapman and Hall/CRC. ISBN: 978-0367352090.
3. Pandit, A., & Wu, W. B. (2019). Applied Time Series Analysis. CRC Press. ISBN: 978-0367242483.

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

Paper Code: BDS603 & BDS693

Paper Name: *Data Visualization*

Credits: 2+2

Module 1: Introduction to Data Visualization

- - Understanding the importance of data visualization
- - Principles and guidelines for effective data visualization
- - Overview of data visualization process
- - Introduction to data visualization tools

Module 2: Data Types and Visualization Techniques

- - Types of data and their visual representations
- - Categorical data visualization
- - Numerical data visualization
- - Time-series data visualization
- - Geospatial data visualization

Module 3: Visualization Tools and Technologies

- - Introduction to popular data visualization tools (e.g., Tableau, Power BI, D3.js)
- - Exploring features and capabilities of visualization tools
- - Hands-on exercises with visualization tools

Module 4: Data Visualization Design Principles

- - Understanding human perception and cognition in visualization
- - Color theory and effective use of colors in visualization
- - Designing visually appealing and informative visualizations
- - Interaction and user experience design in visualization

Module 5: Advanced Data Visualization Techniques

- - Multivariate data visualization
- - Network visualization
- - Hierarchical visualization
- - 3D and interactive visualizations

Module 6: Storytelling with Data Visualization

- - Creating compelling narratives through visualizations
- - Techniques for presenting data stories effectively
- - Ethical considerations in data visualization
- - Case studies and examples of data storytelling

Module 7: Data Visualization Applications

- - Visualizing data for exploratory analysis
- - Visual analytics and decision support systems
- - Data visualization for business intelligence
- - Data visualization in scientific research

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly known as West Bengal University of Technology)
Syllabus of *B.Sc. in Data Science*
Effective from academic session 2023-2024

Module 8: Data Visualization Project

- - Working on a data visualization project from conception to execution
- - Data collection and preprocessing
- - Designing and implementing visualizations
- - Presenting and documenting the project

References:

1. *Data Visualization: Storytelling Using Data* by Sharada Sringswara; Purvi Tiwari & U. Dinesh Kumar , Wiley
2. "Storytelling with Data: A Data Visualization Guide for Business Professionals" by Cole Nussbaumer Knaflic