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 - II

M.Sc. Data Science

Syllabus

To be implemented from June, 2023 onward

- **OBJECTIVES:**

1. To create post-graduates with sound knowledge of Data Science , who can contribute towards recent advances in technology
2. To provide advanced and in-depth knowledge of data 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 Data Science 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 faced in Computer Science industry. To produce computer science professionals who can 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.

- **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 Data Science and AI in real life.
3. Student should avail detail knowledge of Data Science, Artificial Intelligence, Machine Learning, and Big 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.

- **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:

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

Courses Structure for Postgraduate Programme to be implemented from Academic year 2022-23 for Data 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
CCS (Any One)	Fundamentals of Data Science	4	4	-	60	40	100
MDST 104	Introduction to Programming	4	4		60	40	100
MDSP 105	Lab I: Statistical Foundation for Data Science and Programmingusing R	4	-	12	60	40	100
MDSP 106	Lab II: Based on 104 and 105	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 Analysis and Visualization	4	4	-	60	40	100
MDST 204	Digital Systems for Data Science	4	4	-	60	40	100
CCS (Any One)	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 andAI for Data Science and Data Preparation Analysis	4	-	12	60	40	100
MDSP 207	Lab IV: Based on 204 and 205	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 Processing	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	
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)

Department of Computer Science (Entire)

Scheme of Credit for

M.Sc. Data 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 Data 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 Data Science.

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 (Full Time)**

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	Data 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. SECOND 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 VII: MDST201	20	20	Practical Paper – XII : MDSP206	12	4
		Paper VIII: MDST202					
		Paper IX: MDST203			Practical Paper –XIII : MDSP207	12	4
		Paper X: MDST204					
		Paper XI: MDST205					

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

- **M. Sc. I**

Semester I

1. Statistical Foundation for Data Science
2. Programming using R
3. Distributed Database
4. Fundamentals of Data Science
4. Introduction of Programming

- **M. Sc. I**

Semester II

1. Mathematical Foundations for Data Science
2. AI for Data Science
3. Data Analysis and Visualization
4. Digital Systems for Data Science
4. Design and Analysis of Algorithm

- **M. Sc. II**

Semester III

1. Big Data Analytics
2.Data Storage Technologies & Networking
3.Image Processing
4.Machine Learning
5. Deep Learning

- **M. Sc. II**

Semester IV

1. GPU computing
2. Recommended System
3. LAB VII: GPU Computing and Recommender System
4. LAB VIII: Internship Program (Industrial Training)

MDST/Pxyz–

M M.Sc.
DS Data Science
T Theory
P Practical
x 1 to 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**

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.
6. Laura Igual and Santi Segui, Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications,Springer; 1st ed. 2017 edition
7. Database Systems: A Practical Approach to Design, Implementation and Management- Thomas Connolly, Carolyn Begg, Pearson Publisher, 4th Edition.
8. Database Management Systems - Raghu Ramakrishnan and Johannes Gehrke, McGraw-Hill Education publisher, illustrated Edition,2003,ISBN0072465638, 9780072465631
9. Carlo Zaniolo, Stefano Ceri, Christos Faloustsos, R.T.Snodgrass, V.S.Subrahmanian, "Advanced Database Systems", Morgan Kaufman, 1997
10. Ken Black, 2013, *Business Statistics*, New Delhi, Wiley.
11. Lee, Cheng. et al., 2013, *Statistics for Business and Financial Economics*, New York: Heidelberg Dordrecht.
12. Anderson, David R., Thomas A. Williams and Dennis J. Sweeney, 2012, *Statistics for Business and Economics*, New Delhi: South Western.
13. Waller, Derek, 2008, *Statistics for Business*, London: BH Publications.
14. Dusty Phillips, "Python 3 Object-oriented Programming Second Edition", Packt Publishing, 2015. Greg
15. Charles Dierbach ,"Introduction to Computer Science Using Python: A Computational Problem-Solving Focus",John Wiley & Sons,2013.
16. Jan van Eijck , Christina Unger, "Computational Semantics with Functional Programming", Cambridge University Press, 2012 .
17. Kenneth C. Louden, "Programming Languages: Principles and Practice", Course Technology Inc., 2011. Richard L. Halterman, "LEARNING TO PROGRAM WITH PYTHON", Southern Adventist University, 2011
18. Luger G.F. and Stubblefield W.A. (2008). Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition
19. Nilsson Nils J , "Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4
20. Patrick Henry Winston, "Artificial Intelligence", Addison-Wesley Publishing Company, ISBN:0-201-53377-4
21. Dusty Phillips, "Python 3 Object-oriented Programming Second Edition", Packt Publishing, 2015. Greg
22. Charles Dierbach ,"Introduction to Computer Science Using Python: A Computational Problem-Solving Focus",John Wiley & Sons,2013.

