Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Semester III

Paper: Time Series Analysis and Forecasting
Code: MDS 301
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:
equip students with various forecasting techniques and knowledge on modern statistical methods for analyzing time series data

UNITS	COURSE CONTENT
1	Econometric time series Different components, illustration, additive and multiplicative models, determination of trend, seasonal and cyclical fluctuations.
2	Time-series as discrete parameter stochastic process, auto covariance and autocorrelation functions and their properties
3	Exploratory time Series analysis, tests for trend and seasonality, exponential and moving average smoothing. Detailed study of the stationary processes: (1) moving average (MA), (2) auto regressive (AR), (3) ARMA and (4) AR integrated MA (ARIMA) model
4	Box-Jenkins models, choice of AR and MA periods. Discussion (without proof) of estimation of mean, auto covariance and autocorrelation functions under large sample theory, estimation of ARIMA model parameters. Spectral analysis of weakly stationary process, periodogram and correlogram analyses, computations based on Fourier transform, non stationary process, introduction to forecasting

References:

1. Introduction to time series and forecasting Book by Rob J. Hyndman
2. Time series analysis, forecasting and control Book by George E.P.Box

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Paper: Data Mining
Code: MDS 302
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:

Understand data and web mining and cluster analysis techniques and their applications

UNITS	COURSE CONTENT
1	Introduction: Data Mining Concept, Origin, Process, Applications, Techniques, Challenges Data Preprocessing: Data types, Quality, Descriptive data summarization – central tendency and dispersion measure, Data cleaning, Data integration & transform, Data reduction
2	Association Rule Mining: Market-basket analysis basics, Naïve algorithm, Apriori algorithm, Direct Hashing and Pruning (DHP), Software for Association Rule Mining Classification and Prediction: Decision Tree, Classification by decision tree induction, Bayesian classification, Rule-based classification, Prediction – Linear and Nonlinear Regression, Classification software
3	Cluster Analysis: Types of data in cluster analysis, Partitioning methods, Hierarchical methods, Density-based methods, Quality & Validity of clustering methods Cluster analysis software Web Data Mining: Web content mining, Web usage mining, Web structure mining, Hubs and Authorities, HITS algorithm, Web mining software
4	Text Mining, Support Vector Machine. Data Mining Application & Information Privacy: Applications and trends in data mining such as Web, finance, telecommunication, biology and medicine, science and engineering retail industry etc. Social impacts of data mining, information privacy and data security, IT Act overview

References:

1. Tan, Steinbach and Kumar, Introduction to Data Mining, Pearson
2. Han and Camber, Data Mining: Concepts and Techniques, Morgan Kaufmann
3. Foreman, Data Smart:Using Data Science to Transform Information into Insight, John Wiley
4. Dunham, Data Mining : Introductory and Advanced Topics, Pearson

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Paper: Big Data Technology
Code: MDS 303
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:

Understand data and web mining and cluster analysis techniques and their applications

UNITS	COURSE CONTENT
1	Big data definition, structured and unstructured data
2	Need for analytics, Big data programming (Hadoop, Map-Reduce), Application Data store (NoSQL), OLAP
3	Optimization Techniques, Data flow framework. Programming Map-Reduce, Best practices

References:

1. Handbook of big data technology by Zomaya and Sakr.
2. Real time Big Data Analytics Book by Sumit Gupta

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Paper: Design and Analysis of Algorithm
Code: MDS 304
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:

Understand the theoretical backbone of computer science and are a must in the daily work of the successful programmer. The goal of this course is to provide a solid background in the design and analysis of the major classes of algorithms

UNITS	COURSE CONTENT
1	Basic Computational Model and analyzing Algorithms, Asymptotic Notation and recurrence relations
2	Fundamental design methodologies and their implementations: Dynamics Programming, Greedy algorithms, Divide and Conquer, Branch and Bound, Backtracking, Randomized Techniques
3	Algorithms for set manipulations, their implementations and applications: Union-Find, Priority Queues. Graph Algorithms with implementation issues; Depth-First Search and its applications, minimum Spanning Trees and Shortest Paths
4	Matrix multiplication, Pattern Matching, polynomial arithmetic and FFT. Introduction to the Theory of Lower Bounds, NP-Completeness and Reductions

References:

1. E. Horowitz & S. Sahani : Fundamental of Computer Algorithm (Galgotia)
2. Coreman, Leiserson & Rivest : Introduction to Algorithm (MIT)
3. Brassard & Brately : Algorithm- Theory and Practice (PHI)

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Paper: Machine Learning
Code: MDS 305
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:

state-of-the-art methods and modern programming tools for data analysis using machine learning programs and algorithms

UNITS	COURSE CONTENT
1	Supervised Learning, Decision Tree, Linear Discriminant Functions (SVM) Neural Network, Deep belief network, Density elimination Methods
2	Bayes Decision Theory
3	Expectation and Minimization
4	Expectation and Minimization

References:

1. Machine Learning and Knowledge Discovery edited by Walter Daelemans, Katharina Morik
2. Pattern Recognition and Machine Learning by Christopher Bishop
3. Introduction to Machine learning with python by Andreas C. Müller and Sarah Guido

Maulana Abul Kalam Azad University of Technology, West Bengal
(Formerly West Bengal University of Technology)
Syllabus for M. Sc. In Data Science
(Effective for Academic Session 2019-2020)

Paper: Data Mining Lab
Code: MDS 391
Contacts Hours / Week: 3L+1T
Credits: 4

Manipulating strings, Processing Files, Manipulating Lists, Lists and Strings, Dictionarys, Counting with Dictionaries, Dictionaries and Files, Tuples, Tuples and Sorting, Regular Expressions, Networked programs, Sockets and Applications, parsing HTML with Beautifulsoup, parsing XML by python, REST,JSON and APIs, Extracting data from JSON, Using database by python, Object oriented python, Geocoding, Page rank and web searching, Gmane.

Paper: Big Data Technology and OLTP Lab
Code: MDS 392
Contacts Hours / Week: 1L+3P
Credits: 4

A. NoSQL Lab using (MongoDB/Redis/Cassandra/CouchDB/Hbase using HDFs etc): Introduction to Nosql,Difference between RDBMS to NOSQL,JSON and BSON documents,Introduction to MongoDB/.. and its Features,Database, Collection and Documents,Various Data Types in MongoDB/..,Introduction to mongo/.. shell,CRUD Operations,Database Operations,Read and Write Operations,Aggregation,Data Modeling Introduction,Data Modeling Concept,Storage Engine,Indexing,Replication Concept,Failover & Recovery
B. Multidimensional Data Modeling using OLAP: Introduction of Datawarehousing and OLAP, example of a DataWarehouse and Data mart, Data Cleaning and integration, Data analysis techniques, Transformation algorithms, Integrations.