Kolhapur Institute of Technology's
College of Engineering
(Autonomous), Kolhapur

KOLHAPUR INSTITUTE OF TECHNOLOGY'S
COLLEGE OF ENGINEERING KOLHAPUR
(EMPOWERED AUTONOMOUS)

Department of Computer Science &Engineering
Curriculum and Syllabus for
T.Y.B. Tech. Computer Science & Engineering
Scheme 2024-25 (As Per NEP, 2020)
Emerging Minor Courses

Dr. Lingraj A. Hadimani
Head,
Department of Comp. Sci. & Engg.
KIT's College of Engg. (Autonomous)
Kolhapur

R.S.Sutar

Mrs. Ranjeeta Sutar
Academic Coordinator, CSE

Dr. L. A. Hadimani
Head, CSE

Emerging Minor in Data Science										
Course Code	**Course Name**	**L**	**T**	**P**	**Hrs /Week**	**Credits**	**Evaluation Scheme (Components)**			**Semester**
UCSMN0361	Mathematics for Data Science	3	1		4	4	ESE	100	40	III
UCSMN0461	Introduction to Data Science	3	1		4	4	ESE	100	40	IV
UCSMN0561	Computational Data analytics	3	1		4	4	ESE	100	40	V
UCSMN0661	Web Data Mining	3	1		4	4	ESE	100	40	VI
UCSMN0761	Python for Data Science	2			2	2	ESE	100	40	VII
		14	4		18	18	**Total Marks: 500** **Total Credit: 18**			

Title of the Course: UCSMN0361 **Course Code:** Mathematics for Data Science	**L**	**T**	**P**	**Credits**
	3	1	-	4

Course Pre-Requisite: NIL

Course Description:

This course provides students foundational mathematical concepts essential for data science, including linear algebra, probability, statistics, calculus, and discrete mathematics. The course focuses on analytical thinking and problem-solving through mathematical reasoning, preparing students to understand and support data-driven decision-making.

Course Learning Objectives:

1. Equip students with the mathematical tools required for machine learning and data analysis.
2. Apply mathematical reasoning in data-driven problem solving.
3. Lay the foundation for advanced AI and Data Science courses.

Course Outcomes:

CO	**After Completion of the course, the student should be able to**	**Bloom's Cognitive**	
		Level	**Descriptor**
CO1	Recall fundamental mathematical concepts such as matrices, probability, and statistics used in data science applications.	1	Recall
CO2	Explain statistical data, probability distributions, and visualizations relevant to real-world datasets.	2	Explain
CO3	Apply linear algebra and calculus techniques to solve optimization and modeling problems in data science.	3	Solve
CO4	Analyze datasets using mathematical tools to uncover patterns, relationships, and dimensionality reductions (e.g., PCA).	4	Analyze

CO-PO Mapping:

CO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO 10**	**PO 11**	**PSO 01**	**PSO 02**
CO1	3	1	-	-	-	-	-	-	-	-	2	1	1
CO2	3	2	-	1	-	-	-	-	-	-	2	2	1
CO3	3	3	2	2	2	-	-	-	-	-	2	2	1
CO4	3	3	2	3	2	-	-	-	-	-	2	2	1

Assessment:

Assessment	Marks
ESE	100

ESE: End Semester Evaluation is based on a written exam for 100 marks with 100% course content i.e. questions will be from all the six modules.

Course Contents:

Unit 1: Linear Algebra and Matrix Computations	**8 Hours**

Vectors, vector spaces, linear independence, Matrix operations: addition, multiplication, inverse, Rank, nullity, and solution of linear systems Eigenvalues, eigenvectors, diagonalization, Application: Principal Component Analysis (PCA)

Unit 2: Probability and Random Variables	**7 Hours**

Set theory and basic probability, Conditional probability, independence, cBayes' Theorem and applications, Discrete and continuous random variables, Distributions: Binomial, Normal, Poisson

Unit 3: Descriptive and Inferential Statistics	**8 Hours**

Data types, visualization (histograms, box plots),Measures: mean, median, variance, standard deviation,Sampling and estimation,Hypothesis testing, confidence intervals,Central Limit Theorem,Use case: A/B Testing

