Rayat Shikshan Sanstha's

Yashavantrao Chavan Institute of Science, Satara (Autonomous)

Lead College of Karmaveer Bhaurao Patil University, Satara, Maharashtra

Reaccredited by NAAC with 'A+' Grade

Syllabus for Master of

Science

Part - I

M.Sc. Data Science

Syllabus

to be implemented from June, 2023 onward

(As per NEP -2020 Guidelines)

SYLLABUS FOR MASTER OF SCIENCE (M.Sc.)

1. **Title: Subject: Data Science**

2. **Year of implementation: June 2023 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 andmaking 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 boththe 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 necessarybridge 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 collaborativeand 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 professionaldevelopment. 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 weexplore the fascinating world of data science and unlock the potential of data for a better future.

4. RULES AND REGULATIONS:

1. Any person who has taken the degree of B. Sc. of this Institute or the degree of any other statutory University and has kept four terms in the Institute as post-graduate student be admitted tothe examination for the degree of Master of Science (M. Sc.) in Data Science.

2. A student shall be held eligible for admission to the M. Sc. Data Science course provided s/he has passed the B. Sc. examination with Data Science as a principal subject or with a subsidiary/interdisciplinary/applied/allied subjects (B.Sc. Data Science, B.Sc. Computer Science Entire/Optional/Statistics/Mathematics, B.E. Comp.Sc./B.E.IT,BCA Science, B.Voc (Soft.Dev) andhas passed the entrance examination conducted by the Institute.

3. The students with B. Sc. from other universities shall be eligible if they qualify through the entranceexamination.

4. While preparing the merit list for M. Sc. admission, the performance at the performance at theentrance examination should be considered.

5. The examination shall be split up into four semesters.

6. The commencement and conclusion of each semester shall be notified by the Institute from time totime.

7. A student who has passed in semester examination shall not be allowed to take the examination inthe same semester again.

8. Each theory Course in each semester as well as each practical course shall be treated as separatehead of passing.

9. The result shall be declared at the end of each semester examination as per Institute rules.

5. Program specific objective

1. To create post-graduates with sound knowledge of Data Science, who can contributetowards recent advances in technology.
2. To provide advanced and in-depth knowledge of data science and specialization in oneor two subjects of new era of technology.
3. To prepare Postgraduates who will achieve peer-recognition, as an individual or ina 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. and research.
5. To impact industry knowledge and practical skills of current trends in IT field to thestudents.
6. To develop ability among students to formulate, analyze and solve real life problems faced in Computer Science industry. To produce computer science professionals whocan be directly employed or start his/her own work as
    - Data Scientist.
    - Business Analyst.
    - Data Analytics Manager.
    - Data Architect.
    - Data Administrator.
    - Business Intelligence Manager.
    - Entrepreneur in Computer Science industry.
7. To Develop designing, analyzing and critical thinking skill among students.

6. Program Specific Outcomes: The student will be able to:

1. Avail the skills of Current trends in IT Industries and new Technologies.
2. Apply knowledge of programming platforms in Data Science and AI in real life.
3. Student should avail detail knowledge of Data Science, Artificial Intelligence, Machine Learning, andBig Data etc.
4. Students will demonstrate their ability of advanced programming to design and develop innovative applications.
5. Student will be able to Access, evaluate, understand, and compare digital information from various sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
6. 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. Exemplify generate a solution independently, check and validate it and modify if necessary.
7. Translate project plans, use management skills, and lead a team for planning and execution of a task.
8. Student can start his own business or start up.

.

7. Duration :

The course shall be a full-time course.

•The course shall be of two years, consisting of four semesters.

8. Fee Structure:

- **Entrance Examination fees:** as prescribed by the Institute.
- **Course Fee:** as prescribed by the Institute.

9. Eligibility for Admission:

- As per Rule (2) for graduates of this Institute.
- As per Rule (3) for graduates from other universities and merit of entrance exam.

10. Medium of instruction: English

11. SCOPE:

After Successful completion of two years master's degree in data science, we observed that the students have the ample opportunities in diversified areas such as:

9. Software Industry.
10. Communication Industry
11. Digital Media
12. Agriculture Industry
13. Health and Care.
14. Research Field.
15. Research Institutes

12 **. Structure of Course**

Level	Sem	Major			RM	OJT	RP	Total
		DSC Mandatory		DSE Elective				
		T	P	T				
6	I	12 (3 Papers)	2	4 (1 paper out of two)	4	---	---	22
	II	12 (3 Papers)	2	4 (1 paper out of two)	---	---	4	22
6.5	III	12 (3 Papers)	2	4 (1 paper out of two)	---	---	6	22
	IV	12 (3 Papers)	---	4 (1 paper out of two)	---	4	---	22
Total		48	6	16	4	4	10	88
		70			8		10	

M.Sc. Part I

Semester I

Nature of the Course	Course Code	Name of the Course
Theory	MDST 411	Statistical Foundation for Data Science
	MDST 412	Programming using R
	MDST 413	Fundamentals of Data Science
	MDST 414 E-I	Distributed Database Concept
	MDST 414 E-II	IOT for Data Science I
	MDST 415	Research Methodology
Practical	MDSP 416	Practical Course I: Lab I Based On (MDST 411,412,413)

Semester II

Nature of the Course	Course Code	Name of the Course
Theory	MDST 421	Mathematical Foundations for Data Science
	MDST 422	Python Programming
	MDST 423	Data Preparation Analysis
	MDST 424 E-I	AI for Data Science
	MDST 424 E-II	IOT for Data Science II
	MDST 425	Research Project
Practical	MDST 426	Practical Course II: Lab II Based On (MDST 421,422,423)

Project Academic Project is divided into 4 phases.

