Rayat Shikshan Sanstha's

YASHAVANTRAO CHAVAN INSTITUTE OF SCIENCE, SATARA

(AUTONOMOUS)

Lead college

of

Karmaveer Bhaurao Patil University, Satara

Syllabus For
Master of Science

Part - II

DATA SCIENCE

Syllabus to be implemented w.e.f. June 2024

as Per NEP-2020

Preamble:

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

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

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

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

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

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

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

Credit Framework for M.Sc. II

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

Structure of Course: M.Sc. – II

Semester – III

Nature of the Course	Course Code	Name of the Course
	Discipline Specific Courses (Mandatory)	
Theory	MDST 531	Big Data Analytics
	MDST 532	Data Storage Technologies & Networking
	MDST 533	Image Processing
	Discipline Specific Elective (Choose Any one among two)	
	MDST 534 E-I	Machine Learning
	MDST 534 E-II	Data Engineering
	MDST 535	Research Project
Practical	MDSP 536	Practical Course I: Lab, I Based On(MDST 531,532)
	MDSP 537	Practical Course II: Lab II Based On (MDST 533,534 E1)

Semester IV

Nature of the Course	Course Code	Name of the Course
	Discipline Specific Courses (Mandatory)	
Theory	MDST 541	Deep Learning
	MDST 542	GPU Computing
	MDST 543	Recommender System
	Discipline Specific Elective (Choose Any one among two)	
	MDST 544 E-I	Cloud Computing
	MDST 544 E-II	Social Media Analytics
	MDST 545	OJT
Practical	MDST 546	Practical Course I: Lab, I Based On (MDST 541,542)
	MDST 547	Practical Course II: Lab, II Based On (MDST 543,544)

Project Academic Project is divided into 4 phases.

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

SEMESTER III
PAPER I
MDST 531: Big Data Analytics

Course Objectives: Student will be able to …

1. Understand the big data concepts and big data analytics lifecycle
2. Be aware of the big data analytics algorithms and tools
3. Comprehend the importance of big data visualization tools and techniques
4. Get acquainted with advancements in tools and techniques used for big data analytics.

Credits=4	**MDST 531: Big Data Analytics**	**No. of hours per unit= 60**
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	
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.	
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	
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: After completion of syllabus, student will 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. Indicate 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 :

1. *T. White,* "Hadoop the Definitive Guide", O'Reilly Publications, Fourth Edition, 2015
2. *D. Deroos, P. C. Zikopoulos, R. B. Melnyk, B. Brown,* R. Coss, "Hadoop for Dummies", Wiley Publications, 2014
3. *J. Huruwitz, A. Nugent, F. Halper, M. Kaufman,* "Big data for dummies", John Wiley & Sons, Inc. (2013)
4. *R. D. Schneider*, "Hadoop for Dummies", John Wiley & Sons, Inc. (2012)
5. *P. Zikopoulos*, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill, 2012 Chuck Lam, "Hadoop in Action", Dreamtech Publications, 2010
6. *C. Whitbeck*, "Ethics in Engineering Practice and Research", 2nd Ed., Cambridge University Press; ISBN :978-1-107-66847-8, (1998).

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

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

Paper II

MDST 532: Data Storage Technologies and Networks

Course Objectives: Student will be able to…

1. Understand storage systems
2. Learn data storage technologies
3. Acknowledge storage networking fundamentals
4. Acquire skill in storage networking technologies
5. Acquaint learners with knowledge of how to secure storage infrastructure.

Credits=4	**MDST 532: Data Storage Technologies and Networks**	**No. of hours per unit =60**
UNIT I	**Introduction to storage system**	**15**
	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, Storage, Disk Drive Components, Disk Drive Performance, Storage Design Based on Application, Introduction to Flash Drives	
UNIT II	**Intelligent Storage Systems and Virtualization**	**15**
	ISS- 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, Server and Storage I/O Fundamentals- Server and I/O Architectures, Storage Hierarchy, Disk Storage Fundamentals, Initiators and Targets, How write and read from a Storage Device, Storage Sharing , Data Sharing, I/O Connectivity and Networking Fundamentals, IT Clouds, Virtualization: Servers, Storage, and Networking, Virtualization and Storage Services, Data and Storage Access	

UNIT III	**Storage Networking Technologies**	**15**
	Fiber Channel Storage Area Networks - Fiber Channel: Overview, The SAN and Its Evolution, Components of FC SAN, FC Connectivity, Switched Fabric Ports, Fiber Channel Architecture- Fiber Channel Protocol Stack, Fiber Channel Addressing, 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	
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, Network Security	

Course Outcomes: After completion of syllabus, student will be able to:

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

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

Reference :

1.*P. S. Deshpande (Author), S. C. Sharma (Author), S. K. Peddoju* ,"Security and Data Storage Aspect in Cloud Computing" (Studies in Big Data, 52) 1st ed. 2019 Edition, Springer, 2019

2.*G. Santana*, Data Center Virtualization Fundamentals: Understanding Techniques and Designs for Highly Efficient Data Centers with Cisco Nexus, UCS, MDS, and Beyond, Cisco Press; 1 edition, 2013.

