Karmaveer Bhaurao Patil University, Satara

Syllabus for

M. Sc. II (Data Science)

Under

Faculty of Science and Technology

(As per NEP 2020)

With effect from Academic Year 2025-2026

SYLLABUS FOR MASTER OF SCIENCE DATA SCIENCE

1. Title: Subject: Data Science

2. Year of implementation: June 2025 onward

3. Preamble:

Welcome to the Master of Science in Data Science program! In today's digital age, the world is generating vast amounts of data at an unprecedented rate. Extracting meaningful insights and making informed decisions from this data has become crucial for businesses, organizations, and societies alike. The M.Sc. Data Science program is designed to equip students with the knowledge, skills, and tools necessary to tackle complex data challenges and drive innovation in various domains.

Our program offers a comprehensive curriculum that combines theoretical foundations with practical applications. Through a blend of rigorous coursework, hands-on projects, and industry collaborations, we aim to cultivate a strong foundation in data science principles while emphasizing real-world problem-solving and critical thinking. Students will develop expertise in statistical analysis, machine learning, data visualization, data mining, and other essential areas, gaining proficiency in both the technical and analytical aspects of data science.

One of the unique aspects of our program is its interdisciplinary nature. Data science transcends traditional disciplinary boundaries, and we encourage students from diverse academic backgrounds to join us on this exciting journey. Whether you have a background in computer science, mathematics, statistics, engineering, or any other related field, this program will provide the necessary bridge to advance your skills and thrive in the data-driven landscape.

Our distinguished faculty comprises leading experts in the field of data science, bringing a wealth of industry experience and research expertise. They are committed to fostering a collaborative and engaging learning environment, where students can interact with faculty, fellow students, and industry professionals to gain valuable insights and expand their networks.

Beyond the classroom, we offer numerous opportunities for practical experience and professional development. Students will have access to cutting-edge technologies, state-of-the-art data labs, and industry partnerships, enabling them to work on real-world data problems and gain hands-on experience with industry-standard tools and platforms. Additionally, we organize workshops, seminars,and guest lectures to expose students to the latest trends, emerging technologies, and industry best practices.

Upon successful completion of the program, graduates will be equipped to make a significant impact in various sectors, including finance, healthcare, marketing, social sciences, and more. They will possess the skills to extract actionable insights from complex data sets, build predictive models, and communicate findings effectively to diverse stakeholders.

We are thrilled to embark on this data science journey with you, empowering you to become a competent data scientist capable of driving innovation and making data-driven decisions. Join us as we explore the fascinating world of data science and unlock the potential of data for a better future

Programme Outcomes

PO. NO.	Programme Outcomes After completing M.Sc. programme, the students will be able to...
PO-1	Students will critically evaluate, analyze, and comprehend a scientific problem. Think creatively, experiment and mic research into innovation and creatively design scientific solutions to problems.
PO-2	Translate project plans, use management skills, and lead a team for planning and execution of a task.
PO-3	Exemplify generate a solution independently, check and validate it and modify if necessary.
PO-4	Students will demonstrate their ability of advanced programming to design and develop innovativeapplications.
PO-5	A student can start his own business or start up.

Program Specific Objectives:

The students will be able to …

1. To create post-graduates with sound knowledge of Data Science, who can contribute towards recentadvances in technology.
2. To provide advanced and in-depth knowledge of data science and specialization in one or two subjectsof the new era of technology.
3. To prepare Postgraduates who will achieve peer-recognition, as an individual or in team,through demonstration of good analytical, design, programming and implementation skills.
4. To enable students, pursue a professional career in Data Science in related industry, business andresearch.

Programme Specific Outcomes

PSO. NO.	**Programme Outcomes** **After completing M.Sc. (Data Science) programme the students will be able to..**
PSO-1	Avail yourself of Current trends in IT Industries and new Technologies.
PSO-2	Apply knowledge of programming platforms in Data Science and AI in real life.
PSO-3	Students should avail detailed knowledge of Data Science, Artificial Intelligence, Machine Learning, andBig Data etc.
PSO-4	Students will demonstrate their ability of advanced programming to design and develop innovativeapplications
PSO-5	Student will be able to Access, evaluate, understand, and compare digital information from varioussources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.

M.Sc. -II Semester-III

Level	Semester	Course code	Course Title	No. of hours Per Week	Credits
6	III	MDST 531	Cloud Computing with AWS	4	4
		MDST 532	Advanced Machine Learning	4	4
		MDST 533	Computer Vision	4	4
		MDST 534 E-I	Data Structures and Algorithm using Python	2	2
		MDST 534 E-II	Advanced Data Engineering	2	2
		MDSP 535	Practical Lab V Based On: MDST 531. MDST 532, MDST 533	2	2
		MDSP 536	Practical Lab VI Based On: Research Project	6	6
Total					**22**

M.Sc. -I Semester-IV

Level	Semester	Course code	Course Title	No.of hours Per Week	Credits
6	II	MDST 541	Deep Learning	4	4
		MDST 542	Big Data Analytics	4	4
		MDST 543	Data Processing using MATLAB	4	4
		MDST 544 E-I	Advanced Recommender System	2	2
		MDST 544 E-II	Data Analytics	2	2
		MDST 545	OJT	4	4
		MDST 546	Practical Lab VII Based On: MDST 541. MDST 542, MDST 543	2	2
		MDST 547	Practical Lab VIII	2	2
Total					**22**

M.Sc. II

SEM III

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 531:Cloud Computing with AWS

Course Objectives:

1) Define cloud computing and its key characteristics.
2) Differentiate between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).
3) Compare and contrast public, private, hybrid, and multi-cloud deployment models.
4) Understand the architecture of cloud computing systems.