Phase I: Literature Survey
Phase II: Data Collection & Design
Phase III: Implementation
Phase IV: Publication

SEMESTER I

MDST 411: Statistical Foundation for Data Science

Course Objectives: Student should be able to…

1. Understand the concept of Descriptive statistics.
2. Study correlation and regression.
3. Understand the applications probability theory.
4. Study the probability distribution

Credits=4	SEMESTER-I MDST 411 : Statistical Foundation for Data Science	No. of hours per unit/ credits
Credit –I UNIT I	**Descriptive Statistics**	**(15)**
	Sampling Techniques – Data Classification – Tabulation – Frequency and graphic Representation – Measures of Central Tendency – Measures of Variation – Quartiles and Percentiles – Moments - Skewness and Kurtosis.	
Credit –1 UNIT II	**Correlation and Regression**	**(15)**
	Scatter Diagram – Karl Pearson's Correlation Coefficient – Rank Correlation – Correlation Coefficient for Bivariate Frequency Distribution – Regression Coefficients – Fitting of Regression Lines.	
Credit –1 UNIT III	**Probability Theory**	**(15)**
	Random Experiment – Sample Space – Events – Axiomatic Definition of Probability –Addition Theorem – Multiplication Theorem – Baye's Theorem -Applications. Distribution Function Continuous and Discrete Random Variables – Distribution Function of a Random Variable –Probability Mass Functions and Probability Density Functions – Characteristic Functions –Central Limit Theorems.	
Credit –1 UNIT IV	**Probability Distributions**	**(15)**
	Probability Distributions – Recurrence Relationships – Moment Generating Functions –Cumulate Generating Functions – Continuous Probability Distributions – Rectangular Distribution – Binomial Distribution – Poisson Distribution – Continuous Probability Distributions – Uniform Distribution - Normal Distribution – Exponential Distribution.	

Course Outcomes: Student will be able to…

1. Comprehend the concepts of descriptive statistics.
2. Apply recent concepts in correlation and regression.
3. Utilize probability theory.
4. Imbibe concepts of probability distribution.

Reference Books:

1. Gelman, A., Hill, J., & Vehtari, A., Regression and Other Stories (Cambridge, UK: Cambridge University Press), (2020).
2. Shumway, R. H., & Stoffer, D. S., Time Series Analysis and Its Applications: With R Examples (New York: Springer). (2017).
3. Efron, B., & Hastie, T. Computer Age Statistical Inference: Algorithms, Evidence, and Data Science (Cambridge, UK: Cambridge University Press), (2016).
4. Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning (Cambridge, MA: MIT Press), (2016).
5. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. ,Bayesian Data Analysis (Boca Raton, FL: CRC Press), (2013).
6. Montgomery, D. C., Peck, E. A., & Vining, G. G., Introduction to Linear Regression Analysis (Hoboken, NJ: Wiley), (2012).
7. Wasserman, L. All of Statistics: A Concise Course in Statistical Inference (New York: Springer), (2013).
8. Hastie, T., Tibshirani, R., & Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, andPrediction (New York: Springer), (2009).

MDST 412: Programming using R

Course Objectives: Student should be able to …

1. Master the use of the R and R Studio interactive environment.
2. Explore and understand how to use the R documentation.
3. Read Structured Data into R from various sources.
4. Understand the different data types and data structure in R.

Credits=4	**SEMESTER-I** **MDST 412: Programming using R**	**No. of hours per unit/ credits**
Credit –I UNIT I	**R Introduction**	**(15)**
	Introduction to R – Help Functions in R – Vectors – Vectorized Operations – Functions in R – Packages in R ,Data Types ,Subsetting ,Writing data ,Reading from csv files ,Creating a vector and vector operation, Initializing data frame, Flow control: For loop, If condition, Debugging tools,Re-directing R Output	
Credit –1 UNIT II	**Matrices, Arrays and Lists**	**(15)**
	Matrix Operations, Adding and Deleting Rows and Columns – Higher Dimensional Arrays – Lists – General List Operations – Accessing List Components and Values – Applying functions to Lists.	
Credit –1 UNIT III	**Data Frames**	**(15)**
	Creating Data Frames , Matrix-like Operations on a Data Frame – Merging Data Frames – Applying functions to Data Frames – Factors and Tables – Common Functions used with Factors – Working with Tables	
Credit –1 UNIT IV	**Data manipulation and Visualization**	**(15)**
	List Management ,Data Transformation ,Merging Data Frames ,Outlier Detection,Combining multiple vectors ,Creating bar chart and dot plot, Creating histogram and box plot, Plotting with base graphics,Plotting and coloring in R	

Course Outcomes-

Student will be able to…

1. Install, Code and Use R Programming Language in R Studio IDE to perform basic tasks on Vectors, Matrices and Data frames. Describe key terminologies, concepts and techniques employed in Statistical Analysis.
2. Define, Calculate, Implement Probability and Probability Distributions to solve a wide variety of problems
3. Conduct and interpret a variety of Hypothesis Tests to aid Decision Making.
4. Explain, Analyse, Interpret Correlation and Regression to analyse the underlying relationships between different variables

Reference Books:

1. Matloff, N. Parallel Computing for Data Science: With Examples in R, C++, and CUDA (Boca Raton, FL: CRC Press), (2019).
2. Lander, J. P. R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data & Analytics Series), (2017).
3. Silge, J., & Robinson, D., Text Mining with R: A Tidy Approach (Sebastopol, CA: O'Reilly Media), (2017).
4. Wickham, H., & Grolemund, G., R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Sebastopol, CA: O'Reilly Media), (2016).
5. Peng, R. D., R Programming for Data Science (Boca Raton, FL: CRC Press), (2016).
6. Kabacoff, R. I., R in Action: Data Analysis and Graphics with R (Shelter Island, NY: Manning Publications), (2015).
7. Ken Black, *Business Statistics*, New Delhi, Wiley, (2013).
8. Lee, Cheng. et al., *Statistics for Business and Financial Economics*, New York: HeidelbergDordrecht, (2013).