3.*G. Somasundaram, A Shrivastava*, Information Storage and Management, EMC Education Series, Wiley, Publishing Inc., 2011.

4.*S. N. Piramanayagam (Editor), T. C. Chong*, "Developments in Data Storage: Materials Perspective" 1st Edition, Kindle Edition (2011)

5.*G. Schulz*, "The Green and Virtual Data Center", CRC Press 1st Edition,2009.

6.*U. Troppens, W. Muller-Friedt, R. Erkens, and N. Haustein* "Storage Networks: The Complete Reference" Springer 1st Edition 2003

7.R. Barker, P. Massiglia, Wiley "Storage area network essentials" (2002)

Paper III
MDST 533: Image Processing

Course Objectives: Student will be able to…

1. Review the fundamental concepts of a digital image processing system.
2. Analyse 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. Clarify image segmentation and representation techniques.

Credits=4	**MDST 533: Image Processing**	**No. of hours per unit =60**
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,	
UNIT II	**Filtering and restoration**	**15**
	Filtering in the Frequency Domain: Background, Concepts, Sampling and the Fourier Transform of Sampled Functions, The Discrete Fourier Transform of One Variable, Extensions to Functions of Two Variables, 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 High pass Filters. **Image Restoration and Reconstruction:** A Model of the Image Degradation/Restoration Process, Noise Models, Restoration in the Presence of Noise Only Spatial Filtering, Periodic Noise Reduction Using Frequency Domain Filtering, Inverse Filtering, Minimum Mean, Square Error (Wiener) Filtering, Constrained Least Squares Filtering, Geometric Mean Filter, Image Reconstruction from Projections	
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 **Colour Image Processing:** Colour Fundamentals, Colour Models, Pseudocoelom Image Processing, Full-Colour Image Processing, Colour Transformations, Colour Image Smoothing and Sharpening, Using Colour in Image Segmentation, Noise in Colour Images, Colour Image Compression.	
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, Grayscale Morphology	

Course Outcomes: After completion of syllabus, student 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. Acknowledge 2-D convolution, the 2-D DFT, and have the ability to design systems using these concepts.
4. Have a command of basic image restoration techniques.
5. Know the role of alternative colour spaces, and the design requirements leading to choices of colour space.
6. Appreciate the utility of wavelet decompositions and their role in image processing systems.
7. Recognize the underlying mechanisms of image compression, and the ability to design systems using standard.

Reference Books:

1. *A. F. Villan* "Mastering OpenCV 4 with Python: A comprehensive guide to building computer vision applications with OpenCV and Python" Packt Publishing 2019.
2. Digital Image Processing Gonzalez and Woods Pearson/Prentice Hall Fourth 2018.
3. *S. Dey* "Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data" Packt Publishing 2018.
4. *R. Shanmugamani,* "Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Kera's" Packt Publishing 2018.
5. *W. McKinney* "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and I Python" O'Reilly Media 2017.
6. *A. Rosebrock* "Practical Python and OpenCV: An Introductory, Example Driven Guide to Image Processing and Computer Vision" PyImageSearch 2015.
7. *J. C. Russ*, The Image Processing Handbook CRC Fifth 2010.

PAPER IV
MDST 534: Machine Learning

Course Objectives: Student will be able to…

1. Understand Human learning aspects
2. Learn the primitives in learning process by computer
3. Comprehend nature of problems solved with Machine Learning
4. Acquaint with the basic concepts and techniques of Machine Learning.
5. Aware the means for categorization of the information

Credits=4	MDST 534: Machine Learning	No. of hours per unit =60
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	
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 / Hoefflin 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	
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	
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: After completion of syllabus, student will be able to:

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

Reference Books:

1. M. F. Der, L. K. Saul, S. Savage, and G. M. Voelker 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, (2014).

2. C.M. Bishop, Pattern Recognition and Machine learning, Springer, 1st Edition, ISBN No.: 978- 81-322-0906-5 (2013)

3. Hastie, Tibshirani, Friedman, Introduction to statistical machine learning with applications in R, Springer, 2nd Edition, ISBN No.: 978-1-4614-7138-7, (2013).

4. D.-K. Kim, G. M. Voelker, and L. K. Saul, 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. (2013).

5. J. T. Ma, L. K. Saul, S. Savage, and G. M. Voelker. Learning to detect malicious URLs. ACM Transactions on Intelligent Systems and Technology 2(3), pages 30:1-24, (2011).