Course Outcomes:

1) Define cloud computing and its key characteristics.
2) Understand the service models (IaaS, PaaS, SaaS) and deployment models (public, private, hybrid).
3) Knowledge of major cloud service providers (e.g., Amazon Web Services, Microsoft Azure, Google Cloud Platform) and their offerings.
4) Ability to compare different cloud services.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Overview of Cloud Computing** **1.1** Introduction to Cloud Computing **1.2** Benefits of Cloud Computing, Types of Cloud Computing **1.3** Cloud Services Model, Vertical Scaling **1.4** Different Cloud Platforms.	15
Module -2:	**Module -2: Cloud Service Models**. **2.1** Introduction to Amazon AWS, Introduction to EC2 Service **2.2** AMI, Introduction to EBS, Amazon EBS Encryption, Introduction to EFS **2.3** Amazon S3 Bucket Service, Architecture of S3 Service. **2.4** Virtual Private Cloud (VPC) & Virtual Private Network (VPN), DynamoDB, Benefits of DynamoDB.	15
Module -3:	**Module -3: Cloud Native Technologies** **3.1** Introduction to Amazon Route 53,Types of Classification of Domains **3.2** Relational Database Services (RDS),Identity and Access Management (IAM) **3.3** Introduction to SQS ,Introduction to Over Watch Dashboard **3.4** Cloudtail Service, Overview of Redshift, Elastic Cache.	15
Module -4:	**Module -4: Cloud Computing Architecture** **4.1** Introduction to Lambda Service **4.2** Introduction to Storage **4.3** Getaway **4.4** File Getaway Architecture	15

Reference Books:-

1) Bhowmik Sandeep. Cloud Computing. 1st ed. New Delhi: McGraw Hill Education, 2017.
2) Erl Thomas, Zaigham Mahmood, and Ricardo Puttini. Cloud Computing: Concepts, Technology & Architecture. Boston: Pearson Education, 2013.
3) Foster Ian, and Dennis B. Gannon. Cloud Computing for Science and Engineering. Cambridge, MA: MIT Press, 2018.
4) Henson Charles. Cloud Computing: A Guide for Executives & Business Owners. Charleston, SC:

CreateSpace Independent Publishing Platform, 2021.
5) Mather Tim. The Cloud Computing Book: The Future of Computing Explained. Boca Raton, FL: CRC Press, 2023.
6) Ong Logan. The Self-Taught Cloud Computing Engineer: A Comprehensive Professional Study Guide to AWS, Azure, and GCP. Independently published, 2022.
7) Prakash, Apoorva. Practical Serverless Applications with AWS: Harnessing the Power of Serverless Computing. New York: Apress, 2025.

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 532:Advanced Machine Learning

Course Objectives:

1) Grasp Human learning aspects
2) Learn the primitives in learning process by computer
3) Grasp nature of problems solved with Machine Learning
4) Acquaint with the basic concepts and techniques of Machine Learning.

Course Outcomes:

1) Acquire fundamental knowledge of learning theory.
2) Design and evaluate various machine learning algorithms.
3) Use machine learning methods for multivariate data analysis in various scientific fields.
4) Choose and apply appropriate Machine Learning Techniques for analysis, forecasting, categorization, and clustering of the data.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Concepts of Machine Learning** **1.1** Basic concept Machine Learning and Advanced Machine Learning Models and applications **1.2** Types of learning, Supervised, Unsupervised and semi-supervised,Reinforcement learning techniques **1.3** Models of Machine learning,Geometric model,Probabilistic Models, Grouping and grading models, Parametric and nonparametric models, **1.4** Predictive and descriptive learning,Classification concepts,	15
Module -2:	**Module -2: Learning Theory** **2.1** Feature Extraction,Feature Construction and Transformation, **2.2** Feature Selection, Dimensionality Reduction: Subset selection, the Curse of dimensionality, Principle **2.3** Multidimensional scaling, Linear discriminant analysis, **2.4** VC dimension, Probably Approximately Correct (PAC) learning,	15
Module -3:	**Module -3: Logical, Grouping and Grading Models** **3.1** Decision Tree Representation, Decision tree algorithm,Alternative measures for selecting attributes **3.2** Minimum Description length decision trees ,Ranking and probability estimation trees, **3.3** Regression trees, Clustering trees,Association rule mining **3.4** Distance based clustering K-means algorithm, choosing number of clusters, Clustering around medoids – silhouettes, Hierarchical clustering,	15
Module -4:	**Module -4: Advanced Machine Probabilistic Models** **4.1** Uncertainty, Normal distribution and its geometric interpretations, **4.2** Baye's theorem, Naïve Bayes Classifier, **4.3** Bayesian network, Discriminative learning with maximum likelihood, Hidden Markov model, **4.4** Expectation Maximization methods, Gaussian Mixtures and compression-based models.	15

Reference Books:-

1. Fregly, Chris, and Antje Barth. “Data Science on AWS: Implementing End-to-End, Continuous AI and Machine Learning Pipelines”. Sebastopol, CA: O'Reilly Media, 2021.
2. Gupta, Pramod, and Naresh K. Sehgal. “Introduction to Machine Learning in the Cloud with Python: Concepts and Practices”. Cham, Switzerland: Springer, 2021.
3. Hearty John. “Advanced Machine Learning with Python: Solve Data Science Problems by Mastering Cutting-Edge Machine Learning Techniques in Python”. Birmingham, UK: Packt Publishing, 2016.
4. Körner, Christoph, and Kaijisse Waaijer. “Mastering Azure Machine Learning: Perform Large-Scale End-to-End Advanced Machine Learning in the Cloud with Microsoft Azure Machine Learning. Birmingham”, UK: Packt Publishing, 2020.
5. Prakash, Apoorva. “Practical Serverless Applications with AWS: Harnessing the Power of Serverless Computing”. New York: Apress, 2025
6. Sharma, Avinash Kumar, Nitin Chanderwal, and Amarjeet Prajapati, eds. “Advancing Software Engineering Through AI, Federated Learning, and Large Language Models”. Hershey, PA: IGI Global, 2024.
7. Singh, Himanshu. “Practical Machine Learning with AWS: Process, Build, Deploy, and Productionize Your Models Using AWS”. New York: Apress, 2020

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 533:Computer Vision

Course Objectives:

1) Introduce the fundamental concepts of image processing, including the representation and components of digital images.
2) Provide a deep understanding of digital image representation, sampling, and quantization.
3) Understand and implement methods for image restoration, including noise reduction and filtering techniques.
4) Understand real-world applications of image processing in fields such as computer vision, medical imaging, and object recognition.