23. Jan van Eijck , Christina Unger, "Computational Semantics with Functional Programming", Cambridge University Press, 2012 .
24. Kenneth C. Louden, "Programming Languages: Principles and Practice", Course Technology Inc., 2011. Richard L. Halterman, "LEARNING TO PROGRAM WITH PYTHON", Southern Adventist University, 2011
25. . Ellis Horowitz, Sartaj Sahni &Sanguthevar Rajasekaran, "Computer algorithms",Silicon Pr Publication, 2007.
26. 2. T. Cormen, C. Leiserson, & R. Rivest, "Introduction to Algorithms", MIT Press, 2009.
27. 3. Steven Skiena,"The Algorithm Manual", Springer, 2010.
28. 4. Jungnickel, "Graphs, Networks and Algorithms", Springer, 2012.
29. Luger G.F. and Stubblefield W.A. (2008). Artificial Intelligence: Structures and strategies for Complex Problem Solving. Addison Wesley, 6th edition
30. Nilsson Nils J , "Artificial Intelligence: A new Synthesis, Morgan Kaufmann Publishers Inc. San Francisco, CA, ISBN: 978-1-55-860467-4
31. Patrick Henry Winston, "Artificial Intelligence", Addison-Wesley Publishing Company, ISBN:0-201-53377-4

A. JOURNALS AND 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, LCD Projectors, E- Podium, Smart Board, Document Camera, Visualizer

C. LABORATORY EQUIPMENTS:

1. Soft Computing Tools –SCILAB,MATLAB
2. Tableau Software
3. R Software, Py Charm
4. Anaconda, Hadoop, MongoDB, NOSql

,

SEMESTER III
PAPER I

MDST 301 : Big Data Analytics

Course Objectives: Student will able to :-

1. To understand the big data concepts and big data analytics lifecycle
2. To understand the big data analytics algorithms and tools
3. To understand the importance of big data visualization tools and techniques
4. To get acquainted with advancements in tools and techniques used for big data analytics

Credits=4	**SEMESTER-I** **MDST 301 : Big Data Analytics**	**No. of hours per unit/ credits**
Credit –I UNIT I	Basics of Big Data	**(15)**
	Big data: characteristics, types, sources, architectures, Data analysis process, Data analytics lifecycle, Preprocessing data, Market and Business Drivers for Big Data Analytics, Business Problems Suited to Big Data Analytics	
Credit –1 UNIT II	Technologies for big data analytics Hadoop	**(15)**
	Distributed and Parallel Computing for Big Data, Cloud Computing and Big Data, In-Memory Computing Technology for Big Data, Introduction to Hadoop, HDFS, MapReduce, YARN, HBase, Combining HDFS and HBase Hadoop ecosystem: Sqoop, Impala, Apache Flume, Pig, Hive, Data transformation and analysis using Pig, Data analysis using Hive and Impala, Mahout, Oozie, Zookeeper etc.	
Credit –1 UNIT III	Programming languages for Big data analytics	**(15)**
	Big data analytics with PySpark: Python and Apache Spark Big data analytics with RHadoop: R and Hadoop, Text mining in RHadoop, Data mining in Hive, Data Analysis MapReduce techniques using RHadoop	
Credit –1 UNIT IV	Visualization techniques and tools for big data	**(15)**
	Visualizing Big Data, Importance of data visualization, Challenges, Need for advanced visualization techniques, Tools used in data visualization, Big Data Visualization with R/Python/Tableau/other tools.	

