MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB

(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)
3rd SEMESTER

BSCDA-301: DATA STRUCTURES AND ALGORITHMS

Objectives

To enable the students to:

- Analyze the asymptotic performance of algorithms.
- Write rigorous correctness proofs for algorithms.
- Demonstrate a familiarity with major algorithms and data structures.
- Apply important algorithmic design paradigms and methods of analysis.
- Synthesize efficient algorithms in common engineering design situations.

Units	Course Content
1	-Data Structures Basics: Structure and Problem Solving, Data structures, Data structure Operations, Algorithm: complexity, Time- space tradeoff. -Linked List: Introduction, Linked lists, Representation of linked lists in Memory, Traversing a linked list, Searching a linked list, Memory allocation and Garbage collection, insertion into linked list, Deletion from a linked list, Types of linked list. -Stack and Queue: Introduction, Array Representation of Stack, Linked List Representation of stack, Application of stack, Queue, Array Representation of Queue, Linked List Representation of Queue. -Trees: Definitions and Concepts, Operations on Binary Trees, Representation of binary tree, Conversion of General Trees to Binary Trees, Sequential and Other Representations of Trees, Tree Traversal. -Graphs: Matrix Representation of Graphs, List Structures, Other Representations of Graphs, Breadth First Search, Depth First Search, Spanning Trees. -Directed Graphs Types of Directed Graphs; Binary Relation As a Digraph; Euler's Digraphs; Matrix Representation of Digraphs.

	-Applications of Graphs: Topological Sorting, Shortest-Path Algorithms – Weighted Shortest Paths – Dijkstra's Algorithm, Minimum spanning tree- Prim's Algorithm, Introduction to NP-Completeness. -Searching and Sorting Techniques, Sorting Techniques: Bubble sort, Merge sort, Selection sort', Heap sort, Insertion Sort. Searching Techniques: Sequential Searching, Binary Searching, Search Trees. -Elementary Algorithms: Notation for Expressing Algorithms; Role and Notation for Comments; Example of an Algorithm; Problems and Instances; Characteristics of an Algorithm; Building Blocks of Algorithms; Procedure and Recursion – Procedure, Recursion; Outline of Algorithms; Specification Methods for Algorithms. -Mathematical Functions and Notations Functions and Notations; Modular Arithmetic / Mod Function; Mathematical Expectation in Average Case Analysis; Efficiency of an Algorithm; Well Known Asymptotic Functions and Notations; Analysis of Algorithms – Simple Examples; Well Known Sorting Algorithms – Insertion sort, Bubble sort, Selection sort, Shell sort, Heap sort. -Divide and Conquer Divide and Conquer Strategy; Binary Search; Max. And Min.; Merge sort; Quick sort. -Greedy Method Greedy Method Strategy; Optimistic Storage on Tapes; Knapsack Problem; Job Sequencing with Deadlines; Optimal Merge Pattern; Single Source Shortlist Paths. Dynamic Programming Dynamic Programming Strategy; Multistage Graphs; All Pair Shortest Paths; Travelling Salesman Problems -Backtracking Strategy, 8-Queens Problem, Sum of Subsets, Knapsack Problem.

References

1. Data Structures and Algorithms made easy, By Narsimha Karumanchi
2. Data Structures and Algorithms, By Aho, Hopcroft and Ulman
3. Data Structures & Algorithms Using C, R.S. Salaria

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB

(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA- 302: DATA WAREHOUSING AND MINING

Objectives

To enable the students to:

- To understand data warehouse concepts, architecture, business analysis and tools
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools

Units	Course Content
1	**Data Warehousing, Business Analysis and On-line Analytical processing (OLAP):** Basic Concepts – Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors – Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP. **Data Mining, Introduction:** Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.
2	**Data Mining, frequent patten Analysis:** Mining Frequent Patterns, Associations and Correlations – Mining Methods- Pattern Evaluation Method – Pattern Mining in Multilevel, Multi Dimensional Space – Constraint Based Frequent Pattern Mining, Classification using Frequent Patterns **Classification and Clustering:** Decision Tree Induction – Bayesian Classification – Rule Based Classification – Classification by Back Propagation – Support Vector Machines — Lazy Learners – Model Evaluation and Selection-Techniques to improve Classification Accuracy. Clustering Techniques – Cluster analysis-Partitioning Methods – Hierarchical Methods – Density Based Methods – Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods.
3	**Weka tool:** Datasets – Introduction, Iris plants database, Breast cancer database, Auto imports database – Introduction to WEKA, The Explorer – Getting started, Exploring the explorer, Learning algorithms, Clustering algorithms, Association–rule learners.