Course Outcomes:

1) Different techniques used for enhancement and segmentation
2) Process color images in different color spaces
3) Understand different image registration techniques
4) Understand basic of stenography.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Fundamentals of Image Processing** **1.1** Introduction of Image Processing, Digital Image Representation, **1.2** Components of Image Processing System, pixel operations, Arithmetic Operations, **1.3** Application of Image Processing, Formation of image model, **1.4** Image transforms- DCT, Haar, Hadamard, PCA.	15
Module -2:	**Module -2: Digital Image Fundamentals and Image Enhancement** **2.1** Image sampling and quantization, Color models (RGB, CMYK, HSV, etc.), **2.2** Basic relationships between pixels (Neighbors, connectivity), **2.3** Spatial domain techniques: contrast stretching, histogram equalization. **2.4** Filtering in the spatial domain: smoothing and sharpening filters.	15
Module -3:	**Module -3: Image Restoration and Image Compression** **3.1** Image Restoration, Noise reduction using spatial and frequency domain filters, Inverse filtering, Wiener filtering, **3.2** Fundamentals of image compression Lossy compression techniques: JPEG, MPEG, Thresholding methods, Edge detection (Sobel, Prewitt, Canny), **3.3** Region-based segmentation, Contours, shape descriptors, **3.4** Texture analysis, Face detection and recognition, Object detection.	15
Module -4:	**Module -4: Stereo correspondence and Computer Vision** **4.1** Epipolar geometry Rectification, Plane sweep, Sparse correspondence **4.2** 3D curves and profiles, Dense correspondence, **4.3** Similarity measures, Local methods **4.4** Sub-pixel estimation and uncertainty,Application: Stereo-based head tracking, multi-view stereo, Volumetric and 3D surface reconstruction	15

Reference Books:-

1. Ballard, Dana H., and Christopher M. Brown. "Computer Vision." Englewood Cliffs, NJ: Prentice Hall, 1982.
2. Forsyth, David A., and Jean Ponce. "Computer Vision: A Modern Approach." Upper Saddle River, NJ: Pearson, 2003.
3. Marr, David. "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information." Cambridge, MA: MIT Press, 1982.
4. Prince, Simon J. D. "Computer Vision: Models, Learning, and Inference." Cambridge, UK: Cambridge University Press, 2012.
5. Shapiro, Linda G., and George C. Stockman. "Computer Vision." Upper Saddle River, NJ: Pearson, 2001.
6. Solem, Jan Erik. "Programming Computer Vision with Python: Tools and Algorithms for Analyzing Images." Sebastopol, CA: O'Reilly Media, 2012.
7. Szeliski, Richard. "Computer Vision: Algorithms and Applications." London: Springer, 2011.

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: E-I	**Marks:** 30
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 534:Data Structures and Algorithms using Python

Course Objectives:

1) The systematic way of solving problem.
2) Interpret the standard and abstract data representation methods.
3) Grasp the memory requirement for various data structures.
4) Operate on the various structured data.

Course Outcomes:

1) Discriminate the usage of various structures in approaching the problem solution.
2) Design the algorithms to solve the programming problems.
3) Use effective and efficient data structures in solving various Computer Engineering domain problems.
4) Annalise the problems to apply suitable algorithm and data structure

Module	Title and Contents	Hrs
Module -1:	**Module -1: Foundations of Algorithm and Data Structures** **1.1** Need of Data Structure, Definitions - Data and information, Data Structure, Types of Data Structures, **1.2** Introduction to Algorithms, Characteristics of algorithms, **1.3** Algorithm design tools: Pseudo code and flowchart, **1.4** Complexity of algorithms- Space complexity, Time complexity, Asymptotic notation- Big-O, Theta and Omega,	15
Module -2:	**Module -2: Non-Linear Data Structures** **2.1** Trees, Binary Trees (Traversal: Inorder, Pre order, Post order), **2.2** Binary Search Trees (BST), **2.3** Tree Operations (Insert, Delete, Search), **2.4** Graphs, Graph Representation (Adjacency List & Matrix), Graph Traversals (BFS, DFS)	15
Module -3:	**Module -3: Linked List** **3.1** List as a Data Structure, differences with array, **3.2** Dynamic implementation of Linked List, **3.3** internal and external pointers, **3.4** Time complexity of operations, Applications of Linked List – polynomial representation,	15
Module -4:	**Module -4: Hashing, Dictionaries, Searching and Sorting Algorithms** **4.1** Hash Functions, **4.2** Collision Resolution Techniques (Chaining, Open Addressing), **4.3** Dictionaries in Python, **4.4** Dictionary Operations (Insert, Update, Delete, Search),	15

Reference Books:-

1. Agarwal, Dr. Basant. "Hands-On Data Structures and Algorithms with Python." Birmingham, UK: Packt Publishing, 2022.
2. Bhargava, Aditya. "Grokking Algorithms: An Illustrated Guide for Programmers and Other Curious People." San Francisco, CA: Manning Publications, 2016.
3. Goodrich, Michael T., Roberto Tamassia, and Michael H. Goldwasser. "Data Structures and Algorithms in Python." Hoboken, NJ: Wiley, 2013.

4. Karumanchi, Narasimha. "Data Structure and Algorithmic Thinking with Python: Data Structure and Algorithmic Puzzles." Hyderabad, India: CareerMonk Publications, 2011.
5. Lee, Kent D., and Steve Hubbard. "Data Structures and Algorithms with Python: With an Introduction to Multiprocessing." Cham, Switzerland: Springer, 2024.
6. Miller, Bradley N., and David L. Ranum. "Problem Solving with Algorithms and Data Structures Using Python." Franklin, MA: Franklin, Beedle & Associates, 2005.
7. Necaise, Rance D. "Data Structures and Algorithms Using Python." Hoboken, NJ: Wiley, 2011.

Evaluation Pattern:

Total Marks: 50

Internal Continuous Evaluation (20 Marks):	**End Semester Examination (30 Marks):**
• CCE - I : 10 Marks: Objective • CCE - II: 10 Marks: Objective • Mid Semester Exam: 20 Marks: Subjective (20 Marks converted to 10 marks) • Activity: 10 marks	• Question -1: Solve the following questions (5 questions of 2 Marks) • Attempt any two questions from Q. 2 to Q. 4 (three questions of 10 marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: E-II	**Marks:** 30
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 534:Advanced Data Engineering

Course Objectives:

1) Understand the Fundamentals of Data Engineering.
2) Explore Modern Data Storage and Processing Technologies.
3) Design and Implement Scalable Data Pipelines.
4) Apply Best Practices for Data Management.

