WALCHAND INSTITUTE OF TECHNOLOGY, SOLAPUR
(AN AUTONOMOUS INSTITUTE)

Affiliated to
Punyashlok Ahilyadevi Holkar Solapur University, Solapur

Department of Information Technology

CHOICE BASED CREDIT SYSTEM (CBCS)

Structure and Syllabus
for
Honors Degree in
Data Science

S.Y. B. Tech. Information Technology W.E.F. 2023-24
T.Y. B. Tech. Information Technology W.E.F. 2024-25
Final Year B. Tech Information Technology W.E.F. 2025-26

HEAD
Information Technology Dept.
W.I.T. Solapur

HEAD
Information Technology Deptt.
W.I.T., Solapur.

W I T
SOLAPUR
(Autonomous)

Dr. Mrs. M. A. Nirgude
Dean Academics

Department of Information Technology

Department Vision

To be a frontier in Information Technology, to produce globally competent engineers with an aptitude for leadership and research, who will be instrumental in continuous socio-economic development.

Department Mission

M1: To impart quality education in Information Technology in accordance with the needs of the society through blended mode.

M2: To inculcate critical thinking and creativity for identifying various issues and to provide sustainable solutions by becoming a lifelong learner.

M3: To enhance career opportunities through academia-industry interaction and research, while embodying professional ethics.

Department of Information Technology
Under Graduate Program
Program Educational Objectives (PEOs)

1. Graduates will exhibit strong fundamental knowledge and skills in the field of Information Technology to pursue successful professional careers, higher studies and research.
2. Graduates will exhibit capabilities to understand and resolve the various issues through their problem solving skills.
3. Graduates will be sensitive to ethical, societal and environmental issues while serving at their professional work and society.

Program Outcomes (POs)

The program outcomes of B. Tech. Information Technology Program are summarized as following:

1. **Engineering Knowledge: Apply** the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem Analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences and engineering sciences.
3. **Design/Development of Solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct Investigations of Complex Problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

6. **The Engineer and Society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities, relevant to the professional engineering practice.

7. **Environment and Sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and Team Work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project Management and Finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long Learning:** Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Program Specific Outcomes (PSOs)

1. Students will be able to apply fundamentals of mathematics, algorithms and computational systems to Information technology.

2. Students will be able to provide a solution to the problem in the areas of Networking, Database management, System Software, Web Technology, Information Security and Thrust areas.

3. Students will be able to design and develop IT solution for societal problem/s while encouraging usage of Free and Open Source Software (FOSS)

Legends used–

L	Lecture Hours / week
T	Tutorial Hours / week
P	Practical Hours / week
FA	Formative Assessment
SA	Summative Assessment
ESE	End Semester Examination
ISE	In Semester Evaluation
ICA	Internal Continuous Assessment
POE	Practical and Oral Exam
OE	Oral Exam
F.Y.	First Year
S.Y.	Second Year
T.Y.	Third Year
B. Tech.	Bachelor of Technology

Course Code Format for Honors:

2	1	I	T	U/P	2	H	A	1	T/L
Year of Syllabus revision		Program Code		U-Under Graduate, P-Post Graduate	Semester No. / Year 1/2/3/…8	Honors Code		Course Serial No. 1-9	T-Theory, L-Lab session

Program Code	
IT	Information Technology
Honors Code	
HD	Honors in Data Science

Sample Course Code:

22ITU4HD1T	Mathematics for DataScience

Walchand Institute of Technology, Solapur
Information Technology
Honors in Data Science

Structure of S. Y. B. Tech. Information Technology
(W.E.F. 2023-2024)

Semester- IV

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total
22ITU4HD1T	Mathematics for DataScience	3	1	-	4	60	40	25	125
	Grand Total	**3**	**1**		**4**	**60**	**40**	**25**	**125**

Walchand Institute of Technology, Solapur
Information Technology
Honors in Data Science

Structure of T. Y. B. Tech. Information Technology
(W.E.F. 2024-2025)

Semester- V

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
22ITU5HD1T	Data Preprocessing andVisualization	3	-	-	3	60	40	-	100	
	Laboratory:									
22ITU5HD1L	Data Preprocessing andVisualization	-	-	2	1	-	-	25	25	
	Grand Total	**3**		**2**	**4**	**60**	**40**	**25**	**125**	

Structure of T. Y. B. Tech. Information Technology
(W.E.F. 2024-2025)

Semester- VI

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
22ITU6HD1T	Machine Learning	3	-	-	3	60	40	-	100	
	Laboratory:									
22ITU6HD1L	Machine Learning	-	-	2	1	-	-	25	25	
	Grand Total	**3**		**2**	**4**	**60**	**40**	**25**	**125**	