6. Parag Kulkarni, Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEE Press, 978-0-470-91999-6, (2012).

7. M. Bozorgi, L. K. Saul, S. Savage, and G. M. Voelker. 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, (2010).

8. Tom Mitchell, Machine Learning, McGraw Hill, 0-07-042807-7, (1997).

9. *P. 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

10. *E. Alpaydin*, Introduction to Machine Learning, PHI, 2nd edition, 2013, 978-0-262-01243-0

11. *K. Murphy*, Machine Learning: a Probabilistic Approach, MIT Press, 1st Edition, 2012, ISBN No.: 978-0262-30616-4

MDST 534: Data Engineering

Course Objectives: Student will be able to…

1. Develop the skills of managing the data with respect to knowledge generation.
2. Provide the ability to design the data engineering process.
3. Propose the data reliability models
4. Define how to use Machine learning models.

Credits=4	MDST 534: Data Engineering	No. of hours per unit =60
UNIT I	**Data Engineering Concepts**	**15**
	Selecting Appropriate Storage Technologies: From Buisness Requirements to Storage Systems, Technical Aspects of Data, Types Of Structure, Schema Design Consideration. **Building and Operationalizing Storage Systems:** Cloud SQL, Cloud Spanner, Cloud Bigtable, Cloud Firestore, BigQuery, Cloud Memorystore, Cloud Storage, Unmanaged Databases.	
UNIT II	**Data Pipelines and Processing**	**15**
	Designing Data Pipelines: Overview Of Data Pipelines, GCP Pipeline Components, Migrating Hadoop and Spark To GCP. **Designing a Data Processing Solution:** Designing Infrastructure, Designing for Distributed Processing, Migrating a Data Warehouse. **Building and Operationalizing Processing Infrastructure:** Provisioning and Adjusting Processing Resources, Monitoring Processing Resources.	
UNIT III	**Designing for Security and Compliance, Reliability**	**15**
	Designing for Security and Compliance: Identity and Access Management with Cloud IAM, Using IAM with Storage and Processing Services, Data Security, Ensuring Privacy with the Data Loss Prevention API, Legal Compliance **Designing Databases for Reliability, Scalability, and Availability:** Designing Cloud Bigtable Databases for Scalability and Reliability, Designing Cloud Spanner Databases for Scalability and Reliability, Designing BigQuery Databases for Data Warehousing	
UNIT IV	**Deploying Machine Learning Pipelines**	**15**

	Understanding Data Operations for Flexibility and Portability: Cataloging and Discovery with Data Catalog, Data Preprocessing with Data prep, Visualizing with Data Studio, Exploring Data with Cloud Data lab, Orchestrating Workflows with Cloud Composer. **Deploying Machine Learning Pipelines:** Structure of ML Pipelines, GCP Options for Deploying Machine Learning Pipeline **Choosing Training and Serving Infrastructure:** Hardware Accelerators, Distributed and Single Machine Infrastructure, Edge Computing with GCP. **Measuring, Monitoring, and Troubleshooting Machine Learning Models:** Three Types of Machine Learning Algorithms, Deep Learning, Engineering Machine Learning Models, Common Sourcesof Error in Machine Learning Models.	

Course Outcomes: After completion of syllabus, student will be able to:

1. Building the storage system with appropriate data technologies
2. Designing the data pipelines and data flow
3. Processing the data infrastructure Investigate possible diagnostics by designing Databases for Reliability, Scalability, and
4. Availability, Understanding Data Operations for Flexibility
5. Training and measuring the serving Infrastructure for Machine Learning Models

Reference :

1. *L. Jimmy.* Data Engineering for Dummies. Wiley, 2018.
2. *W. Tom.* Hadoop: The Definitive Guide. O'Reilly Media, 2015.
3. *R. Eric, and R. Jim Wilson.* Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement. Pragmatic Bookshelf, 2012.
4. *Kelleher, D. John., and B Tierney*. Data Science. MIT Press, 2018.
5. *O. Cathy, and R. S. John.* Data on the Web: From Relations to Semi structured Data and XML. Morgan Kaufmann, 2006.
6. *K. Ralph, J. Caserta, M. Ross, and B. Becker.* The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley, 2013.
7. *M Nathan, and J. Warren.* Big Data: Principles and Best Practices of Scalable Realtime Data Systems. Manning Publications, 2015.
8. *G. Mathew, and A. Shikiar*. Data Engineering on Azure: Simplifying Big Data Architectures. A press, 2020.

MDST 535: Research Project

Course Objectives: Student will be able to …

1. Develop proficiency in data collection, cleaning, and preprocessing.
2. Apply statistical and machine learning techniques to analyze datasets.
3. Communicate findings effectively through data visualization and storytelling.
4. Gain experience in collaborative research and project management.
5. Enhance critical thinking and problem-solving skills in the context of data science.