Course Outcomes:

1) Understanding the fundamental role of data engineering in modern data ecosystems
2) Design the algorithms to solve the programming problems.
3) Apply batch and stream processing frameworks to process large-scale and real-time data
4) Build and manage ETL pipelines for integrating and transforming data

Module	Title and Contents	Hrs
Module -1:	**Module -1: Concepts of Advanced Data Engineering.** **1.1** Overview of Data Engineering: Role, importance, and challenges. **1.2** Data Engineering vs. Data Science vs. Data Analytics. **1.3** Lifecycle of Data Engineering: Ingestion, storage, **1.4** processing, analysis and visualization.	15
Module -2:	**Module -2: Data Storage Technologies** **2.1** Relational Databases: SQL databases, schema design, normalization, and indexing. **2.2** NoSQL Databases: Document, key-value, column-family, and graph databases. **2.3** Data Warehouses: Concepts, star and snowflake schemas, OLAP, and data marts. **2.4** Distributed File Systems: Hadoop HDFS, Google Cloud Storage, Amazon S3.	15
Module -3:	**Module -3: Data Processing Technologies** **3.1** Batch Processing: MapReduce, Apache Spark, data parallelism **3.2** Stream Processing: Apache Kafka, Apache Flink **3.3** event-driven architectures. **3.4** In-Memory Computing: Redis, Apache Ignite, Memcached	15
Module -4:	**Module -4: Data Integration and ETL** **4.1** Data Integration Tools: Apache Nifi, Talend, Informatica. **4.2** Data Transformation: Data cleaning, enrichment, aggregation, and denormalization. **4.3** design efficient schemas, schedule, monitor, and manage dependencies between tasks, **4.4** unit testing and integration testing of data pipelines.	15

Reference Books:-

1. Gilchrist, Alasdair. "Google Cloud Platform for Data Engineering: Learn Fundamental to Advanced Data Engineering Concepts and Techniques Using 30+ Real-World Use Cases." Kindle Edition, 2023.
2. Mada, Kameshwar. "Azure Data Engineer: The Powerful Comprehensive eBook on Advanced Performance Tuning with PySpark Databricks in Azure Cloud/AWS/GCP Cloud Environment with 200 Petabyte DB Size." Kindle Edition, 2023.
3. Malhotra, Mayank. "Ultimate Data Engineering with Databricks." Kindle Edition, 2023.
4. Paulraj, Balachandar. "Advanced Data Engineering Techniques for Gaming and Telecom Industries." Mumbai: Bright Ink Publishing, 2025.
5. Ramanujakootam, Adithyan. "Data Engineering with Advanced Python: Learn to Build Production Data Applications Using Modern Cloud Data Tools." Kindle Edition, 2025.
6. Verma, Sumit. "Data Engineering with Databricks." Kindle Edition, 2023.
7. Verma, Sumit. "Data Engineering with Databricks." Kindle Edition, 2023.

Evaluation Pattern:

Total Marks: 50

Internal Continuous Evaluation (20 Marks):	**End Semester Examination (30 Marks):**
• CCE - I : 10 Marks: Objective • CCE - II: 10 Marks: Objective • Mid Semester Exam: 20 Marks: Subjective (20 Marks converted to 10 marks) • Activity: 10 marks	• Question -1: Solve the following questions (5 questions of 2 Marks) 5 Attempt any two questions from Q. 2 to Q. 4 (three questions of 10 marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - III
Type: Major	**Marks:** 50
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 535:Practical V

Course Objectives:

1) Define cloud computing and its key characteristics.
2) Introduce the fundamental concepts of image processing, including the representation and components of digital images.
3) Understand nature of problems solved with Machine Learning
4) Learn storage networking technologies

Course Outcomes:

1) Understand different image registration techniques
2) Use machine learning methods for multivariate data analysis in various scientific fields.
3) Understand the relevant aspects of digital image representation and their practical implications.
4) Have a command of basic image restoration techniques.

Practical's	Title and Contents	Hrs
Practical:	1. Sign up to AWS Console. 2. Launch a Virtual Server (EC2) 3. Host a Static Website with S3 4. Set Up IAM Users & Permissions 5. Create a Simple RDS Database 6. Implement a k-Nearest Neighbors classifier for predicting iris flower species. 7. Apply k-means clustering to segment customer groups 8. Train and visualize a Decision Tree classifier using ID3 or CART. 9. Implement a Naïve Bayes classifier for spam email detection. 10. Apply Principal Component Analysis (PCA) for dimensionality reduction and visualization. 11. Train a logistic regression model to predict whether a student will pass/fail. 12. Apply different feature selection methods and observe model accuracy changes. 13. Reading and Displaying Images 14. Image Resizing and Cropping 15. Image Rotation and Translation 16. Drawing Shapes and Text on Images 17. Blurring and Smoothing 18. Image Conversion: Grayscale and Binary	60

	19. Increase or decrease the brightness of an image. 20. Detect edges in an image using methods like Canny edge detection.	

Reference Books:-

1) Bhowmik Sandeep. Cloud Computing. 1st ed. New Delhi: McGraw Hill Education, 2017.
2) Erl Thomas, Zaigham Mahmood, and Ricardo Puttini. Cloud Computing: Concepts, Technology & Architecture. Boston: Pearson Education, 2013.
3) Foster Ian, and Dennis B. Gannon. Cloud Computing for Science and Engineering. Cambridge, MA: MIT Press, 2018.
4) Hearty John. "Advanced Machine Learning with Python: Solve Data Science Problems by Mastering Cutting-Edge Machine Learning Techniques in Python". Birmingham, UK: Packt Publishing, 2016.
5) Körner, Christoph, and Kaijisse Waaijer. "Mastering Azure Machine Learning: Perform Large-Scale End-to-End Advanced Machine Learning in the Cloud with Microsoft Azure Machine Learning. Birmingham", UK: Packt Publishing, 2020.
6) Marr, David. "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information." Cambridge, MA: MIT Press, 1982.
7) Singh, Himanshu. "Practical Machine Learning with AWS: Process, Build, Deploy, and Productionize Your Models Using AWS". New York: Apress, 2020

Evaluation Pattern:	
Total Marks: 50	
	End Semester Examination: • Question -1 (10 Marks) • Question -2 (10 Marks) • Question -3 (10 Marks) • Activity 20 Marks.