Unit 4: Differential Calculus and Optimization	**7 Hours**

Functions and limits review,Derivatives and rules,Partial derivatives and gradients,Optimization concepts,Gradient Descent Algorithm,Application: Cost function minimization in ML

Unit 5: Discrete Mathematics for Data Modeling	**7 Hours**

Logic: Propositions, logical operators, truth tables,Sets and relations,Functions and mappings,Graph theory basics: trees, networks,Applications: Graph algorithms in social networks

Unit 6: Foundations of Machine Learning Mathematics	**8 Hours**

Norms, inner product spaces,Orthogonality and projections,Convex sets and functions,Regularization in models (L1, L2),Introduction to loss functions,Application: Ridge and Lasso Regression

Textbooks:
1. Mathematics for Machine Learning, Cambridge University Press (Open Access) 2. Introduction to Probability, Joseph K. Blitzstein and Jessica Hwang, 3. MIT OCW Video Lectures & Book References:https://ocw.mit.edu
Reference Books:
1. Probability and Statistics for Engineers and Scientists, Pearson

Title of the Course: UCSMN0461 **Course Code:** Introduction to Data Science	**L**	**T**	**P**	**Credits**
	3	1	-	4

Course Pre-Requisite: Basic Knowledge in Computers and Mathematics

Course Description:

This course is to get basic knowledge about data science and its processes. It provides different statistical methods to work on data. It also enlightens machine learning approaches in data science.

Course Learning Objectives:

1. To introduce students to the foundational concepts, scope, and interdisciplinary nature of data science in real-world applications.
2. To enable learners to interpret various types of data and understand basic techniques for data collection, preprocessing, and statistical analysis.
3. To provide a conceptual understanding of sampling, inferential statistics, and theoretical aspects of machine learning models.
4. To develop awareness of ethical, legal, and governance challenges in data usage, with a focus on privacy and fairness in the Indian context.

Course Outcomes:

CO	**After Completion of the course, the student should be able to**	**Bloom's Cognitive**	
		Level	**Descriptor**
CO1	Illustrate the fundamental concepts and lifecycle of data science.	2	Understand
CO2	Apply basic statistical concepts to summarize and interpret data.	3	Apply
CO3	Examine theoretical aspects of machine learning and data science applications.	4	Analyse
CO4	Analyse data privacy, ethics, and interdisciplinary approach from Data Science.	4	Analyse

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PSO 01	PSO 02
CO1	3	1								1		1	
CO2	3	1		1						1		2	
CO3	2	2	1	3						2	2	3	2
CO4	1	2	1	3	3	1	3			2	2	3	2

Assessment:

Assessment	Marks
ESE	100

ESE: End Semester Evaluation is based on a written exam for 100 marks with 100% course content i.e. questions will be from all the six modules.

Course Contents:

Unit 1: Introduction	**6 Hours**

What is Data, Types of Data, AI vs ML vs DS, Introduction to Data Science, The Data Science Lifecycle (Problem definition to decision making), Evolution of Data Science, Data Science Roles, Applications of Data Science in various fields, Web Scrapping overview

Unit 2: Data Collection and Data Pre-Processing	**7 Hours**

Data Collection Strategies, Data Pre-Processing, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Representation: Tabular Data, Time Series, Graphs, Text

Unit 3: Statistics & Data Understanding	**8 Hours**

Descriptive statistics: Mean, median, mode, range, variance, standard deviation
Data distributions: Normal, skewed, bimodal
Graphical summaries: Histograms, boxplots, scatter plots (conceptual)
Interpreting visual data: spotting trends and outliers, Population vs Sample
Sampling methods: random, stratified, systematic

Unit 4: Machine Learning Concepts	**7 Hours**

What is a model? Concept of learning from data, Supervised vs Unsupervised Learning (conceptual), Regression vs Classification examples, Model performance metrics: accuracy, precision, recall, F1-score, Cross-validation and overfitting/underfitting

Unit 5: Data Science Applications & Emerging Technologies	**7 Hours**

Applications of Data Science in Smart Infrastructure (Cities, Transport), Predictive Analytics in Industrial Automation and Manufacturing, Overview of Edge Computing and IoT in real-time data collection, Case Studies from Indian sectors: Healthcare, Education, Agriculture