Course Outcomes: Student should be able to :-

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

Reference Books:

Text Books:

1. David V Thiel, "Research Methods- for Engineers", Cambridge University Press, ISBN:978-1-107-61019-4

2. Kothari C.R., "Research Methodology. New Age International, 2004, 2nd Ed; ISBN:13: 978-81- 224-1522-3**.**

Reference Books:

1. Caroline Whitbeck, "Ethics in Engineering Practice and Research", 2nd Ed., Cambridge University Press; ISBN :978-1-107-66847-8

2. Gordana DODIG-CRNKOVIC, "Scientific Methods in Computer Science", Department of Computer Science Malardalen University, Vasteas, Sweden; ISBN: 91-26-97860-1

Paper II

MDST 302: Data Storage Technologies and Networks

Course Objectives: Student should will be able to

1. To understand storage systems
2. To learn data storage technologies
3. To understand storage networking fundamentals
4. To learn storage networking technologies
5. To acquaint learner with knowledge of how to secure storage infrastructure

Credits=4	SEMESTER-I MDST 302: Data Storage Technologies and Networks	No. of hours per unit/ credits
Credit –I UNIT I	Introduction to storage system	(15)
	Introduction to Information Storage - Information Storage, Data, Types of Data, Big Data, Information, Storage, Evolution of Storage Architecture, Data Center Infrastructure- Core Elements of a Data Center, Key Characteristics of a Data Center, Managing a Data Center,Data Center Environment – Application, Database Management System (DBMS), Host (Compute), Connectivity, Storage, Disk Drive Components, Disk Drive Performance, Host Access to Data, Direct-Attached Storage, Storage Design Based on Application, Disk Native Command Queuing, Introduction to Flash Drives	
Credit –1 UNIT II	Intelligent Storage Systems and Virtualization	(15)
	Intelligent Storage System- Front end, Cache, Back End, Physical disk, Storage Provisioning- Traditional Storage Provisioning, Comparison between Virtual and Traditional Storage Provisioning, Types of Intelligent Storage Systems High-End Storage Systems, Midrange Storage Systems, RAID Server and I/O Architectures, Storage Hierarchy, From Bits to Bytes, Disk Storage Fundamentals, How write and read from a Storage Device, Storage Sharing vs. Data Sharing Different Types of Storage, I/O Connectivity and Networking Fundamentals, IT Clouds, Virtualization: Servers, Storage, and Networking, Virtualization and Storage Services, Data and Storage Access	
Credit –1 UNIT III	Storage Networking Technologies	(15)
	Fibre Channel Storage Area Networks - Fibre Channel: Overview, The SAN and Its Evolution, Components of FC SAN, FC Connectivity, Switched Fabric Ports, Fibre Channel Architecture- Fibre Channel Protocol Stack, Fibre Channel Addressing, World Wide Names, FC Frame, Structure and Organization of FC Data, Flow Control, Classes of Service, Zoning, FC SAN Topologies IP SAN – iSCSI- Components of iSCSI, iSCSI Host Connectivity, iSCSI Topologies, iSCSI	

	Protocol Stack, Stack,Introduction to NAS, Benefits, File Systems and Network File Sharing- Accessing a File System, Network File Sharing, Components of NAS- , NAS I/O Operation, NAS Implementations- Unified NAS, Unified NAS Connectivity, Gateway NAS, Gateway NAS Connectivity, Scale-Out NAS Connectivity, NAS File-Sharing Protocols – NFS, CIFS	
Credit –1 UNIT IV	Securing the Storage Infrastructure	**(15)**
	Information Security Framework, Risk Triad, Storage Security Domains – Securing application access domain, securing management access domain, Security Implementations in Storage Networking- FC SAN, NAS, IP Scan, Securing Storage Infrastructure in Virtualized and Cloud Environments – Security concerns, Security measures, Data Security	

Course Outcomes-

After the successful completion of this module, students will be able to:

1. Describe storage system architecture, its elements, and characteristics.
2. Compare the intelligent storage systems and select one for a storage application.
3. Demonstrate storage virtualization using Xen or KVM
4. Demonstrate the functioning of SAN and NAS using open-source simulators.
5. Describe the mechanisms to secure storage infrastructure.