Credits=4	MDST 535: Research Project	No. of hours per unit =60
UNIT I	**Exploratory Data Analysis (EDA)**	**15**
	Introduction to Exploratory Data Analysis (EDA), Descriptive Statistics for EDA, Univariate Analysis, Bivariate Analysis, Time Series Analysis for EDA, Data Transformation and Feature Engineering	
UNIT II	**Statistical Analysis and Hypothesis Testing**	**15**
	Formulating Research Hypotheses, Descriptive Statistics in Research, Probability Distributions in Research, One-Sample Hypothesis Testing, Two-Sample Hypothesis Testing, Analysis of Variance (ANOVA) in Research, Non-parametric Tests in Research, Regression Analysis in Research	
UNIT III	**Machine Learning for Data Exploration**	**15**
	Supervised Learning for Research Exploration, Unsupervised Learning Techniques for Research, Feature Engineering and Selection in Research, Model Evaluation and Validation in Research, Advanced Methods for Research Exploration, Natural Language Processing (NLP) for Research, Interpretability and Explain ability in Research Models, Real-world Applications in Research, Research Project Proposal	
UNIT IV	**Data Visualization & Report Writing and Communication**	**15**
	Principles of Effective Data Visualization, Choosing the Right Visualization Techniques, Tools for Data Visualization, Advanced Data Visualization Techniques, Storytelling with Data, Principles of Effective Report Writing in Research, Incorporating Visualizations into Reports, Ols for Report Writing, eating Effective Presentation Slides	

Course Outcomes: After completion of syllabus, student will be able to:

1. Articulate clear and well-defined research questions that can be addressed through data analysis.
2. Identify hypotheses to be tested or explored in the research project.
3. Apply appropriate data cleaning and preprocessing techniques to handle missing values, outliers, and inconsistencies.
4. Design and generate clear and compelling data visualizations that enhance understanding.
5. Communicate complex insights through graphical representation.

Reference Book:

1. *Q. Li,* Using R for Data Analysis in Social Sciences: A Research Project-Oriented Approach, UP USA,5 July 2018
2. *J. Weiner*, Why AI/Data Science Projects Fail: How to Avoid Project Pitfalls (Synthesis Lectures on Computation and Analytics) Springer International Publishing AG,18 December 2020
3. *N. Karacapilidis*, Mastering Data-Intensive Collaboration and Decision Making: Research and practical applications in the Dicode, Springer,3 September 2016
4. *R. Verma*, Data Diving: Beginner's Guide to Research Projects,22 December 2020

MDSP 536: Lab Course Based on
Big Data Analytics, Data Storage Technologies, and Networks & Image Analytics

Course Objectives: Student will be able to…

1. Understand the big data analytics algorithms and tools
2. Acknowledge the importance of big data visualization tools and techniques
3. Learn data storage technologies
4. Acquire a knowledge of storage networking technologies

Credits= 4	SEMESTER- III MDSP 536: MDST 531, MDST 532	No. of hours per unit =60
	Group A MDST 531 1. Predictive Maintenance in Manufacturing Using IoT Sensor Data. 2. Optimizing Supply Chain Logistics Using Web and Clickstream Data. 3. Real-Time Data Processing Using Apache Kafka and Hadoop. 4. Interactive Querying with Apache Impala. 5. Machine Learning Model Training with Apache Mahout. 6. Predictive Analytics with PySpark. 7. Data Mining with Hive and RHadoop. 8. Data Analysis with MapReduce Techniques using RHadoop. 9. Geospatial Visualization with Python for Big Data Analysis. 10. Network Analysis with Gephi for Big Data Visualization.	
	Group B MDST 532 1. Practical on Introduction to Storage System and Data Center Infrastructure 2. Experiment on Storage Design Based on Application 3. Practical on Introduction to Intelligent Storage Systems (ISS) and Storage Provisioning 4. Experiment on Server and Storage I/O Fundamentals 5. Practical on Storage Provisioning and Data Access 6. Experiment on Fiber Channel Storage Area Networks (SAN) 7. Practical on iSCSI Storage Networking 8. Experiment on NAS Implementations and Connectivity 9. Practical on Storage Security Testing and Assessment 10. Experiment on Security Implementations in Storage Networking	

Course Outcomes: After completion of syllabus, student 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. *T. White*, "Hadoop the Definitive Guide", O'Reilly Publications, Fourth Edition, 2015
2. *D. Deroos, P. C. Zikopoulos, R. B. Melnyk, B. Brown, R. Coss,* "Hadoop for Dummies", Wiley Publications, 2014
3. *J. Huruwitz, A. Nugent, F. Halper, M. Kaufman,* "Big data for dummies", John Wiley & Sons, Inc. (2013)
4. *S. N. Piramanayagam (Editor), T. C. Chong,* "Developments in Data Storage: Materials Perspective" 1st Edition, Kindle Edition (2011).

5.*G. Schulz,* "The Green and Virtual Data Center", CRC Press 1st Edition,2009.

6.*U. Troppens, W. M.-F., R. Erkens, and N. Haustein* "Storage Networks: The Complete Reference" Springer 1st Edition 2003

MDSP 537: Lab Course Based on
Image Analytics & Machine Learning

Course Objectives: Student will be able to…

1. Review the fundamental concepts of a digital image processing system.
2. Analyse images in the frequency domain using various transforms.
3. Understand nature of problems solved with Machine Learning
4. Acquaint yourself with the basic concepts and techniques of Machine Learning.