Unit 6: Ethical & Future Directions	**7 Hours**

Introduction to AI ethics, algorithmic bias, and social impact, Interdisciplinary Data Science: Civil, Mech, EEE, ECE, Biotech use-cases, Importance of Responsible AI: Explainability, Fairness, and Transparency, Future of data science: skilling, research, entrepreneurship

Textbooks:

1. V.K. Jain – Fundamentals of Data Science (Khanna Book Publishing)
2. Cathy O'Neil – Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishing)
3. Foster Provost & Tom Fawcett – Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking (O'Reilly Media)

Reference Books:

1. Rafael A. Irizarry, Introduction to Data Science: Data Analysis and Prediction Algorithms with R, CRC Press. (Early chapters provide strong theoretical foundations)
2. John D. Kelleher, Brendan Tierney – Data Science (MIT Press Essential Knowledge Series)

Title of the Course: UCSMN0561 **Course Code:** Computational Data analytics	**L**	**T**	**P**	**Credits**
	3	1	-	4

Course Pre-Requisite: Mathematics for Data Science, Introduction to Data Science, Basic Programming (preferably in Python), Understanding of Probability and Statistics.

Course Description:

This course focuses on the computational techniques and analytical tools used for data-driven decision-making. It equips students with algorithmic approaches and programming skills necessary to process, analyse, and visualize complex datasets. Emphasis is placed on practical applications involving real-world datasets across domains.

Course Learning Objectives:

1. To introduce the concepts and computational frameworks used in data analytics.
2. To equip students with the ability to perform advanced data wrangling and feature engineering techniques.
3. To enable students to evaluate and interpret machine learning models using modern explainable AI (XAI) tools such as SHAP and LIME.
4. To develop competency in handling large-scale datasets using Big Data tools such as Hadoop and PySpark for scalable analytics.
5. To apply analytical techniques to real-world problems involving time series and textual data, and build complete data analytics pipelines through hands-on case studies.

Course Outcomes:

CO	**After Completion of the course, the student should be able to**	**Bloom's Cognitive**	
		Level	**Descriptor**
CO1	Understand the role of computation and tools in data analytics	2	Understand
CO2	Apply advanced data wrangling, feature selection, and data processing techniques.	3	Apply
CO3	Analyse model behaviour using interpretability and explainable AI tools.	4	Analyse
CO4	Evaluate end-to-end data analytics pipelines using real-world datasets.	5	Evaluate

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PSO 01	PSO 02
CO1	3	2		3	3			1		2	3	3	3
CO2	3	2	1	2	3			1		2	3	3	3
CO3	3	2	1	3	3			1		2	3	3	3
CO4	3	2	1	3	3			1		2	3	3	2

Assessment:

Assessment	Marks
ESE	100

ESE: End Semester Evaluation is based on a written exam for 100 marks with 100% course content i.e. questions will be from all the six modules.

Course Contents:

Unit 1: Introduction to Computational Data Analytics	**6 Hours**

Role of computation in data science, Types of data analytics: Descriptive, Predictive, Prescriptive, Structured vs Unstructured data, Overview of data analytics lifecycle, Tools and technologies: Python, R, Jupyter, Pandas, NumPy

Unit 2: Data Wrangling and Feature Selection	**8 Hours**

Advanced feature engineering: polynomial features, interaction terms, Feature selection techniques: Filter, Wrapper, Embedded methods, Handling imbalanced datasets: SMOTE, undersampling

Unit 3: Model Interpretability and Explainable AI (XAI)	**8 Hours**

Importance of model interpretability, SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), Partial dependence plots and feature importance, Case studies in healthcare and finance.

Unit 4: Big Data Analytics and Scalable Computing	**7 Hours**

Introduction to Big Data ecosystem, Hadoop and Spark architecture, PySpark for large-scale data analysis, DataFrames and RDDs, Real-time data streaming (basic concepts)

Unit 5: Graph Analytics and Network Data Analysis	**8 Hours**

Introduction to graph theory and its relevance in data analytics, Representation of data as networks: nodes, edges, and adjacency structures, Applications of graph analytics in social

networks, biological networks, and recommendation systems, Community detection algorithms and clustering in graphs.

