MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
(Formerly West Bengal University of Technology)
Syllabus of B.Sc. In IT (Data Science)
(Effective from 2023-24 Academic Sessions)

Curriculum Structure

SEM-II

Sl.	Course Type	Code	Course Name	Credits			Total Credits
				L	T	P	
1	Major	BSCITDSM201 BSCITDSM291	Introduction to Data Science Introduction to Data Science Lab	3	0	2	5
2	Major	BSCITDSM202	Study of Computer Architecture and organization	4	1	0	5
3	Minor	MIM201	Organization Behavior	3	0	0	3
4	GE		Any one from GE Basket B or E	3	0	0	3
5	AECC	AECC201	Modern Indian Languages and Literature	2	0	0	2
6	SEC	SEC201	IT Skills	2	0	0	2
7	VAC	VAC281A VAC281B VAC281C VAC281D	Critical Thinking NSS Mental Health Environmental Studies	2	0	0	2
Total Credits							22

Course: Introduction to Data Science/ Introduction to Data Science Lab
Credits: 3L + 2P
Course Code: BSCITDSM201/ BSCITDSM291
COURSE OBJECTIVE:
This Course provides the knowledge and expertise to become a proficient data scientist. It demonstrate an understanding of statistics and machine learning concepts that are vital for data science. Further, it produces Python code to statistically analyse a dataset and critically evaluate data visualisations based on their design and use for communicating stories from data.

COURSE OUTCOME	
CO1	Understand the core concepts and technologies of data science for data collection, management and data storage
CO2	Apply various data analysis techniques on data sets.
CO3	Examine the various data visualization types and identify the type to be applied.
CO4	Investigate the applications of data science in various domains.
CO5	Examine recent trends in data collection and analysis techniques.

Module 1:
Introduction: What is Data Science? Big Data and Data Science – Datafication - Current landscape of perspectives - Skill sets needed; Matrices - Matrices to represent relations between data, and necessary linear algebraic operations on matrices -Approximately representing matrices by decompositions (SVD and PCA); Statistics: Descriptive Statistics: distributions and probability - Statistical Inference: Populations and samples - Statistical modeling - probability distributions - fitting a model - Hypothesis Testing - Intro to R/ Python.

Module 2:
Data preprocessing: Data cleaning - data integration - Data Reduction Data Transformation and Data Discretization.Evaluation of classification methods – Confusion matrix, Students T-tests and

ROC curves-Exploratory Data Analysis - Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA - The Data Science Process.

Module 3:
Basic Machine Learning Algorithms: Association Rule mining - Linear Regression- Logistic Regression - Classifiers - k-Nearest Neighbors (k-NN), k-means -Decision tree - Naive Bayes- Ensemble Methods - Random Forest. Feature Generation and Feature Selection - Feature Selection algorithms - Filters; Wrappers; Decision Trees; Random Forests.

Module 4:
Clustering: Choosing distance metrics - Different clustering approaches - hierarchical agglomerative clustering, k-means (Lloyd's algorithm), - DBSCAN - Relative merits of each method - clustering tendency and quality.

Module 5:
Data Visualization: Basic principles, ideas and tools for data visualization.

List of Practical
1. **Exploring Datasets:**
Choose a dataset from sources like Kaggle, UCI Machine Learning Repository, or government datasets. Load the dataset using Python libraries (Pandas) and explore its contents.
2. **Data Cleaning and Preprocessing:**
Handle missing values by imputing or removing them. Clean and transform data (e.g., converting categorical variables, standardizing numerical variables).
3. **Basic Data Visualization:**
Create bar charts, histograms, scatter plots, and line charts to visualize data using libraries like Matplotlib and Seaborn.
4. **Descriptive Statistics:**
Calculate basic summary statistics (mean, median, mode, standard deviation, etc.). Generate box plots and violin plots to visualize data distributions.
5. **Simple Linear Regression:**
Implement a simple linear regression model using Scikit-Learn. Plot the regression line and interpret the model's coefficients.
6. **Classification Basics:**
Learn about binary classification and implement logistic regression for predicting binary outcomes.
7. **Time Series Analysis:**
Analyze time series data (e.g., stock prices) using Pandas. Plot time series data and calculate rolling averages.
8. **Data Visualization Projects:**
Create an infographic or data story using visualizations to communicate insights from a dataset.
9. **Predictive Modeling Projects:**
Choose a dataset, preprocess it, and build predictive models for regression or classification tasks.
10. **Mini Capstone Project:**
Combine multiple skills to tackle a small-scale data science project of your own choice.

Text Books:

1. Doing Data Science, Straight Talk From The Frontline - Cathy O'Neil and Rachel Schutt, O'Reilly, 2014.
2. Data Mining: Concepts and Techniques - Jiawei Han, Micheline Kamber and Jian Pei, Third Edition. ISBN 0123814790, 2011.

