Rayat Shikshan Sanstha's

YASHAVANTRAO CHAVAN INSTITUTE OF SCIENCE, SATARA

(An Autonomous College)

Reaccredited by NAAC with 'A+' Grade

New Syllabus For

Master of Science

Part - I

M.Sc. Data Science

Syllabus

To be implemented from June, 2022 onward

- **OBJECTIVES:**

1. To create post-graduates with sound knowledge of Computer Science , who can contribute towards recent advances in technology
2. To provide advanced and in-depth knowledge of computer science and specialization in one or two subjects of new era of technology.
3. To prepare Post Graduates who will achieve peer-recognition, as an individual or in a team, through demonstration of good analytical, design, programming and implementation skills.
4. To enable students, pursue a professional career in Artificial Intelligence and IOT in related industry, business and research.
5. To impact industry knowledge and practical skills of current trends in IT field to the students.
6. To develop ability among students to formulate, analyze and solve real life problems facedin Computer Science industry.
7. To produce computer science professionals who can be directly employed or start his/her own work as Freelance Software Developer, IT consultant, Software Tester, Service engineer, Project Manager and even an entrepreneur in Computer Science industry.
8. To Develop designing, analyzing and critical thinking skill among students.

- **OUTCOMES:**

After completing this courses students shall be expert in following things:

1. Avail the skills of Current trends in IT Industries and new Technologies.
2. Apply knowledge of programming platforms in IOT and AI in real life.
3. Student should avail detail knowledge of Artificial Intelligence, IOT, and Networking 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 generate a solution independently, check and validate it and modify if necessary.
7. Translate academic research into innovation and creatively design scientific solutions to problems. Exemplify 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.

- **SCOPE:**

  After Successful completion of two years Master's Degree in Computer Science, we observed that the students have the ample opportunities in diversified areas such as:

  1. Software Industry.
  2. Communication Industry
  3. Digital Media
  4. Agriculture Industry
  5. Health and Care.
  6. Research Field.
  7. IoT Architects

Courses Structure for Postgraduate Programme to be implemented from Academic year 2022-23 for Computer Science

Course Structure

Course Code	Title of the Course	Credits	Teaching Scheme (h/w)		Evaluation Scheme (marks)		Total
			L	P	ESE	ISE	
M.Sc. Part I - Semester I							
MDST 101	Statistical Foundation for Data Science	4	4	-	60	40	100
MDST 102	Programming using R	4	4	-	60	40	100
MDST 103	Distributed Database	4	4	-	60	40	100
MDST 104	Fundamentals of Data Science	4	4	-	60	40	100
MDSP 105	Lab I: Statistical Foundation for Data Science and Programming using R	4	-	12	60	40	100
MDSP 106	Lab II: Database Technologies and Fundamentals of Data Science	4	-	12	60	40	100
	Total	**24**	**16**	**24**	**360**	**240**	**600**
M.Sc. Part I - Semester II							
MDST 201	Mathematical Foundations for Data Science	4	4	-	60	40	100
MDST 202	AI for Data Science	4	4	-	60	40	100
MDST 203	Data Preparation Analysis	4	4	-	60	40	100
MDST 204	Design and Analysis of Algorithm	4	4	-	60	40	100
MDST 205	Python Programming	4	4	-	60	40	100
MDSP 206	Lab III: Mathematical Foundations and AI for Data Science and Data Preparation Analysis	4	-	12	60	40	100
MDSP 207	Lab IV: Design and Analysis of Algorithm and Python Programming	4	-	12	60	40	100
	Total	**28**	**20**	**24**	**420**	**280**	**700**
M.Sc. Part II - Semester III							
MDST 301	Big Data Analytics	4	4	-	60	40	100
MDST 302	Data Storage Technologies & Networking	4	4	-	60	40	100
MDST 303	Image Analytics	4	4	-	60	40	100
MDST 304	Machine Learning	4	4	-	60	40	100
MDST305	Deep Learning	4	4	-	60	40	100
MDSP 306	Lab V: Big Data Analytics, Data Mining and Data Storage Technologies & Networking and Image Analytics	4	-	12	60	40	100
MDSP 307	Lab VI: Machine Learning and Deep Learning	4	-	12	60	40	100
	Total	**28**	**20**	**24**	**420**	**280**	**700**
M.Sc. Part II - Semester IV							
MDST 401	GPU Computing	4	4	-	60	40	100
MDST 402	Recommender System	4	4	-	60	40	100