MDST 413: Fundamentals of Data Science

Course Objectives: Student should be able to…

1. understand the recommendation system and two basic architectures for a recommendation system.
2. develop the fundamental knowledge and understand concepts to become a data science professional.
3. learn statistical methods and machine learning algorithms required for Data Science.
4. visualize data and use for communicating stories from data.

Credits=4	**SEMESTER-I** **MDST 413 : Fundamentals of Data Science**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Introduction to Data Science**	**(15)**
	What is Data Science , importance of data science, Big data and data Science, The current Scenario, Industry Perspective Types of Data: Structured vs. Unstructured Data, Quantitative vs. Categorical Data, Big Data vs. Little Data, Data science process, Role Data Scientist	
Credit –1 UNIT II	**Machine Learning Algorithms**	**(15)**
	Machine Learning Algorithms: Linear Regression, K-nearest Neighbors(k-NN), K-mean, Spam Filters, Naive Bayes, and Wrangling : Naive Bayes, Comparing Naive Bayes to k-NN, Scraping the Web: APIs and Other Tools	
Credit –1 UNIT III	**Data Visualization**	**(15)**

	Data visualization: Introduction, Types of data visualization, Data for visualization : Data types, Data encodings, Retinal variables, Map ping variables to encodings, Visual encodings	
Credit –1 UNIT IV	**Social Network Analysis**	**(15)**
	Social Networks as Graphs, Varieties of Social Networks, Graphs With Several Node Types, Clustering of Social-Network Graphs: Distance Measures for Social-Network Graphs, Applying Standard Clustering Methods, Betweenness, The Girvan-Newman Algorithm, Using Betweenness to Find Communities	

Course Outcomes:

Student will be able to…

1. Apply data science processes to an e-commerce data and demonstrate the use of estimation methods for analyzing this data.
2. Compare and apply appropriate machine learning algorithms for classification.
3. Compare and choose one data visualization method for effective visualization of data.
4. Apply standard clustering methods to analyze social network graph.

Text Books:

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly.
Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.

Reference Books:

1. Laura Igual and Santi Segui, Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications,Springer; 1st ed. 2017 edition.
2. Cao, L., Data Science: An Introduction (Cambridge, MA: MIT Press), (2018).
3. Donoho, D. L. ,50 Years of Data Science (New York: Springer), (2017).
4. Wickham, H., ggplot2: Elegant Graphics for Data Analysis (New York: Springer), (2017).
5. VanderPlas, J., Python Data Science Handbook: Essential Tools for Working with Data (Sebastopol, CA: O'Reilly Media), (2016).
6. Zaki, M. J., & Meira Jr., W., Data Mining and Analysis: Fundamental Concepts and Algorithms (Cambridge, UK: Cambridge University Press), (2014).
7. Provost, F., & Fawcett, T., Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking (Sebastopol, CA: O'Reilly Media), (2013).
8. James, G., Witten, D., Hastie, T., & Tibshirani, R., An Introduction to Statistical Learning: With Applications in R (New York: Springer), (2013).
9. O'Neil, C., & Schutt, R., Doing Data Science: Straight Talk from the Frontline (Sebastopol, CA: O'Reilly Media), (2013).

MDST 414 E1: Distributed Database Concepts

Course Objectives: Student should be able to…

1. Understand the various aspects in Distributed Data.
2. Understand query processing and optimization in Distributed Database.
3. Management of distributed data with different levels of transparency.
4. Understand how to use database management tools in resolving deadlock situations.

Credits=4	SEMESTER-I MDST 414: Distributed Database Concepts	No. of hours per unit/ credits
Credit –I UNIT I	**Overview of Distributed Database Design**	**(15)**
	What is Distributed Database System (DDBS), Features of DDBS, Design issue in DDBS, Distributed DBMS architecture:- Client/server System, Peer-to-Peer, Multi-Database system, Levels of distribution transparency : Reference Architecture for Distributed Databases, Types of Data Fragmentation, Integrity Constraints in Distributed Databases, Framework of Distributed Databases Design, Design of Database Fragmentation, Allocation of fragments, Transparencies in Distributed Database Design. Structured vs. Unstructured Data, Quantitative vs. Categorical Data, Big Data vs. Little Data	
Credit –1 UNIT II	**Distributed Query Processing , Optimization , Transactions Management**	**(15)**
	Concept, objective, and phases of distributed query processing, Translation of global queries to fragment queries, Query optimization in centralized databases, framework for query optimization in Distributed databases, join queries, general queries. TRANSLATION OF GLOBAL QUERIES TO FRAGMENT QUERIES: Equivalence Transformations For Queries, Transforming Global Queries into Fragment Queries, Distributed Grouping and Aggregate Function Evaluation, Parametric Queries. THE MANAGEMENT OF DISTRIBUTED TRANSACTIONS: A Framework for Transaction Management, Supporting Atomicity of Distributed Transactions, Concurrency Control for Distributed Transactions, Architectural Aspects of Distributed Transactions, Transaction Schedules in Distributed databases	
Credit –1 UNIT III	**Concurrency Control in DDBMS**	**(15)**
	Concurrency Control Based on Timestamps, Optimistic Methods for Distributed Concurrency Control. Introduction to Deadlock, Distributed Deadlock prevention, avoidance, detection and recovery, Two-Phase and Three-Phase Commit Protocol.	
Credit –1 UNIT IV	**Heterogeneous Database**	**(15)**

	Architecture of Heterogeneous Database, Interface Standards for Relational Database :ODBC ODBC architecture, functionality and usage of ODBC Database Integration:- Schema Translation and schema Integration, Query processing issues in Heterogeneous database.	