Unit 6: End-to-End Analytics Projects and Case Studies	**8 Hours**

Designing a data analytics pipeline, Case studies: e.g., stock prediction, churn analysis, sentiment analysis, Tools: Jupyter, scikit-learn, matplotlib, seaborn, Students work on a mini-project with presentation

Textbooks:

1. Cathy O'Neil and Rachel Schutt, Doing Data Science, O'Reilly Media, 2013.
2. Wes McKinney, Python for Data Analysis, O'Reilly, 2nd Edition, 2017.
3. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly, 2nd Edition, 2019.

Reference Books:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer.
2. Anurag Bhardwaj et al., Practical Data Science Cookbook, Packt Publishing.
3. Jules S. Damji et al., Learning Spark: Lightning-Fast Data Analytics, O'Reilly, 2nd Edition.
4. Christoph Molnar, Interpretable Machine Learning, 2022 (open source book).

Title of the Course: UCSMN0661 **Course Code:** Web Data Mining	**L**	**T**	**P**	**Credits**
	3	1	-	4

Course Pre-Requisite: Basic concepts in data structure, database engineering, web technologies, data mining and statistics.

Course Description:

This course focuses on the principles and techniques used to extract meaningful knowledge from web data. It covers the three main areas of web mining: web content mining, web structure mining, and web usage mining. Students will learn how to analyze and process web-based data using various types of data mining techniques.

Course Learning Objectives:

1. Understand fundamentals of data mining and web mining for discovering frequent patterns in data.
2. Analyze and evaluate information retrieval models and web search techniques.
3. Apply link analysis and social network mining methods such as PageRank, HITS, and community detection.
4. Design and evaluate web crawling strategies using universal, focused, and topical crawler algorithms with ethical considerations.
5. Develop and maintain wrappers for structured data extraction from web pages using DOM trees and wrapper induction techniques.
6. Model and mine web usage data, query logs, and user behavior for pattern discovery and computational advertising .

Course Outcomes:

COs	**After Completion of the course, the student should be able to**	**Bloom's Cognitive**	
		Level	**Descriptor**
CO1	Explain the fundamental concepts of data mining and web mining	2	Explain
CO2	Apply algorithms for social network analysis and structured web crawling.	3	Apply
CO3	Make use of key techniques in information retrieval and web search.	3	Make use of
CO4	Build solutions for web usage mining and structured data extraction using wrapper generation and DOM tree models.	3	Build

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10	PO 11	PSO 01	PSO 02
CO1	1	1										1	2
CO2	2	2	2	2	2	1	3					2	2
CO3	2	3	2	2	2	1	3					2	2
CO4	2	3	2	3	2	1	3					2	2

Assessment:

Assessment	Marks
ESE	100

ESE: End Semester Evaluation is based on a written exam for 100 marks with 100% course content i.e. questions will be from all the six modules.

Course Contents:

Unit 1: Introduction to Data Mining	**7 Hours**

What is data mining? What is web mining? Basic Concepts of Association Rules. Apriori Algorithm for data mining.Data Formats for Association Rule Mining.Mining with Multiple Minimum Supports.

Unit 2: Information Retrieval and Web Search	**8 Hours**

Basic Concepts of Information Retrieval , Information Retrieval Models, Evaluation Measures, Text and Web Page Pre-Processing,Inverted Index and Its Compression. Web Search: Meta-Search: Combining Multiple Rankings,Web Spamming.

Unit 3: Social Network Analysis	**7 Hours**

Social Network Analysis, Co-Citation and Bibliographic Coupling, PageRank, HITS, Community Discovery.

Unit 4: Web Crawling	**7 Hours**

A Basic Crawler Algorithm, Implementation Issues,Universal Crawlers, Focused Crawlers,Topical Crawlers, Evaluation, Crawler Ethics and Conflicts.

Unit 5: Structured Data Extraction: Wrapper Generation	**8 Hours**

Two Types of Data Rich Pages,Data Model. Wrapper Induction:Extraction from a Page, Learning Extraction Rules, Identifying Informative Examples , Wrapper Maintenance. Building DOM Trees.