Walchand Institute of Technology, Solapur
Information Technology
Honors in Data Science

Structure of Final Year B. Tech. Information Technology
(W.E.F. 2025-2026)

Semester- VII

Course Code	Theory Course Name	Engagement Hours			Credits	FA	SA			
		L	T	P		ESE	ISE	ICA	Total	
22ITU7HD1T	Predictive Analytics	3	-	-	3	60	40	-	100	
	Laboratory:					**OE**				
22ITU7HD1L	Predictive Analytics	-	-	2	1	-	-	25	25	
22ITU7HD2L	Mini project	-	-	4	2	50	-	50	100	
	Grand Total	**3**		**6**	**6**	**110**	**40**	**75**	**225**	

Note: -These courses are to be completed by the student in addition to the courses of B. Tech Information Technology

Walchand Institute of Technology, Solapur
Honors Degree in Data Science
S.Y. B. Tech. (Information Technology), Semester - IV
22ITU4HD1T: MATHEMATICS FOR DATA SCIENCE

Teaching Scheme	**Examination Scheme**
Lecture: 3 Hours /Week, 3 Credits	ESE – 60 Marks
Tutorial: 1 Hour /Week, 1 Credit	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 introduces the fundamentals of mathematics which are used by data scientists to solve the real time problems.

Prerequisite: Basics of Linear algebra, Probability, Statistics

----

COURSE OUTCOMES:

Students will be able to:
1. Use the properties of Linear Maps in solving problems on Linear Algebra.
2. Build a strong statistical foundation and learn how to 'infer' insights from a huge population using a small sample.
3. Demonstrate various random variables, discrete and continuous distributions and their usage.
4. Use optimization techniques and formulate hypotheses for a population to solve real-life business problems

----

SECTION-I

Unit 1–Basics of Data Science: (04)

Introduction to data science, Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective, Structured thinking for solving data science problems.

Unit 2–Linear Algebra (10)

Vectors, Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigen values and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyper planes; half-planes.

Unit 3- Statistics (08)

Describing a Single Set of Data, Central Tendencies, Dispersion, Correlation, Simpson's Paradox, Correlation & causation.

SECTION-II

Unit 4 – Probability (08)

Dependence and Independence, Conditional Probability, Bayes's Theorem, Random Variables Continuous Distributions, The Normal Distribution, The Central Limit Theorem.

Unit 5- Hypothesis and Inference (06)
Statistical Hypothesis Testing, Confidence Intervals, P- hacking, Bayesian Inference

Unit 6: Optimization (08)
Unconstraint optimization ,necessary and sufficiency conditions for optima, gradient descent methods, constraint optimization, KKT condition, Introduction to non-gradient techniques, Introduction to least squares optimization,Optimization view of machine learning,

Internal Continuous Assessment (ICA):
ICA should consist of Solving 8- 10 assignments on above units.

Text Books :
1. Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media
2. David G. Luenberger (1969), Optimization by Vector Space Methods, John Wiley & Sons(NY).
3. G. Strang (2016). Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition,USA.

Reference Books:
1. Bendat, J. S. and A. G. Piersol (2010). Random Data: Analysis and Measurement Procedures, 4th Edition, John Wiley & Sons, Inc., NY, USA.
2. Montgomery, D. C. and G. C. Runger (2011). Applied Statistics and Probability for Engineers, 5th Edition, JohnWiley & Sons, Inc., NY,USA.
3. Cathy O'Neil and Rachel Schutt (2013). Doing Data Science, O'Reilly Media
4. Data Sciences, Jain V.K., Khanna Publishing House, Delhi

Walchand Institute of Technology, Solapur
Honors Degree in Data Science
T.Y. B. Tech. (Information Technology), Semester - V
22ITU5HD1T: DATA PRE-PROCESSING AND VISUALIZATION

Teaching Scheme
Lecture: 3 Hours /Week, 3 Credits
Practical: 2 Hours /Week, 1 Credits

Examination Scheme
ESE- 60 Marks
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

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 Style sheets, 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)

Walchand Institute of Technology, Solapur
Honors Degree in Data Science
T.Y. B. Tech. (Information Technology), Semester - VI
22ITU6HD1T: MACHINE LEARNING

Teaching Scheme
Lecture: 3 Hours /Week, 3 Credits
Practical: 2 Hours /Week, 1 Credit

Examination Scheme
ESE – 60 Marks
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, and 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, KeepingModel 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).