Course Outcomes:

Student will be able to:

1. Design distributed database for any real world application.
2. Write query for data manipulation on Distributed Database.
3. Manage Transaction using fragmentation.
4. Handle deadlock situation in Distributed Database.
5. Apply security policies on Distributed Databases, Manage data from Heterogeneous databases.

Text Books:

1. Distributed Databases principles & systems by Stefano Ceri, Giuseppe Pelagatti, 2nd edition, McGraw-Hill, New York, 1985, ISBN 0-07-010829-3.
2. N. TamerOzsu, Patrick Valduriez, "Principles of Distributed Database Systems", 2nd , IllustratedEdition, Prentice Hall International Inc., 1999, ISBN 0136597076, 9780136597070.
3. Database system Concept by Silberschatz And Korth 6th Edition,Tata Mcgraw Hill Education Private Limited, ISBN - 9789332901384

Reference Books:

1. Database Systems: A Practical Approach to Design, Implementation and Management- Thomas Connolly, Carolyn Begg, Pearson Publisher, 4th Edition.
2. Database Management Systems - Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill Education publisher, illustrated Edition,2003,ISBN0072465638, 9780072465631
3. Carlo Zaniolo, Stefano Ceri, Christos Faloustsos, R.T.Snodgrass, V.S.Subrahmanian, "Advanced Database Systems", Morgan Kaufman, 1997

MDST- 414 E2: IOT for Data Science I

Learning Objectives: The students should be able to…

1. Explain basic functioning of sensors and display units.
2. Familiarize the concepts of signal processing and converting elements.
3. Understand basic operation in digital systems.
4. Acquire the knowledge of microcontrollers and communication Systems.

Credits (Total Credits 2)	**SEMESTER-I** **MDST- 414 E2: IOT for Data Science I**	**No. of hours per unit/ credits**
Unit I :	**Sensors**	**(15)**
	Sensors and Transducers – Classification, Potentiometer, Strain Gauge, Piezoelectric Sensor, Linear Variable Differential Transformer (LVDT), Resistance temperature detectors (RTD), Thermocouples, Displays - LCD, Light Emitting Diode (including OLED) displays.	
Unit II :	**Signal Conditioning And Data Logging**	**(15)**
	Operational Amplifiers- Inverting, Non-Inverting, Instrumentation Amplifier, Active filters: - Low pass, High pass filter, Analog to Digital Converter – Successive Approximation, Digital to Analog Converter – R-2R ladder, Data Logging.	
Unit III :	**Digital Systems And Data Processing Devices**	**(18)**
	Analog and Digital signals, Number systems - Decimal, Binary, Hexadecimal. Logic gates-AND, OR, NOT, NOR, NAND, EX-OR.	

	Comparison of Microcontroller & Microprocessor, 8-Bit, 16-Bit microcontrollers, Study of 8051 and its Family (89C51, DS5000, 89C51VRD2). Architecture of 8051, Memories, I/O Ports.	
Unit IV :	**Data Communication System**	**(12)**
	Wireless Technology: Bluetooth, Wi-Fi, Wi-Max. Communication System: Global Positioning System, Mobile Communication, Satellite Communication (Qualitative Idea).	

Course Outcomes: -

Students will be able to:

1. Explain the operation/working of Transducer and sensor.
2. Describe Signal conditioning and Operational amplifier.
3. Explain number system and Design and construct logic gates.
4. Discuss basics of 8051microcontroller, wireless technology and communicationsystems.

Reference Books:

1. The 8051 Microcontroller and Embedded Systems, Muhammad A. Mazidi, J.G. Mazidi, R.D. Mckinlay, Pearson India Education Services Pvt. Ltd., Seventeenth Edition, 2017
2. Communication Systems, Simon Heykin, Wiley Pvt. Ltd., Third Edition, 2016.
3. The 8051 Microcontroller, Kenneth Ayala, Cengage Learning India Pvt. Ltd., Third Edition, 2014.
4. Electronic Instrumentation, H.C. Kalsi, McGraw Hill (India) Pvt. Ltd. New Delhi, Twelfth Edition,2014
5. OP-AMP and Linear Integrated Circuits, Ramakant A. Gayakwad, PHI Learning Pvt. Ltd. Delhi, Fourth Edition, 2014
6. Digital System Design, M. Morris Mano, Pearson Education Delhi, 2010

MDST- 415: Research Methodology

Learning Objectives:

The students should be able to:

1. Understand the fundamentals of research and Develop research skills
2. Explore different research methodologies and data collection and analysis techniques
3. Enhance critical thinking and problem-solving skills
4. Apply research methodologies in practical settings

Credits (Total Credits 4)	**SEMESTER-I** **MDST- 415:Research Methodology**	**No. of hours per unit/ credits**
Unit I :	**Foundations of Research**	**(15)**
	Meaning, Objectives, Motivation, Utility. Concept of theory, empiricism, deductive and inductive theory. Characteristics of scientific method – Understanding the language of research – Concept, Construct, Definition, Variable. Research Process	
Unit II :	**Problem Identification & Formulation**	**(15)**
	Research Question, Investigation Question, Measurement Issues, Hypothesis, Qualities of a good Hypothesis, Null Hypothesis & Alternative Hypothesis. Hypothesis Testing , Logic & Importance	
Unit III :	**Research Design**	**(18)**
	Concept and Importance in Research, Features of a good research design, Exploratory Research Design, concept, types and uses, Descriptive Research Designs, concept, types and uses. Experimental Design: Concept of Independent & Dependent variables.	