M.Sc. II
SEM
IV

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 541:Deep Learning

Course Objectives:

1) Introduce major deep learning algorithms
2) Propose optimization techniques to training deep neural networks
3) Learn regularization techniques to train deep neural networks.
4) Introduce Convolution Neural Networks and its applications

Course Outcomes:

1) Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains.
2) Implement optimization techniques to training deep neural networks
3) Apply regularization techniques to improve the performance of deep learning algorithms.
4) Implement deep learning algorithms and solve real-world problems in computer vision.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Introduction to Deep Learning** **1.1** Overview of Deep Learning, Definition and significance, Historical background and evolution, Deep learning vs. traditional machine learning, **1.2** Applications of Deep Learning, Image and speech recognition, Natural language processing (NLP), **1.3** Basics of Neural Networks, Biological inspiration, **1.4** Perceptron model, Activation functions (Sigmoid, ReLU, Tanh, etc.)	15
Module -2:	**Module -2: Optimization and Deep Learning** **2.1** Learning Parameters of a feedforward neural network, The vanishing gradient problem, Multi-Layer Perceptron (MLP), **2.2** Feedforward networks, Backpropagation algorithm, Loss functions and gradient descent, Key Concepts in Deep Learning, **2.3** Epochs, batches, learning rate, Underfitting vs overfitting, Tools and Libraries for Deep Learning, TensorFlow, Keras, PyTorch, **2.4** Challenges in Deep Learning, Data requirements, Computational complexity, Model interpretability.	15
Module -3:	**Module -3: Regularization Techniques and Convolutional Neural Networks** **3.1** Bias Variance Tradeoff, L2 regularization, Early stopping, **3.2** Dataset augmentation, Parameter sharing and tying, **3.3** Injecting noise at input, Ensemble methods, Dropout. **3.4** Convolutional Neural Networks, Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream,	15
Module -4:	**Module -4: Deep Unsupervised Learning and Sequence Models** **4.1** Autoencoders: standard, sparse, denoising, contractive, Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM **4.2** RNN, LSTM, GRU models, Application to NLP, language models, **4.3** machine translation, image captioning, video processing, visual question answering, video processing, learning from descriptions, **4.4** Attention Mechanism, Attention over images.	15

Reference Books:-

1. Chollet, François, and J.J. Allaire. "Deep Learning with R." Shelter Island, NY: Manning Publications, 2018.
2. Chollet, François. "Deep Learning with Python." Shelter Island, NY: Manning Publications, 2017.
3. Géron, Aurélien. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems." Sebastopol, CA: O'Reilly Media, 2017.
4. Gibson, Adam, and Josh Patterson. "Deep Learning: A Practitioner's Approach." Beijing: O'Reilly Media, 2017.
5. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. "Deep Learning." London: MIT Press, 2016.
6. Shanmugamani, Rajalingappaa. "Deep Learning for Computer Vision." Birmingham, UK: Packt Publishing, 2018.
7. Zhang, Aston, Zachary C. Lipton, Mu Li, and Alexander J. Smola. "Dive into Deep Learning." Cambridge: Cambridge University Press, 2020.

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 542:Big Data Analytics

Course Objectives:

1) Perceive the big data concepts and big data analytics lifecycle
2) Understand the big data analytics algorithms and tools
3) Recognize the importance of big data visualization tools and techniques
4) Acquainted with advancements in tools and techniques used for big data analytics.

Course Outcomes:

1) Design the data analytics life cycle for selected problem statement.
2) Develop insights into the big data and present results for selected problem statement through visualization techniques.
3) Demonstrate the use of Hadoop and its ecosystem elements to analyze big data.
4) Demonstrate use of advanced FOSS computing environments for big health care data

Module	Title and Contents	Hrs
Module -1:	**Module -1: Overview to Big Data** **1.1** Introduction to Big Data Analytics. **1.2** What are the challenges, Benefits of Big Data Technology, **1.3** Introduction to Hadoop, Introduction to HDFS, **1.4** HDFS Architecture: Datanode, Name Node, Blocks.	15
Module -2:	**Module -2: Technologies for big data analytics** **2.1** Distributed and Parallel Computing for Big Data, **2.2** In-Memory Computing Technology for Big Data, Introduction to MapReduce, **2.3** Architecture of MapReduce Technology, Introduction to YARN, Architecture of YARN, **2.4** Introduction to Hive, Architecture of Hive.	15
Module -3:	**Module -3: Programming languages for big data analytics** **3.1** Introduction to Apache Spark/ PySpark, Spark Application concepts, **3.2** Apache Spark Structured APIs, Apache Spark DataFrame vs Dataset vs RDD, **3.3** Apache Spark: Working with external data sources, **3.4** Introduction NoSQL Database (MongoDB), Introduction to Apache Cassandra.	15
Module -4:	**Module -4: Machine Learning Technology using Big Data technology** **4.1** Introduction to Data Source and Sinks, **4.2** Architecture of Data Lakes, Introduction Lake house, **4.3** Machine Learning using Apache Spark MLlib, **4.4** Evaluating PySpark Machine Learning Model, PySpark Machine Learning Model Management	15

Reference Books:-

1. Corea, Francesco. "Big Data Analytics: A Management Perspective." Cham: Springer, 2016.
2. Demirbaga, Ümit, Gagangeet Singh Aujla, Anish Jindal, and Oğuzhan Kalyon. "Big Data Analytics: Theory, Techniques, Platforms, and Applications." Cham: Springer, 2024.
3. Kulkarni, Parag, Sarang Joshi, and Meta S. Brown. "Big Data Analytics." New Delhi: PHI Learning, 2016.
4. Pourroostaei Ardakani, Saeid, and Ali Cheshmehzangi. "Big Data Analytics for Smart Urban Systems." Singapore: Springer, 2023.
5. Prabhu, C. S. R., Aneesh Sreevallabh Chivukula, Aditya Mogadala, Rohit Ghosh, and L. M. Jenila Livingston. "Big Data Analytics: Systems, Algorithms, Applications." Singapore: Springer, 2019.
6. Raj, Pethuru, Anupama Raman, and Dhivya Nagaraj. "High-Performance Big-Data Analytics: Computing Systems and Approaches." Cham: Springer, 2015.
7. Upadhyay, Nitin. "Big Data: Management and Analytics." New Delhi: Cengage India, 2018.

