BON SECOURS COLLEGE FOR WOMEN (AUTONOMOUS)
Accredited with 'A++' Grade by NAAC in Cycle II
Affiliated to Bharathidasan University, Tiruchirappalli
UGC Recognized 2(f) and 12(B) Institution
Vilar Bypass, Thanjavur – 613 006

DEPARTMENT OF COMPUTER SCIENCE
B. Sc., Data Science

SEMESTER-I

Core Course-1 – PRINCIPLES OF PROGRAMMING WITH C
Course Code: 24UCP401

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS

UNIT I : Principles of Computer

Principles of Computers: Introduction to computer –Classification of Computers-Basic Anatomy of a Computer System-Input Devices- -Output Devices-Memory Management –Programming Languages- Algorithms – Flow charts – Developing algorithms and flowcharts for solving simple problems using sequential, selection and iterative programming Structures.

UNIT II : Overview and Decision Making, Looping

History of C and its importance – Structure of a C program – Data Types – Constants and Variables – Operators and Expressions – Control structures – Looping structures.

UNIT III : Arrays & Functions

Arrays – single dimensional arrays – declaration –memory representation– initialization and access. 2D arrays and multidimensional arrays- Strings – defining strings, initializing, accessing, character handling functions- Functions – definition-declaration-prototypes and function call- actual and formal arguments-types of functions-call by value-call by reference-nesting of functions-recursive functions

UNIT – IV :Pointers

Introduction – Pointer Expressions – Chain of Pointers –Pointers and Arrays – Array of Pointers – Pointers as function arguments – Functions returning Pointers – Pointers to Functions – Function pointer – Pointers and Structures.

UNIT V: Structure & Files

Structures: Introduction – Defining a structure – Declaration of structure – Accessing Structures members – Array of Structures – Structures within structures – Structures and functions – Structures and Pointers – Union. Files: Opening and closing files – Operations on files.

TextBooks:

1. E. Balagurusamy, “Programming in ANSI C”, McGraw Hill, 8th Edition, 1st January 2024.

ReferenceBooks:

1. JyotiPMirani&Vikas B. Agarwal, “ComputerFundamentals”, 2nd Edition, NiraliPrakashan, 2019.
2. YashavantKanetkar, “Let Us C”, 18th Edition, BPB Publications, 2021.
3. Szuhay, Jeff, and Szuhay, Jeff, “Learn C Programming: A Beginner's Guide to Learning C Programming the Easy and Disciplined Way”, Packt Publishing, 2020.

E-Resources:

1. https://www.w3resource.com/c-programming-exercises/
2. https://www.javatpoint.com/c-programming-language-tutorial
3. https://www.geeksforgeeks.org/best-c-programming-courses-for-beginners/

Core Course-II DIGITAL COMPUTER FUNDAMENTAL
Course Code: 24UCP402

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS

UNIT 1: Number Systems

Digital Systems-Binary numbers- Number base conversions-Octal, Hexadecimal -one's complement and 10's complement, Binary Arithmetic, Fixed point representation and Floating-point representation of-Signal -Binary Numbers-Binary Codes.

UNIT 11: Boolean Algebra & Logic Gates

Boolean Algebra-Basic definitions Axiomatic Definition of Boolean Algebra -Basic Theorems and Properties of Boolean Algebra-Boolean Functions- Canonical and standard forms -The AND gate, The OR The NAND gate, NOR gate, The Exclusive-OR gate and Exclusive-NOR gato- Boolean Algebra and Logic Simplification DeMorgan's Theorems, Boolean Expressions and Truth Tables, The Karnaugh Map - SOP, POS, Don't care conditions.

UNIT-III: Combinational Logic Circuits

Basic combinational Logic Circuits, Implementing Combinational Logic. The Universal Property of NAND and NOR Gates. Functions of Combinational Logic Adders, Subtractors, Decoders, Encoders, Code Converters, Maltiplexers, Demultiplexers

UNIT-IV: Sequential Circuit Design

Flip flops (RS, Clocked RS. D, JK, JK Master Slave, T, Counters, Shift registers and their types, Counters: Synchronous and Asynchronous counters

UNIT-V: Memory and Programmable Logic

Introduction, Random access memory, memory decoding, error detection and correction, read only memory, Flash Memory Special Types of Memories, Magnetic and Optical Storage programmable logic array, soquential programmable devices.

Text Book :

1. M. Morris Mano, Michael D. Ciletti, “Digital Design”, 6th Edition, Pearson Education, 2018.

Reference Book :

1. Donald P leach, Albert Paul Malvino, GoutamSaha, Digital Principles and Application, 8thEdition., McGraw Hill education (India) Private Limited, 2015.
2. Prof. Satish Jain Shashi Singh, “Digital Computer Fundamentals”, 1st Edition, BPB Publications, 2017.
3. Ronald Tocci, Neal Widmer , Gregory Moss,” Digital Electronics Principles and applications,12th Edition, Pearson, 2017.
4. Digital Electronics Circuits and Systems, BishnuCharanSarkar and Suvra Sarkar,1st Edition, Sad-Ul-Tan Prakasani, Burdwan, 2018.
5. Roger L. Tokheim, Patrick E., “Digital Electronics: Principles and Applications”, 9th Edition, McGraw-Hill Education, 2022.

E-Resources:

1. http://nptel.ac.in/courses.php
2. http://jntuk-coeerd.in
3. http://electrical4u.com/

SEMESTER-II

Core Course-1 – INTRODUCTION TO DATA SCIENCE
Course Code: 24UDS501

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS

UNIT I - INTRODUCTION
Introduction, Toolboxes: Python, fundamental libraries for data Scientists-Integrated development environment (IDE)-Data operations: Reading, selecting, filtering, manipulating, sorting, grouping, rearranging, ranking, and plotting.