Credits= 4	SEMESTER-III MDSP 536: MDST 533, MDST 534 E1	No. of hours per unit =60
	Group A MDST 533 1. Practical on Understanding Digital Image Processing Basics. 2. Experiment on Applications and Examples of Digital Image Processing. 3. Practical on Experimenting with Image Sampling and Quantization 4. Experiment on Frequency Domain Filtering for Image Smoothing. 5. Practical on Inverse Filtering for Image Restoration 6. Experiment on Image Compression using Wavelet Transforms. 7. Practical on Edge Detection with Hear Transform 8. Experiment on Image Watermarking using Fourier-Related Transforms. 9. Practical on Digital Image Watermarking using DCT-based Embedding. 10. Experiment on Image Compression using Predictive Coding	
	Group B MDST 534 E1 1. Practical on Introduction to Machine Learning and Classification Concepts 2. Experiment on Machine Learning Applications and Probabilistic Models 3. Practical on Feature Engineering and Dimensionality Reduction 4. Experiment on Learning Theory and Model Evaluation 5. Practical on Model Selection and Bias/Variance Tradeoff 6. Experiment on Decision Trees and Logical Models 7. Practical on Clustering Techniques 8. Experiment on Ensemble Methods: Bagging and Boosting 9. Practical on Naïve Bayes Classifier and Bayesian Network 10. Experiment on Hidden Markov Models (HMMs) and Expectation Maximization (EM) Methods	

Course Outcomes: After completion of syllabus, student will be able to:

1. Understand the relevant aspects of digital image representation and their practical implications.
2. Have a command of basic image restoration techniques.
3. Use machine learning methods for multivariate data analysis in various scientific fields.
4. Choose and apply appropriate Machine Learning Techniques for analysis, forecasting, categorization, and

clustering of the data.

Reference :

1. *M. F. Der, L. K. Saul, S. Savage, and G. M. Voelker* 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, (2014).

2. *C.M. Bishop*, Pattern Recognition and Machine learning, Springer, 1st Edition, ISBN No.: 978- 81-322-0906-5 (2013)

3. *H. T. Friedman*, Introduction to statistical machine learning with applications in R, Springer, 2nd Edition, ISBN No.: 978-1-4614-7138-7, (2013).

4. *A. F. Villan* "Mastering OpenCV 4 with Python: A comprehensive guide to building computer vision applications with OpenCV and Python" Packt Publishing 2019

5. *S. Dey* "Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data" Packt Publishing 2018

SEMESTER IV

MDST 541: Deep Learning

Course Objectives: Student will be able to…

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

Credits=4	**MDST 541- Deep Learning**	**No. of hours per unit=60**
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	
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	

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	
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	

Course Outcomes: After completion of syllabus, student will be able to:

1. Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains.
2. 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. Autoencoders to solve real world problems.
6. Implement deep learning algorithms and solve real-world problems in Natural Language Processing

Reference Books:

1. *R. L. LaRose* "Discovering Statistics Using R" SAGE Publications Ltd (2016).
2. *H. Wickham and G. Grolemund* "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data" O'Reilly Media (2016).
3. *K. Black*, 2013, Business Statistics, New Delhi, Wiley,(2013).
4. *L. Cheng.* et al., Statistics for Business and Financial Economics, New York: Heidelberg Dordrecht, (2013)
5. *G. James, D. Witten, Tr. Hastie, and R. Tibshirani* "An Introduction to Statistical Learning: With Applications in R" Springer (2013).
6. *A. R. David., A. T. Williams and D. J. Sweeney*, Statistics for Business andEconomics, New Delhi: South Western, (2012).
7. *W. Derek*, Statistics for Business, London: BH Publications, (2008).

MDST 542: GPU computing

Course Objectives: Student will be able to…

1. Understand the different approaches of parallel programming.
2. Study massively parallel computing hardware and programming models.
3. Be conversant with GPGPU programming with CUDA.
4. Develop parallel programs in heterogeneous environments with OpenCL.
5. comprehend machine learning using GPU.

Credits=4	**MDST 542: GPU computing**	**No. of hours per unit =60**
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.	
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	
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.	
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: After completion of syllabus, student will be able to:

1. Analyse 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.

Textbooks:

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) *D. B. Kirk, W. W. Meri. Hwu, and R. Nassiry*, “Programming Massively Parallel Processors: A Hands-on Approach (3rd Edition)" Morgan Kaufmann (2016).
2) *J. Cheng, M. Grossman, and Ty McKercher* "Professional CUDA C Programming". Wiley (2014).
3) *S. Cook*, “CUDA Programming: A Developer's Guide to Parallel Computing with GPUs”, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ISBN: 9780124159884, (2013).
4) *I. Buck and J Clayton* "CUDA Handbook: A Comprehensive Guide to GPU Programming" Addison-Wesley Professional (2013).
5) *B. D. Kirk and Wen-mei W. Hwu* "Programming Massively Parallel Processors: A Hands-on Approach" Morgan Kaufmann, (2012).
6) *S Coo*k "CUDA Programming: A Developer's Guide to Parallel Computing with GPUs" Morgan Kaufmann, (2012).
7) *A. Abdelfattah, A Awad, and M. Zahran* "GPU Computing Gems Emerald Edition" Morgan Kaufmann, (2011).

