Maharashtra Institute of Technology, Aurangabad
(An Autonomous Institute)

MAHARASHTRA INSTITUTE OF TECHNOLOGY, AURANGABAD

An Autonomous Institute Affiliated to Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra (India)

First Year M. Tech. Syllabus Computer Science and Engineering (Data Science and Analytics) 2023-24

First Year M. Tech. Syllabus Structure w.e.f. 2023-24 (Pattern 2021-22)															
First Year M. Tech. Syllabus Computer Science and Engineering (Data Science and Analytics)															
Semester-I															
Course Code	Course Title	Teaching Scheme (Hours/Week)			Examination Scheme and Marks							Credits			
		Theory	Tutorial	Practical	MSE-I	MSE-II	TA	ESE	TW	PR/OR	Total	TH	TUT	TW/PR	Total
MTM101	Research methodology and IPR	3	1	-	15	15	20	50	-	-	100	3	1	-	4
MDA102	Probability and Statistics for Data Science	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA103	Cloud Analytics	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA104	Advanced Data Engineering	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA121-123	Professional Elective-I	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA111	Lab-I: Cloud Analytics	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA 112	Lab-II: Advanced Data Engineering	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA113	Lab-III: Professional Elective-I	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA114	Seminar	-	-	4	-	-	-	-	-	50	50	-	-	2	2
	Total (Semester-I)	15	1	10	75	75	100	250	75	50	625	15	1	5	21
Semester-II															
Course Code	Course Title	Teaching Scheme (Hours/Week)			Examination Scheme and Marks							Credits			
		Theory	Tutorial	Practical	MSE-I	MSE-II	TA	ESE	TW	PR/OR	Total	TH	TUT	TW/PR	Total
MDA141	Exploratory Data Analysis	3	1	-	15	15	20	50	-	-	100	3	1	-	4
MDA142	Big Data Analytics	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA143	Data Visualization and Interpretation	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA144	Data Science and Security	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA161-163	Professional Elective-II	3	-	-	15	15	20	50	-	-	100	3	-	-	3
MDA151	Lab-I: Exploratory Data Analysis	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA152	Lab-II: Big Data Analytics	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA153	Lab-III: Data Visualization and Interpretation	-	-	2	-	-	-	-	25	-	25	-	-	1	1
MDA154	Minor Project	-	-	4	-	-	-	-	-	50	50	-	-	2	2
	Total (Semester-II)	15	1	10	75	75	100	250	75	50	625	15	1	5	21
	Grand Total				150	150	200	500	150	100	1250	30	2	10	42

L-Lecture, T- Tutorial, P- Practical, MSE- Mid Semester Exam, CIE- Continuous In-semester Evaluation, TA-Teacher Assessment, ESE-End Semester Examination

Master Copy

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Dean
Academics
Maharashtra Institute of Technology
Aurangabad.

Chairman Academic Council
MIT Aurangabad
(An Autonomous Institute)

Professional Elective Courses – I

Group A	Group B	Group C
MDA121 Deep Learning	MDA122 Natural Language Processing	MDA123 Advance Business Intelligence

Professional Elective Courses – II

Group A	Group B	Group C
MDA161 Generative AI and Prompt Engineering	MDA162 Digital Forensics	MDA163 MOOC Courses
MDA201* (MOOC Course) I - Social Network Analysis II - Computer Vision		

Semester-III

Course Code	Course Name	Teaching Scheme (Hours/week)			Examination Scheme and Marks					Credits			
		Theory	Tutorial	Practical	Igeative AISE-II	TA	ESE	PR/OR	Total	TH	TUT	TW/PR	Total
MDA201	MOOC course	3	-		-	-	100	-	100	3	-	-	3
MDA211	Dissertation-I	-	-	18	-	-	-	100	150	-	-	9	9
	Total (Semester-III)	3	-	18	-	-	100	100	250	3	-	9	12

Semester-IV

Course Code	Course Name	Teaching Scheme (Hours/week)			Examination Scheme and Marks					Credits			
		Theory	Tutorial	Practical	MSE-II	TA	ESE	PR/OR	Total	TH	TUT	TW/PR	Total
MDA251	Dissertation-II	-	-	24	-	-	-	100	200	-	-	12	12
	Total (Semester-III)	-	-	24	-	-	-	100	200	-	-	12	12

M. Tech (Second Year)

	Grand Total	-	-	-	-	-	100	200	450	3	-	21	24

M. Tech (Computer Science and Engineering (Data Science and Analytics))

	Grand Total (MTech)	-	-	-	150	200	600	300	1700	33	2	31	66

Master Copy

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Dean
Academics
Maharashtra Institute of Technology
Aurangabad.

Chairman Academic Council
MIT Aurangabad
(An Autonomous Institute)

Faculty of Science & Technology Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I	
Course Code: MTM101 Course: Research Methodology & IPR Teaching Scheme: Theory: 2 Hrs/week Tutorial: 1 Hr/week	Credits: 3-1-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 hrs

Prerequisite	Basic concepts of Research Methodology & IPR.
Objectives	The objectives of the course are. 1. To introduce various research problems. 2. To understand the research process. 3. Learn efficient Report Writing and Patents.
Unit-I	**Research Problems and Research Design** Meaning of research, types of research, steps in involved in research process, criteria of good research, importance of ethics in research, codes and policies for research ethics, Selection of research problem, steps involved in defining research problem, need for research design, types of research designs, basic principles of experimental design, formal and informal experimental design. (5 Hrs)
Unit-II	**Sampling Design** Need for sampling, steps in sampling design, different types of sampling designs, sampling distributions, concept of central limit and standard error, sources of errors, population mean and proportion, sample size calculations, tests of measurements for validity, reliability and practicality. (5 Hrs)
Unit-III	**Data collection, Processing and Analysis** Methods for collection of data, selection of data collection method, data processing

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	operations, statistics in research, confidence level, measures of central tendency, dispersion, asymmetry and relationship, Spearman's and Pearson's coefficient of correlation, simple & multiple regression analysis, analysis of variance (ANOVA), factor analysis methods. (8 Hrs)
Unit-IV	**Hypothesis Test and Report Writing** Concept of research hypothesis, concept of testing of hypothesis, Parametric tests (z, t, F and chi- square tests), Hypothesis testing of means and correlation coefficient, Nonparametric tests, significance of research report writing, types of reports, structure of the research report, steps in (5 Hrs)
Unit-V	**Introduction to IPR** Origin and evolution of IPR to its present form and use, Different Tools ofIPR and what is the nature of these rights, Balancing Rights and Responsibilities, Societal implications of IPR. (5 Hrs)
Unit-VI	**Patents** Concept of inventions/discoveries, patents protect; benchmarks for patent ability of inventions; Exceptions to patent ability; Patenting issues in Biotechnology and computer based inventions, process to apply for patents in India and in other countries around the world, The steps to granting of a patent; Opposing grant of a patent; term of a patent; rights of a patent holder; challenging validity of a patent licensing of patent rights; using patent rights in the market place, compulsory license. (6 Hrs)