Walchand Institute of Technology, Solapur
Honors Degree in Data Science
Final Year B. Tech. (Information Technology), Semester - VII
22ITU7HD1T: PREDICTIVE ANALYTICS

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:

Predictive Analytics aims to build computational abilities, inferential thinking, and practical skills for tackling core data scientific challenges and how predictive analytics tools can be used to analyze real-life business problems.

Prerequisite:

Before proceeding with this course, students should have an understanding of the basic mathematical modeling, linear algebra, statistical analysis and programming.

COURSE OUTCOMES:

At the end of the course the student will able to

1. Explore the predictive analytics and identify common use cases.
2. Analyze and interpret data.
3. Identify and apply appropriate algorithms for developing predictive model.
4. Make the decision using data classification algorithms.

SECTION-I

Unit 1 – Entering the Arena (06)
Exploring predictive analysis, adding business values, starting predictive analytic project, ongoing predictive analysis, surveying the marketplace,

Unit 2 – Predictive analytics in the wild (04)

Online marketing and retail, implementing a recommender system, Target marketing, personalization.

Unit 3 – Data types and associated techniques (06)

Recognizing your data types, Identifying data categories, generating predictive analytics, connecting to related disciplines,

Unit 4 – Complexities of data (06)

Finding value in your data, constantly changing your data, complexities in searching your data, exploration of raw data.

SECTION-II

Unit 5 – Applying Models   (06)

Modeling data: Models and simulations, categorizing model. Healthcare analytics case studies: Google flu trends, Cancer survivability predictors, Social and marketing analytics case studies: Target store predicts pregnant women, Twitter-based predictors of earthquakes, Twitter-based predictors of political campaign outcomes, Twitter-based predictors of the stock market.

Unit 6 – Identifying similarities in data   (06)

Explaining data clustering, Converting raw data into matrix: Creating a matrix of terms in documents, term selection, identifying groups in data: K-means clustering algorithm, clustering by nearest neighbors, Finding Association in data items.

Unit 7 –Predicting the future using data classification   (06)

Explaining data classification, introducing data classification to business, exploring the data classification process, using data classification predict the future: Decision tree, algorithm for generating decision trees, support vector machines.

Unit 8 Developing a Road map for predictive analytics model   (07)

Ensemble methods to boost prediction accuracy: naïve bayes classification algorithms, the markov model, linear regression, neural networks, Listing the objectives, Processing the data: Identifying the data, cleaning the data, generating the derive data, reducing the dimensionality of data, applying principal component analysis, structuring the data: extracting, transforming and loading the data, keeping the data up to date, outlining testing and test

----

Internal Continuous Assessment (ICA) :

Develop a predictive analysis model using the any dataset for the real time case studies.
Tasks to be performed are:

 Data exploration and analysis

1. Data preprocessing improve the quality of data for model building.
2. Train classification algorithm or consider any machine learning algorithm.
3. Split data into training and testing data.
4. Evaluate the performance of the model based on prediction on testing data sets.
5. Select the best suitable model for the given data set based on evaluation result
6. Tune the selected model to improve its performance.
7. Use hyper parameter tuning process using grid search technique.
8. Develop the interface for prediction of new data.

Text Book :

1. Predictive Analytics for dummies A Wiley brand: Anasse Bari, Mohamed Chaouchi, Tommy Jung. 2nd edition, published by John Wiley & Sons.

Reference Books :

1. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die 2nd Edition, Kindle Edition by Eric Siegel (Author) Format: Kindle Edition
2. Olivia Parr Rud "Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management", Wiley,2001.
3. Risk Management in Global Supply Chains", Wiley,2012.

Walchand Institute of Technology, Solapur
Honors Degree in Data Science
Final Year B. Tech. (Information Technology), Semester - VII
22ITU7HD2P: MINI PROJECT

Teaching Scheme
Practical: 4 Hours/Week, 2 Credits

Examination Scheme
POE – 50 Marks
ICA - 50 Marks

Introduction:
Project based learning is a paradigm which is becoming time-honored now a days. To keep abreast with this, Project course is included in the curriculum which is spread over both semesters of final year. For this course, students carry out a project as a team that allows them to demonstrate their abilities and to develop skills within their chosen area of interest. Hardware realization as well software projects with focus on design, development and research aspects are accepted. Also communicating effectively, both in oral and written form are an important skill for engineering graduates in many different contexts. This course also aims to foster these skills.

COURSE OUTCOMES:
1. Analyze technological alternatives for developing IT solution with relevance to environment and sustainability.
2. Explore state-of-art tools and FOSS alternatives to develop solutions meeting societal and professional needs.
3. Develop a system through Software Development Life Cycle.
4. Demonstrate ability to engage in teamwork while observing professional ethics.
5. Write and present a well organized project report
6. Inculcate habit of self study and lifelong learning.