UNIT II – DESCRIPTIVE STATISTICS
Descriptive statistics, data preparation-Exploratory Data Analysis data summarization, data distribution, measuring asymmetry-Sample and estimated mean, variance and standard score-Statistical Inference frequency approach, variability of estimates, hypothesis testing using confidence intervals, using p values.

UNIT III - SUPERVISED LEARNING
Supervised Learning: First step, learning curves, training-validation and test-Learning models generalities, support vector machines, random forest, examples.

UNIT IV – REGRESSION
Regression analysis, Regression: linear regression simple linear regression-Multiple & Polynomial regression-Sparse model-Unsupervised learning, clustering, similarity and distances, quality measures of clustering, case study.

UNIT V – NETWORK ANALYSIS
Network Analysis-Graphs-Social Networks-centrality-drawing centrality of Graphs-Page Rank-Ego-Networks-community Detection

TEXTBOOK:
1. IntroductiontoDataScienceaPythonapproachtoconcepts,TechniquesandApplications,Igual,L;Seghi‘,
S. Springer,ISBN:978-3-319-50016-4
2. DataAnalysiswithPythonAModernApproach,DavidTaieb,PacktPublishing,ISBN-9781789950069
3. PythonDataAnalysis,SecondEd. ,Armand of andango,PacktPublishing,ISBN:9781787127487

Reference Books

1. **Cathy O'Neil & Rachel Schutt**, *Doing Data Science*, O'Reilly Media.
2. **Trevor Hastie, Robert Tibshirani & Jerome Friedman**, The Elements of Statistical Learning, Springer.
3. **Gareth James et al.**, An Introduction to Statistical Learning, Springer.
4. **Aurélien Géron**, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, O'Reilly.
5. **Albert-László Barabási**, Network Science, Cambridge University Press.

E-Resources

1. Python Official Documentation
2. NumPy Documentation
3. Pandas Documentation
4. Scikit-Learn Documentation

CORE LABORATORY – II: DATA SCIENCE LAB
COURSECODE: 24UCS551

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Download, install and explore the features of NumPy, SciPy, Jupyter, Stats models and Pandas packages
2. Working with Numpy arrays
3. Working with Pandas data frames
4. Reading data from text files ,Excel and the web and exploring various commands for doing descriptive analytics on the Iris data set.
5. Use the diabetes dataset from UCI and PimaIndians Diabetes dataset for performing the following:
    a. Univariate analysis Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis.
    b. Bivariate analysis: Linear and logistic regression modeling
    c. Multiple Regression analysis
    d. Also compare the results of the above analysis for the two data sets
6. Apply and explore various plotting functions on UCI datasets.
    a. Normal curves
    b. Density and contour plots
    c. Correlation and scatterplots
    d. Histograms DataScience Laboratory Lab Manual
    e. Three dimensional plotting
7. Visualizing Geographic Data with Basemap

SEMESTER-III

Core Course – IV: PYTHON PROGRAMMING
Course Code: 24UCP403

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus

Unit I

Introducing Python - Python on Different Operating Systems: Python on Windows- Python on macOS - Python on Linux -Variables and Assignment Statements – Arithmetic – Function print and an Introduction to Strings.

Unit II

Decision Making: The if Statement and Comparison Operators - Objects and Dynamic Typing-Control Statements: If Statement-If Else and If Elif Else Statements -While Statement, For Statement -Iterables, Lists and Iterators -Builtin Range Function -Augmented Assignments –Sequences: Lists and Tuples-Dictionaries and Set-controlled Iteration; Formatted Strings –Sentinel controlled Iteration - Builtin Function Range-Break and Continue Statements.

Unit III

Functions: Defining Functions-Functions with Multiple Parameters- Python Standard Library -math Module Functions -Using IPython Tab Completion for Discovery - Default Parameter Values -Keyword Arguments Arbitrary Argument Lists-Methods: Functions That Belong to Objects- Scope rules- Recursion- Files: Text File Processing - Updating Text Files - Handling Exceptions- finally Clause -Explicitly Raising an Exception.

Unit IV

Object Orientation: Custom Class Account- Definition-Composition: Object References as Members of Classes Controlling Access to Attributes-Properties for Data Access- Account Class Definition-Composition: Object References as Members of Classes -Controlling Access to Attributes-Properties for Data Access- Duck Typing and Polymorphism- Operator Overloading- Namespaces and Scopes.

Unit V

NumPy Basics: Arrays and Vectorized Computation: The NumPy ndarray: A Multidimensional Array Object-Pseudorandom Number Generation- Universal Functions: Fast Element-Wise Array Functions- Array-Oriented Programming with Arrays- File Input and Output with Arrays- Linear Algebra- Getting Started with pandas: Introduction to pandas Data Structures- Essential Functionality.

Text Book:

1. Paul Deitel and Harvey Deitel, “Python for Programmers”, Pearson Education, 1st Edition, 2021(Unit-I-IV)

2. Python for Data Analysis Data Wrangling with pandas, NumPy, and Jupyter, Wes McKinney, oreilly ,3rd Edition,2022(Unit-V)

Reference Books:

3. Python Crash Course by Eric Matthes, 3rd Edition, No Starch Press, January 2023

4. Martin C. Brown, Python: The Complete Reference, McGraw Hill Education (India), 2024.

5. Jeeva Jose and P. SojanLal, “Introduction to Computing and Problem Solving with PYTHON”, Khanna Book Publishing Co, 2016.