References

Sr.No.	Title	Author	Publication	Edition
1.	Research Methodology: Methods and Techniques	C. R. Kothari and G. Garg	New Age International	2019

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	2.	Research Methodology	R. Pannerselvam	PHI Leaming, 2014	2014
	3.	Research Methodology- As Theoretical Approach	D. Napolean& B. Narayan	Laxmi Publications,2014.	2014
	4.	Research Methods and Statistics	Bernard C. Beins& Maureen, A. McCarthy	Pearson Education Inc,2012.	2012
	5.	Research Methods Handbook, CLES	Stuart MacDonald & Nicola Headlam.		2009
	6	Intellectual Property Rights- Unleashing the Knowledge Economy	Ganguli Prabuddha.	Tata Mc Graw Hill,2001.	2001

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Faculty of Science & Technology

Syllabus of M. Tech. (Computer Science and Technology) Semester-I

Course Code: MDA102 Course: Probability and Statistics for Data Science **Teaching Scheme:** Theory: 3 Hrs. / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs.

Prerequisite	Basic Mathematics
Objectives	1. To learn the probability and probabilistic models of data science 2. To learn the basic statistics and testing hypothesis for specific problems 3. To solve probabilistic problems in engineering and applied science 4. To understand the concept of random processes in engineering disciplines
Unit-I	**Basic Probability Theory and Random Variables** Basic Probability Theory: Probability spaces, Conditional probability Independence, Random Variables: Definition, Discrete random variables, Continuous random variables, Conditioning on an event, Functions of random variables, Generating random variables **(6 Hrs.)**
Unit-II	**Linear Algebra and Set Theory for Data Engineering** Set Theory: Definitions, set operations, Linear Algebra: Vector spaces, Inner product and norm, Orthogonality, Projections, Matrices, Eigen decomposition, Eigen decomposition of symmetric matrices. **(6 Hrs.)**
Unit-III	**ETL and Random Processes** ETL: ETL Basics, ETL pipeline, Challenges with ETL, Expectation operator, Mean and Variance, Covariance, Conditional expectation, Random Processes: Definition, Mean and autocovariance functions, Independent identically distributed sequences, Gaussian process, Poisson process, Random walk. **(6 Hrs.)**

Unit-IV	**Random Processes Convergence** Types of convergence, Law of large numbers, Central limit theorem, Monte Carlo simulation, Markov Chains: Time-homogeneous discrete-time Markov chains, Recurrence, Periodicity, Convergence, Markov-chain Monte Carlo (06 Hrs)
Unit-V	**Statistics for Data Engineering** Descriptive Statistics: Histogram, Sample mean and variance, Order statistics, Sample covariance, Sample covariance matrix, Frequentist Statistics: Independent identically distributed sampling, Mean square error, Consistency, Confidence intervals, Nonparametric model estimation, Parametric model estimation, Bayesian Statistics: Bayesian parametric models, Conjugate prior, Bayesian estimators. (06 Hrs)
Unit-VI	**Hypothesis Testing** The hypothesis-testing framework, Parametric testing, Nonparametric testing: The permutation test, Multiple testing, Linear Regression: Linear models, Least-squares estimation, Overfitting (06 Hrs)

References

Sr. No.	Title	Author	Publication	Edition
1	Data Engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python	Paul Crickard	Packt Publishing Limited	2020
2	Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Grayscale Indian Edition)	Mat Housley, Joe Reis	Shroff/O'Reilly	First Edition (27 June 2022)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	3	The Data Engineering Cookbook	Andreas Kretz		
	4	Data Pipelines Pocket Reference: Moving and Processing Data for Analytics	James Densmore	O'Reilly Media, Inc, USA	(31 March 2021)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(An Autonomous Institute)

Faculty of Science & Technology Syllabus of M. Tech. (Computer Science and Technology) Semester-I	
Course Code: MDA103 Course: Cloud Analytics Teaching Scheme: Theory: 03 Hrs / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs
Prerequisite	Database Management System
Objectives	1.To learn and understand basic concepts of Cloud Computing & its Models. 2.To design, develop and deploy Cloud applications. 3.To introduce basic principles, concepts and applications of data warehousing. 4. To understand the fundamental concepts of big data and analytics.
Unit-I	**Introduction to Cloud Computing** Parallel & Distributed Computing, Cluster Computing, Grid Computing, Definition and Evolution of Cloud Computing, the Vision of Cloud Computing, Cloud Deployment Models, Cloud Service Models, Key Characteristics, Benefits, Risks & Challenges in Cloud Computing. (6 Hrs)
Unit-II	**Cloud Computing Architecture** Introduction, Cloud Reference Model, Architecture, Infrastructure / Hardware as a Service, Platform as a Service, Software as a Service, Types of Clouds, Public Clouds, Private Clouds, Hybrid Clouds, Community Clouds, Economics of the Cloud, Open Challenges, Cloud, Interoperability and Standards. (06 Hrs)
Unit-III	**Enabling Cloud Technologies** Web services: XML, SOAP, REST, Virtualization: Introduction to virtualization, Hypervisor: Type-I & Type II, Types of Virtualization, Pros and cons of virtualization, Virtualization applications in enterprises: Server virtualization, Desktop and Application, Virtualization, Storage and Network Virtualization.

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	(06 Hrs)
Unit-IV	**Cloud Applications** Scientific Applications – Health care, Business and Consumer, Applications-CRM and ERP, Social Networking. Cloud Platforms in Industry: Amazon Web Services- Compute Services, Storage Services, Communication Services and Additional Services. Google AppEngine-Architecture and Core Concepts. Microsoft Azure- Azure Core Concepts, SQL Azure. (06 Hrs)
Unit-V	**Data Warehouse & ETL Process** Data Warehouse: Basic Concepts, A Multitiered Architecture, Enterprise Warehouse, Data Mart, Extraction, Transformation, and Loading, Metadata Repository.Introduction to ETL, ETL Process in Data Warehouse, ETL Tools. (06 Hrs)
Unit-VI	**Cloud for Big Data Storage** Introduction to Big Data , Big data characteristics, Challenges in Big Data, Hadoop: Definition, Architecture, Introduction to Storage Systems, Cloud Storage Concepts, Distributed File Systems (HDFS, Ceph FS),Cloud Databases (HBase, MongoDB, Cassandra, DynamoDB), Cloud Object Storage (Amazon S3, OpenStack Swift,) NoSQL (Document-based, Key-Value, Column-based, Graph-based) (06 Hrs)

References

Sr. No.	Title	Author	Publication	Edition
1	Enterprise Cloud Computing: Technology, Architecture, Applications	Gautam Shroff	Cambridge University Press	
2	Cloud Computing Implementation, Management, and Security	Barrie Sosinsky	CRC Press	

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	3	Cloud computing Bible	Andreas Kretz	Wiley India Pvt Ltd	2011
	4	Mastering Cloud Computing	kumar Buyya, Christian Vecchiola, S. Thamarai Selvi		

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Faculty of Science & Technology Syllabus of M. Tech. (Computer Science and Technology) Semester-I	
Course Code: MDA104 Course: Advanced Data Engineering **Teaching Scheme:** Theory: 3 Hrs. / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs.