MDSP 403	LAB VII: GPU Computing and Recommender System	4	-	12	60	40	100
MDSP 404	LAB VIII: Internship Program (Industrial Training)	4	-	12	60	40	100
	Total	**16**	**8**	**24**	**240**	**160**	**400**
	Grand Total	**96**	**64**	**96**	**1440**	**960**	**2400**

Project Academic Project is divided into 4 phases.

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

Evaluation Pattern PG: M.Sc. I
Semester-I
Theory: Practical (60: 40)
ESE: ISE (60: 40)

Class	Semester	Paper Name	Theory			Practical			Total
			ESE	ISE	Total	ESE	ISE	Total	
M.Sc. I	I	Paper I: MDST101	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Book Review)	100	-	-	-	**100**
		Paper II: MDST 102	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Home Assignments)	100	-	-	-	**100**
		Paper III: MDST 103	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Survey/Seminar)	100	-	-	-	**100**
		Paper IV: MDST 104	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Group Discussion/Innovative Idea Presentation)	100	-	-	-	**100**
		Practical P-I: MDSP 105	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Case Study/Survey Report	100	**100**
		Practical P-II: MDSP106	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Model Presentation/ Project Part I)	100	**100**
		Total	240	160	**400**	120	80	**200**	**600**

Class	Semester	Paper Name	Theory			Practical			Total
			ESE	ISE	Total	ESE	ISE	Total	
M.Sc. I	II	Paper V: MDST201	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Book Review)	100	-	-	-	**100**
		Paper VI: MDST202	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Home Assignments)	100	-	-	-	**100**
		Paper VII: MDST203	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Survey/Seminar)	100	-	-	-	**100**
		Paper VIII: MDST204	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Group Discussion/ Innovative Idea Presentation)	100	-	-	-	**100**
		Paper IX: MDST205	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (MOOC/Open Book Test)	100	-	-	-	**100**
		Practical P-III: MDSP206	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Case Study/Survey Report)	100	**100**
		Practical P-IV: MDSP207	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Model Presentation/ Project Part II)	100	**100**
		Total	300	200	**500**	120	80	**200**	**700**

.

Class	Semester	Paper Name	Theory			Practical			Total
			ESE	**ISE**	**Total**	**ESE**	**ISE**	**Total**	**Total**
M.Sc. II	III	Paper X: MDST301	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Book Review/ Innovative Idea Presentation)	100	-	-	-	**100**
		Paper XI: MDST302	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Home Assignments)	100	-	-	-	**100**
		Paper XII: MDST303	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Seminar)	100	-	-	-	**100**
		Paper XIII: MDST304	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Open Book Test)	100	-	-	-	**100**
		Paper XIV: MDST305	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (MOOC/Group Discussion)	100	-	-	-	**100**
		Practical P-V: MDSP306	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Case Study/Survey Report)	100	**100**
		Practical P-VI: MDSP307	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Model Presentation/ Project Part III)	100	**100**
		Total	300	200	**500**	120	80	**200**	**700**

Class	Semester	Paper Name	Theory			Practical			Total
			ESE	ISE	Total	ESE	ISE	Total	
M.Sc. II	IV	Paper XV: MDST401	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (Paper Presentation/ Webinar Participation)	100	-	-	-	**100**
		Paper XVI: MDST402	60	**40** ISE I: 10 ISE II: 10 (online) Activity: 20 (MOOC/Open Book Test)	100	-	-	-	**100**
		Practical P-VII: MDSP403	-	-	-	60	**40** Journal: 10 Day to Day Performance: 10 Activity: 20 (Case Study/Survey Report)	100	**100**
		Internship	-	-	-	60 Internship: (Report Submission: 30 Presentation and Viva: 30)	**40** Internship certificate: 10 Day to Day Performance: 10 Activity: 20 (Model Presentation/ Project Part IV)	100	**100**
		Total	120	80	**200**	120	80	**200**	**400**

Note:

The strength of the student per batch is as per university norms.