Unit IV :	Qualitative and Quantitative Research	(12)
	Qualitative research, Quantitative research, Concept of measurement, causality, generalization, replication. Merging the two approaches, Measurement, Sampling , Data analysis	

Course Outcomes: -

The students will be able to:

1. Explain principles and concepts of research
2. Apply different research methodologies
3. Collect and analyze data and Interpret research findings
4. Evaluate research studies and Apply research skills in practical settings

Reference Books:

1. Donald Cooper & Pamela Schindler, "Business Research Methods" ,TMGH, 9th edition
2. Alan Bryman & Emma Bell ,"Business Research Methods" , Oxford University Press
3. Kothari C.R." Research Methodology "

MDSP 416: Lab Course MDST411, MDST412, MDST413

Course Objectives: Student should be able to…

1. Master the use of the R and R Studio interactive environment.
2. Study correlation and regression, and applications probability theory.
3. To learn statistical methods and machine learning algorithms required for Data Science.
4. To visualize data and use for communicating stories from data.

Credits= 2	SEMESTER-I MDSP 416: Lab Course MDST411, MDST412, MDST413	No. of hours per unit/ credits
	1. Diagrammatic and Graphical Representation 2. Measures of central tendency and measures of dispersion 3. Correlation and Regression 4. Applications of probability 5. Fitting of discrete and continuous distribution 6. Creating and displaying Data. 7. Matrix manipulations , Creating and manipulating a List and an Array 8. Creating a Data Frame and Matrix-like Operations on a Data Frame 9. Merging two Data Frames and Applying functions to Data Frames 10.Using Functions with Factors, Accessing the Internet and String Manipulations 11.Histograms and Density Charts 12.Visualization Effects, Plotting with Layers 13.Overriding Aesthetics and Histograms and Density Charts 14.Simple Linear Regression – Fitting, Evaluation 15.Simple Linear Regression – Fitting, Evaluation 16.Case Study On Big Data 17.Case Study on Applications of Data Science 18.Case Studies on Supervised 19.Case Studies Unsupervised learnings	8

	20.Case Study on Data Visualization	

Course outcomes –

Student will be able to…

1. Install, Code and Use R Programming Language in R Studio IDE to perform basic tasks on Vectors, Matrices and Data frames.
2. Describe key terminologies, concepts and techniques employed in Statistical Analysis.
3. Define, Calculate, Implement Probability and Probability Distributions to solve a wide variety of problems
4. Conduct and Interpret a variety of Hypothesis Tests to aid Decision Making.

Reference Books:

1. Ken Black, 2013, *Business Statistics*, New Delhi, Wiley.
2. Lee, Cheng. et al., 2013, *Statistics for Business and Financial Economics*, New York: Heidelberg Dordrecht.
3. Anderson, David R., Thomas A. Williams and Dennis J. Sweeney, 2012, *Statistics for Business and Economics*, New Delhi: South Western.
4. Waller, Derek, 2008, *Statistics for Business*, London: BH Publications.

SEMESTER II

MDST 421: Mathematical Foundations for Data Science

Course Objectives:

Student should be able to…

1. Understand the various aspects in district mathematics in data science.
2. Understand Data Analysis and probability theory.
3. Study the concept of Linear Algebra and Calculus
4. Understand how to use Regression model.

Credits=4	**SEMESTER-I** **MDST 421: Mathematical Foundations for Data Science**	**No. of hoursper unit/ credits**
Credit –I UNIT I	**Discrete mathematics for Data Science:**	**(15)**
	Concept of set, cardinality of set, finite, infinite and uncountably infinitesets, Basic set operations, Principal of inclusion Exclusion, Graph: Basic terminologies, representation of graph, path and circuit, graph traversal, travelling salesperson problem, Trees: Basic terminologies, search tree: Binary & M-ary tree.	
Credit –1 UNIT II	**Data Analysis & Probability Theory**	**(15)**
	Data Representation, Average, Spread, Experiments, Outcomes, Events, Probability, Permutations and Combinations, Random Variables, Probability Distributions, Mean and Variance of a Distribution, Binomial, Poisson, and Hyper geometric Distributions, Normal Distribution, Distributions of Several Random Variables.	
Credit –1 UNIT III	**Linear Algebra and Calculus**	**(15)**
	Linear Algebra: Matrix and vector algebra, systems of linear equations using matrices, linear independence, Matrix factorization concept/LU decomposition, Eigen values and eigenvectors, Understanding of calculus: concept of function and derivative, Multivariate calculus: concept, Partial Derivatives, chain rule, the Jacobian and the Hessian	
Credit –1 UNIT IV	**Regression Model**	**(15)**
	Introduction, types of regression. Simple regression- Types, Makingpredictions, Cost function, Gradient descent, Training, Model evaluation. Multivariable regression: Growing complexity, Normalization, making predictions, initialize weights, Cost function, Simplifying with matrices, Bias term, Model evaluation	

Course Outcomes: Student will be able to…

1. Design discrete mathematics for any real world application.
2. Handle data analysis and probability theory.
3. Apply concept of linear algebra and calculus.
4. Manage regression model.

Reference Books:

1. Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger, 2018
2. Goodfellow, I., Bengio, Y., & Courville, A., Deep Learning (Cambridge, MA: MIT Press), (2016).
3. Vanderbei, R. J., Linear Programming: Foundations and Extensions (New York: Springer), (2014).
4. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux, RémyDrouilhet, Benoit Liquet, Springer 2013

5. Wasserman, L. All of Statistics: A Concise Course in Statistical Inference (New York: Springer), (2013).
6. Hastie, T., Tibshirani, R., & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, andPrediction (New York: Springer), (2009).
7. Bertsekas, D. P., & Tsitsiklis, J. N., Introduction to Probability (Belmont, MA: Athena Scientific), (2008).
8. Bishop, C. M. Pattern Recognition and Machine Learning (New York: Springer), (2006).
9. Boyd, S., & Vandenberghe, L., Convex Optimization (Cambridge, UK: Cambridge University Press). (2004).
10. Trefethen, L. N., & Bau, D. Numerical Linear Algebra (Philadelphia, PA: Society for Industrial and AppliedMathematics), (1997).