Prerequisite	Programming language, Mathematics and Statistics
Objectives	1. To give a hands-on experience with real-world data analysis 2. To construct Data engineering infrastructure 3. To work with data pipeline
Unit-I	**Introduction to Data Engineering** Definition and Overview of Data Engineering, Raw Data, Data Engineering Roles, Data Engineering Process, The Modern Data Stack, Data Engineering Vs Data Science, Data Preprocessing, Role of data engineer, Data engineering tools **(6 Hrs.)**
Unit-II	**Data Engineering Pipelines and Big Data Engineering** Data Pipeline, ETL pipeline, Data pipeline challenges, Data warehouse Architecture, Data warehouse Storage, Metadata, Data warehouse access tools, Data warehouse management tools, Data Marts Definition and Types, OLAP and OLAP cubes, Big Data Engineering: Data Lake, Hadoop and its ecosystem **(6 Hrs.)**
Unit-III	**Advanced SQL** Into to Postgres and psql, SQL Basics, Types of Joins, Advanced SQL Features - subqueries, CTE's, and Window functions, Common Table Expressions (CTEs), Recursive CTEs, Temporary Functions, Pivoting Data With Case When, Calculating Delta Values, Calculating Running Totals, Date-Time Manipulation,

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	Manipulation, (06 Hrs)
Unit-IV	**Source Systems and Data Ingestion** Data Lake, Data Warehouse, Data Lakehouses, Source Systems, Replication of source data, Batch Processing, Data Ingestion: Introduction to Data Ingestion, Types of Data Ingestion, Data Ingestion vs. ETL, Data Ingestion tools, Data Ingestion Best Practices and Challenges, Streaming Bulk ingestion using the Copy command (06 Hrs)
Unit-V	**Data Cleansing, Validation and Modeling** Data Quality of Source Systems, Statistical validation, Rule-based validation, Normalization, Dimensional Modeling, Creating Tables, Schema Migration. (06 Hrs)
Unit-VI	**Data Presentation and Visualization** Business Intelligence Tools, Introduction to Superset, Creating visualizations (06 Hrs)

References

Sr. No.	Title	Author	Publication	Edition
1	Data Engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python	Paul Crickard	Packt Publishing Limited	2020
2	Fundamentals of Data Engineering: Plan and Build Robust Data Systems (Grayscale Indian Edition)	Mat Housley, Joe Reis	Shroff/O'Reilly	First Edition (27 June 2022)
3	The Data Engineering Cookbook	Andreas Kretz		
4	Data Pipelines Pocket Reference: Moving and Processing Data for	James Densmore	O'Reilly Media, Inc, USA	(31 March

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

		Analytics			2021)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of T. Y. B. Tech. (AIDS) (Semester V)		
Course Code: MDA121 Course: Professional Elective-I Deep Learning Teaching Scheme: Theory: 03 Hrs/week		Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs
Prerequisites	Machine Learning	
Objectives	1. Understand the context of neural networks and deep learning 2. To introduce techniques used for training artificial neural networks 3. To enable design and deployment of deep learning models for machine learning problems	
Unit-I	**Foundations of Deep learning** What is machine learning and deep learning?, Supervised and Unsupervised Learning, bias variance tradeoff, hyper parameters, under/over fitting regularization, Limitations of machine learning, History of deep learning, Advantage and challenges of deep learning. Learning representations from data , Understanding how deep learning works in three figures, Common Architectural Principles of Deep Network, Architecture Design, Applications of Deep learning, Introduction and use of popular industry tools such as TensorFLow, Keras, PyTorch, Caffe, Shogun. **(6 Hrs)**	

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-II	**Deep Neural Networks(DNNs)** Introduction to Neural Networks :The Biological Neuron, The Perceptron, Multilayer Feed-Forward Networks , Training Neural Networks :Backpropagation and Forward propagation Activation Functions :Linear ,Sigmoid, Tannh, Hard Tanh, Softmax, Rectified Linear, Loss Functions :Loss Function Notation , Loss Functions for Regression , Loss Functions for Classification, Loss Functions for Reconstruction, Hyperparameters : Learning Rate, Regularization, Momentum, Sparsity, Deep Feedforward Networks – Example of Ex OR, Hidden Units, cost functions, error backpropagation, Gradient-Based Learning, Implementing Gradient Descent, vanishing and Exploding gradient descent, Sentiment Analysis, Deep Learning with Pytorch, Jupyter, colab. (07 Hrs)
Unit-III	**Convolution Neural Network(CNN)** Introduction, CNN architecture overview, The Basic Structure of a Convolutional Network- Padding, Strides, Typical Settings, the ReLU layer, Pooling, Fully Connected Layers, The Interleaving between Layers, Local Response Normalization, Training a Convolutional Network (07 Hrs)
Unit-IV	**Recurrent Neural Network(RNN)** Recurrent and Recursive Nets: Unfolding Computational Graphs, Recurrent Neural Networks, Bidirectional RNNs, Encoder-Decoder Sequence-to-Sequence Architectures, Deep Recurrent Networks, Recursive Neural Networks, The Challenge of Long-Term Dependencies, Echo State Networks, Leaky Units and Other Strategies for Multiple Time Scales, The Long Short-Term Memory and Other Gated RNNs, Optimization for Long-Term Dependencies, Explicit Memory. Practical Methodology: Performance Metrics, Default Baseline Models, Determining Whether to Gather More Data, Selecting Hyper parameters. (07 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-V	**Deep Generative Models** Introduction to deep generative model, Boltzmann Machine, Deep Belief Networks, Generative adversarial network (GAN), discriminator network, generator network, types of GAN, Applications of GAN networks (07 Hrs)
Unit-VI	**Reinforcement Learning** Introduction of deep reinforcement learning, Markov Decision Process, basic framework of reinforcement learning, challenges of reinforcement learning, Dynamic programming algorithms for reinforcement learning, Q Learning and Deep Q-Networks, Deep Q recurrent networks, Simple reinforcement learning for Tic-Tac-Toe. (07 Hrs)

Textbooks / Reference Books

Sr. No.	Title	Author	Publication	Edition
1.	Deep Learning	Goodfellow, I., Bengio, Y.,,Courville	MIT Press	2016
2.	Deep Learning	Josh Patterson & Adam Gibson		
3.	Neural Networks and deep learning	Charu Agarwal		
4.	Reinforcement Learning: An Introduction	Richard S. Sutton and Andrew G. Barto		

.Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I	
Course Code: MDA122 Course: Professional Elective-I Natural Language Processing Teaching Scheme: Theory: 03 Hrs / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisite	Machine Learning
Objectives	1. Learn leading trends and systems in natural language processing. 2. To describe the application based on natural language processing and to show the points of syntactic, semantic, and pragmatic processing. 3. Understand text processing techniques
Unit-I	**Introduction to NLP** Natural Language Processing - Problems and perspectives, Introduction/Recall to/of probability calculus, N-grams and Language Models, Markov Models, Introduction to Machine Learning and Deep Learning, Recurrent Neural Network Language Models, The evaluation of NLP applications (6 Hrs)
Unit-II	**Computational Phonetics and Speech Processing** Speech samples: properties and acoustic measures, Analysis in the frequency domain, Spectrograms, Applications in the acoustic phonetic field, Speech recognition with HMM and Deep Neural Networks, Tokenization and Sentence splitting, Computational Morphology, Morphological operations, Static lexica, Two-level morphology (6 Hrs)
Unit-III	**Computational Syntax** Computational Syntax: Part-of-speech tagging, Grammars for natural language,

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	Natural language Parsing, Supplementary worksheet: formal grammars for NL, Formal languages, and Natural languages. Natural language complexity, Phrase structure grammars, Dependency Grammars, Treebanks, Modern formalisms for parsing natural languages. (6 Hrs)
Unit-IV	**Computational Semantics** Lexical semantics: WordNet and FrameNet, Word Sense Disambiguation, Distributional Semantics & Word-Space models, Logical approaches to sentence semantics. (6 Hrs)
Unit-V	**Information Extraction and Machine Translation** Named entity recognition and relation extraction, IE using sequence labeling, Basic issues in MT, Statistical translation, word alignment, phrase-based translation, and synchronous grammars. (6 Hrs)
Unit-VI	**Applications and Case studies** Applications and Case studies: Solving Downstream Tasks: Document classification, Sentiment Analysis, Named Entity Recognition, Semantic Textual Similarity, Prompting Pre-Trained Language Models, Network Embedding (6 Hrs)

References

Sr. No.	Title	Author	Publication	Edition
1	Foundations of Natural Language Processing	Christopher D. Manning and Hinrich Schutze	MIT Press	6th
2	Speech and Language Processing	Daniel Jurafsky and James H. Martin	Prentice Hall, 2009.	3rd

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	3	Handbook of Natural Language Processing	NitinIndurkhya, Fred J. Damerau	CRC Press	2^{nd}
	4	Natural Language Understanding	James Allen	Natural Language Understanding	8^{th}

Chairman Board of Stu...
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

<table>
<tr><td colspan="3" align="center">(Faculty of Science & Technology)
Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I</td></tr>
<tr><td colspan="2">Course Code: MDA122
Course: Professional Elective-I Advance Business Intelligence
Teaching Scheme:
Theory: 03 Hrs / week</td><td>Credits: 3-0-0
Mid Semester Examination-I: 15 Marks
Mid Semester Examination-II: 15 Marks
Teacher Assessment: 20 Marks
End Semester Examination: 50 Marks
End Semester Examination (Duration): 02 Hrs</td></tr>
<tr><td>Prerequisites</td><td colspan="2">No Prerequisites</td></tr>
<tr><td>Objectives</td><td colspan="2">1. Learn structured approach to business problem-solving.
2. To study data analytics practices executed in the business world
3. To learn analytics frameworks and methods to support reporting and decision making.</td></tr>
<tr><td>Unit-I</td><td colspan="2">Understanding Business Intelligence
The Challenge of Decision Making, What Is Business Intelligence?, The Business Intelligence Value Proposition, The Combination of Business and Technology
(6 Hrs)</td></tr>
<tr><td>Unit-II</td><td colspan="2">Business Intelligence Technology Counterparts
Data Warehousing: What Is a Data Warehouse?, Data Marts and Analytical Data, Organization of the Data Warehouse
Enterprise Resource Planning: Distributing the Enterprise, First ERP, then Business Intelligence, The Current State of Affairs
Customer Relationship Management: CRM, ERP, and Business Intelligence
Customer Decisions, Decisions About Customers, Business Intelligence and Financial Information (6 Hrs)</td></tr>
<tr><td>Unit-III</td><td colspan="2">The Spectrum of Business Intelligence
Enterprise and Departmental Business Intelligence, Strategic and Tactical Business Intelligence, Power and Usability in Business Intelligence, Finding</td></tr>
</table>

Chairman Board of
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	the Right Spot on the Continuum, Business Intelligence: Art or Science? (6 Hrs)
Unit-IV	**Business Intelligence User Interfaces** Querying and Reporting, Reporting and Querying Toolkits, Basic Approaches: Building Ad-Hoc Queries, Building On-Demand Self-Service Reports, Enhancing and Modifying, Data Access: Pull-Oriented Data Access, Push-Oriented Data Access, Dashboards: EIS Is the Engine, Metric System and KPIs, Business Intelligence Dashboards (6 Hrs)
Unit-V	**On-Line Analytical Processing (OLAP)** OLAP:OLAP and OLTP, Operational Data Stores, Variations in Data and Approach, OLAP Applications and Functionality, Multi-Dimensions: Thinking in More Than Two Dimensions, What Are the Possibilities?, Drilling and Pivoting, OLAP Architecture: Cubism, Tools, ROLAP, MOLAP, HOLAP, Data Mining (6 Hrs)
Unit-VI	**Visualization, Guided Analysis and** Visualization: The Basics, Unconstrained Views, Guided Analysis: The Business Intelligence Two-Step, How to Guide the Guides, Handling Unstructured Data (6 Hrs)

Textbooks / Reference Books	Sr. No.	Title	Author	Publication	Edition
	1	Decision Support and Business Intelligence Systems	Efraim Turban, Ramesh Sharda, Jay Aronson, David King	Pearson Education, 2009.	9th
	2	The Savy Manager's Guide Getting Onboard with Emerging IT,	David Loshin, Business Intelligence	Morgan Kaufmann Publishers.	2009

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I	
Course Code: MDA111 Course: Lab-I Cloud Analytics Teaching Scheme: Practical: 02 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks

Objectives	1. To learn the fundamental ideas behind Cloud Computing, the evolution of the paradigm, its applicability. 2. To understand the basic ideas and principles in data center design, cloud management techniques and cloud software deployment considerations. 3. To explore tools and practices for working with big data.
List of Practical	1. Develop any application in Google colab (SaaS) using C and Python. 2. Develop any one application in Codeanywhere using C, C++, Java or Python (PaaS) 3. Creating a Warehouse application in SalesForce.com. 4. Adding Master Detail and Lookup Relationship to the objects using Salesforce.com. 5. Implementation of Web services in SOAP for JAVA Applications 6. Implement and use sample cloud services with the help of Microsoft Azure. 7. Categorize Amazon Web Service (AWS) and implement its various cloud entities using its Cloud Toolbox support. 8. Write a Map Reduce program to count words from a given text file 9. Case study: Study & Installation of Cloudera CDH