The duration of practical examination for M.Sc. Semester I, II, III and IV should be 2 days of 12 hours excluding inspection day.

Rayat Shikshan Sanstha‟s

Yashavantrao Chavan Institute of Science, Satara (Autonomous Institute)

Department of Computer Science Scheme of Credit for M.Sc. Computer Science Under Choice Based Credit System (CBCS)

W e f (June 2022-23)

1. **SUBJECT: Computer Science**

2. **YEAR OF IMPLEMENTATION:** New Syllabi for the M.Sc. I Computer Science will be implemented from June 2022 onwards.

3. **PREAMBLE:**

Master of Science is an integrated academic degree in faculty of Science. The faculty is not ignoring the developments in the field of Computer Science. The revision of existing syllabus of 6 Computer Science subject in science faculty is essential. This is a humble endeavor to initiate the process towards an era of knowledge. The students from science faculty should also be competent for this change in the technology. In this year, a student will able to understand Computer languages and technologies to build software with confidence. In the subject, the student will also get a basic and proper knowledge in the field of Artificial Intelligence and IOT.

4. **GENERAL OBJECTIVES OF THE COURSE:**
   1. To create post-graduates with sound knowledge of fundamentals of Computer Science, who can contribute towards advancing science and technology.
   2. To create post-graduates with sufficient capabilities in Computer Science who can become researchers and developers to satisfy the needs of the core Computer Science Industry.
   3. To develop ability among students to formulate, analyze and solve real life problems.
   4. To provide opportunity to students to learn the latest trends in Computer Science and make them ready for life-long learning process.
   5. To make the students aware of professional ethics of the industry, and prepare them with basic soft skills essential for working in community and professional teams.
   6. To prepare the students for postgraduate studies through competitive examinations, enabling them to reach higher echelons of excellence.
   7. To produce Computer Science professionals who can be directly employed or start his/her own work as Software Developer, Data Scientist, testing professional, Network engineer and even an entrepreneur in IT industry.

5. **DURATION:02Years (FullTime)**

6. **PATTERN: SEMESTER EXAM(CBCS)**

7. **MEDIUM OF INSTRUCTIONS :ENGLISH**

8. STRUCTURE OF COURSE:

1. FIRST SEMESTER

Sr. No.	SUBJECT TITLE	Theory			Practical		
		PAPER NO & Paper Code	No. of lectures per week	Credits		No. of lectures Per week	Credits
1	Computer Science	Paper I: MDST101	16	16	Practical Paper – V : MDSP106	12	4
		Paper II: MDST102					
		Paper III: MDST103			Practical Paper –VI : MDSP107	12	4
		Paper IV: MDST104					

2. FOURTH SEMESTER

Sr. No.	SUBJECT TITLE	Theory			Practical		
		PAPER NO & Paper Code	No. of lectures per week	Credits		No. of lectures Per week	Credits
1	Computer Science	Paper I: MDST101	20	20	Practical Paper – V : MDSP106	12	4
		Paper II: MDST102					
		Paper III: MDST103			Practical Paper –VI : MDSP107	12	4
		Paper IV: MDST104					
		Paper V: MDST105					

3. Structure and Title of Papers of M. Sc. Course:

- **M. Sc. I**

Semester I

- **M. Sc. I**

Semester II

MDST/Pxyz–

M M.Sc.
DS Data Science
T Theory
P Practical
x 1to 4 :Semester number
yz 1to 7 :course number

Rules and Regulations:

1. Core courses will be offered only to the students of M.Sc. Data Science.
2. The pre-requisites for electives courses will be decided by the departmental committee and Certificate and diploma program will be mandatory for all students.
3. Electives will be offered for minimum 08 and maximum 12students in view of the infrastructure of the department. Electives to be offered or otherwise will be at the sole discretion of the departmental committee.
4. Minimum attendance required to appear for semester-end examination will be 75 % for each credit course.