Reference Books:

Textbooks:

1. “Information storage and management”, EMC Education Services, 2nd edition, SAGE Publication

2. “Cloud and Virtual Data Storage Networking”, Greg Schulz, CRC Press Reference Books: 1. “Storage Networks:

Reference Book

1. The Complete Reference, Robert Spalding”, Publisher: McGraw-Hill Osborne Media ISBN: 0072224762, 9780072224764 2.

2. “Storage area network essentials”, Richard Barker, Paul Massiglia, Wiley

Paper III

MDST 303: Image Processing

Course Objectives: Student should will be able to

1. Review the fundamental concepts of a digital image processing system.
2. Analyze images in the frequency domain using various transforms.
3. Evaluate the techniques for image enhancement and image restoration.
4. Categorize various compression techniques.
5. Interpret Image compression standards.
6. Interpret image segmentation and representation techniques.

Credits=4	**SEMESTER-I** **MDST 303:** Image Processing	**No. of hours per unit/ credits**
Credit –I UNIT I	**Introduction**	**(15)**
	Introduction: Digital Image Processing, Origins of Digital Image Processing, Applications and Examples of Digital Image Processing, Fundamental Steps in Digital Image Processing, Components of an Image Processing System, **Digital Image Fundamentals:** Elements of Visual Perception, Light and the Electromagnetic Spectrum, Image Sensing and Acquisition, Image Sampling and Quantization, Basic Relationships Between Pixels, Basic Mathematical Tools Used in Digital Image Processing,	
Credit –1 UNIT II	**Filtering and restoration**	**(15)**
	Filtering in the Frequency Domain: Background, Preliminary Concepts, Sampling and the Fourier Transform of Sampled Functions, Properties of the 2-D DFT and IDFT, Basics of Filtering in the Frequency Domain, Image Smoothing Using Lowpass Frequency Domain Filters, Image Sharpening Using Highpass Filters, Selective Filtering, Fast Fourier Transform **Image Restoration and Reconstruction:** A Model of the Image Degradation/Restoration Process, Noise Models, Spatial Filtering, Periodic Noise Reduction Using Frequency Domain Filtering, Linear, Inverse Filtering, Minimum Mean ,Square Error (Wiener) Filtering, Constrained Least Squares Filtering, Geometric Mean Filter, Image Reconstruction from Projections	
Credit –1 UNIT III	**Wavelet and Other Image Transforms**	**(15)**
	Wavelet and Other Image Transforms: Preliminaries, Matrix-based Transforms, Correlation, Basis Functions in the Time-Frequency Plane, Basis Images, Fourier-Related Transforms, Walsh-Hadamard Transforms, Slant Transform, Haar Transform, Wavelet Transforms **Color Image Processing:** Color Fundamentals, Color Models, Pseudocolor Image Processing, Full-Color Image Processing, Color Transformations, Color Image Smoothing and Sharpening, Using Color in Image Segmentation, Noise in Color Images, Color Image Compression.	
Credit –1 UNIT IV	**Image Compression ,Watermarking and Morphological Image Processing**	**(15)**

	Image Compression and Watermarking: Fundamentals, Huffman Coding, Golomb Coding, Arithmetic Coding, LZW Coding, Run-length Coding, Symbol-based Coding, 8 Bit-plane Coding, Block Transform Coding, Predictive Coding, Wavelet Coding, Digital Image Watermarking **Morphological Image Processing:** Preliminaries, Erosion and Dilation, Opening and Closing, The Hit-or-Miss Transform, Morphological Algorithms, Morphological Reconstruction¸ Morphological Operations on Binary Images	

Course Outcomes:

Student should will be able to

1. Understand the relevant aspects of digital image representation and their practical implications.
2. Have the ability to design pointwise intensity transformations to meet stated specifications.
3. Understand 2-D convolution, the 2-D DFT, and have the abitilty to design systems using these concepts.
4. Have a command of basic image restoration techniques.
5. Understand the role of alternative color spaces, and the design requirements leading to choices of color space.
6. Appreciate the utility of wavelet decompositions and their role in image processing systems.
7. Have an understanding of the underlying mechanisms of image compression, and the ability to design systems using standard

Reference Books:

1. Digital Image Processing Gonzalez and Woods Pearson/Prentice Hall Fourth 2018
2. Fundamentals of DigitalImage Processing A K. Jain PHI
3. The Image Processing Handbook J. C. Russ CRC Fifth 2010