Paper VII

MDST 543: Recommended System

Course Objectives: Student will be able to…

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	MDST 543: Recommender System	No. of hours per unit = 60
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.	
UNIT II	**Collaborative Filtering**	15
	User-based nearest neighbour recommendation, Item-based nearest neighbour recommendation, Model based and pre-processing based approaches, Attacks on collaborative recommender systems.	
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.	
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: After completion of syllabus, student 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
3. Social media platforms such as Facebook and Twitter
4. Explain a variety of approaches for building recommender systems.
5. Describe system evaluation methods from both algorithmic and users' perspectives

Reference Books:

1. *C. C. Aggarwal* "Recommender Systems: The Textbook" Springer, (2016).
2. *S. Berkovsky, T. Kuflik, F. Ricci* "Recommender Systems Handbook" Publisher: Springer, (2015).
3. *N. Manouselis, H. Drachsler, K. Verbert, E. Duval,* Recommender Systems for Learning, Springer, 2013
4. *J. A. Konstan, J. Riedl,*"Recommender Systems: An Introduction" Cambridge University Press, (2012).
5. *F.Ricci, L.Rokach, D. Shapira, B.P. Kantor,* Recommender Systems Handbook, Springer, 2011.
6. *D. Jannach, M. Zanker, A. Felfernig, G. Friedrich* "Recommender Systems: An Introduction" Cambridge University Press, (2010).
7. *H. R. Varian, P. Resnick, D. M. Pennock* "Recommender Systems" MIT Press 2010
8. *Alexander Tuzhilin, Daniel Billsus,* "Practical Approaches to Recommender Systems" Springer (2010).
9. *Michael J. Pazzani, Daniel Billsus* "Content-Based Recommendation Systems" Morgan & Claypool Publishers, (2007).
10. *F. Ricci, L. Rokach, and B. Shapira,* eds. Recommender Systems Handbook, 2nd edition. Spring US, 2015.
11. *D . Jannach, M. Zanker, A. F. Fering,* Recommender Systems: An Introduction, Cambridge University Press, 2011.

SEMESTER IV
MDST 544: Cloud Computing

Course Objectives: Student will be able to …

1. Define cloud computing and its key characteristics.
2. Differentiate between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).
3. Compare and contrast public, private, hybrid, and multi-cloud deployment models.
4. Understand the architecture of cloud computing systems.
5. Explore the roles and functions of key components such as virtualization, hypervisors, and containers.
6. Identify common security concerns in cloud computing.

Credits=4	MDST 544: Cloud Computing	No. of hours per unit=60
UNIT I	**Overview of Cloud Computing**	**15**
	Definition and Characteristics of Cloud Computing, Historical Evolution of Cloud Computing, Key Players in the Cloud Computing Industry, Benefits and Challenges of Cloud Computing	
UNIT II	**Cloud Service Models**	**15**
	Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Function as a Service (FaaS) Cloud Deployment Models Public Cloud, Private Cloud, Hybrid Cloud, Community Cloud Amazon EC2 (Elastic Compute Cloud), Amazon ECS (Elastic Container Service), AWS Lambda for Serverless Computing, Auto Scaling and Load Balancing	
UNIT III	**Cloud Native Technologies**	**15**
	Cloud Native Application Development, DevOps Practices in the Cloud, Continuous Integration and Continuous Deployment (CI/CD) Infrastructure as Code (IaC)	
UNIT IV	**Cloud Computing Architecture**	**15**
	Components of Cloud Architecture, Virtualization in Cloud Computing, Cloud Security and Compliance, Case Studies: Real-world examples of cloud architecture implementations	

Course Outcomes: After completion of syllabus, student will be able to:

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

6. Understanding compliance requirements and how they relate to cloud services.

Reference Book:

1. *Srinivasan*,Cloud Computing: A Practical Approach for Learning and Implementation, Pearson Education India,1 January 2014
2. *S. Singh*,Cloud Computing,Oxford University Press,1 June 2018
3. *Buyya*,MASTERING CLOUD COMPUTING,McGraw Hill Education,1 July 2017
4. *P. Sharma*, Applications of Cloud Computing: Approaches and Practices (Chapman & Hall/CRC Distributed Sensing and Intelligent Systems Series) [Hardcover] Sharma, Prerna; Sharma, Moolchand and Elhoseny, Mohamed,Chapman and Hall/CRC,13 November 2020

