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 - 3

Paper Code: BDS301
Paper Name: *Statistics for Data Science - 1*
Credits: 5

Module I: Introduction to Statistics

- Definition and scope of statistics
- Role of statistics in data science
- Types of data: categorical and numerical
- Descriptive statistics vs. inferential statistics
- Appropriate laboratory practice in R

Module II: Correlation & Simple Linear Regression

- Bivariate data: Definition, scatter diagram, Correlation & Rank correlation. Simple linear regression, principle of least squares and fitting of polynomials and exponential curves.
- Theory of attributes, consistency of data, independence and association of attributes, measures of association and contingency
- Assumptions of linear regression
- Estimating regression coefficients
- Residual analysis and model diagnostics
- Appropriate laboratory practice in R

Module III: Sampling Techniques

- Population vs. sample
- Probability and Non-probability Sampling
- Bias and sampling error
- Sample size determination
- Appropriate laboratory practice in R

Module IV: Probability Theory

- Basic probability concepts: events, sample space, probability axioms

- Conditional probability and independence
- Bayes' Theorem
- Joint distribution & Random variables
- Expectation & Variance
- Appropriate laboratory practice in R

Module V: Probability Distributions

- Discrete probability distributions: Binomial, Poisson, Geometric
- Continuous probability distribution: Uniform, Exponential, Normal, Normal approximation of Binomial and Poisson
- Sampling distribution and central limit theorem
- Appropriate laboratory practice in R

Module VI: Statistical Inference - Hypothesis Testing 1

- Null and alternative hypotheses
- Type I and Type II errors, power of a test
- Test statistics and p-values
- Appropriate laboratory practice in R

References:

1. *The Practice of Business Statistics By Manish Sharma, Khanna Publishing House (AICTE Recommended Textbook)*
2. *Statistical Methods* by N. G. Das, Combined edition volume 1 & 2, McGraw Hill Education.
3. *Statistical Programming in R* by K.G. Srinivasa, G.M. Siddesh, Chetan Shetty , Sowmya B.J. , Oxford University Press
4. *A First Course in Statistical Programming with R* (3rd ed.) by Braun, W., & Murdoch, D. (2021), Cambridge University Press.

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: BDS302 & BDS392
Paper Name: *Artificial Intelligence & Machine Learning - 1 & AI ML-1 Lab*
Credits: 5 (3Th+2P)

Module 1: Introduction to Artificial Intelligence
- Definition and scope of Artificial Intelligence
- Historical overview and milestones in AI
- AI applications in various fields
- Agents and environments

Module 2: Problem Solving and Searching
- Problem-solving agents
- State space representation
- Uninformed search algorithms (breadth-first search, depth-first search)
- Informed search algorithms (heuristic search, A* algorithm)
- Appropriate laboratory work in Python/R

Module 3: Knowledge Representation and Reasoning
- Propositional logic
- First-order logic and predicate calculus
- Knowledge representation using logic
- Inference rules and resolution

Module 4: Expert Systems
- Introduction to expert systems
- Knowledge engineering process
- Rule-based systems
- Forward chaining and backward chaining

Module 5: Introduction to Statistical Learning
- Overview of statistical learning and its applications
- Supervised vs. unsupervised learning

- Bias-variance tradeoff
- Performance evaluation and cross-validation

Module 6: Linear Regression
- Simple linear regression
- Multiple linear regression
- Model assessment and selection
- Polynomial regression and model flexibility
- Appropriate laboratory work in Python/R

Module 7: Classification
- Logistic regression
- Linear discriminant analysis (LDA)
- K-nearest neighbors (KNN) classification
- Model evaluation and validation techniques
- Appropriate laboratory work in Python/R

Module 8: Resampling Methods
- Cross-validation
- Bootstrap methods for estimating standard errors
- Model selection using cross-validation and bootstrapping
- Appropriate laboratory work in Python/R

Module 9: Dimensionality Reduction
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)
- Applications of dimensionality reduction
- Appropriate laboratory work in Python/R

