

VEERMATA JIJABAI TECHNOLOGICAL INSTITUTE
(VJTI)
MATUNGA, MUMBAI 400 019

(Autonomous Institute affiliated to University of Mumbai)



Curriculum
(Scheme of Instruction & Evaluation and Course contents)

For
Multi-Disciplinary Minor
in
Data Science

Implemented for the batch admitted in Academic Year 2023-24

	V J T I Veermata Jijabai Technological Institute (Central Technological Institute, Maharashtra State, INDIA) H. R. Mahajani Marg, Matunga, Mumbai 400019 Tel.No. +91 22 24198101-02 Fax +91 22 24102874 www.vjti.ac.in
---	--

(Autonomous Institute affiliated to University of Mumbai)

Curriculum

(Scheme of Instruction & Evaluation and Course contents)

For

Multi-Disciplinary Minor

in

Data Science

	V J T I Veermata Jijabai Technological Institute (Central Technological Institute, Maharashtra State, INDIA) H. R. Mahajani Marg, Matunga, Mumbai 400019 Tel.No. +91 22 24198101-02 Fax +91 22 24102874 www.vjti.ac.in
---	---

Credit Framework for
Multi-Disciplinary Minor
in
Data Science

Sr.	Semester	Course Code	Course Name	L	T	P	Hr	Cr	Examination Weightage in %		
									TA	MST	ESE
1	III	R5CO2201T	Introduction to Data Science	2	0	0	2	2	20	30	50
2	IV	R5CO2205T	Data Visualization, Regression and Prediction	2	0	0	2	2	20	30	50
3	V	R5CO3201T	Introduction to ML with python	3	0	0	3	3	20	30	50
4	VI	R5CO3205T	Data Visualization and Processing Tools	3	0	0	3	3	20	30	50
5	VII	R5CO4201T	Web Data Mining	3	1	0	4	4	20	30	50
Total								14			

Abbreviations: **L** Lecture, **T** Tutorial, **P** Practical, **TA** Teacher Assessment / Term work Assessment, **MST** Mid Semester Test, **ESE** End Semester Written Examination,

Programme Name		Semester	III
Course Code	R5CO2201T		
Course Title	Multi-disciplinary Minor-I: Introduction to Data Science		
Course Scheme	(L-T-P: 2-0-0) (TA-MST-ESE: 20-30-50)		
Credits	2		
Course Type	Multi-disciplinary Minor-I (Data Science)		
Prerequisites: Nil			
Course Outcomes: At the end of the course student will be able to: CO1. To Provide the knowledge and expertise to become a proficient data scientist. CO2. Demonstrate an understanding of statistics and machine learning concepts that are vital for data science. CO3. Produce Python code to statistically analyse a dataset.			

Course Contents		CO Mapped	Teaching Hours/Week
1	Introduction to Data Science Different Sectors using Data science, Purpose and Components of Python in Data Science.	1	4
2	Data Analytics Process Knowledge Check, Exploratory Data Analysis (EDA), EDA- Quantitative technique, EDA- Graphical Technique, Data Analytics Conclusion and Predictions.	2	6
3	Feature Generation and Feature Selection Extracting Meaning from Data- Motivating application: user (customer) retention- Feature Generation (brainstorming, role of domain expertise, and place for imagination)- Feature Selection algorithms.	3	6
4	Data Visualization Basic principles, ideas and tools for data visualization, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset.	3	8

Text Books	
1	Business Analytics: The Science of Data - Driven Decision Making, U Dinesh Kumar, John Wiley & Sons.
2	Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, John Wiley & Sons.
Recommended Reading	
1	Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher.

Programme Name		Semester	IV
Course Code	R5CO2205T		
Course Title	Multi-disciplinary Minor-II Data Visualization, Regression and Prediction		
Course Scheme	(L-T-P: 2-0-0) (TA-MST-ESE: 20-30-50)		
Credits	2		
Course Type	Multi-disciplinary Minor-II (Data Science)		
Prerequisites: Nil			
Course Outcomes: At the end of the course student will be able to: CO1. Demonstrate a strong understanding of core Python concepts. CO2. Utilizing powerful Python libraries like Pandas (Data Frames) and NumPy (multidimensional arrays) to load, clean, explore, analyse, and transform datasets of varying sizes and complexities.			

Course Contents		CO Mapped	Teaching Hours/Week
1	Basics of Python data types, operators, variables, and expressions. Control structures using sample dataset, objects and functions. Python sequence data structures including String, Array, List, Tuple, Set, and Dictionary, hashes.	1	4
2	Data Analysis libraries Using Pandas, Data Frames, Numpy multidimensional arrays, and SciPy libraries to work with a various dataset. Machine Learning in detail.	1	4
3	Exploratory data analysis Data pre-processing (data loading, dealing with missing values and outliers, data wrangling). ANN,SVM,KNN,DT etc.	2	4
4	Exploratory data analysis Data pre-processing (Data Normalization, Data Formatting, data cleaning). More APIs and library relevant to Data Processing with examples.	2	4

Text Books	
1	Data Visualization with Python and JavaScript, Kyran Dale, Shroff Publisher/O'Reilly Publisher Publication
2	Data Science Using Python and R by Chantal D. Larose and Daniel T. Larose, Wiley Publication.
Recommended Reading	
1	Python for Data Science and Visualization –Beginners to Pro, Udemy.