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: Major	**Marks:** 60
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 543: Data Processing using MATLAB

Course Objectives:

1) Understand the fundamentals of MATLAB environment and basic programming techniques.
2) Develop the ability to import, visualize, and preprocess data using MATLAB tools.
3) Enable students to apply MATLAB functions for statistical analysis and signal processing.
4) Build skills for automating data workflows and generating reports from processed data

Course Outcomes:

1) Navigate MATLAB environment and write basic scripts and functions.
2) Load, clean, transform, and visualize data effectively.
3) Apply statistical and basic signal processing techniques to real-world datasets.
4) Design simple automated data analysis workflows using MATLAB.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Basics of Data Handling and Overview of MATLAB** **1.1** Overview of MATLAB Interface (Command Window, Editor, Workspace), **1.2** Basic MATLAB syntax: variables, operators, control flow (if, for, while), **1.3** Scripts vs Functions, Data Types: Arrays, Matrices, Tables, Structures, **1.4** Importing data from files: .csv, .txt, Excel, Basic plotting (line plot, scatter plot, bar plot).	15
Module -2:	**Module -2: Data Preprocessing and Cleaning** **2.1** Handling missing data (fillmissing, missing), **2.2** Data transformation (normalization, standardization), **2.3** Data smoothing and filtering, Resampling and interpolation, **2.4** String operations and data parsing, Working with categorical data	15
Module -3:	**Module -3: Data Analysis and Visualization** **3.1** Descriptive statistics (mean, median, mode, variance, std dev), **3.2** Correlation and covariance analysis, **3.3** Histograms, Boxplots, Heatmaps, Advanced plotting: subplot, multi-axis plot, 3D plots, **3.4** Data fitting: linear and polynomial regression, Overview of curve fitting toolbox.	15

Module -4:	**Module -4: Advanced Data Processing and Automation** **4.1** Basics of signal processing (filtering, FFT), **4.2** Time-series data handling, Batch data processing using loops and functions, **4.3** generating reports with Live Scripts, **4.4** Saving and exporting processed data, Introduction to App Designer for creating simple data apps.	15

Reference Books:-

1. Chapman, Stephen J. "MATLAB Programming for Engineers." Boston: Cengage Learning, 2022
2. Gonzalez, Rafael C., Richard E. Woods, and Steven L. Eddins. "Digital Image Processing Using MATLAB." Upper Saddle River, NJ: Pearson Prentice Hall, 2003.
3. Schilling, Robert J., and Sandra L. Harris. "Fundamentals of Digital Signal Processing Using MATLAB." Boston: Cengage Learning, 2011.
4. Şen, Zekâi. "Earth Systems Data Processing and Visualization Using MATLAB." Cham: Springer, 2020.
5. Siciliano, Antonio. "MATLAB: Data Analysis and Visualization." Singapore: World Scientific Publishing, 2008.
6. Umbaugh, Scott E. "Digital Image Processing and Analysis: Applications with MATLAB and CVIPtools." Boca Raton, FL: CRC Press, 2017.
7. Wanhammar, Lars, and Tapio Saramäki. "Digital Filters Using MATLAB." Cham: Springer, 2020.

Evaluation Pattern:

Total Marks: 60

Internal Continuous Evaluation:	**End Semester Examination:**
• CCE-I – 10 Marks • CCE-II – 10 Marks • Mid Semester – 20 Marks • Activity – 10 Marks	• Question -1 (02 Marks = 2 X 6=12 Marks) • Question -2 (06 Marks = 6 X 2=12 Marks) • Question -3 (06 Marks = 6 X 2=12 Marks) • Question -4 (06 Marks = 6 X 2=12 Marks) • Question -5 (04 Marks = 4 X 3=12 Marks) • Question -6 (04 Marks = 4 X 3=12 Marks) • Question -7 (04 Marks = 4 X 3=12 Marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: E-I	**Marks:** 30
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 544:Advanced Recommender System

Course Objectives:

1) Introduction and understanding of basic terms of recommender system.
2) Identify different approaches of recommendation system.
3) Hands-on application with basic logic of recommendation system.
4) Awareness of real-world implementation issues of recommendation system.

Course Outcomes:

1) Familiarize with recommender systems and their applications.
2) Analyze the different approaches towards recommendation.
3) Evaluate the effectiveness of the recommendation system.
4) Design recommender system.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Basics of Recommender System** **1.1** Recommender Systems Function, Applications of recommendation systems, **1.2** Issues with recommender system, Advantages and Drawbacks of Content-based Filtering, **1.3** Item Representation, Methods for Learning User Profiles. **1.4** Linear Algebra notation: Matrix addition, Multiplication, transposition, and inverses; covariance matrices. High-Level Architecture of Content-based Systems	15
Module -2:	**Module -2: Collaborative Recommendation System:** **2.1** Collaborative Recommendation: User-based nearest neighbour recommendation, **2.2** Item-based nearest neighbour recommendation, model-based and preprocessing-based approaches. **2.3** Attacks on collaborative recommender systems: Attack dimensions, Attack types, **2.4** Countermeasures, Privacy aspects – distributed collaborative filtering..	15
Module -3:	**Module -3: Knowledge-based recommendation:** **3.1** Knowledge representation and reasoning, interacting with constraint-based recommenders, **3.2** Interacting with case-based recommenders.Hybrid recommendation approaches: Opportunities for hybridization, **3.3** Monolithic hybridization design, **3.4** Parallelized hybridization design, Pipelined hybridization design.	15
Module -4:	**Module -4: Evaluating recommended systems:** **4.1** General properties of evaluation research, Popular evaluation designs, **4.2** Evaluation on historical datasets, Alternate evaluation designs. Community-Based Web Search- The Collaborative Web Search System. Shared Web Search - The	15

	HeyStaks System, The HeyStaks Recommendation Engine. **4.3** Social Tagging Recommenders Systems- Folksonomy, The Traditional Recommender Systems Paradigm, BibSonomy as Study Case, **4.4** Tag Acquisition. Trust and Recommendations- Computational Trust, Trust-Enhanced Recommender Systems	

Reference Books:-

1) Aggarwal, Charu C. "Recommender Systems: The Textbook." Cham: Springer, 2016.

2) Falk, Kim. "Practical Recommender Systems." Shelter Island, NY: Manning Publications, 2015.

3) Hassanien, Aboul Ella, and Sachi Nandan Mohanty, eds. "Recommender System with Machine Learning and Artificial Intelligence." Hoboken, NJ: Wiley, 2020.

4) Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." New York: Springer, 2011.

5) Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." 2nd ed. New York: Springer, 2015.

6) Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." 3rd ed. New York: Springer, 2022.

7) Zhen, Lu. "Knowledge Recommender Systems: An Advanced Mode of Knowledge Management Tools." Saarbrücken: LAP Lambert Academic Publishing, 2010.