Module 10: Introduction to Machine Learning
- Overview of different machine learning algorithms
- Decision Trees and Random Forests
- Support Vector Machines (SVM)
- Ensemble methods (bagging and boosting)
- Appropriate laboratory work in Python/R

Module 11: Unsupervised Learning

- K-means clustering
- Hierarchical clustering
- Association rules and market basket analysis
- Anomaly detection
- Appropriate laboratory work in Python/R

References:

1. M.C. Trivedi (2023). A Classical Approcah to Artificial Intelligence, Khanna Publishing House (AICTE Recommended Textbook)
2. Dan W. Patterson (1990). *Introduction to artificial intelligence and expert systems*. Englewood Cliffs, N.J : Prentice Hall
3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* with applications in R *(2nd ed.)*. Springer.
4. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An introduction to statistical learning* with applications in Python *(1st ed.)*. Springer.
5. M.C. Trivedi (2024). Artificial Intelligence & Machine Learning, Khanna Publishing House

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 - 4

Paper Code: BDS401
Paper Name: *Statistics for Data Science - 2*
Credits: 4

Module 1: Multiple Linear Regression

- Multiple regression model and interpretation of coefficients
- Model selection techniques: forward selection, backward elimination, etc.
- Assumptions and diagnostics in multiple regression: Collinearity, Heteroscedasticity & Autocorrelation
- Appropriate laboratory practice in R

Module 2: Statistical Inference - Hypothesis Testing 2

- Tests for means
- Tests for proportions
- Tests for variances
- Testing of correlation & regression coefficients
- Appropriate laboratory practice in R

Module 3: ANOVA & DOE

- One-way classified data, two-way classified data with equal number of observations per cell
- Concepts of multiple correlation and partial correlation coefficients.
- Two-way classified data with equal number of observations per cell
- Factorial experiments and interaction effects
- Analysis of variance (ANOVA) for experimental data
- Appropriate laboratory practice in R

Module 4: Nonparametric Methods

- Chi-square test
- Wilcoxon signed-rank test and Mann-Whitney U test
- Kruskal-Wallis test and Friedman test
- Rank correlation: Spearman's rank correlation coefficient
- Appropriate laboratory practice in R

References:

1. The Practice of Business Statistics by Manish Sharma, Khanna Publishing House (AICTE Recommended Textbook)
2. Beginner's Guide for Data Analysis using R Programming, Jeeva Jose, Khanna Publishing House (AICTE Recommended Textbook)
3. *Statistical Methods* by N. G. Das, Combined edition volume 1 & 2, McGraw Hill Education.
4. *Statistical Programming in R* by K.G. Srinivasa, G.M. Siddesh, Chetan Shetty , Sowmya B.J. , Oxford University Press
5. *A First Course in Statistical Programming with R* (3rd ed.) by Braun, W., & Murdoch, D. (2021), Cambridge University Press.

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: BDS402 & BDS492
Paper Name: *Artificial Intelligence & Machine Learning - 2 & AI ML-2 Lab*
Credits: 4 (2Th+2P)

1. **Introduction to Artificial Neural Networks**
   - Historical overview and motivations
   - Biological inspiration and neuron model
   - Basic components of neural networks
   - Types of neural networks (e.g. feedforward, recurrent)
2. **Perceptrons and Activation Functions**
   - McCulloch - Pitt Neuron & Single-layer perceptron model
   - Activation functions and their properties
   - Multilayer perceptron architecture
   - Training and learning algorithms for perceptrons
   - Appropriate laboratory work in Python/R
3. **Feedforward Neural Networks**
   - Feedforward network structure and architecture
   - Forward propagation and activation functions
   - Backpropagation algorithm for training
   - Overfitting, underfitting, and regularization techniques
   - Appropriate laboratory work in Python/R
4. **Radial Basis Function Networks**

- Introduction to radial basis functions
- RBF network architecture and training
- Applications of RBF networks in pattern recognition
- Appropriate laboratory work in Python/R

5. Self-Organizing Maps

- Introduction to Kohonen Self-Organizing Maps (SOMs)
- SOM architecture and learning algorithm
- Clustering and visualization using SOMs
- Applications of SOMs in data analysis
- Appropriate laboratory work in Python/R