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I	
Course Code: MDA112 Course: Lab-I Advanced Data Engineering Teaching Scheme: Practical: 02 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks

Objectives	1. To give a hands-on experience with real-world data analysis 2. To construct Data engineering infrastructure 3. To work with data pipeline
List of Practical	1. Build the data engineering infrastructure. 2. Build the data pipelines to work with database. 3. Perform data cleaning, data transformation and data enriching. 4. Build a 311 data pipeline. 5. Build idempotent and atomic data pipeline. 6. Implement version control with NiFi registry and monitor the data pipeline. 7. Build and deploy production data pipeline. 8. Stream data with Apache Kafka

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I	
Course Code: MDA113 Course: Lab-III Professional Elective – I Deep Learning Teaching Scheme: Practical: 02 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks

Objectives	1. To analyse the given dataset for designing a neural network based solution. 2. To use deep learning models or develop new architectures to solve practical real-world problems such as computer vision and natural language processing. 3. To become familiar with deep learning programming frameworks based on Python to solve real world problems.
List of Practical	1. Implement image Classification using CNN. 2. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets. 3. Implement Decision tree using sklearn and its parameter tuning. 4. Use TensorFlow for Music Generation with RNNs 5. Implement Deep Neural Network With L – Layers. 6. Implement Text Generation using Recurrent Long Short Term Memory Network. 7. Implement training Neural Networks with Validation using PyTorch. 8. Implement generative adversarial network.

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology)

Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I

Course Code: MDA114 Course: Professional Elective-I Natural Language Processing Teaching Scheme: Practical: 02 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks

Objectives	1. Demonstrate the state-of-the-art algorithms and techniques for text-based processing of natural language. 2. Perform POS tagging for a given natural language and select a suitable language modelling technique based on the structure of the language. 3. Explore Real World Applications based on NLP
List of Practical	Practical may be implemented using programming Python / Java. 1. Preprocessing of text (Tokenization, Filtration, Script Validation, Stop Word Removal, Stemming) 2. Tokenizing Text and WordNet basics 3. Develop Word Embeddings in Python 4. Implement N-gram model 5. perform POS tagging on sample text data 6. Implement model to identify the misspelled word 7. Train a neural network with GLoVe word embeddings to perform sentiment analysis of tweets 8. Case Study/ Mini Project

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

<table>
<tr><td colspan="2">(Faculty of Science & Technology)
Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I</td></tr>
<tr><td>Course Code: MDA113
Course: Lab-III Professional Elective – I Advanced Business Intelligence
Teaching Scheme:
Practical: 02 Hrs/week</td><td>Credits: 0-0-1
Teacher Assessment: 25 Marks</td></tr>
<tr><td>Objectives</td><td>1. To build conceptual models of business and simple database models
2. Practice data extraction using SQL, apply predictive and prescriptive analytics to business problems,
3. Develop models for decision making, interpret the software output</td></tr>
<tr><td>List of Practical</td><td>1. Study of Decision Support Systems
2. Experiment on Creating Cube
3. Creating reference, fact and many to many dimensions
4. Retrieve data or values from cube
5. Retrieving Data from Analysis Services using Excel with Analysis Services
6. Study of different data mining tools
7. Crete Reports Using SQL Server
8. Creation of Prediction queries</td></tr>
</table>

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

<table>
<tr><td colspan="2" align="center">(Faculty of Science & Technology)
Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-I</td></tr>
<tr><td>Course Code: MDA114
Course: Lab-IV: Seminar
Teaching Scheme:
Practical: 02 Hrs/week</td><td>Credits: 02
ESE/Oral: 50 Marks</td></tr>
<tr><td>Objectives</td><td>• To encourage the students to study advanced engineering developments.
• To develop skills in doing literature survey, technical presentation and report preparation.
• To prepare and present technical reports.
• To encourage the students to use various teaching aids such as power point presentation and demonstrative models.</td></tr>
</table>

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA141 Course: Exploratory Data Analytics Teaching Scheme: Theory: 03 Hrs / week Tutorial : 01 Hr / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisite	Data Engineering
Objectives	1. Understand the fundamentals of exploratory data analysis. 2. Implement the data visualization using Matplotlib. 3. Perform univariate data exploration and analysis. 4. Apply bivariate data exploration and analysis. 5. Use Data exploration and visualization techniques for multivariate and time series data.
Unit-I	**Introduction to EDA** Introduction to exploratory data analysis and data visualization: Perception, Continuous variables, Discrete variables, Dependency relationships, Multivariate categorical variables, Temporal data, Spatial data, Data Science Pipeline: Collect, Import, Clean, Transform, Visualize, Model, Communicate (6 Hrs)
Unit-II	**Univariate Analysis** Introduction to Single variables: Distribution Variables, Numerical Summaries of Level and Spread, Scaling and Standardizing, Inequality. (06 Hrs)
Unit-III	**Multivariate and Time Series Analysis** Introducing a Third Variable, Causal Explanations, Three-Variable Contingency Tables and Beyond, Fundamentals of TSA, Characteristics of time series data, Data Cleaning, Time-based indexing, Visualizing, Grouping, Resampling. (06 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-IV	**Bivariate Analysis** Relationships between Two Variables, Percentage Tables, Analyzing Contingency Tables, Handling Several Batches, Scatterplots and Resistant Lines **(6 Hrs.)**
Unit-V	**EDA Using Python** Data Manipulation using Pandas, Pandas Objects, Data Indexing and Selection, Operating on Data, Handling Missing Data, Hierarchical Indexing, combining datasets, Concat, Append, Merge and Join, Aggregation and grouping, Pivot Tables, Vectorized String Operations **(6 Hrs.)**
Unit-VI	**EDA Tools and Techniques** Correlation Analysis, Summary Statistics, Missing data Analytics, Outlier Detection, Categorical data Exploration, Dimensionality Reduction, Domain Knowledge Integration **(6 Hrs.)**

References

Sr. No.	Title	Author	Publication	Edition
1	Hands-On Exploratory Data Analysis with Python	Suresh Kumar Mukhiya, Usman Ahmed	Packt Publishing	2020
2	Python Data Science Handbook: Essential Tools for Working with Data	Jake Vander Plas	O Reilly	First Edition
3	Exploring Data: An Introduction to Data Analysis for Social Scientists	Catherine Marsh, Jane Elliott	Wiley Publications	2008