MDST 422: Python Programming

Course Objectives: Student should be able to…

1. Understand python programming.
2. Imbibe build and package python modules for reusability.
3. Understand suitable knowledge about their implementation.
4. Understand various file handling techniques and database interactions.

Credits=4	**SEMESTER-I** **MDST 422: Python Programming**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Introduction To Python**	**(15)**
	Introduction, Various IDEs, Numeric data types: int, float, complex, String, list and list slicing, Tuples, Control Flow Conditional blocks using if, else and elif Simple for and while loops in python For loop using ranges, string, list and dictionaries Loop manipulation using pass, continue, break and else	
Credit –1 UNIT II	**Functions and Packages**	**(15)**
	Functions Arguments, Lambda Expressions, Function Annotations, Modules Organizing python projects into modules Importing own module as well as external modules, Packages, Programming using functions, modules and external packages	
Credit –1 UNIT III	**Data Structures and Python File Operations**	**(15)**
	Lists as Stacks, Queues, Comprehensions, Tuples and sequences, Sets, Dictionaries, reading config files in python, Writing log files in python, Understanding read functions, read (), redline () and readlines (), Understanding write functions, write() and writelines(), Manipulating file pointer using seek, Programming using file operations	
Credit –1 UNIT IV	**Database Interaction SQL**	**(15)**
	Database connection using python, Creating and searching tables, Reading and storing config information on database, Programming using database connections	

Course outcomes –Student will be able to…

1) Explain principles python programming.
2) Implement clear and effective python code.
3) Discuss Data Analysis using python libraries.
4) Try to understand Error handling.

References:

1. McKinney, W., Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (Sebastopol, CA: O'Reilly Media), (2017).
2. Dusty Phillips, "Python 3 Object-oriented Programming Second Edition", Pack Publishing, Greg, 2015.
3. VanderPlas, J." Python Data Science Handbook: Essential Tools for Working with Data (Sebastopol, CA: O'Reilly Media)", (2016).
4. Charles Dierbach," Introduction to Computer Science Using Python: A Computational Problem-Solving Focus", John Wiley & Sons,2013.
5. Jan van Eijck , Christina Unger, "Computational Semantics with Functional Programming", Cambridge University Press, 2012 .
6. Kenneth C. Louden, "Programming Languages: Principles and Practice", Course Technology Inc., 2011. Richard L. Halterman, "LEARNING TO PROGRAM WITH PYTHON", Southern Adventist University, 2011
7. Lutz, M., Learning Python (Sebastopol, CA: O'Reilly Media), (2013).
8. Sweigart, A., Automate the Boring Stuff with Python: Practical Programming for Total Beginners (San Francisco, CA: No Starch Press), (2015).
9. Ramalho, L., Fluent Python (Sebastopol, CA: O'Reilly Media), (2015).

MDST 423: Data Preparation Analysis

Course Objectives: Student should be able to…

1. understand the importance of data and data preprocessing
2. understand data cleaning and conditioning
3. understand an ETL – Extract, Transform and Load – process and ETL tools
4. get acquainted with data visualization techniques for exploratory analysis

Credits=4	**SEMESTER-I** **MDST 423: Data Preparation Analysis**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Data Gathering and Data Discovery**	**(15)**
	Identifying potential data sources, Gathering data, Data discovery-understanding the data, assessing data, data formats, Parsing, Selecting features, Transformation, Scalability and real-time issues	
Credit –1 UNIT II	**Cleaning and Conditioning Data**	**(15)**
	Data Preparation Basic Models: Data Integration, Data Cleaning, Data Normalization, Min-Max Normalization, Z-score Normalization, Decimal Scaling Normalization, Consistency checking, Heterogeneous and missing data, Dealing with missing values, Duplicate values, Noise, Inconsistent data, Outliers.	
Credit –1 UNIT III	**Exploratory Analysis**	**(15)**

	Formulating Hypothesis, Data Terminology, Data Exploration, Data Exploration through Summary Statistics, Data Exploration through Plots, Feature Engineering, Feature selection, Feature transformation, Dimensionality reduction	
Credit –1 UNIT IV	**Data Visualization and Advanced Tools for Data Preparation**	**(15)**
	Visualization techniques, Different types of plots, Designing visualizations, Time series, Geolocated data, Correlations and connections, Hierarchies and networks, Interactivity.Web scraping, Data from social networks, Open-source tools for data preparation: Open Refine, R/Python libraries for data preparation and visualization	

Course Outcomes: Student will be able to …

1. Design Data Gathering and Data Discovery
2. Handle Cleaning and Conditioning Data
3. Manage Exploratory Analysis
4. Apply Data Visualization and Advanced Tools for Data Preparation

Reference Books:

1. Bateman, S., & D'Ignazio, C., Data Feminism (Cambridge, MA: MIT Press), (2020).
2. Wickham, H., & Grolemund, G., R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Sebastopol, CA: O'Reilly Media), (2016).
3. Bramer, M., Principles of Data Mining (London, UK: Springer), (2016).
4. Swayne, D. F., Wickham, H., & Cook, D., Interactive and Dynamic Graphics for Data Analysis: WithR and GGobi (New York: Springer), (2015).
5. Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (Cambridge, MA: MIT Press), (2015).
6. Franks, B., Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics (Indianapolis, IN: Wiley), (2012).
7. Luger G.F. and Stubblefield W.A.,Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition, (2008).
8. Berson, A., Smith, S. J., & Thearling, K., Building Data Mining Applications for CRM (New York: McGraw-Hill Education), (1999).