4. OTHERFEATURES:

A. LIBRARY:

- **REFERENCE BOOKS**

- **JOURNALSAND PERIODICALS**

1.Acta Informatics. 0.900 Impact Factors 2019. 2.AI and Ethics.
3. AI & SOCIETY.
4. Algorithmic. 0.650 Impact Factors 2019.
5. Annals of Mathematics and Artificial Intelligence. 0.778 Impact Factor 2019. 6.Applicable Algebra in Engineering, Communication and **Computing**. ...
7. Applied Intelligence. ...
8. International journal of computer vision 9.Expert Systems with applications 10.IEEE Transactions on Image Processing

B. SPECIFIC EQUIPMENTS:

Computers, Laptops, Printers, Scanners,LCDProjectors, E-Podium,SmartBoard,DocumentCamera, Visualizer

C. LABORATORY EQUIPMENTS:

1. Soft Computing Tools –SCILAB,MATLAB
2. Tableau Software
3. R Software,PyCharm
4. Anaconda

SEMESTER I
PAPER I

MDST 101 : Statistical Foundation for Data Science

Course Objectives: Student will 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 101 : 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 should 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. James, G., Witten, D., Hastie, T.J., Tibshirani, R. and Friedman, J. (2013). *An Introduction to Statistical Learning with Applications in R* . Springer, New York.
2. Hastie, T.J., Tibshirani, R. and Friedman, J. (2009). *The elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed). Springer, New York.
3. Buehlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, New York.
4. Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity*. CRC press, New York.
5. Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint.* Cambridge University Press.

Paper II

MDST 102: Programming using R

Course Objectives:

1. Master the use of the R and RStudio interactive environment.
2. Expand R by installing R packages.
3. Explore and understand how to use the R documentation.
4. Read Structured Data into R from various sources.
5. Understand the different data types in R.
6. Understand the different data structures in R.

Credits=4	**SEMESTER-I** **MDST 102: 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-

After the successful completion of this module, students 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.
5. Understand, Analyse, Interpret Correlation and Regression to analyse the underlying relationships between different variables

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.

Paper III

MDST 103 : Fundamentals of Data Science

Course Objectives:

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

Credits=4	**SEMESTER-I** **MDST 103 : 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 scien ce 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 visualisation: Introduction, Types of data visualisation, 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:

On completion of the course, learner 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. Design a model of recommendation system based on the content of the data.
5. 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

PAPER IV

MDST 104: Distributed Database Concepts

Course Objectives:

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 104: 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, Data science process, Role Data Scientist	

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 DIST RIBUTED 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:

On completion of the course, learner 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.
6. 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 , Illustrated Edition, 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

Paper V

MDSP 105: Lab Course on Statistical Foundation for Data Science and Programming using R

Course Objectives:

1. Master the use of the R and R Studio interactive environment.
2. Expand R by installing R packages.
3. Explore and understand how to use the R documentation.
4. Read Structured Data into R from various sources.
5. Understand the different data types in R.
6. Understand the different data structures in R.

Credits=	SEMESTER-I **MDSP 105: Lab Course on Statistical Foundation for Data Science and Programming using R**	No. of hours per unit/ credits
	Group A 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	
	Group B 1. Creating and displaying Data. 2. Matrix manipulations , Creating and manipulating a List and an Array 3. Creating a Data Frame and Matrix-like Operations on a Data Frame 4. Merging two Data Frames and Applying functions to Data Frames 5. Using Functions with Factors, Accessing the Internet and String Manipulations 6. Histograms and Density Charts 7. Visualization Effects, Plotting with Layers 8. Overriding Aesthetics and Histograms and Density Charts 9. Simple Linear Regression – Fitting, Evaluation 10. Simple Linear Regression – Fitting, Evaluation	

Course outcomes –

After the successful completion of this module, students 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.

Understand, Analyse, Interpret Correlation and Regression to analyse the underlying relationships between different variables

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.