PAPER IV

MDST 304 : Machine Learning

Course Objectives: Student should will be able to

1. To understand Human learning aspects
2. To learn the primitives in learning process by computer
3. To Understand nature of problems solved with Machine Learning
4. To acquaint with the basic concepts and techniques of Machine Learning.
5. To learn the means for categorization of the information

Credits=4	**SEMESTER-I** **MDST 304 :** Machine Learning	**No. of hours per unit/ credits**
Credit –I UNIT I	Machine Learning Concepts	**(15)**
	Introduction to Machine Learning, Machine Learning applications, Types of learning: Supervised, Unsupervised and semi-supervised, reinforcement learning techniques, Models of Machine learning: Geometric model, Probabilistic Models, Logical Models, Grouping and grading models, Parametric and nonparametric models, Predictive and descriptive learning, Classification concepts, Binary and multi-class classification	
Credit –1 UNIT II	Learning Theory	**(15)**
	Features: Feature Extraction, Feature Construction and Transformation, Feature Selection, Dimensionality Reduction: Subset selection, the Curse of dimensionality, Principle Components analysis, Independent Component analysis, Factor analysis, Multidimensional scaling, Linear discriminant analysis, Bias/Variance tradeoff, Union and chernoff / Hoeffding bounds, VC dimension, Probably Approximately Correct (PAC) learning, Concept learning, the hypothesis space, Least general generalization, Internal disjunction, Paths through	

	the hypothesis space, model Evaluation and selection	
Credit –1 UNIT III	Logical, Grouping And Grading Models	**(15)**
	Decision Tree Representation, Alternative measures for selecting attributes, Decision tree algorithm: ID3, Minimum Description length decision trees, Ranking and probability estimation trees, Regression trees, Clustering trees, Rule learning for subgroup discovery, Association rule mining, Distance based clustering K-means algorithm, Choosing number of clusters, Clustering around medoids – silhouettes, Hierarchical clustering, Ensemble methods: Bagging and Boosting	
Credit –1 UNIT IV	Probabilistic Models	**(15)**
	Uncertainty, Normal distribution and its geometric interpretations, Baye's theorem, Naïve Bayes Classifier, Bayesian network, Discriminative learning with maximum likelihood, Probabilistic models with hidden 30 variables, Hidden Markov model, Expectation Maximization methods, Gaussian Mixtures and compression based models	

Course Outcomes:

Student should will be able to

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

Text Books:

1. Peter Flach, Machine Learning: The Art and Science of Algorithms that make sense of data, Cambridge University Press, 1st Edition, 2012, ISBN No.: 978-1-316-50611-0
2. Ethem Alpaydin, Introduction to Machine Learning, PHI, 2nd edition, 2013, 978-0-262-01243-0
3. Kevin Murphy, Machine Learning: a Probabilistic Approach, MIT Press, 1st Edition, 2012, ISBN No.: 978-0262-30616-4

Reference Books:

1. C.M. Bishop, Pattern Recognition and Machine learning, Springer, 1st Edition, 2013, ISBN No.: 978- 81-322-0906-5
2. Hastie, Tibshirani, Friedman, Introduction to statistical machine learning with applications in R, Springer, 2nd Edition, 2013, ISBN No.: 978-1-4614-7138-7
3. Tom Mitchell, Machine Learning, McGraw Hill, 1997, 0-07-042807-7
4. Parag Kulkarni, Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEE Press, 2012, 978-0-470-91999-6
5. M. F. Der, L. K. Saul, S. Savage, and G. M. Voelker (2014). Knock it off: profiling the online storefronts of counterfeit merchandise. In Proceedings of the Twentieth ACM Conference on Knowledge Discovery and Data Mining (KDD-14), pages 1759-1768. New York, NY.
6. J. T. Ma, L. K. Saul, S. Savage, and G. M. Voelker (2011). Learning to detect malicious URLs. ACM Transactions on Intelligent Systems and Technology 2(3), pages 30:1-24.
7. D.-K. Kim, G. M. Voelker, and L. K. Saul (2013). A variational approximation for topic modeling of hierarchical corpora. To appear in Proceedings of the 30th International Conference on Machine Learning (ICML-13). Atlanta, GA.
8. M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker (2010). Beyond heuristics: learning to classify vulnerabilities and predict exploits. In Proceedings of the Sixteenth ACM Conference on Knowledge Discovery

and Data Mining (KDD-10), pages 105-113. Washington, DC

PAPER V

MDST 305 : Deep Learning

Course Objectives:

