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

Semester II

Paper: Modeling Techniques
Code: MDS 201
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:
To enable the students to: to understand various system modeling and simulation techniques, and highlight their applications in different areas. It includes modeling, design, simulation, planning, verification and validation

UNITS	COURSE CONTENT
1	Non-parametric tests- Goodness of fit, Test of independence, sign, run, Wilcoxon, Mann-Whitney, Wald-Wolfowitz. Linear Modelling: Theory of linear estimation, Gauss-Markov linear models, estimable functions, error and estimation space, normal equations and least square estimators, estimation of error variance, estimation with correlated observations, properties of least square estimators, generalized inverse of a matrix and solution of normal equations, variances and covariances of least square estimators
2	Stochastic models One way and two-way classifications, fixed, random and mixed effects models. Analysis of variance and Design of Experiment, Multivariate Analysis. Elements of Decision Theory and Bayesian approach

References:
1. Basic Business statistics Book by David Levine, Mark Berenson and Timothy C Krehbiel 2. Business Statistics : Contemporary Decision making Book by Jacquelyn G Black 3. Complete Business Statistics Book By Amir Aczel

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

Paper: **Advanced Programming in Python**
Code: MDS 202
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:
To enable the students to: learn how to analyze data in Python using multi-dimensional arrays in numpy, manipulate DataFrames in pandas, use SciPy library of mathematical routines,

UNITS	COURSE CONTENT
1	**Python Basic:** Python fundamental, working with data
2	**Importing Dataset** Domain, Dataset, Package for Data Science, Importing/Exporting Data, Insight from Dataset
3	**Cleaning and Preparing the Data** Identify and Handle Missing Values, Data Formatting, Normalisation, Binning
4	**Summarising the Data Frame** Descriptive Statstics, Grouping, ANOVA, Correlation
5	**Model Development** Linear Regression, Prediction and Decission making
6	**Data Vizualization** Introduction to Matplotlib, Basic plotting, Charts
7	**Artificial Libraries in Python** Introduction to TensorFlow, Sciket-learn and Theano

References:

1. Advanced Machine Learning with Python, By John Hearty

2. Introduction to Machine Learning using Python, Jeeva Jose, Khanna Publishing House

3. Taming Python by Programming, Jeeva Jose, Khanna Publishing House

Paper: Data Warehousing
Code: MDS 203
Contacts Hours / Week: 3L+1T
Credits: 4

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, Khanna Publishing House

6. Introduction to Machine Learning using Python, Jeeva Jose, Khanna Publishing
   House

Paper: Data Visualisation using Business Intelligence
Code: MDS 204
Contacts Hours / Week: 3L+1T
Credits: 4

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 organization's complex problems

UNITS	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 • Knowledge discovery, data mining, data warehousing • Business analytics, OLAP analysis, metadata • Data visualisation, visualisation techniques, dashboard • 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 • Privacy, ethical, legal issues associated with BI Implementation • BI, Decision Support Systems, Expert Systems and Executive Information Systems • Data modelling, star schemas • 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, Khanna Publishing House

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

Paper: Cloud Computing
Code: MDS 205
Contacts Hours / Week: 3L+1T
Credits: 4

Objectives:

To enable the students to:

provide with the comprehensive and in-depth knowledge of Cloud Computing concepts, technologies, architecture and applications by introducing and researching state-of-the-art in Cloud Computing fundamental issues, technologies, applications and implementations

UNITS	COURSE CONTENT
1	Introduction: Cloud computing definition, reference model, Characteristics, Benefits, Challenges, Distributed Systems, Virtualization, Service-oriented computing, Utility-oriented computing, Overview on computing platforms & technologies – AWS, Google AppEngine, MS Azure, Hadoop, saleforce.com, Manjrasoft Aneka
2	Parallel & Distributed Computing: Parallel vs. Distributed computing, Elements of parallel computing, Parallel processing - hardware architecture & approaches, Concept & Component of Distributed Computing, RPC, Service-oriented computing Virtualization: Cloud reference model – IaaS, PaaS, SaaS, Types of clouds – Public, Private, Hybrid, Community, Cloud interoperability & standards, scalability & fault tolerance, Security, trust & privacy
3	Concurrent Computing, High-throughput Computing and Data-Intensive Computing: Programming applications with Threads, Thread API, Parallel computation with Threads, Task computing, Frameworks for Task computing, Task-based application model, Data-intensive computing, characteristics, technology
4	Cloud Platforms and Applications: Overview on Amazon Web Services, Google AppEngine and Microsoft Azure, Cloud applications in scientific, business and consumer Domain

References:

1. Buyya, Vecciola and Selvi, Mastering Cloud Computing: Foundations and Applications Programming, Tata McGraw Hill
2. Rittinghouse and Ransome, Cloud Computing: Implementation, Management, and Security, CRC Press
3. Aravind Doss, Cloud Computing, Tata McGraw Hill
4. Kris Jamsa, Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More, Jones & Bartlett Learning

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

Paper: Advanced Python Lab
Code: MDS 291
Contacts Hours / Week: 1L+1T+2P
Credits: 4

Python fundamental, working with data, Domain, Dataset, Package for Data Science, Importing/Exporting Data, Insight from Dataset, Identify and Handle Missing Values, Data Formatting, Normalisation, Binning, Descriptive Statstics, Grouping, ANOVA, Correlation, Introduction to Matplotlib, Basic plotting, Charts, Introduction to TensorFlow, Sciket-learn and Theano

Paper: Analytics Lab II
Code: MDS 292
Contacts Hours / Week: 2L+2P
Credits: 4

Functions, Some guideline of writing good functions, functional programming, debugging and profiling, Object oriented programming, Linear modeling, ANOVA and DOE, Tiem series Analysis.