MDST 424 E I: AI for Data Science

Course Objectives: Student should be able to…

1. Understand the various aspects in intelligent agents.
2. Understand problem solving methods.
3. Study the concept Knowledge, reasoning and planning
4. Study the applications.

Credits=4	SEMESTER-II MDST 424: AI for Data Science	No. of hours per unit/ credits
Credit –I UNIT I	**Introduction and Intelligent Agents:**	**(15)**
	Introduction: What is AI, Foundations History of Artificial Intelligence, The State of the Art Intelligent Agents: Agents and Environments, Good Behavior: The Concept of Rationality, The Nature of Environments, and The Structure of Agents	
Credit –1 UNIT II	**Problem-solving:**	**(15)**
	Solving Problems by Searching: Problem-Solving Agents, Uninformed Search Strategies, Informed (Heuristic) Search Strategies, Heuristic Functions, Beyond Classical Search Local Search Algorithms and Optimization Problems, Local Search in Continuous Spaces, Searching with Nondeterministic Actions, Searching with Partial Observations, Online Search Agents and Unknown Environment.	
Credit –1 UNIT III	**Knowledge, reasoning, and planning**	**(15)**
	Knowledge based Agents, First-Order Logic and Its Inference, Classical Planning, Planning and Acting in the Real World, Knowledge Representation, Quantifying Uncertainty, Probabilistic Reasoning, Probabilistic Reasoning over Time, Making Simple Decisions, Making Complex Decisions	
Credit –1 UNIT IV	**Learning and Applications with case studies**	**(15)**
	Learning from Examples, Knowledge in Learning, Learning Probabilistic Models, Reinforcement Learning, AI Applications in various fields in marketing, healthcare, banking, finance, etc. Case Studies: Credit card Fraud Analysis, Sentiment Analysis, Recommendation Systems and Collaborative filtering, Uber Alternative Routing	

Course Outcomes:

Student will be able to…

1. Design intelligent agent.
2. Handle problem solving methods.
3. Manage Knowledge, reasoning and planning.
4. Apply the Knowledge in Applications

Reference Books:

1. Chollet, F., Deep Learning with Python (Shelter Island, NY: Manning Publications), (2017).
2. Russell, S. J., & Norvig, P., Artificial Intelligence: A Modern Approach (New York: Pearson), (2016).
3. Goodfellow, I., Bengio, Y., & Courville, A., Deep Learning (Cambridge, MA: MIT Press), (2016).
4. Marsland, S., Machine Learning: An Algorithmic Perspective (Boca Raton, FL: CRC Press), (2014).
5. Luger G.F. and Stubblefield W.A. ,Artificial Intelligence: Structures and strategies for Complex

Problem Solving. Addison Wesley, 6th edition, (2008).

6. Bishop, C. M., Pattern Recognition and Machine Learning (New York: Springer), (2006).

7. Nilsson Nils J, "Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. SanFrancisco, CA, ISBN: 978-1-55-860467-4, (1998).

MDST 424 E II: IOT for Data Science II

Learning Objectives: Student should be able to…

1. gain knowledge on basic functioning of Arduino with sensors.
2. familiarize the concepts of Arduino programming with sensors.
3. understand basic of Raspberry Pi and to acquire the knowledge of Interfacing with Raspberry Pi.
4. familiarize the applications of IOT.

Credits (Total Credits 2)	**SEMESTER-II** **MDST- 424 EII: IOT for Data Science II**	**No. of hours per unit/ credits**
Unit I :	**Arduino in IOT**	**(15)**
	Characteristics of IoT, Physical design of IoT, Difference between IoT and M2M, Interoperability in IoT, Introduction to Arduino Programming, Interfacing of Temperature, Humidity, Motion, Light and Gas Sensor with Arduino.	
Unit II :	**Introduction of Raspberry Pi**	**(15)**
	Introduction to Python programming, Introduction to Raspberry Pi, Interfacing Raspberry Pi with basic peripherals, Implementation of IoT with Raspberry Pi.	
Unit III :	**Implementation of IOT**	**(16)**
	Implementation of IoT with Raspberry Pi, Introduction to Software defined Network (SDN), SDN for IoT, Data Handling and Analytics	
Unit IV :	**Applications of IOT**	**(14)**
	Cloud Computing, Cloud services - SaaS, PaaS, IaaS, Sensor-Cloud, Smart Cities and Smart Homes, Connected Vehicles, Smart Grid, Industrial IoT.,Case Study: Agriculture, Healthcare, Activity Monitoring.	

Course Outcomes: - Students will be able to:

1. Explain the Arduino with sensors.
2. Describe concepts of Arduino with programming.
3. Explain the raspberry pi.
4. Discuss interfacing of raspberry pi.
5. Explain the various applications of raspberry pi.

TEXT BOOKS:

1. "The Internet 'of Things: Enabling Technologies, Platforms, and Use Cases", by Pethuru Raj and Anupama C. Raman (CRC Press)

2. "Make sensors": Terokarvinen, kemo, karvinen and villey valtokari, 1st edition, maker media,2014.