6. Hopfield Networks

- Introduction to associative memory
- Hopfield network architecture and energy function
- Pattern storage and retrieval in Hopfield networks
- Applications of Hopfield networks in optimization problems
- Appropriate laboratory work in Python/R

7. Neural Networks in Pattern Recognition

- Feature extraction and selection techniques
- Handwritten digit recognition using neural networks
- Face recognition using neural networks
- Applications of neural networks in image classification
- Appropriate laboratory work in Python/R

References:

1. "Neural Networks and Learning Machines" by Simon Haykin. Publisher: Pearson Date of Publication: 2021
2. “Deep Learning & Neural Networks” by M.C. Trivedi, Khanna Publishing House: 2024.
3. "Pattern Recognition and Machine Learning" by Christopher Bishop Publisher: Springer, 2006
4. “Neural Networks and Deep Learning” by Charu C. Aggarwal, Springer
5. “Principles of Soft Computing. 2nd Edition” by Sivanandam, S.N. and Deepa, S.N. (2009), John Wiley & Sons, Inc.

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: BDS403 & BDS493
Paper Name: *Python for Data Science & Python for Data Science Lab*
Credits: 5 (3Th+2P)

Module 1: Emerging Trends in Computing and Role of Python

- Artificial Intelligence (AI)
- Big Data
- Internet of Things (IoT)
- Cloud Computing
- Grid Computing
- Blockchains

Module 2: Overview of Python

- Introduction to Python
- Python Keywords and Identifiers
- Variables and Data Types
- Operators and Expressions
- Input and Output
- Debugging
- Functions
- `if..else` Statements
- `for` Loops, `while` Loops and Nested Loops

Module 3: Working with Lists and Dictionaries

- Introduction to List
- List Operations
- Traversing a List
- List Methods and Built-in Functions
- List Manipulation
- Introduction to Dictionaries
- Traversing a Dictionary
- Dictionary Methods and Built-in Functions
- Manipulating Dictionaries

Module 4: Understanding Data under Python

- Introduction to Data
- Data Collection
- Data Storage
- Data Processing
- Statistical Techniques for Data Processing

Module 5: Introduction to NumPy

- Introduction
- Array and NumPy Array
- Indexing and Slicing
- Operations on Arrays
- Concatenating Arrays
- Reshaping Arrays
- Splitting Arrays
- Statistical Operations on Arrays
- Loading Arrays from Files
- Saving NumPy Arrays in Files on Disk

Module 6: Data Handling Using Pandas - I

- Introduction to Python Libraries
- Series
- DataFrame
- Importing and Exporting Data between CSV
- Files and DataFrames
- Pandas Series Vs NumPy *ndarray*

Module 7: Data Handling Using Pandas - II

- Introduction to Descriptive Statistics
- Data Aggregations
- Sorting a DataFrame
- GROUP BY Functions
- Altering the Index
- Other DataFrame Operations
- Handling Missing Values
- Import and Export of Data between Pandas and MySQL

Module 8: Plotting Data Using Matplotlib

- Introduction
- Plotting using Matplotlib
- Customisation of Plots
- The Pandas Plot Function (Pandas Visualisation)

Module 9: Minor Project using Python

- Introduction to Project Based Learning
- Approaches for Solving Projects
- Importance of Teamwork
- Project Planning, Execution and Reporting

References:

1. Introduction to Computing & Problem Solving With PYTHON, Jeeva Jose, Khanna Publishing House (AICTE Recommended Textbook)
2. Taming PYTHON By Programming, Khanna Publishing House (AICTE Recommended Textbook)
3. Introduction To Python Programming, Venkatesh, Nagaraju Y, Khanna Publishing House
4. *Python Data Science Handbook: Essential Tools for Working with Data* - Jake VanderPlas, O'Reilly Inc
5. *Mastering Python for Data Science* (1st ed.) - Samir Madhavan (2015), Packt Publishing
6. *Python Official Documentation* - The Python Software Foundation, http://www.python.org