6. Phill Vega., "Python Programming 2020: Guide to Learn Python, Step by Step, with Practical Exercises, Tips and Tricks", Independently Published, 2020.
7. Reema Thareja,"Python Programming: Using Problem Solving Approach", Oxford University Press, 2019.
Kanetkar Aditya Yashavant, Kanetkar, Let Us Python, 2nd Edition, BPB Publications, 2020.

E-Resources:

https://www.tutorialspoint.com/python/index.htm
https://www.w3schools.com/python/default.asp
https://www.learnpython.org/
https://nptel.ac.in/courses/106106145
https://www.tpointtech.com/python-tutorial

Core Course – V : DATA STRUCTURES AND ALGORITHMS
Course Code: 24UCP404

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus

Unit I

Introduction of Algorithms, Analysing Algorithms: Theoretical Foundation, Algorithms and it's Specification, Random Access Machine Model ,Counting Primitive Operations, Notion of best case, average case and worst case, Characterizing Run Time, Use of asymptotic notation, Big-Oh Notation, Little-Oh, Omega and Theta Notations.

UNIT II

Arrays: Sparse Matrices–Representation of Arrays- Stacks and Queues - Fundamentals–Evaluation of Expression Infix to Postfix Conversion - Multiple Stacks and Queues - Linked List: Singly Linked List - Linked Stacks and Queues – Polynomial Addition- More on Linked Lists – Sparse Matrices – Doubly Linked List and Dynamic – Storage Management –Garbage Collection and Compaction.

UNIT III

Basic Terminology – Binary Trees – Binary Tree Representations – Binary Trees-Traversal-More on Binary Trees – Threaded Binary Trees – Binary Tree.Representation of Trees – Counting Binary Trees. Graphs: Terminology and Representations-Traversals, Connected Components and Spanning Trees, Shortest Paths and Transitive Closure.

UNIT – IV

Introduction to Algorithms - Priority Queues - Heaps – Heap Sort – Divide and Conquer Algorithms - Binary Search -Merge Sort – Quick Sort — Finding the Maximum and Minimum.

UNIT - V

Greedy Method: The General Method – Knapsack Problem – Minimum Cost Spanning Trees– Optimal Merge Patterns- Single-Source Shortest Paths. Back Tracking: The General Method – The 8-Queens Problem – Sum of Subsets – Graph Coloring – Hamiltonian Cycles.

Text Book :

1. Ellis Horowitz, Sartaj Sahni, "Fundamentals of Data Structure", Galgotia Publications (UNIT I-III)

2. Ellis Horowitz, Sartaj Sahni and Sangu Thevar Rajasekaran, "Computer Algorithms", University Press. (UNIT IV&V)

Reference Book :

1. Seymour Lipschutz, "Data Structures", Tata McGraw Hill, Schaum's Outline Series, 2014.

2. Micheal T. Goodrich and Roberto Tamassia: Algorithm Design: Foundations, Analysis and Internet examples (John Wiley & Sons, Inc., 2006).

3. ThomasH.Cormen, ChalesE.Leiserson, RonaldL.Rivest, Clifford Stein,–Introduction to Algorithms‖,3rdEdition, McGraw-Hill, 2010.

E-Resources:

- https://www.geeksforgeeks.org/dsa-tutorial-learn-data-structures-and-algorithms/
- https://www.google.com/search?q=HackerRank&rlz=1C1ONGR_enIN961IN962&oq=HackerRank&aqs=chrome..69i57.1016j0j4&sourceid=chrome&ie=UTF-8
- https://www.edx.org/masters/micromasters/ucsandiegox-algorithms-and-data-structures

Core Course – III: PYTHON PROGRAMMING LAB
Course Code: 24UCP452

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Write a Python program to simulate an ATM using conditional statements.
2. Write a Python program for a BMI calculator with health condition output.
3. Write a Python program to count word frequency in a sentence.
4. Write a Python program to compute the factorial of a number using both loops and recursion.
5. Write a Python program for an inventory management system using dictionary.
6. Write a Python program for an employee attendance tracker using dictionary.
7. Write a Python program to add two matrices of the same size and display the result.
8. Write a Python program to find the transpose of a matrix.
9. Write a Python program for a bank transaction system using functions.
10. Write a Python program for a to-do list application using file handling.
11. Create a program to count the number of times a specific word appears in a given text file.
12. Implement a BankAccount class with methods to deposit, withdraw, and display the balance.
13. Create a Vehicle class and derive Car and Bike subclasses from it.
14. Write a Python program Basic Array Operations using NumPy
15. Write a Python program to create and display a DataFrame using Pandas.
16. Write a Python program to implement a Stack and queue data structure using a list.
17. Write a Python program to implement a singly linked list with the following operations:
    Insertion at the end
    Deletion of a node by value
    Display the list
18. Write a Python program to implement the following sorting algorithm using Bubble Sort.
19. Write a Python program to implement Linear Search and Binary Search.
20. Write a Python program to implement graph traversal using Breadth-First Search (BFS) and Depth-First Search (DFS).

SEMESTER-I

Core Course – II: R PROGRAMMING FOR STATISTICS
Course Code: 24UDS502

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus

UNIT I: Introducing to R:
Introducing to R – R Data Structures – Help Functions in R – Vectors – Scalars – Declarations – Recycling – Common Vector Operations – Using all and any – Vectorized operations – Filtering – Victoriesed if-then else – Vector Element names.

UNIT II: Matrices:
Creating matrices – Matrix Operations – Applying Functions to Matrix Rows and Columns – Adding and deleting rows and columns - Vector/Matrix Distinction – Avoiding Dimension Reduction – Higher Dimensional arrays – lists – Creating lists – General list operations – Accessing list components and values – applying functions to lists – recursive lists.

