Walchand Institute of Technology, Solapur
Computer Science and Engineering
Honors in Data Science

Structure of S. Y. B. Tech. Computer Science and Engineering
(W.E.F. 2022-2023)

Semester- IV

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
21CSU4HD1T	Mathematics for Data Science	3	1	-	4	60	40	25	125	
	Grand Total	**3**	**-**	**2**	**4**	**60**	**40**	**25**	**125**	

Walchand Institute of Technology, Solapur
Computer Science and Engineering
Honors in Data Science

Structure of T. Y. B. Tech. Computer Science and Engineering
(W.E.F. 2023-2024)

Semester-V

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
21CSU5HD2T	Data Preprocessing and Visualization	3	--	--	3	60	40	--	100	
	Laboratory									
21CSU5HD2L	Data Preprocessing and Visualization	--	--	2		--	--	25	25	
	Grand Total	**3**	**-**	**2**	**4**	**60**	**40**	**25**	**125**	

Structure of T. Y. B. Tech. Computer Science and Engineering

(W.E.F. 2023-2024)

Semester-VI

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
21CSU6HD3T	Machine Learning	3	--	--	3	60	40	--	100	
	Laboratory									
21CSU6HD3L	Machine Learning	--	--	2	1	--	--	25	25	
	Grand Total	**3**	**-**	**2**	**4**	**60**	**40**	**25**	**125**	

Walchand Institute of Technology, Solapur
Computer Science and Engineering
Honors in Data Science

Structure of Final Year B. Tech. Computer Science and Engineering
(W.E.F. 2024-2025)

Semester-VII

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA		
		L	T	P		ESE	ISE	ICA	Total
21CSU7HD4T	Predictive Analytics	3	--	--	3	60	40	--	100
	Laboratory								
21CSU7HD4L	Predictive Analytics	--	--	2	1	--	--	25	25
21CSU7HD5P	Mini Project	--	--	4	2	50	--	50	100
	Grand Total	**3**	**-**	**6**	**6**	**60**	**40**	**25**	**125**

Note: -These courses are to be completed by the student in addition to the courses of B.Tech Computer Science and Engineering

WALCHAND INSTITUTE OF TECHNOLOGY
(AUTONOMOUS INSTITUTE)

Name of the Faculty: Science and Technology

CHOICE BASED CREDIT SYSTEM

Structure and Syllabus : Computer Science & Engineering

Name of the Course: T.Y. B. Tech. Sem V and Sem VI

Hons. Degree : Data Science

(Syllabus to be implemented from July 2023)

WALCHAND INSTITUTE OF TECHNOLOGY, SOLAPUR
Faculty of Science and Technology
Third Year B. Tech. (CSE)
Honors Degree in Data Science
SEMESTER - V
21CSU5HD2T : DATA PRE-PROCESSING AND VISUALIZATION

Teaching Scheme	Examination Scheme
Lecture: 3 Hours /Week, 3 Credits	ESE- 60 Marks
Practical :2 Hours /Week, 1 Credits	ISE – 40 Marks
	ICA -25 Marks

Introduction:

Data science is a field of study and application that has been growing rapidly for the past several decades. As a growing field, it is gaining a lot of attention in both the media as well as in the job market. This course will introduce students to data pre-processing and visualization techniques and tools.

Prerequisite:

Fundamentals of Python Programming

COURSE OUTCOMES:

At the end of this course, the student will be able to -

1. Identify the different types of data
2. Transform raw data into understandable format
3. Use python libraries for data pre processing and visualization
4. Represent the data in various graphical forms.

SECTION - I

Unit 1 - Introduction and Describing Data (6)

Overview, Sources of Data ,Process for Making Sense of Data, Observations and Variable , Types of Variables, Central Tendency, Distribution of the Data, Confidence Intervals, Hypothesis Tests

Unit 2 - Preparing Data Tables (8)

Overview, Cleaning the Data, Removing Observations and Variables, Generating Consistent Scales Across Variables, New Frequency Distribution, Converting Text to Numbers, Converting Continuous Data to Categories, Combining Variables, Generating Groups, Preparing Unstructured, Data Visualizing Relationships between Variables, Calculating Metrics about Relationships.

Unit 3 - Introduction to NumPy (8)

Understanding Data Types in Python, The Basics of NumPy Arrays, Computation on NumPy Arrays: Universal Functions, Aggregations: Min, Max, and Everything in Between, Computation on Arrays: Broadcasting, Comparisons, Masks, and Boolean Logic, Fancy Indexing ,Sorting Arrays, Structured Data: NumPy's Structured Arrays

SECTION-II

Unit 4 - Data Manipulation with Pandas (7)

Installing and Using Pandas, Introducing Pandas Objects, Data Indexing and Selection, Operating on Data in Pandas, Handling Missing Data, Hierarchical Indexing, Combining Datasets: Concat and Append, Combining Datasets: Merge and Join, Aggregation and Grouping, Pivot Tables, Vectorized String Operations, Working with Time Series, High-Performance Pandas.