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA142 Course: Big Data Analytics Teaching Scheme: Theory: 03 Hrs / week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisite	Machine Learning Basic Concepts
Objectives	1. Understand the Big Data Analytics concepts and its applications in business. 2. Understand the functions and components of Map Reduce Framework and HDFS. 3. Discuss Data Management concepts in NoSQL environment. 4. Explain process of developing Map Reduce based distributed processing applications. 5. Explain process of developing applications using HBASE, Hive, Pig etc
Unit-I	**Introduction to Big Data** Types of digital data, history of Big Data innovation, introduction to Big Data platform, drivers for Big Data, Big Data architecture and characteristics, 5 Vs of Big Data, Big Data technology components, Big Data importance and applications. Big Data features – security, compliance, auditing and protection, Big Data privacy and ethics, Big Data Analytics, Challenges of conventional systems, intelligent data analysis, nature of data, analytic processes and tools, analysis vs reporting, modern data analytic tools. (6 Hrs)
Unit-II	**Hadoop** History of Hadoop, Apache Hadoop, the Hadoop Distributed File System, components of Hadoop, data format, analyzing data with Hadoop, scaling out, Hadoop streaming, Hadoop pipes, Hadoop Echo System. Map Reduce: Map Reduce framework and basics, how Map Reduce works, developing a Map

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	Reduce application, unit tests with MR unit, test data and local tests, anatomy of a Map Reduce job run, failures, job scheduling, shuffle and sort, task execution, Map Reduce types, input formats, output formats, Map Reduce features, Real-world Map Reduce (06 Hrs)
Unit-III	**HDFS (Hadoop Distributed File System)** Design of HDFS, HDFS concepts, benefits and challenges, file sizes, block sizes and block abstraction in HDFS, data replication, how does HDFS store, read, and write files, Java interfaces to HDFS, command line interface. Hadoop file system interfaces, data flow, data ingest with Flume and Scoop, Hadoop archives, Hadoop I/O: compression, serialization, Avro and file-based data structures. Hadoop Environment: Setting up a Hadoop cluster, cluster specification, cluster setup and installation, Hadoop configuration, security in Hadoop, administering Hadoop, HDFS monitoring & maintenance, Hadoop benchmarks, Hadoop in the cloud (06 Hrs)
Unit-IV	**Hadoop Eco System and YARN** Hadoop ecosystem components, schedulers, fair and capacity, Hadoop 2.0 New Features: Name Node high availability, HDFS federation, MRv2, YARN, Running MRv1 in YARN. NoSQL Databases: Introduction to NoSQL MongoDB: Introduction, data types, creating, updating and deleing documents, querying, introduction to indexing, capped collections Spark: Installing spark, spark applications, jobs, stages and tasks, Resilient Distributed Databases, anatomy of a Spark job run, Spark on YARN SCALA: Introduction, classes and objects, basic types and operators, built-in control structures, functions and closures, inheritance. (07 Hrs)
Unit-V	**Hadoop Eco System Frameworks** Applications on Big Data using Pig, Hive and HBase Pig - Introduction to PIG,

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators (07 Hrs)
Unit-VI	**Hive** Apache Hive architecture and installation, Hive shell, Hive services, Hive metastore, comparison with traditional databases, HiveQL, tables, querying data and userdefined functions, sorting and aggregating, Map Reduce scripts, joins & subqueries. HBase: Hbase concepts, clients, example, Hbase vs RDBMS, advanced usage, schema design, advance indexing, Zookeeper: how it helps in monitoring a cluster, how to build applications with Zookeeper. IBM Big Data strategy, introduction to Infosphere, BigInsights and Big Sheets, introduction to Big SQL. (07 Hrs)

References

Sr. No.	Title	Author	Publication	Edition
1	Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses	Michael Minelli, Michelle Chambers, and Ambiga Dhiraj	Wiley	
2	Big-Data Black Book		DT Editorial Services, Wiley	
3	Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data	Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch	McGrawHill	
4	Big Data Analytics	Raj Kamal, Preeti Saxena	McGrawHill	

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA143 Course: Data Visualization and Interpretation **Teaching Scheme:** Theory: 03 Hrs/week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisites	AI Basics
Objectives	1. Understand fundamentals of Data Visualization. 2. Understand the advanced concepts of data interpretation.
Unit-I	**Introduction to Data Visualization** How visualization affects data interpretation, Role of visualization in data science, Two flavors of data visualization: exploratory and communicative, Data Visualization Tools (6 Hrs)
Unit-II	**Visualization design principles** Data and task abstraction, Best practices for encoding, Marks and channels, Effectiveness and expressiveness, How to critique visualizations, Design problems and consequences, How not to cause misinterpretation (6 Hrs)
Unit-III	**Exploratory Visualization techniques** Handling high-dimensional data, Comparison techniques, Small multiples, Handling uncertainty, Depicting time (6 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-IV	**Interactive Visualization** Why interactivity is needed, handling multiple views, Brushing and Linking, Data Interpretability : Causes of misinterpretation, Role of communicative visualization in sciences, Ginterpretability: metrics and approaches. **(6 Hrs)**
Unit-V	**Interpretation of Data** What Is Data Interpretation? How To Interpret Data?, Why Data Interpretation Is Important?, Data Analysis & Interpretation Problems, Data Interpretation Techniques & Methods, The Use of Dashboards For Data Interpretation **(6 Hrs)**
Unit-VI	**Visualization in data science: methods and examples** Explaining machine learning models, Interpretability challenges and solutions, Transparency and human-machine trust, Impactful case studies: biology, healthcare, cyber security, climate science, social science **(6 Hrs)**

Textbooks / Reference Books

Sr. No.	Title	Author	Publication	Edition
1	Effective Data Visualization: The Right Chart for the Right Data (Old Edition)	Dr. Stephanie D. H. Evergreen	SAGE Publications	**2016**
2	Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures	Claus O. Wilke	O'Reilly	**2019**
3	Storytelling With Data: A Data Visualization Guide For Business Professionals	Nussbaumer Knaflic, Cole	Wiley	1st

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology)

Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II

Course Code: MDA144 Course: Data Science and Security **Teaching Scheme:** Theory: 03 Hrs/week Tutorial: 01Hr/week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisites	No Prerequisites
Objectives	1. To give introduction to data security and related software 2. To learn information security, risk management 3. To study Protecting and securing data/Information, backup
Unit-I	**Introduction** International Laws and Standards, Data Protection Act, Trusted Computing Group, Data Breach, Data Remanence, Data Theft, Wireless Identity Theft (6 Hrs)
Unit-II	**Information Security** Data and Information, terminology of information security, asset, compartmentalization, end-point security, and information sensitivity (6 Hrs)
Unit-III	**Risk Management** Risk Management, Types of Controls, Administrative Control, Logical Controls, Physical Controls, Potential Risk Treatments, Single Loss Expectancy (6 Hrs)
Unit-IV	**Protecting and securing data/Information** Protecting and Securing data/Information, Active Hard Drive Protection, Anomaly Detection, Back Up, Cloud Service Gateway, Firewall, types of Firewall, Security of ATM's (6 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-V	**Backup** Repository Models of backup, Storage Media, Managing the Data Repository, Live Data, Manipulation, Limitations (6 Hrs)
Unit-VI	**Protecting and securing data through software** Introduction of different software for Protecting and securing data, BC Wipe, Data Shredder and Database Monitoring, DBAN, Disk Utility, DHDerase and MyDLP, Shred etc (6 Hrs)