UNIT III : Data Frames:
Creating Data Frames – Matrix-like operations in frames – merging Data frames – Applying functions to Data Frames – Factors and Tables – Factors and levels – Common Functions used with factors – Working with tables – Other factors and table related functions – Control statements – Arithmetic and Boolean operators and values – Default Values for arguments – Returning Boolean Values – Functions are objects – Recursion

UNIT IV : Classes:
S3 Classes – S4 Classes – Managing your objects – Input/output – accessing keyboard and monitor – reading and writing files – accessing the internet – String Manipulation – Graphics – Creating Graphs– Customizing Graphs – Saving Graphs to files – Creating Three-Dimensional plots

UNIT V: Modelling in R:
Interfacing R to other languages – Parallel R – Basic Statistics – Linear Model–Generalized Linear models – Non-linear Models – Time Series and Auto-Correlation – Clustering.

Text Book

- Norman Matloff, —TheArt of R Programming: A Tour of Statistical Software Design‖, NoStarchPress,2011.
- JaredP. Lander, —R for Everyone: Advanced Analytics and Graphics‖, Addison-Wesley Data & Analytics Series, 2013.

Reference Book

1. Mark Gardner, —Beginning–TheStatisticalProgrammingLanguage‖,Wiley,2013.
2. Robert Knell, —Introductory R: A Beginner‘s Guide to Data Visualization, Statistical Analysis and programming in R‖ ,AmazonDigitalSouthAsiaServicesInc,2013.RichardCotton(2013).LearningR,O‘Reilly Media.
3. GarretGrolemund(2014).Hands-onProgrammingwithR.O‘ReillyMedia,Inc.
4. RogerD.Peng(2018).RProgrammingforDataScience.Lean Publishing

E-Resources:

- https://www.tutorialspoint.com/r/index.htm
- https://rstudio-education.github.io/hopr/
- https://www.geeksforgeeks.org/r-tutorial/
- https://www.coursera.org/learn/r-programming

Core Course – IV: R PROGRAMMIG LAB
Course Code: 24UDS552

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus:

1. Program to convert the given temperature from Fahrenheit to Celsius and vice versa depending upon user‘s choice.
2. Program, to find the area of rectangle, square, circle and triangle by accepting suitable input parameters from user.
3. Write a program to find list of even numbers from 1 to nusing R-Loops.
4. Create a function to print squares of numbers in sequence.
5. Write a program to join columns and rows in a data frame using cbind() and rbind() in R.
6. Implement different String Manipulation functions in R.
7. Implement different data structures in R (Vectors, Lists, Data Frames)
8. Write a program to read a csv file and analyze the data in the file in R.
9. Create piechart and barchart using R.
10. Create a data set and do statistical anal
11. Write a program for R Expressions and Data Structures
12. Write a program to Manipulate of vectors and matrix
13. Write a program Operators on Factors in R
14. Create a Data Frames in R
15. Implement Lists and Operators
16. Use R to build machine learning models.
17. Use R to deploy models in production
18. Collaborate with others on data analysis projects using R and Power BI

SEMESTER-V

Core Course – III: LARGE LANGUAGE MODEL

Course Code: 24UDS50

Contact Hours: 4 **Internal Marks: 25**

Credits: 4 **External Marks: 75**

Syllabus:

Unit I – Introduction to NLP and LLMs

Natural Language Processing Overview– Text Preprocessing– Tokenization– Language Models– N-grams– Word Embedding's– Word2Vec, GloVe– Evolution of LLMs– BERT, GPT Models

Unit II – Neural Architectures for LLMs

Neural Networks Basics– Recurrent Neural Networks – LSTM and GRU– Attention Mechanism– Transformer Architecture – Encoder-Decoder Models – Self-Attention

Unit III – Training and Optimization of LLMs

Pre-training Concepts – Fine-tuning – Transfer Learning – Prompt Engineering – Few-shot and Zero-shot Learning – Reinforcement Learning from Human Feedback

Unit IV – Applications of LLMs

Chatbots and Virtual Assistants– Machine Translation – Text Summarization – Question Answering Systems – Code Generation – Content Creation

Unit V – Ethics, Limitations and Future Trends

Bias in AI – Fairness and Accountability – Hallucination in LLMs – Privacy Concerns – Model Limitations – Future Trends

Text Books

1. Speech and Language Processing – Daniel Jurafsky & James H. Martin, Pearson.
2. Natural Language Processing with Transformers – Lewis Tunstall, Leandro von Werra & Thomas Wolf, O'Reilly, 2022.
3. Transformers for Natural Language Processing – Denis Rothman, Packt Publishing.
4. Deep Learning – Ian Goodfellow, Yoshua Bengio & Aaron Courville, MIT Press.

Reference Books

1. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow – Aurélien Géron, O'Reilly.
2. Natural Language Processing in Action – Hobson Lane, Cole Howard & Hannes Hapke, Manning.
3. Generative AI with Large Language Models – Amazon Web Services & DeepLearning.AI, O'Reilly.

E-Resources

1. Hugging Face Documentation
2. OpenAI Documentation
3. TensorFlow Tutorials
4. PyTorch Tutorials

Core Course – VI: AI AND CYBER SECURITY
Course Code: 24UDS504

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus:

Unit I – Introduction to AI and Cyber Security
Introduction to Artificial Intelligence – History and Applications – Intelligent Agents – Problem Solving – Introduction to Cyber Security – CIA Triad– Types of Cyber Threats – Security Goals and Principles.