545: Social Media Analytics

Course Objectives: Student will be able to…

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

Credits=4	**MDST 544: Social Media Analytics**	**No. of hours per unit =60**
UNIT I	**Social Media Basics**	**15**
	Users: The Who of social media. Measuring Variations in User Behavior in Wikipedia, Long Tails Everywhere: The 80/20 Rule (p/q Rule), Online Behavior on Twitter. **Networks: The How of social media.** Types and Properties of Social Networks, Visualizing Networks, Degrees: The Winner Takes All, Capturing Correlations: Triangles, Clustering, and Assortativity. **Temporal Processes: The When of social media**. What Traditional Models Tell You About Events in Time, Inter-Event, Bursty Activities of Individuals, Forecasting Metrics in Time.	
UNIT II	**Content and large dataset of social media**	**15**
	Content: What of social media. Defining Content: Focus on Text and Unstructured Data, Using Content Features to Identify Topics, Extracting Low-Dimensional Information from High-Dimensional Text. **Processing Large Datasets.** MapReduce: Structuring Parallel and Sequential Operations, Multi-Stage MapReduce Flows, Patterns in MapReduce Programming, Sampling and Approximations: Getting Results with Less Computation, Sampling and Approximations: Getting Results with Less Computation, Bloom Filter, Count-Min Sketch, Executing on a Hadoop Cluster	

	(Amazon EC2).	
UNIT III	**Learn, Map, and Recommend.**	**15**
	Social Media Services Online, Problem Formulation, Learning and Mapping, Prediction and Recommendation. Social Media Data, From Data to Insights, Luis Madureira, Analytics in social media, Dedicated vs. Hybrid Tools. Alexander and Frederik Peiniger, Social Network Landscape, Tam Su, The Analytics Process, Armando Terribili, Metrics, Dashboards.	
UNIT IV	**Creating final Reports**	**15**
	Reports, Milan Veverka, Strategy, Tactics, Michael Wu, Prescriptive Analytics, The Future of Social Media Analytics.	

Course Outcomes: After completion of syllabus, student will be able to:

1. Understand and deal with any social media network, strategy, or campaign.
2. Social media analytics integrates with and affects other areas of business.
3. Give real-world context and insight.
4. Present decisions.
5. Learn and think in the field and reach a point where we can effortlessly approach any project with a sharp analytical mind.

Reference:

1. *B Mary*. Social Media Analytics in Data Science: Strategies and Tools for Analyzing Social Media Data. Chicago: Wiley, 2018.
2. *S John*. Social Media Analytics for Data Science: Methods and Techniques. Chicago: University Press, 2020.
3. *J Alice, and B Williams*. Data-Driven Social Media Analytics: A Practical Guide for Data Scientists. Chicago: Academic Press, 2019.
4. *G Szabo, G Polatkan, O Boykin, A John* Wiley Social Media Data Minning and Analytics , & Sons 2019.
5. *A. Goncalves* Social Media Analytics Strategy Apress 2017.

MDSP 546: Lab Course Based on
Deep Learning, GPU computing

Course Objectives: Student will be able to…

1. Introduce major deep learning algorithms
2. Present optimization techniques to training deep neural networks
3. Learn regularization techniques to train deep neural networks.
4. Understand the different approaches of parallel programming.
5. Study massively parallel computing hardware and programming models.
6. Be conversant with GPGPU programming with CUDA.

Credits= 4	MDSP 546: MDST 541, MDST 542	No. of hours per unit=60
	Group A MDST 541 1. Image Classification using Feed forward Neural Networks 2.Natural Language Processing (NLP) using Recurrent Neural Networks (RNNs) 3.Stochastic Gradient Descent (SGD) with Momentum 4. Adam Optimization Algorithm for Recurrent Neural Networks (RNNs) in Natural Language. 5.Hyperparameter Optimization using Grid Search or Random Search 6.Dropout Regularization in Image Classification with CNNs 7.L2 Regularization for Object Detection with CNNs 8.Batch Normalization for Semantic Segmentation with CNNs 9.Variational Auto encoder (VAE) for Image Generation 10. Sequence-to-Sequence (Seq2Seq) Model for Machine Translation	
	Group B MDST 442 1. Practical on Introduction to GPU Architectures and Parallelism 2. Experiment on Parallel Programming with GPUs using CUDA 3. Practical on Introduction to Grids, Blocks, and Threads 4. Experiment on Data Parallelism and CUDA C 5. Practical on CUDA Memory Management	

	6. Experiment on Synchronization Across CPU and GPU 7. Practical on Floating-Point Considerations in Parallel Computing 8. Experiment on Understanding CUDA Memory Handling 9. Practical on MPI-CUDA Programming in a Heterogeneous Computing Cluster 10. Practical on Multi-GPU Processing and Peer Access	

Course Outcomes: After completion of syllabus, student will be able to:

1. Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains.
2. Apply optimization techniques to training deep neural networks
3. Concern regularization techniques to improve the performance of deep learning algorithms.
4. Analyse and measure performance of modern parallel computing systems.
5. Design and implement parallel programs on GPUs.
6. Develop a high-performance parallel application in CUDA.