1. To introduce major deep learning algorithms
2. To introduce optimization techniques to training deep neural networks
3. Learn regularization techniques to training deep neural networks
4. To introduce Convolution Neural Networks and its applications
5. Learn auto encoders and its applications 6. Learn deep recurrent and memory networks

Credits=4	**SEMESTER-I** **MDST 305 -** Deep Learning	**No. of hours per unit/ credits**
Credit –I UNIT I	Introduction to deep learning	**(15)**
	Perceptron's, Perceptron Learning Algorithm, Sigmoid Neuron, Shallow neural networks, Deep neural networks, Feedforward Neural networks, Gradient descent and the backpropagation algorithm	
Credit –1 UNIT II	Optimization and Deep Learning	**(15)**
	Learning Parameters of a feedforward neural network, the vanishing gradient problem, and ways to mitigate it, RelU Heuristics for avoiding bad local minima, Heuristics for faster training, Nesterov Accelerated GD, Stochastic GD, AdaGrad, RMSProp, Adam, Momentum.Adagrad, Principal Component Analysis and its interpretations, Singular Value Decomposition	
Credit –1 UNIT III	Regularization Techniques and Convolutional Neural Networks	**(15)**
	Bias Variance Tradeoff, L2 regularization, Early stopping, Dataset augmentation, Parameter sharing and tying, Injecting noise at input, Ensemble methods, Dropout Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream, Deep Art, Fooling Convolutional Neural Networks	
Credit –1 UNIT IV	Deep Unsupervised Learning and Sequence Models:	**(15)**
	Autoencoders: standard, sparse, denoising, contractive, Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM RNN, LSTM, GRU models, Application to NLP, language models, machine translation, image captioning, video processing, visual question answering, video processing, learning from descriptions, Attention Mechanism, Attention over images, Self-attention	

Course Outcomes:

On completion of the course, learner will be able to–

1. Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains
2. To apply optimization techniques to training deep neural networks

3. Apply regularization techniques to improve the performance of deep learning algorithms.
4. Implement deep learning algorithms and solve real-world problems in computer vision.
5. Apply autoencoders to solve real world problems.
6. Implement deep learning algorithms and solve real-world problems in Natural Language Processing

Text Books:

1. **Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.**

Reference Books:

1. **Neural Networks: A Systematic Introduction, Raúl Rojas, 1996**
2. **Pattern Recognition and Machine Learning, Christopher Bishop, 2007**

MDSP 306: Lab Course Based on Big Data Analytics, Data Storage Technologies and Networks & Image Analytics

Course Objectives: Student should will be able to

1. To understand the big data analytics algorithms and tools
2. To understand the importance of big data visualization tools and techniques
3. To learn data storage technologies
4. To learn storage networking technologies
5. Review the fundamental concepts of a digital image processing system.
6. Analyze images in the frequency domain using various transforms

Credits= 4	**SEMESTER-I** **MDSP 306: MDST 301,MDST 302, MDST 303**	**No. of hours per unit/ credits**
	Group A MDST 301 **1. Assignment 1.** Demonstrate application of Apache spark to analyse streaming data from social media. (Installation of multi-node Hadoop as well as Spark is to be done by students.) **2. Assignment 2.** Take any text or image dataset (e.g. Stanford Sentiment Treebank, Sentiment140, Amazon Product data) and perform analysis on it. Useful links: 1. https://nlp.stanford.edu 2. http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip 3. https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews 4. https://snap.stanford.edu/data/amazon/productGraph/	
	Group B MDST 302 1. Demonstration of SAN 2. Demonstration of NAS 3. Demonstration of iSCSI 4. Demonstration of OSD	
	Group B MDST 303 **1.** Computation of Mean, Standard Deviation, Correlation coefficient of the given Image **2.** Implementation of Image Smoothening Filters(Mean and Median filtering of an	

	Image) 9. Implementation of image sharpening filters and Edge Detection using Gradient Filters **3.** Implementation of image restoring techniques	

Course outcomes –

After the successful completion of this module, students will be able to:

1. Demonstrate the use of Hadoop and its ecosystem elements to analyze big data.
2. Describe storage system architecture, its elements, and characteristics
3. Understand the relevant aspects of digital image representation and their practical implications
4. Have a command of basic image restoration techniques

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.