Unit II – AI Techniques
Search Algorithms – Uninformed Search – Informed Search – Heuristics – Knowledge Representation – Logical Reasoning – Machine Learning Basics – Supervised and Unsupervised Learning

Unit III – Cryptography and Network Security
Introduction to Cryptography – Symmetric Encryption – Asymmetric Encryption – Hash Functions– Digital Signatures – Network Security Basics – Firewalls – Intrusion Detection Systems.

Unit IV – Cyber Threats and Defense Mechanisms
Malware Types – Phishing and Social Engineering– Vulnerability Assessment – Risk Management – Security Policies – Authentication Mechanisms – Access Control.

Unit V – Ethics, Laws and Future Trends
Cyber Laws – Ethical Hacking – Privacy Issues – AI Ethics – Bias in AI – Future Trends in AI and Security

Text Books

1. Artificial Intelligence: A Modern Approach – Stuart Russell & Peter Norvig, Pearson.
2. Cyber Security Essentials – Charles J. Brooks, Christopher Grow, Philip Craig & Donald Short, Wiley.
3. Cryptography and Network Security – William Stallings, Pearson.
4. Machine Learning – Tom M. Mitchell, McGraw-Hill.

Reference Books

1. Introduction to Artificial Intelligence – Eugene Charniak & Drew McDermott.
2. Computer Security: Principles and Practice – William Stallings & Lawrie Brown.
3. The Web Application Hacker's Handbook – Dafydd Stuttard & Marcus Pinto.
4. Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow – Aurélien Géron.

E-Resources

1. NPTEL Artificial Intelligence Course
2. OWASP Foundation
3. CISA Cybersecurity Resources
4. Khan Academy Cryptography Course

Core Course – VI: DATA WRANGLING
Course Code: 24UDS505

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS

UNIT I: Introduction to Data Wrangling and Data Preparation

Introduction to Data Wrangling – Importance of Data Wrangling in Data Science – Data Preparation Process – Data Cleaning and Preprocessing Concepts – Setting up Python Environment – Introduction to Jupyter Notebook – Introduction to Pandas and NumPy – Loading and Exploring Datasets .

UNIT II: Data Cleaning and Transformation

Handling Missing Data – Removing Duplicates – Treating Inconsistent Data – Outlier Detection and Correction – Data Transformation Techniques – Normalization and Standardization – Encoding Techniques – String Manipulation – Regular Expressions .

UNIT III: Data Integration and Reshaping

Combining Multiple Data Sources – Merging and Joining Datasets – Concatenation Operations – Data Reshaping Techniques – Pivot Tables – Melt and Crosstab Functions – Group By and Aggregation Operations – Feature Extraction and Derived Attributes.

UNIT IV: Time Series and Exploratory Data Analysis (EDA)

Working with Date and Time Data – Time Series Parsing and Indexing – Time Series Manipulation – Exploratory Data Analysis Techniques – Data Visualization Methods – Descriptive Statistics – Identifying Data Patterns and Trends.

UNIT V: Data Wrangling Pipeline and Case Studies

Building End-to-End Data Wrangling Pipelines – Data Integration with Visualization Tools – Exporting Data into CSV, Excel and JSON Formats – Case Studies using Real-World Datasets – Data Wrangling in Finance, Healthcare and E-commerce – Mini Project on Data Cleaning and Transformation Workflow.

Text Books

1. **Wes McKinney**, *Python for Data Analysis: Data Wrangling with Pandas, NumPy and Jupyter*, 3rd Ed., O'Reilly, 2022.
2. **Jacqueline Kazil & Katharine Jarmul**, *Data Wrangling with Python*, O'Reilly Media.
3. **Jake VanderPlas**, *Python Data Science Handbook*, O'Reilly Media.
4. **Pratap Dangeti**, *Statistics for Machine Learning*, Packt Publishing.

Reference Books

1. **Mark Fenner**, *Machine Learning with Python for Everyone*, Pearson.
2. **Joel Grus**, *Data Science from Scratch*, O'Reilly Media.
3. **Laura Igual & Santi Seguí**, *Introduction to Data Science: A Python Approach*, Springer.
4. **Aurélien Géron**, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, O'Reilly.

E-Resources

1. Pandas Documentation
2. NumPy Documentation
3. Jupyter Documentation
4. Kaggle Learn – Data Cleaning

Core Course – III: SECURITY WRANGLING LAB
Course Code: 24UDS553

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Install and configure Python, Anaconda, and Jupyter Notebook environment
2. Demonstrate basic operations and commands in Jupyter Notebook.
3. Create NumPy arrays and perform basic operations such as indexing, slicing, and aggregation.
4. Create and manipulate Pandas Series and DataFrames.
5. Load datasets from CSV, Excel, and JSON files using Pandas.
6. Handle missing data using techniques such as removal, imputation, and interpolation.
7. Identify and remove duplicate records from datasets.
8. Clean datasets by handling inconsistent and noisy data.
9. Apply data transformation techniques such as scaling and encoding.
10. Perform normalization and standardization on datasets.
11. Perform string manipulation operations on dataset attributes.
12. Implement Regular Expressions for pattern matching and data cleaning.
13. Merge, join, and concatenate multiple datasets.
14. Perform data reshaping using pivot tables, melt, and crosstab functions.
15. Apply GroupBy operations and aggregation functions.
16. Perform Exploratory Data Analysis (EDA) and visualize data using plots and charts.

Core Course – VI:
SUPERVISED AND UNSUPERVISED LEARNING FOR DATA NETWORK SCIENCE
Course Code: 24UDS506

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus:

UNIT I: Introduction to Machine Learning and Data Networks
Introduction to Machine Learning – Types of Machine Learning – Supervised vs Unsupervised Learning-Applications in Data Network Science – Data Representation in Networks – Graph-based Data Structures-Feature Extraction from Network Data .