Programme Name	
Course Code	R5CO3201T
Course Title	Introduction to ML with python
Course Type	MDM
Prerequisites: Nil	
Course Outcomes: At the end of the course student will be able to: CO1. Learn the basics of Python used for Machine Learning. CO2. Understand different models of Machine Learning. CO3. Apply the knowledge to solve Machine Learning problems.	

	Course Contents	Hrs.	CO
1.	Introduction: Types of Learning, Essential Libraries and Tools - scikit-learn, NumPy, SciPy, matplotlib, pandas, and editors	4	CO1,2
2.	Data Pre-processing: Data cleaning and formatting, Feature extraction for supervised and unsupervised learning Supervised Learning: Regression - Linear Regression, Logistic Regression, Classification - Nearest Neighbours, Decision Trees, Naive Bayes, SVM	6	CO3
3.	Unsupervised Learning and Pre-processing: Challenges in Unsupervised Learning, Pre-processing and Scaling, Different Kinds of Pre-processing, Applying Data Transformations, Scaling Training and Test Data Clustering Algorithms: Evaluation of algorithms	7	CO4
4.	Evaluating Machine Learning Algorithms and Model Selection: Cross Validation, Cross-Validation in scikit-learn, Types and Benefits of Cross Validation	7	CO5,6

Text Books	
1.	Joel Grus, “Data Science from Scratch: First Principles with Python”, 2nd Edition, O’Reilly Publication.
2.	Andreas C. Müller and Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”.

Course Code	R5CO3205T
Course Title	Data visualization & Processing Tool
Course Type	MDM
Prerequisites: Nil	
Course Outcomes: At the end of the course student will be able to:	
CO1. Learn the basics of Python for data visualization.	
CO2. Understand different models and techniques used for visualization.	
CO3. Apply the knowledge to solve various data visualization problems.	

	Course Contents	Hrs.	CO
1.	Data Import and Visualization: Introduction to various plots, Implementing data visualization techniques using Python libraries such as Matplotlib and Seaborn	4	CO1,2
2.	Chart and Plot Creation: Line, area, histograms, bar, pie, box, scatter, and bubble charts	6	CO3
3.	Correlation Visualization and Analysis: Dimensionality Reduction Algorithms, PCA Analysis	7	CO4
4.	APIs for Advanced Processing: APIs for various learning models	7	CO5,6

Text Books	
1.	Joel Grus, “Data Science from Scratch: First Principles with Python”, 2nd Edition, O’Reilly Publication.
2.	Andreas C. Müller and Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”.

Programme Name		Semester	VII
Course Code	R5CO4201T		
Course Title	Web Data Mining		
Course Scheme	(L-T-P: 3-1-0) (TA-MST-ESE: 20-30-50)		
Credits	3		
Course Type	Multi-disciplinary Minor - V (Data Science)		
Prerequisite	Programming Fundamentals, Basic understanding of Data Structures and AI-gorithms, HTML Basics, Logical & Analytical Skills, Introductory Data Han-dling, Mathematical Foundation		
Course Outcomes: At the end of the course student will be able to			
CO1. Understand the fundamentals of web data mining, web architecture, data types, and ethical issues in web data collection.			
CO2. Design and implement web crawling and scraping techniques, including crawler algorithms and handling real-world challenges.			
CO3. Understand web mining techniques such as usage mining, structure mining, and link analysis (e.g., PageRank) to discover patterns and insights.			
CO4. Apply web mining techniques to real-world applications such as sentiment analysis and search systems.			

	Course Contents	Hrs.	COs
1.	Introduction to Web Data Mining & Foundations: Overview of Data Mining and Web Mining, Types of Web Mining: Content, Structure, Usage Mining, Types of Data: Structured, Semi-Structured, Unstructured, Role of Web Data in Data Science, Web Anatomy: HTTP Request/Response, HTML, CSS, DOM, Basics of Web Crawling and Web Scraping.	6	1
2.	Ethical Web Scraping & Crawling Fundamentals: Ethical and Legal Issues in Web Data Collection, Crawling Restrictions, Web Crawlers: Breadth First (BFS) Crawlers and Preferential Crawlers, Universal Crawlers vs Focused Crawlers, Implementation Issues: Fetching and parsing, Stopword removal, Link extraction and Spider traps, Page repository management.	7	1,2
3.	Web Usage Mining: Web Usage Mining Overview, Data Collection and Preprocessing, Sources of Web Usage Data (logs, clicks, sessions), Data Modeling for Usage Mining, Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Sequential and Navigation Pattern Mining.	6	3,4

4.	Web Structure Mining: Web as a Graph Structure, Link Analysis Fundamentals, PageRank Algorithm, Introduction to Social Network Analysis (SNA), Centrality Measures (Degree, Betweenness, Closeness – basic idea), PageRank Algorithm: Basic formula, Damping factor concept, Iterative computation, Simple numerical examples.	7	3,4
5.	Information Retrieval: Concepts of Information Retrieval (IR), IR Models: Boolean Model, Vector Space Model, Web Page Preprocessing, Query Processing and Retrieval, Web Search Engines and Meta Search Engines.	5	4
6	Applications: Duplicate Detection Techniques, Opinion Mining, Sentiment Classification, Web Usage Pattern Analysis, Web Search and Ranking.	5	4

TEXTBOOKS	
1.	Bing Liu, <i>Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data</i> .
REFERENCE BOOKS	
1.	Soumen Chakrabarti, <i>Mining the Web: Discovering Knowledge from Hypertext Data</i> .
2.	George Chang et al., <i>Mining the World Wide Web: An Information Search Approach</i> .