Reference :

1. *R. L. LaRose* "Discovering Statistics Using R" SAGE Publications Ltd (2016).
2. *H. Wickham and G. Grolemund* "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data" O'Reilly Media (2016).
3. *K. Black, Business Statistics*, New Delhi, Wiley, (2013).
4. *L. Cheng*. et al., *Statistics for Business and Financial Economics*, New York: HeidelbergDordrecht, (2013).
5. *D. B. Kirk, W.mei* W. Hwu, and Ryan Nassiry, "Programming Massively Parallel Processors: A Hands-on Approach (3rd Edition)" Morgan Kaufmann (2016).
6. *J Cheng, M Grossman, and T. McKercher* "Professional CUDA C Programming". Wiley (2014).
7. *S Cook,* "CUDA Programming: A Developer's Guide to Parallel Computing with GPUs", Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ISBN: 9780124159884, (2013).
8. *I. Buck and J. Clayton* "CUDA Handbook: A Comprehensive Guide to GPU Programming" Addison-Wesley Professional (2013).

MDSP 547: Lab Course Based on
Recommender System, Cloud Computing

Course Objectives: Student will be able to…

1. Define cloud computing and its key characteristics.
2. Differentiate between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).
3. Compare and contrast public, private, hybrid, and multi-cloud deployment models.
4. Describe the purpose of recommendation systems.
5. Understand the components of a recommendation system including candidate generation, scoring, and re-ranking.
6. Familiarity with linear algebra

Credits= 4	**MDSP 547: MDST 543, MDST 544**	**No. of hours per unit =60**
	Group A MDST 543 1. Practical on Matrix Operations for Recommender Systems and Covariance Matrices in Collaborative Filtering 2. Experiment on Applications of Recommendation Systems. 3. Practical on User-based and Item-based Nearest Neighbor Recommendation. 4. Experiment on Model-based Approaches and Pre-processing Based Approaches. 5. Practical on Attacks on Collaborative Recommender Systems. 6. Experiment on Advantages and Drawbacks of Content-Based Filtering. 7. Practical on Representing Item Profiles and Methods for Learning User Profiles. 8. Experiment on Evaluation of Historical Datasets for Predictive Modeling. 9. Practical on User-Centered Evaluation of Social Tagging Recommender Systems. 10. Experiment on Group Recommender Systems for Collaborative Decision Making.	
	Group B MDST 544 1. Practical on Historical Evolution of Cloud Computing 2. Experiment on Cloud Computing Use Cases and Implementation Planning 3. Practical on Introduction to Cloud Service Models 4. Experiment on Amazon EC2 and Elastic Container Service (ECS) 5. Practical on AWS Lambda for Serverless Computing 6. Experiment on Cloud Native Application Development 7. Practical on DevOps Practices in the Cloud 8. Experiment on Continuous Integration and Continuous Deployment	

	(CI/CD) 9. Practical on Understanding Components of Cloud Architecture 10. Experiment on Virtualization in Cloud Computing	

Course Outcomes: After completion of syllabus, student 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
3. Social media platforms such as Facebook and Twitter
4. Explain a variety of approaches for building recommender systems.
5. Define cloud computing and its key characteristics.
6. Understand the service models (IaaS, PaaS, SaaS) and deployment models (public, private, hybrid).
7. Knowledge of major cloud service providers (e.g., Amazon Web Services, Microsoft Azure, Google Cloud Platform) and their offerings.
8. Ability to compare different cloud services.

Reference :

1. *A. Srinivasan*, Cloud Computing: A Practical Approach for Learning and Implementation, Pearson Education India,1 January 2014
2. *S. Singh*, Cloud Computing, Oxford University Press,1 June 2018
3. *Buyya*, MASTERING CLOUD COMPUTING, McGraw Hill Education,1 July 2017
4. *P. Sharma*, Applications of Cloud Computing: Approaches and Practices (Chapman & Hall/CRC Distributed Sensing and Intelligent Systems Series) [Hardcover] Sharma, Prerna; Sharma, Boolchand and Elhassan, Mohamed, Chapman and Hall/CRC,13 November 2020
   Charu C. Aggarwal, Recommender Systems: The Textbook, 1/e, Springer, 2016.
5. *C. Charu Aggarwal* "Recommender Systems: The Textbook" Springer, (2016).
6. *S. Berkovsky, T. Kuflik, F. Ricci* "Recommender Systems Handbook" Publisher: Springer, (2015).
7. *N. Manouselis, H. Drachsler, K. Verbert, E. Duval*, Recommender Systems for Learning, Springer, 2013
8. *J. A. Konstan, J. Riedl*,"Recommender Systems: An Introduction" Cambridge University Press, (2012).