Unit 6: Web Usage Mining	**8 Hours**
Data Collection and Pre-Processing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web Usage Patterns,Query Log Mining, Computational Advertising.	
Textbooks:	
1. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data- Bing Liu-2th Edition (Springer)	
Reference Books:	
1. Data Mining: Concepts and Techniques- Jiawei Han, Micheline Kamber, Jian Pei(Morgan Kaufmann) 2. Introduction to Information Retrieval-Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze(Cambridge University Press) 3. Networks, Crowds, and Markets: Reasoning About a Highly Connected World-David Easley and Jon Kleinberg(Cambridge University Press) 4. Foundations of Web Crawlers-Olston & Najork 5. Introduction to Information Retrieval – Manning, Raghavan, Schütze	

Course Code: UCSMN0761	**L**	**T**	**P**	**Credits**
Course Name: Python for Data Science	2	-	-	2

Course Pre-Requisite:

Basic mathematical concepts, Data Structures,

Course Description:

This course introduces third-year engineering students to Python programming with an emphasis on data science applications. Students will learn essential data analysis tools, including NumPy, Pandas, Matplotlib, and Scikit-learn, and will work on real-world datasets to develop skills in data preprocessing, visualization, and basic machine learning.

Course Learning Objectives:

1. Apply Python programming and relevant libraries (NumPy, Pandas, Matplotlib, Seaborn) for data manipulation, analysis, and visualization.
2. Implement data preprocessing and exploratory data analysis (EDA) techniques to understand and prepare datasets.
3. Understand and apply fundamental supervised and unsupervised machine learning algorithms using Scikit-learn.
4. Analyze real-world problems using data science methodologies and interpret the results.

Course Outcomes:

CO	**After Completion of the course, the student should be able to**	**Bloom's Cognitive**	
		Level	**Descriptor**
CO1	Identify and explain fundamental concepts in Python programming, data handling, and basic machine learning. (Remembering, Understanding)	1	Identify
CO2	Utilize Python libraries to process, explore, and visualize data, and implement basic machine learning algorithms. (Applying)	2	Utilize
CO3	Analyze datasets using exploratory data analysis techniques and interpret the results of basic machine learning models. (Analyzing)	3	Analyze
CO4	Evaluate the suitability of different data science techniques for specific problems and propose solutions using a structured approach. (Evaluating, Creating)	4	Evaluate

CO-PO Mapping:

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO01	PSO02
CO1	3	1								1		1	
CO2	3	1		1						1		2	
CO3	2	2	1	3						2	2	3	2
CO4	1	2	1	3	3	1	3			2	2	3	2

Assessment:

Assessment	Marks
End Semester Examination (ESE)	100

Course Contents:

Unit 1: Introduction to Python & Data Science:	**5 Hours**
Overview of Data Science and its Applications; Python basics: Variables, Data Types, Operators; Control Structures: If-else, Loops; Functions and Modules; File Handling.	
Unit 2: Working with Python Libraries	**5 Hours**
NumPy: Arrays, Operations, Indexing, Broadcasting; Pandas: Series, DataFrames, Reading/Writing CSV, Filtering, Aggregation; Matplotlib and Seaborn: Data Visualization Techniques.	
Unit 3: Data Preprocessing and Cleaning	**5 Hours**
Handling Missing Data; Data Transformation & Normalization; Encoding Categorical Data; Data Merging, Concatenation, and Reshaping.	
Unit 4: Exploratory Data Analysis (EDA)	**5 Hours**
Univariate, Bivariate, Multivariate Analysis; Correlation and Covariance; Outlier Detection and Handling; Feature Selection Basics.	
Unit 5: Introduction to Machine Learning with Scikit-learn:	**5 Hours**
Overview of Supervised and Unsupervised Learning; Model Building Workflow; Linear Regression, Logistic Regression; Decision Trees, K-Nearest Neighbors; Clustering: K-Means Basics.	
Unit 6: Real-World Applications & Mini Projects	**5 Hours**
Mini Project: End-to-end Data Analysis using a Public Dataset; Application Areas: Finance, Healthcare, Manufacturing, social media, Basics of Model Evaluation and Validation**.**	

Textbooks:
1. Python for Data Analysis by Wes McKinney
Reference Books:
1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron 2. Online Platforms: Kaggle, UCI Machine Learning Repository