Unit 5 - Data Visualization (6)

Overview, Visualization Design Principles, Tables, Univariate Data Visualization, Multivariate Data Visualization, Visualizing Groups, Dynamic Techniques

Unit 6 - Visualization with Matplotlib and Seaborn (9)

General Matplotlib Tips, Two Interfaces for the Price of One, Simple Line Plots, Visualizing Errors, Density and Contour Plots, Histograms, Binnings, and Density, Customizing Plot Legends, Customizing Colorbars, Multiple Subplots, Text and Annotation, Customizing Matplotlib: Configurations and Stylesheets, Three-Dimensional Plotting in Matplotlib, Geographic Data with Basemap, Introduction to Seaborn: Seaborn functionalities and usage, Spatial Visualizations and Analysis in Python with Folium, Case Study.

Internal Continuous Assessment (ICA):

ICA should consist of Solving 8- 10 practical assignments on above units.

Text Book:

1. Glenn J. Myatt, Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, John Wiley Publishers, 2014. (Unit- I and II)
2. Glenn J. Myatt, Making sense of Data: A practical Guide to Data Visualization, Advanced Data Mining Methods and Applications, John Wiley Publishers, 2009.(Unit-V)
3. Python Data Science Handbook – Essential Tools for working with Data : Jake VanderPlas, O'rielly (Unit III, IV, VI)

Teaching Scheme	**Examination Scheme**
Lecture: 3 Hours /Week, 3 Credits	ESE – 60 Marks
Practical: 2 Hours /Week, 1 Credit	ISE – 40 Marks
	ICA - 25 Marks

Introduction :
Machine learning is the science of getting computers to act without being explicitly programmed. This course provides a broad introduction to machine learning and its mathematical foundation. It includes the types and the best practices in machine learning along with its real world applications.

Pre-requisite:
Knowledge of Probability & Statistics with a basic course in Python.

COURSE OUTCOMES:
At the end of the course students will be able to
1. Demonstrate types of machine learning algorithms.
2. Design a model by selecting appropriate machine learning algorithm for a given problem.
3. Validate designed machine learning model.
4. Evaluate and tune machine learning model based on various parameters.
5. Design various applications using machine learning algorithm.

SECTION I

Unit 1: Introduction to Machine learning (08)
Understanding Machine Learning: What Is Machine Learning?, Leveraging the Power of Machine Learning,The Roles of Statistics and Data Mining with Machine Learning, Putting Machine Learning in Context, Typesof machine Learning, Applications of Machine Learning.
Applying Machine Learning: Getting Started with a Strategy, Applying Machine Learning to Business Needs, Understanding Machine Learning Techniques, Tying Machine Learning Methods to Outcomes.

Unit 2: Offerings of Machine learning (05)
Looking Inside Machine Learning: The Impact of Machine Learning on Applications, Data Preparation, The Machine Learning Cycle.
Getting Started with Machine Learning: Understanding How Machine Learning Can Help, Focus on the Business Problem, Requirement of Collaboration in Machine Learning, Executing a Pilot Project, Determining the Best Learning Model.

Unit 3: Basic mathematics for Machine Learning (10)
Getting Started with The Math Basics, Working with Data, Exploring the World of Probabilities, Describing the Use of Statistics, Interpreting Learning As Optimization, Exploring Cost Functions, Descending the Error Curve, Updating by Mini-Batch and Online.

SECTION II

Unit 4: Validating Machine Learning Models **(10)**

Validating Machine Learning: Checking Out-of-Sample Errors, Getting to Know the Limits of Bias, Keeping Model Complexity in Mind and Solutions Balanced, Training, Validating, and Testing, Resorting to Cross-Validation. Looking for Alternatives in Validation. Optimizing Cross-Validation Choices, Avoiding Sample Bias and Leakage Traps, Discovering the Incredible Perceptron

Simplest learning strategies to learn from Data: Discovering the Incredible Perceptron, Growing GreedyClassification Trees, Taking a Probabilistic Turn

Unit 5: Improving Machine Learning Models **(08)**

Improving Machine Learning Models, Studying Learning Curves, Using Cross-Validation Correctly,Choosing the Right Error or Score Metric, Searching for the Best Hyper-Parameters, Testing Multiple Models,Averaging Models, Stacking Models, Applying Feature Engineering, Selecting Features and Examples,Looking for More Data.

Unit 6: Applications of Machine Learning **(04)**

Applying Learning to Real Problems**,** Classifying Images, Scoring Opinions and Sentiments, Recommending Products and Movies, Using Machine Learning to Provide Solutions to Business Problems, Future of MachineLearning.

Internal Continuous Assessment (ICA):

Student should implement the following:

1. Basic mathematics for Machine Learning –
    Simulating solutions using Python to
    - I. Matrix operations
    - II. Problems using Probability
    - III. Statistical Estimations.
2. Introduction to Jupyter Notebook and Colab.
3. Working with data.
4. Data Exploration and Preprocessing.
5. Linear Regression
6. Introduction to Dimensionality Reduction
7. Logistic Regression
8. Decision Trees
9. Ensemble Models
10. Clustering (Unsupervised Learning)

Text Books:

1. Machine Learning For Dummies, IBM Limited Edition by Judith Hurwitz, Daniel Kirsch (Published byWiley, First edition).
2. Machine Learning For Dummies by John Paul Mueller, Luca Massaron (Published by For Dummies; Firstedition).

Reference Books :

1. Introduction to Machine Learning (Second Edition) by Ethem Alpaydın (published by The MIT PressCambridge, Massachusetts London, England
2. Machine Learning by Tom M. Mitchell (Publisher: McGraw Hill Education; First edition + New Chaptersfrom Second edition).