Textbooks / Reference Books

Sr. No.	Title	Author	Publication	Edition
1	Data Science for Cyber-security	Nick Heard, Niall Adams, Patrick Rubin-Delanchy	World Scientific Publishing Europe Ltd	2018
2	Secure Data Science: Integrating Cyber Security and Data Science	Bhavani Thuraisingha, Latifur Khan, Murat Kantarcioglu	CRC Press	1st
3	Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies	Alessandro Parisi	Packt Publishing Limited	2019

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA161 Course: Generative AI and Prompt Engineering **Teaching Scheme:** Theory: 03 Hrs/week Tutorial: 01Hr/week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs

Prerequisites	No Prerequisites
Objectives	1.To give introduction to Generative AI and GAN 2.To learn Technology behind Generative AI and GAN 3.To Explore the concept of prompt engineering
Unit-I	**Introduction** The concept of Generative AI, Discriminative vs. Generative AI, Broad Application Fields and Potential, GAI in Top Strategic, Example: Face Generation, GAI-generated Art AND the Interlock with Crypto & NFTs, GAI-generated Art AND the Interlock with Crypto & NFTsGAI-generated Art AND the Interlock with Crypto & NFTs, The Evolution of Deep Generative Models (6 Hrs)
Unit-II	**Generative Architecture and Generative Adversarial Networks (GAN's)** Generator and discriminator, Cross Entropy, Equation to calculate discriminator loss and generator loss, hyperparameters and Data Loader, generator class, optimizer and testing the generator, loss values of generator and discriminator, training loop, GANs are Powerful, Working of GAN, and Excursion: Neural Networks, The implementation of simple GAN. (6 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II		
Course Code: MDA161 Course: Generative AI and Prompt Engineering **Teaching Scheme:** Theory: 03 Hrs/week Tutorial: 01Hr/week		Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs
Prerequisites	No Prerequisites	
Objectives	1.To give introduction to Generative AI and GAN 2.To learn Technology behind Generative AI and GAN 3.To Explore the concept of prompt engineering	
Unit-I	**Introduction** The concept of Generative AI, Discriminative vs. Generative AI, Broad Application Fields and Potential, GAI in Top Strategic, Example: Face Generation, GAI-generated Art AND the Interlock with Crypto & NFTs, GAI-generated Art AND the Interlock with Crypto & NFTsGAI-generated Art AND the Interlock with Crypto & NFTs, The Evolution of Deep Generative Models (6 Hrs)	
Unit-II	**Generative Architecture and Generative Adversarial Networks (GAN's)** Generator and discriminator, Cross Entropy, Equation to calculate discriminator loss and generator loss, hyperparameters and Data Loader, generator class, optimizer and testing the generator, loss values of generator and discriminator, training loop, GANs are Powerful, Working of GAN, and Excursion: Neural Networks, The implementation of simple GAN. (6 Hrs)	

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-III	**Applications and Benefits of Generative AI** 3D-Object Generation, Interactive Image Generation, Conditional GAN (cGAN), Data Augmentation Techniques , Data Augmentation with a GAN , ChatGPT, Evolutionary strategies, Reinforcement learning, AI towards our own human nature, capable of generating, imagining and creating (6 Hrs)
Unit-IV	**Introduction to Prompt Engineering and Language Model Architectures and Techniques** Fundamentals of prompt engineering, importance of prompt design in shaping language model behavior, drafting effective prompts for specific tasks and use cases, The process of drafting and refining prompts, types of Prompting, Priming Prompt, Task Decomposition, Exploring techniques for controlling model behavior through prompt engineering, Ensuring fairness and ethical considerations in prompt design, Leveraging prompt engineering for generating desired outputs, Fine-tuning language, Adapting language models through prompt customization. (6 Hrs)
Unit-V	**Evaluating and Debugging Prompt Responses** Analyzing prompt-response pairs for quality assessment, Identifying and addressing common issues and biases in model outputs, Debugging and refining prompts to improve the reliability of language models (6 Hrs)
Unit-VI	**Multi-modal prompt engineering and Ethical Considerations in Prompt Engineering** Multi-modal prompt engineering, Contextual adaptation and dynamic prompts for improved performance, Cutting-edge research and emerging techniques in prompt engineering, Potential ethical, Fairness, transparency, and responsible AI through prompt design, Mitigation of biases and potential harm in language model outputs (6 Hrs)

Textbooks / Reference	Sr. No.	Title	Author	Publication	Edition
	1	Generative Deep Learning	David Foster	O'Reilly	2nd

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Books				Media	
	2	GANs in Action: Deep learning with Generative Adversarial Networks	Jakub Langr and Vl adimir Bok	Packt Publishing	2019
	3	Generative AI for Business The Essential Guide for Business Leaders	Dr. Gleb Tsipursky		2023
	4	Modern Generative AI with ChatGPT and OpenAI Models	Valentina Alto		2023

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA162 Course: Digital Forensics **Teaching Scheme:** Theory: 03 Hrs/week Tutorial: 01Hr/week	Credits: 3-0-0 Mid Semester Examination-I: 15 Marks Mid Semester Examination-II: 15 Marks Teacher Assessment: 20 Marks End Semester Examination: 50 Marks End Semester Examination (Duration): 02 Hrs
Prerequisites	No Prerequisites
Objectives	1. To introduce the principle and concepts of digital forensic 2. To detail about the various investigation procedures like data acquisition and evidence gathering
Unit-I	**Basics of Digital Forensics** Digital Forensics: Introduction, Objective and Methodology, Rules of Digital Forensics, Good Forensic Practices, Daubert's Standards, Principles of Digital Evidence. Overview of types of Computer Forensics: Network Forensics, Mobile Forensics, Social Media Forensics and E-mail Forensics. Services offered by Digital Forensics. First Responder: Role, Toolkit and Do's and Don'ts (6 Hrs)
Unit-II	**Cyber Crime Investigation** Introduction to Cyber Crime Investigation, Procedure for Search and seizure of digital evidences in cyber-crime incident- Forensics Investigation Process- Pre-search consideration, Acquisition, Duplication & Preservation of evidences, Examination and Analysis of evidences, Storing of Evidences, Documentation and Reporting, Maintaining the Chain of Custody. (6 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