MDSP 307: Lab Course Based on Machine Learning and Deep Learning

Course Objectives: Student should will be able to

1. To introduce major deep learning algorithms
2. To introduce optimization techniques to training deep neural networks
3. To Understand nature of problems solved with Machine Learning
4. To acquaint with the basic concepts and techniques of Machine Learning

Credits=4	SEMESTER-III MDSP 307: Machine Learning and Deep Learning	No. of hours per unit/ credits
	Group A: Machine Learning **Assignment 1.** a) Using appropriate dataset from UCI machine learning repository design a decision tree. Implement two different decision tree algorithms. Find the root node of the decision tree. b) Extract confusion matrix from the test results. Compare the performance of the two decision tree algorithms in terms of at least six relevant measures. c) Now, classify the data using three decision tree algorithms from the Weka tool and compare the performance of your implementations with the results from the Weka tool. **Assignment 2.** 1. Implement k-NN classifier to classify a standard dataset (from UCI machine learning repository). Use Java/Python/R for implementation. Test the performance for various values of k. Now, classify the same dataset using distance-weighted k-NN and Locally weighted averaging methods. Compare the performance on at least six standard performance measures.	
	Group B: 1. Demonstration of Perceptron Learning Algorithm 2. Demonstration of Gradient descent algorithm 3. Demonstration of backpropagation algorithm	

	4. Demonstration of Accelerated GD algorithm 5. Demonstration of Convolutional Neural Networks 6. Demonstration of Autoencoders	

Course outcomes –

After the successful completion of this module, students will be able to:

1) Acquire fundamental knowledge of learning theory
2) Design and evaluate various machine learning algorithms
3) Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains
4) To apply optimization techniques to training deep neural networks

SEMESTER IV
Paper VI
MDST 401: GPU computing

Course Objectives:

1. To understand the different approaches of parallel programming.
2. To study massively parallel computing hardware and programming models.
3. To be conversant with GPGPU programming with CUDA.
4. To develop parallel programs in heterogeneous environments with OpenCL.
5. To understand machine learning using GPU.

Credits=4	**SEMESTER-VI** **MDST 401: GPU computing**	**No. of hours per unit/ credits**
Credit –I UNIT I	**Understanding Parallelism with GPUs.**	**(15)**
	Review of traditional computer architecture – basic five stage RISC pipeline, cache memory, register file, SIMD instructions, and GPU architectures - streaming multi processors, cache hierarchy, the graphics pipeline, parallel programming languages and models. Understanding Parallelism with GPUs.	
Credit –1 UNIT II	**Grids, Blocks, and Threads**	**(15)**
	Grids, Blocks, and Threads Introduction to Data Parallelism and CUDA C, Data-Parallel Execution Model, CUDA Memories-Memory types and memory Access Efficiency, Performance Considerations-Warps ,Thread Execution, Global Memory Bandwidth, Dynamic Partitioning of Execution Resources, Instruction Mix and Thread Granularity, the CUDA extensions to the C language, and the basic programming/debugging tools	
Credit –1 UNIT III	**Memory Handling and Synchronization**	**(15)**
	Memory Handling with CUDA- The basic CUDA memory/threading model, floating-point considerations in parallel computing and common data-parallel programming patterns needed to develop a high-performance parallel application.Programs for concurrent Data Structure such as Worklists, Linked-lists. Synchronization across CPU and GPU.	
Credit –1 UNIT IV	**Designing GPU-Based Systems.**	**(15)**
	Parallel Programming and Computational Thinking, MPI-CUDA programming in a heterogeneous computing cluster. Dynamic parallelism, Unified Virtual Memory, CPU vs GPU, GPU hardware overview, GPU memory architecture, GPU properties, compute capability of GPU, multi- GPU solution. Multi-GPU processing, Peer access, Heterogeneous processing	

Course Outcomes: Student should will be able to

1. **Analyze** and measure performance of modern parallel computing systems.
2. **Design and Implement** parallel programs on GPUs.
3. **Develop** a high-performance parallel application in CUDA.
4. **Build** parallel programming logic on current system architectures using OpenCL.
5. **Implement** machine learning using GPU .