Paper VI
MDSP 106: Lab Course on Fundamentals of Data Science and Distributed Database Systems

Credits=	SEMESTER-I MDSP 106: Lab Course on Fundamentals of Data Science and Distributed Database Systems	No. of hours per unit/ credits
	Group A: 1. Implementation of Supervised Learning Algorithms - Regression. 2. Implementation of Supervised Learning Algorithms - Classification. 3. Implementation of Unsupervised Learning Algorithms - Clustering. 4. Implementation of Unsupervised Learning Algorithms - Assocation. **5.** Graphical representation of Data using charts and plots Data Visualization	
	Group B: List of Experiments: 1. Create two databases either on single DBMS and Design Database to fragment and share the fragments from both database and write single query for creating view. 2. Create two databases on two different computer systems and create database view to generate single DDB. 3. Create various views using any one of examples of database and Design various constraints. 4. Write and Implement algorithm for query processing using any of Example in either 5. C /C++ /Java / .NET 6. Using any of example, write various Transaction statement and show the information about concurrency control [i.e. various lock's from dictionary] by executing multiple update and queries. 7. Using Transaction /commit rollback, Show the transaction ACID properties. 8. Write java JDBC program and use JTA to show various isolation level's in transaction. 9. Implement Two Phase Commit Protocol 10. Case study on noSQL 11. Case study on Hadoop	

SEMESTER II
Paper VI
MDST 201: Mathematical Foundations for Data Science

Course Objectives:

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 201: Mathematical Foundations for Data Science**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Discrete mathematics for Data Science:**	**(15)**
	Concept of set, cardinality of set, finite, infinite and uncountably infinite sets, 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, Making predictions, 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:

On completion of the course, learner 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, Wiley (Low price edition available)
2. Introduction to. Mathematics. Statistics. Robert V. Hogg. Allen T. Craig, Low price Indian edition by Pearson Education
3. Probability and Statistics for Engineers. Richard A. Johnson, Irwin Miller, John Freund
4. Mathematical Statistics with Applications. Irwin Miller, Marylees Miller, Pearson Education
5. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoit Liquet, Springer 2013

Paper VII
MDST 202: AI for Data Science

Course Objectives:

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-I** **MDST 202: 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:

On completion of the course, learner 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. Luger G.F. and Stubblefield W.A. (2008). Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition
2. Nilsson Nils J , "Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4
3. Patrick Henry Winston, "Artificial Intelligence", Addison-Wesley Publishing Company, ISBN:0-201-53377-4

Paper VIII
MDST 203: Data Preparation Analysis

Course Objectives:

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

Credits=4	**SEMESTER-I** **MDST 203: 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 missi ng 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:
On completion of the course, learner 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:**

4. Luger G.F. and Stubblefield W.A. (2008). Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition
5. Nilsson Nils J , "Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4

Patrick Henry Winston, "Artificial Intelligence", Addison-Wesley Publishing Company, ISBN:0-201-53377-4

Paper IX
MDST 204: Design and Analysis of Algorithm

Course Objectives:

1. To learn the algorithms and to learn basic Algorithm Analysis techniques and understand the use of asymptotic notation.
2. To understand different design strategies and Greedy Method.
3. Understand classical problem and solutions and Learn a variety of useful algorithms
4. Understand classification of problems

Credits=4	**SEMESTER-I** **MDST 204: Design and Analysis of Algorithm**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Introduction**	**(15)**
	Definition of Algorithm & its characteristics, Recursive and Non-recursive Algorithms, Time & Space Complexity, Definitions of Asymptotic Notations, Insertion Sort (examples and time complexity), Heaps & Heap Sort (examples and time complexity) ,Divide and Conquer Strategy, Binary Search (recursive), Quick Sort, Merge sort	
Credit –1 UNIT II	**Greedy Method**	**(15)**
	Fractional Knapsack problem, Optimal Storage on Tapes, Huffman codes, Concept of Minimum Cost Spanning Tree, Prim's and Kruskal's Algorithm	

Credit –1 UNIT III	**Dynamic Programming**	**(15)**
	The General Method, Principle of Optimality, Matrix Chain Multiplication, 0/1 Knapsack Problem, Concept of Shortest Path, Single Source shortest path, Dijkstra's Algorithm, Bellman Ford Algorithm, Floyd- Warshall Algorithm, Travelling Salesperson Problem	
Credit –1 UNIT IV	**P Problem Classification**	**(15)**
	Basic Concepts: Deterministic Algorithm and Non deterministic, Definitions of P,NP, NP-Hard, NP-Complete problems, Cook's Theorem (Only Statement and Significance)	

Course: Students are able to

1. Understanding Algorithmic complexity and analyzing the same
2. Developing an understanding of various techniques and methods to design algorithms
3. Skill to make the algorithm and solve real-world problems
4. Analysis of traditional algorithms and apply to various problems.