Unit-III	**Data Acquisition and Evidence Gathering** Data Acquisition of live system, Shutdown Systems and Remote systems, servers. E-mail Investigations, Password Cracking. Seizing and preserving mobile devices. Methods of data acquisition of evidence from mobile devices. Data Acquisition and Evidence Gathering from social media, Performing Data Acquisition of encrypted systems. Challenges and issues in cyber-crime investigation. (6 Hrs)
Unit-IV	**Analysis of Digital Evidence** Search and Seizure of Volatile and Non-volatile Digital Evidence, Imaging and Hashing of Digital Evidence, Introduction to Deleted File Recovery, Steganography and Steganalysis, Data Recovery Tools and Procedures, Duplication and Preservation of Digital Evidence, Recover Internet Usage Data, Recover Swap files/Temporary Files/Cache Files. Software and Hardware tools used in cybercrime investigation: Open Source and Proprietary tools. Importance of Log Analysis in forensic analysis. Understanding Storage Formats for Digital Evidence: Raw Format, Proprietary Formats, Advanced Forensic Formats. (6 Hrs)
Unit-V	**Windows and Linux Forensics** Windows Systems Artifacts: File Systems, Registry, Event logs, Shortcut files, Executables. Alternate Data Streams (ADS), Hidden files, Slack Space, Disk Encryption, Windows registry, startup tasks, jump lists, Volume Shadow, shell bags, LNK files, Recycle Bin Forensics (INFO, $i, $r files) (6 Hrs)
Unit-VI	**Forensic Analysisof the Registry** Use of registry viewers, Regedit. Extracting USB related artifacts and examination of protected storages. Linux System Artifact: Ownership and Permissions, Hidden files, User Accounts and Logs. (6 Hrs)

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	Sr. No.	Title	Author	Publication	Edition
Textbooks / Reference Books	1	Cyber Security: Understanding , Computer Forensics and Legal Perspectives	Nina Godbole and Sunit Belapore	Wiley Publications	2011
	2	Guide to Computer Forensics and Investigations	Bill Nelson, Amelia Phillips and Christopher Steuart	Cengage	3rd
	3	All in One CISSP Guide	Shon Harris	McGraw Hill	6th 2013
	4	The Best Damn Cybercrime and Digital Forensic Book	Anthony Reyes, Jack Wiles	Syngress	2007

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA151 Course: Lab- : Exploratory Data Analytics **Teaching Scheme:** Practical: 2 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks
Objectives	1. Understand the basic concepts of Data analytics. 2. Use data analytic techniques for decision making.
List of Practical	1. Introduction to R, Rstudio, and RMarkdown and perform Data cleaning and aggregation. 2. Create visualizations and dashboards with spreadsheets. 3. Perform fundamental data wrangling tasks that, together, form the pre-processing phase of data analysis. These tasks include handling missing values in data, formatting data to standardize it and make it consistent, normalizing data, grouping data values into bins, and converting categorical variables into numerical quantitative variables. 4. perform computations on the data to calculate basic descriptive statistical information and use the Pearson correlation method to compare two continuous numerical variables, use the Chi-square test to find the association between two categorical variables and how to interpret them. 5. Develop the model. a. define the explanatory variable and the response variable. b. evaluate a model using visualization. c. use the R-squared and the mean square error measures to perform in-sample evaluations to numerically evaluate the model. d. Implement Prediction and Decision Making to check the model accuracy 6. Model Evaluation a. identify overfitting and underfitting in a predictive model. b. using Ridge Regression to regularize and reduce standard errors to prevent overfitting a regression model c. use the Grid Search method to tune the hyperparameters of an estimator

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	7. Implement following probability distributions a. Binomial Distribution b. Poisson Distribution c. Normal Distribution 8. Collect and analyze sample data avoiding sampling bias and accurate estimates by using sampling distributions. 9. Mini Project

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

<table>
<tr><td colspan="2" align="center">(Faculty of Science & Technology)
Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II</td></tr>
<tr><td>Course Code: MDA152
Course: Lab- : Big Data Analytics
Teaching Scheme:
Practical: 2 Hrs/week</td><td>Credits: 0-0-1
Teacher Assessment: 25 Marks</td></tr>
<tr><td>Objectives</td><td>1. Demonstrate knowledge of Big Data Analytics concepts and its applications in business.
2. Demonstrate functions and components of Map Reduce Framework and HDFS. Discuss Data Management concepts in NoSQL environment.
3. Explain process of developing Map Reduce based distributed processing applications.
4. Explain process of developing applications using HBASE, Hive, Pig etc.</td></tr>
<tr><td>List of Practical</td><td>1. Install Apache Hadoop
2. Develop a MapReduce program to calculate the frequency of a given word in each file.
3. Perform File Management tasks in Hadoop.
4. Develop Word Count Map Reduce program to understand Map Reduce Paradigm
5. Weather Report POC-Map Reduce program to analyse time-temperature statistics and generate report with max/min temperature.
6. Implement Matrix Multiplication with Hadoop Map Reduce
7. Write queries to sort and aggregate the data in a table using HiveQL.
8. Develop an application to find the maximum temperature using Spark.</td></tr>
</table>

Chairma… …oard of Studies
Computer …ence & Engineering
MI… Aurangabad
(An Autonomous Institute)

Master Copy

(Faculty of Science & Technology) Syllabus of M. Tech. CSE (Data Science and Analytics) Semester-II	
Course Code: MDA153 Course: Lab-I Data Visualization and Interpretation Teaching Scheme: Practical: 02 Hrs/week	Credits: 0-0-1 Teacher Assessment: 25 Marks
Objectives	To acquire practical knowledge of data visualization and interpretation.
List of Practical	The laboratory work includes. 1 Introduction to Data Visualization and Data Visualization Tools 2 Create a diverse range of plots using Python (Matplotlib and Seaborn) and R Programming. 2.1 Area Plots 2.2 Histograms 2.3 Bar Charts 2.4 Pie Charts 2.5 Box Plots 2.6 Scatter Plots 3 Creating Dashboards with Plotly and Dash 3.1 Introduction to Plotly 3.2 Introduction to Dash 3.3 Make Dashboards Interactive 4 Data Visualization and Dashboards with Excel 4.1 Create basic visualizations such as line graphs, bar graphs, and pie charts using Excel spreadsheets. 4.2 Explain the important role charts play in telling a data-driven story. 4.3 Construct advanced charts and visualizations such as Treemaps, Sparklines, Histogram, Scatter Plots, and Filled Map Charts. 4.4 Build and share interactive dashboards using Excel

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy

	5 Fundamentals of Visualization with Tableau 5.1 Install Tableau Public Software and create a visualization 5.2 Examine and navigate the Tableau Public workspace 5.3 Practice and connect to different data sources 6 Visual Analytics with Tableau 6.1 Create a chart using Tableau 6.2 Create dates using calculated fields 6.3 Customize table calculations 6.4 Customize and create dual layer maps 7 Creating Dashboards and Storytelling with Tableau 7.1 Combine the data and follow the best practices to present your story 7.2 Create calculated fields for KPIs to build a figure that will be used to measure progress in the data 7.3 Assemble a dashboard 8 Introduction to Power BI 9 Data Visualization with Power BI 10 Mini Project

Chairman Board of Studies
Computer Science & Engineering
MIT Aurangabad
(An Autonomous Institute)

Master Copy