Text Books:

1. "Programming Massively Parallel Processors" - David Kirk and Wen-meiHwu
2. " Heterogeneous Computing with OpenCL" -- Benedict Gaster,Lee Howes, David R. Kaeli
3. Hands-On GPU Computing with Python: (Kindle Edition) by Bandyopadhyay, Avimanyu

Reference Books:

1) Shane Cook, "CUDA Programming: A Developer's Guide to Parallel Computing with GPUs", Morgan Kaufmann Publishers Inc. San Francisco, CA, USA 2013 ISBN: 9780124159884

2) CUDA BY EXAMPLE by Jason Sanders,Edvard Kandrot

Paper VII

MDST 402: Recommended System

Course Objectives:

1. Describe the purpose of recommendation systems.
2. Understand the components of a recommendation system including candidate generation, scoring ,and re-ranking.
3. Familiarity with linear algebra

.

Credits=4	**SEMESTER-I** **MDST 402:** Recommender System	**No. of hours per unit/ credits**
Credit –I UNIT I	**Introduction**	**(15)**
	Recommender system functions, Linear Algebra notation: Matrix addition, Multiplication, transposition, and inverses; covariance matrices, Understanding ratings, Applications of recommendation systems, Issues with recommender system.	
Credit –1 UNIT II	**Collaborative Filtering**	**(15)**
	User-based nearest neighbour recommendation, Item-based nearest neighbor recommendation, Model based and pre-processing based approaches, Attacks on collaborative recommender systems.	
Credit –1 UNIT III	**Content & knowledge based recommendation**	**(15)**
	High level architecture of content-based systems, Advantages and drawbacks of content based filtering, Item profiles, Discovering features of documents, Obtaining item features from tags, Representing item profiles, Methods for learning user profiles, Similarity based retrieval, Classification algorithms. Knowledge based recommendation: Knowledge representation and reasoning, Constraint based recommenders, Case based recommenders.	
Credit –1 UNIT IV	**Evaluating Recommender System and communities**	**(15)**
	Introduction, General properties of evaluation research, Evaluation designs, Evaluation on historical datasets, Error metrics, Decision-Support metrics, User-Centred metrics. Communities, collaboration and recommender systems in personalized web search, Social tagging recommender systems, Trust and recommendations, Group recommender systems	

Course Outcomes:

On completion of the course, learner will be able to

1:Predict the "rating" or "preference" a user would give to an item

2:Product recommenders for services such as Amazon, or content recommenders for social media platforms such as Facebook and Twitter

3:Explain a variety of approaches for building recommender systems

4:Describe system evaluation methods from both algorithmic and users' perspectives

Text Books:

1. Francesco Ricci, Lior Rokach, and Bracha Shapira, eds. Recommender Systems Handbook, 2nd edition.Spring US, 2015.

2. Jannach D., Zanker M., Fel FeringA., Recommender Systems: An Introduction, Cambridge University Press, 2011.

Reference Books:

1. Charu C. Aggarwal, Recommender Systems: The Textbook, 1/e, Springer, 2016.

2. Ricci F., Rokach L., Shapira D., Kantor B.P., Recommender Systems Handbook,Springer, 2011.

3. Manouselis N., Drachsler H., Verbert K., Duval E., Recommender Systems for Learning, Springer, 2013

MDSP 403: Lab Course based on GPU Computing and Recommender System

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

MDST 403: Lab Course on GPU Computing and Recommender System

Credits=	SEMESTER-IV **Lab Course on** GPU Computing and Recommender System	No. of hoursper unit/ credits
	Group A GPU computing 1. Demonstration of Architecture of CUDA 2. Demonstration of CUDA programming /debugging tools 3. Demonstration of MPI-CUDA	
	Group B Recommender System 1. Case study of E-commerce Recommender System 2. Case study of Social Media Recommender System 3. Case study of Travel Recommender System 4. Case study of Health Care Recommender System 5. Case study of Education Recommender System	

MDSP 404 LAB VIII: Internship Program (Industrial Training)