Evaluation Pattern:	
Total Marks: 50	
Internal Continuous Evaluation (20 Marks): • CCE - I : 10 Marks: Objective • CCE - II: 10 Marks: Objective • Mid Semester Exam: 20 Marks: Subjective (20 Marks converted to 10 marks) • Activity: 10 marks	**End Semester Examination (30 Marks):** • Question -1: Solve the following questions (5 questions of 2 Marks) • Attempt any two questions from Q. 2 to Q. 4 (three questions of 10 marks)

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: E-II	**Marks:** 30
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 544: Data Analytics

Course Objectives:

1) Develop the skills of managing the data with respect to knowledge generation.
2) Propose the data reliability models
3) Understand all the different parts of a problem and then be able to find improvement points from facts in the past
4) Predict the future outcome of present decisions.

Course Outcomes:

1) Understand and deal with any social media network, strategy, or campaign.
2) Social media analytics integrates with and affects other areas of business.
3) Give real-world context and insight.
4) Present decision.

Module	Title and Contents	Hrs
Module -1:	**Module -1: Fundamentals of Data Analytics and social media** **1.1** Introduction to Data Analytics, Users: The Who of social media, Measuring Variations in User Behavior in Wikipedia, **1.2** Long Tails, everywhere: The 80/20 Rule (p/q Rule), Online Behavior on Twitter, **1.3** Networks: The How of social media, Types and Properties of Social Networks, Visualizing Networks, Degrees: The Winner Takes All, **1.4** Capturing Correlations: Triangles, Clustering, and Assortative, Temporal	15
Module -2:	**Module -2: Content and Large Dataset of social media** **2.1** Content: What of social media, Defining Content: Focus on Text and Unstructured Data, Using Content Features to Identify Topics, **2.2** Extracting Low-Dimensional Information from High-Dimensional Text, Processing Large Datasets. MapReduce: Structuring Parallel and Sequential Operations, Multi-Stage MapReduce Flows, **2.3** Patterns in MapReduce Programming, Sampling and Approximations: Getting Results with Less **2.4** Computation, Sampling	15
Module -3:	**Module -3: Learn, Map, and Recommend.** **3.1** Social Media Services Online, Problem Formulation, **3.2** Learning and Mapping, Prediction and Recommendation, Social Media Data, From Data to Insights, **3.3** Analytics in social media, Dedicated vs. Hybrid Tools. Alexander and	15

	Frederik Peiniger,Social Network Landscape.	
Module -4:	**Module -4: Creating final Reports** **4.1** Reports, Milan Veverka, Strategy **4.2** Tactics, Michael Wu **4.3** Prescriptive Analytics **4.4** The Future of Social Media Analytics	15

Reference Books:-

1. Amit Kumar Tyagi, Data Science and Data Analytics: Opportunities and Challenges, 23 September 2021
2. Brown, Mary. Social Media Analytics in Data Science: Strategies and Tools for Analyzing Social Media Data. Chicago: Wiley, 2018.
3. Dr. Gaurav Aroraa, Data Analytics: Principles, Tools, and Practices: A Complete Guide for Advanced Data Analytics Using the Latest Trends, Tools, and Technologies, 23 January 2022
4. Johnson, Alice, and Bob Williams. Data-Driven Social Media Analytics: A Practical Guide for Data Scientists. Chicago: Academic Press, 2019.
5. Smith, John. Social Media Analytics for Data Science: Methods and Techniques. Chicago: University Press, 2020.
6. Vibrant Publishers, Data Analytics Essentials You Always Wanted To Know (Self-Learning Management Series),Self-Learning Management Series, 29 February 2024
7. Wes McKinney, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Third Edition (Grayscale Indian Edition),Shroff/O'Reilly, 6 September 2022

Evaluation Pattern:

Total Marks: 50

Internal Continuous Evaluation (20 Marks):	**End Semester Examination (30 Marks):**
• CCE - I : 10 Marks: Objective • CCE - II: 10 Marks: Objective • Mid Semester Exam: 20 Marks: Subjective (20 Marks converted to 10 marks) • Activity: 10 marks	• Question -1: Solve the following questions (5 questions of 2 Marks) • Attempt any two questions from Q. 2 to Q. 4 (three questions of 10 marks)

Master of Science (M.Sc.) Part – II
Semester IV

MDSP 545: OJT

OJT will provide the opportunities for internship with local/regional industries, business organization, health and allied areas, local government, etc. so that students may actively engage with the employability opportunities. Students will undergo 4 credit work-based learning/OJT/internship.

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: Major	**Marks:** 50
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 546:Practical VII

Course Objectives:

1) Develop the skills of managing the data with respect to knowledge generation.
2) Propose the data reliability models.
3) Understand the fundamentals of MATLAB environment and basic programming techniques.
4) Develop the ability to import, visualize, and preprocess data using MATLAB tools

Course Outcomes:

1) Navigate MATLAB environment and write basic scripts and functions.
2) Load, clean, transform, and visualize data effectively.
3) Design the data analytics life cycle for selected problem statement.\
4) Develop insights into the big data and present results for selected problem statement through visualization techniques.

Practical's	Title and Contents	Hrs
Practical:	1. Build and train a simple Perceptron model for binary classification. 2. Compare logistic regression and a simple neural network on the same dataset. 3. Build and train a Multi-Layer Perceptron (MLP) on the MNIST dataset. 4. Experiment with epochs, batch sizes, and learning rates. 5. Demonstrate overfitting and apply L2 regularization and dropout. 6. Train a basic CNN for digit recognition (MNIST). 7. Use VGGNet or ResNet for image classification with transfer learning. 8. Installation of Hadoop Ecosystem in VirtualBox 9. Word Count Program in Hadoop MapReduce 10. Analyze Web Server Logs. 11. Load & Query Big CSV Data using Hive 12. Perform Sentiment Analysis on Tweets 13. Big Data Joins using MapReduce or Hive 14. Implementation of MATLAB: Basic Commands and Operations. 15. Importing Data from CSV and Excel Files. 16. Handling Missing Data and Data Cleaning Techniques.	