UNIT II: Supervised Learning Techniques
Regression Concepts and Types – Linear Regression Model – Classification Techniques – Decision Trees-k-Nearest Neighbors – Support Vector Machines – Model Training and Testing – Evaluation Metrics (Accuracy, Precision, Recall) .

UNIT III: Unsupervised Learning Techniques
Clustering Concepts – K-Means Clustering – Hierarchical Clustering – Density-Based Clustering – Association Rule Mining – Dimensionality Reduction – Principal Component Analysis (PCA).

UNIT IV: Network-Based Learning and Analysis
Graph Mining Concepts – Social Network Analysis – Community Detection Algorithms – Link Prediction Techniques – Centrality Measures – Network Visualization – Applications in Real-World Networks.

UNIT V: Advanced Topics and Applications
Ensemble Learning Methods – Random Forest and Boosting – Semi-Supervised Learning – Deep Learning for Network Data – Case Studies in Data Network Science – Mini Project on Network Data Analysis.

Text Books
1. **Tom M. Mitchell**, Machine Learning, McGraw-Hill Education.
2. **Ethem Alpaydin**, Introduction to Machine Learning, MIT Press.
3. **Christopher M. Bishop**, Pattern Recognition and Machine Learning, Springer.
4. **Mark Newman**, Networks: An Introduction, Oxford University Press.

Reference Books
1. **Aurélien Géron**, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, O'Reilly Media.
2. **Charu C. Aggarwal**, Machine Learning for Data Streams, Springer.
3. **Albert-László Barabási**, Network Science, Cambridge University Press.

E-Resources
1. Scikit-Learn Documentation
2. NetworkX Documentation
3. TensorFlow Tutorials
4. Stanford Network Analysis Platform (SNAP)
5. Google Colab

Discipline Centric Elective I: ADVANCED DATA SCIENCE
Course Code: 24UDS701

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus :

UNIT I –Advanced Data Pre-processing and Feature Engineering
Data cleaning for large datasets – Handling missing values – Handling imbalanced data – Feature selection techniques – Feature extraction – Dimensionality reduction (PCA, LDA) – Data transformation and – Data integration.

UNIT II – Advanced Machine Learning Techniques
Ensemble learning – Bagging – Boosting – Random Forest – Gradient Boosting – Support Vector Machines (SVM) – Hyper parameter tuning techniques – Model evaluation metrics (Precision, Recall, F1- score, ROC-AUC) – Cross validation techniques .

UNIT III – Deep Learning for Data Science
Introduction to Deep Learning – Artificial Neural Networks (ANN) – Convolutional Neural Networks (CNN) – Recurrent Neural Networks (RNN) – Activation functions (– Deep learning applications – Introduction to Tensor Flow / Keras .

UNIT IV – Big Data Analytics
Introduction to Big Data – Hadoop ecosystem – HDFS architecture – Map Reduce programming model – Apache Spark basics – Distributed data processing – Big data analytics applications .

UNIT V – Data Science Applications and Visualization
Data visualization techniques – Advanced visualization tools (Tableau / Power BI / Python libraries) – Time series analysis basics – Text analytics – Sentiment analysis – Real-time applications in healthcare analytics, finance, and social media analytics

Text Books
1. **Aurélien Géron**, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, O'Reilly Media.
2. **Ian Goodfellow, Yoshua Bengio & Aaron Courville**, Deep Learning, MIT Press.
3. **Jure Leskovec, Anand Rajaraman & Jeffrey Ullman**, Mining of Massive Datasets, Cambridge University Press.
4. **Seema Acharya & Subhashini Chellappan**, *Big Data Analytics*, Wiley India.

Reference Books
1. **Trevor Hastie, Robert Tibshirani & Jerome Friedman**, The Elements of Statistical Learning, Springer.
2. **Ethem Alpaydin**, Introduction to Machine Learning, MIT Press.
3. **Tom White**, Hadoop: The Definitive Guide, O'Reilly Media.
4. **Foster Provost & Tom Fawcett**, Data Science for Business, O'Reilly Media.

E-Resources
1. TensorFlow Documentation
2. Keras Documentation
3. Scikit-Learn Documentation
4. Apache Hadoop Documentation
5. Apache Spark Documentation

Discipline Centric Elective Laboratory I: ADVANCED DATA SCIENCE LAB
Course Code: 24UDS751

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Load a dataset and clean the data by filling missing values (mean/median/mode).
2. Find and remove outliers using simple methods like IQR or Z-score.
3. Handle imbalanced data using oversampling or undersampling techniques.
4. Select important features using correlation or SelectKBest method.
5. Reduce data size using PCA and understand the result.
6. Build a simple Bagging model using Decision Tree.
7. Implement Random Forest and check its performance.
8. Apply SVM for classification of data.
9. Calculate Accuracy, Precision, Recall, and F1-score for a model.
10. Perform cross-validation and tune model parameters using GridSearchCV.
11. Create a basic ANN model using Keras.
12. Build a simple CNN model for image classification.
13. Perform basic data processing using PySpark.
14. Create simple graphs like bar chart, histogram, and heatmap.

Discipline Centric Elective Laboratory I
CATEGORICAL AND NONPARAMETRIC DATA ANALYTICS
Course Code: 25UDS703

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS

UNIT - I: Foundations of Categorical Data and Statistical Concepts

Introduction – Types of Data: Nominal, Ordinal, Interval, Ratio – Categorical vs Numerical Data – Descriptive Statistics for Categorical Data – Frequency Distribution and Proportions – Basic Probability Concepts – Sampling Distributions – Statistical Inference – Hypothesis Testing: Null and Alternative Hypothesis, p-value, Type I & II Errors.