3. Mining of Massive Datasets. v2.1 - Jure Leskovek, AnandRajaraman and Jeffrey Ullman, Cambridge University Press.

Reference Books:

1. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking - Foster Provost and Tom Fawcett, ISBN 1449361323, 2013
2. Elements of Statistical Learning - Trevor Hastie, Robert Tibshirani and Jerome Friedman, Second Edition. ISBN 0387952845, 2009

Course: Study of Computer Architecture and organization
Credits: 4L + 1T
Course Code: BSCITDSM202
COURSE OBJECTIVE:

- To understand the structure, function and characteristics of computer systems.
- To understand the design of the various functional units and components of computers.
- To identify the elements of modern instructions sets and their impact on processor design.
- To explain the function of each element of a memory hierarchy,
- To identify and compare different methods for computer I/O.

COURSE OUTCOME	
CO1	Understand basic structure of digital computer, instruction set, number system, and arithmetic operations.
CO2	Understand basic structure of stored program concept and different arithmetic and control unit operations.
CO3	Become skilled at about memory hierarchy and mapping techniques.
CO4	Study the techniques that computer use to communicate with peripheral devices.
CO5	Understand parallel architecture, pipelines, and interconnection network.
CO6	Design the non Von-Neumann architectures.

Module 1:
Data Representation: Number Systems – decimal, binary, octal, hexadecimal, alphanumeric representation, 2. Complements – 1's complement, 2' complement, 9's complement, 10' complement, [r-1]'s complement, r's complement, 3. Fixed point representation – Integer representation, arithmetic addition, arithmetic subtraction, overflow, decimal fixed point representation, 4. Floating point representation, 5. IEEE 754 floating point representation

Module 2:
Computer arithmetic: Addition algorithm of sign magnitude numbers, Subtraction algorithm of sign magnitude numbers, Addition algorithm of signed 2's complement data, Subtraction algorithm of signed 2's complement data, Multiplication algorithm, Booth's algorithm, Division algorithm

Module 3:
Register transfer and micro-operations: Register transfer language, Register transfer, Bus system for registers, Memory transfers – memory read, memory write, Micro operations – register transfer micro operations, arithmetic micro operations, logic micro operations, shift micro operations, Binary adder, binary adder subtractor, binary incrementer, arithmetic circuit for

arithmetic micro operations, One stage logic circuit,Selective set, Selective complement, Selective clear, Mask, Insert, Clear.

Module 4:
Basic Computer organization and design: Instruction codes, Direct address, Indirect address & Effective address, List of basic computer registers, Computer instructions: memory reference, register reference & input – output instructions,Block diagram & brief idea of control unit of basic computer, 6. Instruction cycle.

Module 5:
Micro programmed control: Control memory, Address sequencing,Micro program examples.

Module 6:
Central processing unit: General register organization,Stack organization, Register stack, Memory stack, Stack operations – push & pop, Evaluation of arithmetic expression using stack, Instruction format, Types of CPU organization [single accumulator, general register & stack organization] & example of their instructions, Three, two, one & zero address instruction, Definition and example of data transfer, data manipulation & program control instructions, Basic idea of different types of interrupts [external, internal & software interrupts], Difference between RISC & CISC.

Module 7:
Pipeline and vector processing: Parallel processing, Flynn's classification, Pipelining, Example of pipeline, space time diagram, speedup, Basic idea of arithmetic pipeline, example of floating point addition/ subtraction using pipeline.

Module 8:
Input – output organization: Peripheral devices,Input – output interface, Isolated I/O, Memory mapped I/O, Asynchronous data transfer: strobe & handshaking, Programmed I/O, Interrupt initiated I/O, Basic idea of DMA & DMAC, Input – output processor.

Module 9:
Memory organization: Memory hierarchy, Main memory definition, types of main memory, types of RAM, ROM, difference between SRAM & DRAM, Cache memory, Cache memory mapping – Direct, Associative, Set Associative, CAM, hardware organization of CAM, Virtual memory, mapping using pages, page fault, mapping using segments, TLB, Auxiliary memory, diagrammatic representation of magnetic disk & hard disk drive, Definitions of seek time, rotational delay, access time, transfer time, latency.

Text Books:

1. Computer System Architecture - M. Morris Mano, THIRD EDITION, PEARSON.
2. Computer Organization & Architecture – Designing For Performance - William Stallings, 9th Edition, PEARSON.
3. Computer Architecture & Organisation - J.P. Hayes, 3rd Edition, TATA MCGRAW HILL.

Reference Books:

1. Computer Organization and Architecture - T. K. Ghosh, Edition: 3rd, TATA MCGRAW-HILL.
2. Computer Architecture - Behrooz Parhami, SECOND EDITION, OXFORD UNIVERSITY PRESS.