References

1. Jiawei Han and Micheline Kamber, ―Data Mining Concepts and Techniques, Third Edition, Elsevier, 2012.
2. Alex Berson and Stephen J.Smith, ―Data Warehousing, Data Mining & OLAP‖, Tata McGraw – Hill Edition, 35th Reprint 2016.
3. K.P. Soman, Shyam Diwakar and V. Ajay, ―Insight into Data Mining Theory and Practice, Eastern Economy Edition, Prentice Hall of India, 2006.
4. Ian H.Witten and Eibe Frank, ―Data Mining: Practical Machine Learning Tools and Techniques, Elsevier, Second Edition.
5. Machine Learning, Rajiv Chopra, Second Edition, Khanna Publishing House, 2018
6. Machine Learning, Jeeva Jose, Khanna Publishing House

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB

(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science

(Effective for Academic session 2019-20)

BSCDA-303: STATISTICAL THEORY OF HYPOTHESIS TESTING

Objectives

To enable the students to:

- Learn a **statistical hypothesis**, sometimes called **confirmatory data analysis**, which is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables. A **statistical hypothesis test** is a method of statistical inference. Commonly, two statistical data sets are compared, or a data set obtained by sampling is compared against a synthetic data set from an idealized model. A hypothesis is proposed for the statistical relationship between the two data sets, and this is compared as an alternative to an idealized null hypothesis that proposes no relationship between two data sets

Units	Course Content
1	Elements of Hypothesis Testing : Null and Alternative hypotheses, Simple and Composite hypotheses,
2	Critical Region, Type I and Type II Errors, Level of Significance and Size, p-value, Power
3	Tests of Significance related to a single Binomial proportion and Poisson parameter; two Binomial proportions and Poisson parameters;
4	the mean(s) and variance(s) of a single univariate normal distribution,
5	two independent normal distributions and a single bivariate normal distribution;
	Idea of Inference - Point & Interval Estimations and Testing of Hypothesis. Point estimation: Requirements of a good estimator – notions of Mean Square Error, Unbiasedness: Minimum Variance Uunbiasedness and Best Linear Unbisedness, Sufficiency, Factorization Theorem (Discrete case only), Properties of minimum variance unbiased estimators, consistent estimators and asymptotic efficiency, Cramer-Rao lower bound, Rao-Blackwell Theorem. 9 Methods of Estimation – Moment, Least-square, Maximum Likelihood & Minimum $\chi 2$ methods and their properties (excluding proofs of large sample properties).
	References

1. Goon A.M., Gupta M.K. & Dasgupta B. (1994): An Outline of Statistical Theory (Vol-2), World Press
2. Mood A.M., Graybill F. & Boes D.C. (1974): An Introduction to the Theory of Statistics (3rd ed), McGraw Hill 3
3. Rao C.R. (1952): Advanced Statistical Methods in Biometric Research, John Wiley
4. Hogg R.V. & Craig A.T. (1978): Introduction to Mathematical Statistics
5. Rohatgi V.K. (1984): An Introduction to Probability Theory & Mathematical Statistics, John Wiley 6.
6. Stuart G & Ord J.K. (1991): Advanced Theory of Statistics (Vol 2), Charles Griffin 7.
7. Goon A. M., Gupta M. K. and Dasgupta B. (1997): Fundamentals of Statistics (V-1 and 2), World Press 8.
8. Bhattacharya GK & Johnson R. A. (1977): Concepts & Methods of Statistics, John Wiley
9. Manish Sharma & Amit Gupta, The Practice of Business Statistics, Khanna Publishing House
10. Scheffe H. (1959): The Analysis of Variance, John Wiley

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB

(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-304: DATA VISUALISATION USING BUSINESS INTELLIGENCE

Objectives:

To enable the students to:

Analyse large corporate datasets using Business Intelligence/Business Analytics tools to generate insights and provide alternative solutions to an organisation's complex problems

Unit	Course Content
1	Concept of business value from corporate data, the exploitation of information for advantage, types and sources of information value • Nature and value of business intelligence, the business intelligence environment, and how types of data processing can add value to corporate data sources
2	• Knowledge discovery, data mining, data warehousing • Business analytics, OLAP analysis, metadata • Data visualisation, visualisation techniques, dashboard
3	• The relationship between corporate strategy, IS strategy and business intelligence strategy • BI links to enterprise systems, CRM (Customer Relationships Management), SCM (Supply Chain Management) Structured & unstructured data, content management systems
4	• Privacy, ethical, legal issues associated with BI Implementation • BI, Decision Support Systems, Expert Systems and Executive Information Systems • Data modelling, star schemas
5	• Using select and complementary BI/BA tools to provide insights in decision making scenarios • Use of Power BI functions

References

1. Big Data Visualization, By James W Miller.
2. High Impact Data Visualization with Power View, Power Map, and Power BI, By Adam Aspin
3. Big Data & Hadoop, V.K. Jain

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WB

(Formerly West Bengal University of Technology)

Syllabus for B. Sc. in Data Science
(Effective for Academic session 2019-20)

BSCDA-305: WEB FOUNDATION

Objectives

To enable the students to:

Learn about internet and world wide web, Digital Data

Units	Course Content
1	**Introduction** Introduction to Internet and www, web browsers, web servers, search engine
2	**Internet Commerce**
3	**Digital Marketing**
4	**Web Analytics**
5	**Social Media Analytics**

References

1. Fundamentals of Web Development, By Connolly and Hoar
2. Internet & Web Development, By Soma Das Gupta