UNIT - II: Analysis of Contingency Tables and Association Measures

Contingency Table Analysis – One-way and Two-way Tables – Joint, Marginal and Conditional Distributions – Chi-Square Test for Independence – Measures of Association: Odds Ratio, Relative Risk – Fisher's Exact Test – Introduction to Measures of Agreement .

UNIT - III: Categorical Data Modeling and Logistic Regression

Categorical Data Modelling – Logistic Regression Model (Binary Response) – Logit Function and Odds Interpretation – Parameter Estimation and Inference – Model Fit and Diagnostics – Multiple Logistic Regression – Introduction to Multinomial and Ordinal Logistic Models.

UNIT - IV: Introduction to Nonparametric Methods

Nonparametric Methods – Introduction and Need for Distribution-Free Methods – Advantages and Limitations – Sign Test – Wilcoxon Signed-Rank Test – Mann–Whitney U Test – Kruskal–Wallis Test – Rank Correlation: Spearman's Rho, Kendall's Tau.

UNIT - V: Advanced Nonparametric Techniques

Advanced Nonparametric Techniques – Kolmogorov–Smirnov Test – Runs Test for Randomness – Nonparametric Regression Basics – Comparison of Parametric, Categorical and Nonparametric Methods – Applications and Case Studies – Model Selection and Interpretation.

Text Books

1. Categorical Data Analysis – Alan Agresti, Wiley.
2. An Introduction to Categorical Data Analysis – Alan Agresti, Wiley.
3. Nonparametric Statistical Methods – Myles Hollander, Douglas A. Wolfe & Eric Chicken, Wiley.
4. Applied Logistic Regression – David W. Hosmer, Stanley Lemeshow & Rodney X. Sturdivant, Wiley.

Reference Books

1. Practical Nonparametric Statistics – W.J. Conover, Wiley.
2. Introduction to the Practice of Statistics – David S. Moore, George P. McCabe & Bruce Craig.
3. Biostatistical Analysis – Jerrold H. Zar, Pearson.
4. Discovering Statistics Using IBM SPSS Statistics – Andy Field, Sage Publications.

E-Resources

1. NIST Engineering Statistics Handbook
2. Penn State STAT Online Courses
3. UCLA Statistical Consulting Resources
4. R Project Documentation

Discipline Centric Elective Laboratory I :ANALYTICAL TOOL LAB
Course Code: 24UDS752

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Collect a small dataset (students' gender, marks, age) and classify them into Nominal, Ordinal, Interval, and Ratio data.
2. Create a frequency distribution table for categorical data (e.g., favorite subject of students).
3. Create a simple frequency table for your classmates' favorite color.
4. Calculate proportions and percentages from the dataset.
5. Calculate probability of choosing a student who likes Maths.
6. Construct a one-way and two-way contingency table using sample data.
7. Perform Chi-square test for independence using Excel/Python.
8. Fit a binary logistic regression model using sample dataset (pass/fail, yes/no).
9. Calculate Spearman's Rank Correlation for given data.
10. Compare two groups of students using ranking method.
11. Conduct Wilcoxon Signed-Rank Test for related samples.
12. Demonstrate basic nonparametric regression using sample data.
13. Compare parametric and nonparametric test results using same dataset.
14. Write a sequence of numbers and check if pattern looks random.
15. Draw a simple graph to observe data trend.

Discipline Centric Elective Laboratory I : DATA ARCHITECTURE
Course Code: 24UDS705

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

Syllabus:

UNIT – I Fundamentals of Data Architecture and Database Systems
Introduction - What Is Data Architecture - Importance of Data Architecture - Types of Data - Data Architecture vs Data Engineering - Overview of Data Systems - Basics of Databases - Relational Databases and NoSQL Databases - Introduction to Data Storage Concepts.

UNIT – II Data Modeling and Storage Design
Data Models: Relational Model - Document Model - Key-Value Model - Graph Model - Schema Design- Schema-on-Write vs Schema-on-Read - Data Modeling Concepts - Indexing and Partitioning - Transactions - ACID Properties - Normalization vs De-normalization- Introduction to Data Warehousing

UNIT – III Data Processing and Data Pipeline Architectures
Data Processing Overview - Batch Processing- Stream Processing - Data Pipelines - ETL and ELT Processes- Data Integration Techniques- Message Queues - Event-Driven Architecture- Data Lakes- Data Warehouses- Workflow Orchestration

UNIT – IV Distributed Data Systems and Scalability
Distributed Systems Overview- Replication- Partitioning- Consistency Models- CAP Theorem- Fault Tolerance- Reliability- Scalability Concepts- Vertical Scaling vs Horizontal Scaling- Distributed Databases- Handling Failures in Distributed Systems

UNIT – V Data Governance and Modern Data Architecture Practices
Data Governance- Data Quality Management- Data Security and Privacy- Metadata Management- Modern Data Architectures- Data Mesh- Data Fabric - Cloud Data Architecture - Case Studies in Data Architecture- Designing End-to-End Data Systems

Text Books

1. Fundamentals of Data Engineering – Joe Reis & Matt Housley, O'Reilly Media.
2. Designing Data-Intensive Applications – Martin Kleppmann, O'Reilly Media.
3. Data Management for Researchers – Kristin Briney, Pelagic Publishing.
4. Database System Concepts – Abraham Silberschatz, Henry Korth & S. Sudarshan, McGraw-Hill.

Reference Books

1. Data Warehouse Toolkit – Ralph Kimball & Margy Ross, Wiley.
2. NoSQL Distilled – Pramod J. Sadalage & Martin Fowler, Addison-Wesley.
3. Streaming Systems – Tyler Akidau, Slava Chernyak & Reuven Lax, O'Reilly.
4. Cloud Data Management – Chunming Rong & others.