3. "Internet of Things: A Hands-on Approach", by Arshdeep Bahga and Vijay Madisetti

REFERENCE BOOKS:

1. Antonopoulos, A., & Gillam, L., Internet of Things: Principles and Paradigms (Boca Raton, FL: CRC Press), (2017).
2. Patel, V. M., & Zhang, S, Internet of Things: A Hands-On Approach (Boca Raton, FL: CRC Press).
3. Ray, P. P. (2018). Internet of Things: A Complete Introduction (New York: Springer), (2017).
4. Banerjee, A., & Dasgupta, S., Internet of Things: Challenges and Opportunities (New York: Springer), (2015).
5. Kortuem, G., Kawsar, F., Sundramoorthy, V., & Fitton, D., Smart Objects, and the Internet of Things: AHandbook for IoT Developers (San Francisco, CA: Morgan & Claypool Publishers), (2010).
6. Bandyopadhyay, D., & Sen, J. (2017). Internet of Things: Applications and Challenges in Smart Cities (Boca Raton, FL: CRC Press), (2017).
7. Minoli, D., IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things (Indianapolis, IN: Wiley), (2017).
8. Bhuyan, M. H., Bhattacharyya, D., Kalita, J. K., & Bhattacherjee, V., Internet of Things: Challenges, Advances, and Applications (Cham, Switzerland: Springer), (2019).

MDST- 425: Research Project

Learning Objectives:

Students should be able to…

1. Apply data science concepts and techniques to address a specific research problem.
2. Conduct a thorough literature review and critically analyze existing research in the field.
3. Design and implement a research plan, including data collection and preprocessing.
4. Interpret and evaluate the results of the research project.
5. Communicate the research findings effectively through a written report and an oral presentation.
6. Develop project management and time management skills.

Credits (Total Credits 4)	**SEMESTER-I** **MDST- 425:Research Project**	**No. of hours per unit (60)**
Unit I :	Introduction to Research in Data Science	**(15)**
	Introduction to research methodologies in data science,Identifying research problems and formulating research questions,Reviewing and critiquing existing literature,Ethical considerations in data science research	
Unit II :	Research Planning and Design	**(15)**
	Developing a research plan and timeline, Defining research objectives and hypotheses, Selecting appropriate research methods Data collection strategies and techniques	
Unit III :	Data Collection and Preprocessing , Data Analysis and Modeling	**(18)**
	Data acquisition and cleaning, Exploratory data analysis Feature selection and dimensionality reduction,Handling missing data and outliers, Supervised and unsupervised learning techniques Regression analysis,Classification and clustering algorithms Evaluation metrics for model performance	
Unit IV :	Project Management and Professional Skills	**(12)**

	Time management and project organization ,Collaboration and teamwork, Ethical considerations in data science research Intellectual property and plagiarism	

Course Outcomes: - Student will be able to…

1. Demonstrate a deep understanding of research methodologies in data science and their application to real-world problems.
2. Formulate clear and concise research questions and objectives.
3. Conduct a comprehensive literature review and critically analyze existing research in the field.
4. Design and execute a research plan, including data collection, preprocessing, and analysis

Reference Books:

1. Donald Cooper & Pamela Schindler, "Business Research Methods" ,TMGH, 9th edition
2. Alan Bryman & Emma Bell ,"Business Research Methods" , Oxford University Press
3. Kothari C.R." Research Methodology "

MDSP 426: Lab Course on MDST 421, MDST 422, MDST 423

Course Objectives: Student should be able to…

1. Understand the various aspects in district mathematics in data science.
2. Understand Data Analysis and probability theory. Study the concept of Linear Algebra and Calculus
3. Understand how to use Regression model and Understand python programming
4. To understand the importance of data and data preprocessing, data cleaning

Credits=2	**SEMESTER-II** **MDSP 426: Lab Course on MDST 421,MDST 422,MDST 423**	**No. of hoursper unit/ credits**

	1.Practical based on MDSP 426 2.Practical on Strings and Lists 3.Practical on Conditional statements 4.Practical on Looping statements 5.Practical on Functions 6.Practical on Packages 7.Practical on Stacks, Queues, Tuples, Sets, Dictionaries 8.Practical on File Handling 9.Practical on Database Operations 10. Choose a dataset from UCI Machine Learning repository (e.g. Cleveland). 11. Compute and display summary statistics for each feature available in the	8

	dataset. (e.g. minimum, maximum, mean, range, standard deviation, variance and percentiles). Use a bar-graph to demonstrate your results. 12. Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram. 13. Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single plot. Compare distributions and identify outliers. 2. 14. Take any dataset from UCI repository (like air quality dataset) and perform regression analysis on it. 15. Demonstrate your results using appropriate visualization techniques for numerical and categorical features (e.g., histogram, scatter plot, heat map, box plot). 16. Compute Eigen values and Eigen vectors for dataset in part a. seful links: 1. https://archive.ics.uci.edu/ml/datasets/heart+disease 2.https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original) 3. https://archive.ics.uci.edu/ml/datasets/Air+Quality	

Course Outcomes: - Students will be able to…

1. Apply principles python programming.
2. Implement clear and effective python code.
3. Design discrete mathematics for any real world application. Handle data analysis and probability theory
4. Design Data Gathering and Data Discovery and Handle Cleaning and Conditioning Data

Reference books:

1. Bateman, S., & D'Ignazio, C., Data Feminism (Cambridge, MA: MIT Press), (2020).
2. Wickham, H., & Grolemund, G., R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (Sebastopol, CA: O'Reilly Media), (2016).
3. Bramer, M., Principles of Data Mining (London, UK: Springer), (2016).
4. McKinney, W., Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (Sebastopol, CA: O'Reilly Media), (2017).
5. Dusty Phillips, "Python 3 Object-oriented Programming Second Edition", Packt Publishing, Greg,2015.
6. VanderPlas, J." Python Data Science Handbook: Essential Tools for Working with Data (Sebastopol, CA: O'Reilly Media)", (2016).
7. Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger, 2018
8. Goodfellow, I., Bengio, Y., & Courville, A., Deep Learning (Cambridge, MA: MIT Press), (2016).
9. Vanderbei, R. J., Linear Programming: Foundations and Extensions (New York: Springer), (2014).