Reference Books:

1. Ellis Horowitz, Sartaj Sahni &Sanguthevar Rajasekaran, "Computer algorithms",Silicon Pr Publication, 2007.
2. T. Cormen, C. Leiserson, & R. Rivest, "Introduction to Algorithms", MIT Press, 2009.
3. Steven Skiena,"The Algorithm Manual", Springer, 2010.
4. Jungnickel, "Graphs, Networks and Algorithms", Springer, 2012.

Paper X
MDST 205: Python Programming

Course Objectives:

1. Describe the basics of python programming.
2. Explain programming constructs and apply them to build and package python modules for reusability.
3. Use various data structures to gain suitable knowledge about their implementation.
4. Compare various file handling techniques and database interactions.

Credits=4	**SEMESTER-I** **MDST 205: 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(), readline() 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 a nd searching tables, Reading and storing config information on database, Programming using database connections	

Course outcomes –After successful completion of the course student will be able to do

1)Able to apply the principles python programming.
2)Write clear and effective python code.
3) Data Analysis using python libraries.
4) Error handling.

References:

1. Dusty Phillips, "Python 3 Object-oriented Programming Second Edition", Packt Publishing, 2015. Greg
2. Charles Dierbach ,"Introduction to Computer Science Using Python: A Computational Problem-Solving Focus",John Wiley & Sons,2013.
3. Jan van Eijck , Christina Unger, "Computational Semantics with Functional Programming", Cambridge University Press, 2012 .
4. Kenneth C. Louden, "Programming Languages: Principles and Practice", Course Technology Inc., 2011. Richard L. Halterman, "LEARNING TO PROGRAM WITH PYTHON", Southern Adventist University, 2011

Paper XI

MDSP 206: Lab Course on Mathematical Foundation for Data Science and AI for Data Science and Data Preparation Analysis

Credits=	SEMESTER-II **MDSP 206: Lab Course on Mathematical Foundation for Data Science and AI for Data Science and Data Preparation Analysis**	No. of hours per unit/ credits
	Group A **Group B** 1. Case study on Banking Fraud 2. Case study on Healthcare 3. Case study on Manufacturing Industry 4. Case study on Agriculture **5.** Case study on Marketing	
	Group C **Assignment 1** Choose a dataset from UCI Machine Learning repository (e.g. Cleveland). a) Compute and display summary statistics for each feature available in the	

	dataset. (eg. minimum, maximum, mean, range, standard deviation, variance and percentiles). Use a bar-graph to demonstrate your results. b) Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram. c)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. **Assignment 2** a) Take any dataset from UCI repository (like air quality dataset) and perform regression analysis on it. Demonstrate your results using appropriate visualization techniques for numerical and categorical features (e.g. histogram, scatter plot, heat map, box plot). b) Compute Eigen values and Eigen vectors for dataset in part a. Useful 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	

Paper XII

MDSP 207: Lab Course on Design and Analysis of Algorithms and Python Programming

Group A:

Learning Objectives:

1. To understand how to implement different algorithms.
2. To learn how to use different methods to solve real world Problems.
3. To learn how to program using Python language.
4. To understand implementation of python libraries.

Credits=	**SEMESTER-II** **MDSP 207: Lab Course on Design and Analysis of Algorithms and Python Programming**	**No. of hours per unit/ credits**
	List of Experiments: **GROUP A :** 1. Case Study on Sorting Problem 2. Case Study on Searching Algorithms 3. Case Study on Warshall's Algorithm 4. Case Study on Knapsack Problem 5. Case Study on Shortest Paths Algorithm 6. Case Study on Bellman Ford Algorithm 7. Case Study on Minimum Cost Spanning Tree 8. Case Study on All Pairs Shortest Paths **Group B:** 1. Practical on Strings and Lists 2. Practical on Conditional statements 3. Practical on Looping statements 4. Practical on Functions 5. Practical on Packages 6. Practical on Stacks, Queues, Tuples, Sets, Dictionaries 7. Practical on File Handling 8. Practical on Database Operations	