E-Resources

1. MongoDB Documentation
2. Apache Hadoop Documentation
3. Apache Spark Documentation
4. PostgreSQL Documentation

Discipline Centric Elective Laboratory I : DATA ARCHITECTURE LAB
Course Code: 24UDS754

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus

1. Study and compare RDBMS and NoSQL databases with a simple example (table vs document).
2. Identify different types of data (structured, semi-structured, unstructured) with examples.
3. Create indexes and demonstrate partitioning concepts.
4. Demonstrate transactions and ACID properties.
5. Demonstrate Key-Value and Graph model using simple examples.
6. Build a simple data pipeline using scripts (Python/SQL).
7. Implement ETL (Extract, Transform, Load) process.
8. Demonstrate distributed database concepts using case study.
9. Demonstrate replication techniques (primary-secondary model).
10. Perform partitioning (horizontal/vertical) on sample data.
11. Analyze modern data architectures (Data Mesh, Data Fabric).
12. Perform data quality management (handling missing/duplicate data).
13. Demonstrate data security techniques (user roles and permissions).
14. Compare vertical scaling and horizontal scaling.
15. Design an end-to-end data architecture system (Student/Product system).

Discipline Centric Elective Laboratory I : DATA PRODUCT DESIGN
Course Code: 24UDS707

Contact Hours: 4 **Internal Marks: 25**
Credits: 4 **External Marks: 75**

SYLLABUS:

UNIT – I Foundations of Data Products
Introduction– What is a Data Product– Data vs Information vs Insights– Types of Data Products– Characteristics of Good Data Products– Data Product vs Software Product – Data Product Lifecycle – Role of Data in Decision Making – Introduction to Data Product Thinking

UNIT – II Data Product Design and User-Centric Development

Data Product Design: Design Thinking – Identifying Users and Use Cases – Data Product Canvas – Defining Business Objectives and KPIs– User Experience (UX) for Data Products – Data Visualization Principles – Dashboard Design – Data Storytelling Concepts

UNIT – III Data Product Development and Engineering

Building Data Products: Data Collection Methods – Data Ingestion – Data Cleaning and Preprocessing – Data Transformation – ETL and ELT Pipelines– Data Storage – Introduction to APIs for Data Access – Tools and Technologies

UNIT – IV Product Design, Evaluation and Management
Concept Selection-Concept Testing- Product Architecture –Industrial Design for Environment - Design for Manufacturing - Prototyping - Robust Design - Patents and Intellect Property - Product Development Economics - Managing Projects

UNIT – V Advanced Data Products and Emerging Technologies
Advanced Data Products: AI and Machine Learning in Data Products – Real-Time Data Processing – Personalization and Recommendation Systems – Data Mesh Architecture – Decentralized Data Systems – Case Studies of Data Products – Future Trends in Data Product Design

Text Books

1. Data Product Management in Action – Christopher Bergh, Manning Publications.
2. Lean Analytics – Alistair Croll & Benjamin Yoskovitz, O'Reilly Media.
3. Designing Data-Intensive Applications – Martin Kleppmann, O'Reilly Media.
4. Product Design and Development – Karl T. Ulrich & Steven D. Eppinger, McGraw-Hill.

Reference Books

1. Inspired – Marty Cagan, Wiley.
2. Storytelling with Data – Cole Nussbaumer Knaflic, Wiley.
3. Fundamentals of Data Engineering – Joe Reis & Matt Housley, O'Reilly Media.
4. The Design of Everyday Things – Don Norman, Basic Books.

E-Resources

1. Google Data Analytics Resources
2. Microsoft Power BI Learning
3. Tableau Learning Resources

Discipline Centric Elective Laboratory I :
EXPERIMENTAL DRIVEN PRODUCT LAB
Course Code: 24UDS754

Contact Hours: 4 **Internal Marks: 40**
Credits: 2 **External Marks: 60**

Syllabus:

1. Identify a real-world problem and define it as a data product use case.
2. Analyze a dataset to differentiate data, information, and insights.
3. Study and classify different types of data products (dashboard, ML model, application).
4. Apply Design Thinking to frame a problem statement and solution approach.
5. Create user personas and use cases for a selected data product.
6. Develop a Data Product Canvas including objectives and KPIs.
7. Design a low-fidelity prototype (wireframe) of a data product.
8. Create a dashboard design using visualization principles (Power BI/Tableau).
9. Perform data collection and ingestion using open datasets or APIs.
10. Apply data cleaning and preprocessing techniques on raw data.
11. Implement data transformation and ETL process.
12. Build a basic data pipeline and store processed data.
13. Develop a working prototype of the data product.
14. Conduct user testing and evaluation of the product.
15. Build a simple AI/ML-based data product (e.g., recommendation system) and present final report.

Ability Enhancement

CLOUD COMPUTING TECHNIQUES LAB

Course Code: 24UDS381

Syllabus:

1. Study of cloud computing platforms (AWS/Azure/GCP).
2. Installation of VirtualBox / VMware and OS setup.
3. Creation and management of virtual machines.
4. Cloud storage implementation (Google Drive / AWS S3).
5. Setting up a web server in cloud environment.
6. Deployment of a static website in cloud.
7. Database setup in cloud environment.
8. Managing cloud instances and services.
9. Introduction to load balancing.
10. Cloud security basics (user roles and access control).
11. Installation of Docker.
12. Creation of Docker images and containers.
13. Deploying applications using Docker.
14. Cloud monitoring tools usage.
15. Mini Project – Deploying a cloud-based application.