	17. Data Transformation: Normalization and Standardization. 18. Plotting Basic Graphs: Line, Bar, and Scatter Plots. 19. Exploring Statistical Measures: Mean, Median, Mode, Variance. 20. Correlation and Covariance Analysis between Variables.Blurring and Smoothing	

Reference Books:-

1. Chollet, François, and J.J. Allaire. "Deep Learning with R." Shelter Island, NY: Manning Publications, 2018.
2. Chollet, François. "Deep Learning with Python." Shelter Island, NY: Manning Publications, 2017.
3. Géron, Aurélien. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems." Sebastopol, CA: O'Reilly Media, 2017.
4. Corea, Francesco. "Big Data Analytics: A Management Perspective." Cham: Springer, 2016.
5. Demirbaga, Ümit, Gagangeet Singh Aujla, Anish Jindal, and Oğuzhan Kalyon. "Big Data Analytics: Theory, Techniques, Platforms, and Applications." Cham: Springer, 2024.
6. Kulkarni, Parag, Sarang Joshi, and Meta S. Brown. "Big Data Analytics." New Delhi: PHI Learning, 2016.
7. Chapman, Stephen J. "MATLAB Programming for Engineers." Boston: Cengage Learning, 2022

Evaluation Pattern:	
Total Marks: 50	
	End Semester Examination: • Question -1 (10 Marks) • Question -2 (10 Marks) • Question -3 (10 Marks) • Activity 20 Marks.

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: E-I	**Marks:** 50
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 547: Practical VIII

Course Objectives:

1) Introduction and understanding of basic terms of recommender system.
2) Identify different approaches of recommendation system.
3) Hands-on application with basic logic of recommendation system.
4) Awareness of real-world implementation issues of recommendation system.

Course Outcomes:

1) Familiarize with recommender systems and their applications.
2) Analyze the different approaches towards recommendation.
3) Evaluate the effectiveness of the recommendation system.
4) Design recommender system.

Practical's	Title and Contents	Hrs
Practical:	1. Basic construction of recommender system. 2. Build an item-item similarity matrix using cosine_similarity and generate recommendations. 3. Construction of Weighted Hybrid Based Filtering. 4. Use pandas and sklearn to create a user-user similarity matrix and recommend top-N items to a user. 5. Use the Book-Crossing Dataset to build a recommender system using either collaborative or content-based filtering. 6. Implement Precision, Recall, and RMSE on any model developed in earlier labs. 7. Develop a hybrid model that uses both user-item interaction and content metadata for better recommendations. 8. Book-Crossing Dataset Recommender 9. Evaluate with Precision, Recall, RMSE 10. Hybrid Model with Content Metadata	

Reference Books:-

1. Aggarwal, Charu C. "Recommender Systems: The Textbook." Cham: Springer, 2016.
2. Falk, Kim. "Practical Recommender Systems." Shelter Island, NY: Manning Publications, 2015.
3. Hassanien, Aboul Ella, and Sachi Nandan Mohanty, eds. "Recommender System with Machine Learning and Artificial Intelligence." Hoboken, NJ: Wiley, 2020.
4. Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." New York: Springer, 2011.

5. Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." 2nd ed. New York: Springer, 2015.

6. Ricci, Francesco, Lior Rokach, and Bracha Shapira, eds. "Recommender Systems Handbook." 3rd ed. New York: Springer, 2022.

7. Zhen, Lu. "Knowledge Recommender Systems: An Advanced Mode of Knowledge Management Tools." Saarbrücken: LAP Lambert Academic Publishing, 2010.

Evaluation Pattern:	
Total Marks: 50	
	End Semester Examination: • Question -1 (10 Marks) • Question -2 (10 Marks) • Question -3 (10 Marks) • Activity 20 Marks.

Karmaveer Bhaurao Patil University, Satara

(A State Public University Est. u/s 3(6) of MPUA 2016)

Faculty of Science and Technology

Yashavantrao Chavan Institute of Science, Satara

Board of Studies in Computer Science

Programme: M.Sc.	**Semester** - IV
Type: E-II	**Marks:** 50
Credits: 4	**From:** A. Y. 2025-26

Name of the Course: MDST 547:Practical VIII

Course Objectives:

1) Learn how to extract and analyze large datasets from social media platforms
2) earn to model and visualize social networks, applying concepts like degree centrality, betweenness centrality, and network clustering.
3) Develop the ability to build basic content-based and collaborative filtering recommender systems for social media content.
4) Acquire the ability to process large-scale social media data using methods like MapReduce and sampling techniques.

Course Outcomes:

1) Demonstrate proficiency in extracting, cleaning, and analyzing large datasets from various social media platforms
2) Design and implement social network analysis algorithms to model relationships and influence dynamics on social media.
3) Develop and deploy content-based and collaborative filtering recommender systems for suggesting relevant content or connections on social media.
4) mplement scalable data processing workflows using MapReduce and other parallel computing techniques for handling large volumes of unstructured data.

Practical's	Title and Contents	Hrs
Practical:	1. Use edits history datasets to measure variations in user contributions. 2. Collect Twitter data using Tweepy and plot hashtag frequency. 3. Build and visualize a Twitter retweet or mention network. 4. Use NLP tools to clean, tokenize, and extract topics from tweets. 5. Simulate a MapReduce flow to process a large text corpus (e.g., word count). 6. Implement random and stratified sampling on a large dataset (e.g., Reddit or Twitter). 7. Build a basic content-based or collaborative filtering recommender for posts or users. 8. Analyze tweet sentiment over time for a specific hashtag 9. Study collaborations among YouTubers using metadata 10. Evaluate the impact of a marketing campaign using post engagement metrics	

Reference Books:-

1. Amit Kumar Tyagi, Data Science and Data Analytics: Opportunities and Challenges, 23 September 2021

2. Brown, Mary. Social Media Analytics in Data Science: Strategies and Tools for Analyzing Social Media Data. Chicago: Wiley, 2018.
3. Dr. Gaurav Aroraa, Data Analytics: Principles, Tools, and Practices: A Complete Guide for Advanced Data Analytics Using the Latest Trends, Tools, and Technologies, 23 January 2022
4. Johnson, Alice, and Bob Williams. Data-Driven Social Media Analytics: A Practical Guide for Data Scientists. Chicago: Academic Press, 2019.
5. Smith, John. Social Media Analytics for Data Science: Methods and Techniques. Chicago: University Press, 2020.
6. Vibrant Publishers, Data Analytics Essentials You Always Wanted To Know (Self-Learning Management Series),Self-Learning Management Series, 29 February 2024
7. Wes McKinney, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Third Edition (Grayscale Indian Edition),Shroff/O'Reilly, 6 September 2022

Evaluation Pattern:	
Total Marks: 50	
	End Semester Examination: • Question -1 (10 Marks) • Question -2 (10 Marks) • Question -3 (10 Marks) • Activity 20 Marks.