Scheme of Examinations and Syllabus
M.Sc. in Computer Science (Data Science & Machine Learning) Program
w.e.f. 2022-23

Program Specific Outcomes:

The students upon completion of **M.Sc. in Computer Science (Data Science & Machine Learning) Program** will be able:

PSO1: To become a skilled Data Scientist in industry, academia, or government.

PSO2: To use specialist software tools for data storage, analysis and visualization.

PSO3: To be able to independently carry out research/investigation to solve practical problems.

PSO4: To be able to apply the concepts of Artificial Intelligence and Machine Learning with practical knowledge in analysis, design and development of intelligent systems and applications to multi-disciplinary problems.

PSO5: To provide a concrete foundation to the students in the cutting edge areas of Machine Learning and excelling in the specialized areas like Natural Language Processing, Computer Vision, Reinforcement Learning, Internet of Things, Cloud computing, Data Security and privacy etc.

M.Sc. in Computer Science (Data Science & Machine Learning) First Year
First Semester

Paper Code	Nomenclature	University Exams	Internal Assessment	Total Marks	Credits (L:T:P)
22MDM21C1	Mathematical Foundation for Data Science	80	20	100	4:0:0
22MDM21C2	Information Retrieval and Data Mining	80	20	100	4:0:0
22MDM21C3	Artificial and Computational Intelligence	80	20	100	4:0:0
22MDM21C4	Machine Learning and Python Programming	80	20	100	4:0:0
22MDM 21CL1	Software Lab-I (Based on 22MDM21C2, 22MDM21C3, 22MDM21C4)	100*	-----	100	0:0:4
					Total Credits=20

Second Semester

Paper Code	Nomenclature	University Exams	Internal Assessment	Total Marks	Credits (L:T:P)
22MDM22C1	Introduction to Data Science	80	20	100	4:0:0
22MDM22C2	Big Data Analytics and R Programming	80	20	100	4:0:0
22MDM22C3	Computer Vision and Image Processing	80	20	100	4:0:0
22MDM22C4	Cloud Computing and IoT	80	20	100	4:0:0
22MDM 22CL1	Software Lab-II (Based on 22MDM22C2, 22MDM22C3 and 22MDM22C4)	100*	-----	100	0:0:4
					Total Credits=20
Foundation Elective (F)					
To be chosen from the pool of Foundation Electives provided by the university.					2 Credits
Open Elective (O)					
To be chosen from the pool of Open Electives provided by the University (excluding the Open Elective offered by the Department of Computer Science & Applications)					3 Credits

Total Credits: 25

*20 marks out of 100 will be based on the evaluation/assessment of the candidate in Test(s) and Assignment(s) during the semester, which will be forwarded by the Head of Dept./Director/Principal to the Examiner(s).

M.Sc. in Computer Science (Data Science & Machine Learning) Second Year

Third Semester

Paper Code	Nomenclature	University Exams	Internal Assessment	Total Marks	Credits (L:T:P)
23MDM23C1	Deep Learning	80	20	100	4:0:0
23MDM23C2	Data Visualization and Interpretation	80	20	100	4:0:0
23MDM23DA1/ 23MDM23DA2/ 23MDM23DA3	Natural Language Processing Or Stream Processing and Analytics Or Time Series Analysis	80	20	100	4:0:0
23MDM23DB1/ 23MDM23DB2/ 23MDM23DB3	Reinforcement Learning Or Information Security and Privacy Or Predictive Analytics for IoT	80	20	100	4:0:0
23MDM23CL1	Software Lab-III (Based on 23MDM23C1, C2, 23MDM23DA1/2/3, DB1/2/3)	100*	-----	100	0:0:4 **Total Credits=20**
	Open Elective (O)				
To be chosen from the pool of Open Electives provided by the University (excluding the Open offered by the Department of Computer Science & Applications)					3 Credits

Total Credits : 23

Fourth Semester

Paper Code	Nomenclature	University Exams	Internal Assessment	Total Marks	Credits (L:T:P)
23MDM24C1	Blockchain Technology	80	20	100	4:0:0
23MDM24DA1/ 23MDM24DA2/ 23MDM24DA3	AR/VR Systems and Wearable Computing Or Data Wrangling Or Graph Algorithms and Mining	80	20	100	4:0:0
23MDM24DB1/ 23MDM24DB2/ 23MDM24DB3	Social Network Analysis Or High Dimensional Data Or Artificial Intelligence and Decision Sciences	80	20	100	4:0:0
23MDM24CL1	Software Lab-IV (Based on 23MDM24C1, 23MDM24DA1/2/3, DB1/2/3)	100*	-----	100	0:0:4
23MDM24C3	Project Report/Dissertation	100**	-----	100	0:4:0
					Total Credits=20

Overall Credits : 88

*20 marks out of 100 will be based on the evaluation/assessment of the candidate in Test(s) and Assignment(s) during thesemester, which will be forwarded by the Head of Dept./Director/Principal to the Examiner(s).

**20 marks out of 100 will be based on the progress of the candidate in the Project/Dissertation assigned during the semester, which will be forwarded by the Head of Dept./Director/Principal to the Examiner(s).

M.Sc. in Computer Science (Data Science & Machine Learning) First Year First Semester

SUBJECT: MATHEMATICAL FOUNDATION FOR DATA SCIENCE

PAPER CODE: 22MDM21C1

Course Outcomes:
By the end of the course the students will be able to:

CO1: Apply measures of central tendency to analyze a payroll dataset.
CO2: Apply probabilistic model for credit card fraud detection.
CO3: Evaluate covariance and correlation of between two variables.
CO4: Demonstrate use eigenvalues and eigenvectors for a reducing dimension of a healthcare dataset
CO5: Apply simple regression model to predict the near future sales based on a time series data.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Discrete mathematics for Data Science: Concept of set, cardinality of set, finite, infinite and uncountably infinite sets, Basic set operations, Principal of inclusion Exclusion, Graph: Basic terminologies, representation of graph, path and circuit, graph traversal, travelling salesperson problem, Trees: Basic terminologies, search tree: Binary & M-ary tree.

UNIT-II

Data Analysis & Probability Theory: Data Representation, Average, Spread, Experiments, Outcomes, Events, Probability, Permutations and Combinations, Random Variables, Probability Distributions, Mean and Variance of a Distribution, Binomial, Poisson, and Hyper geometric Distributions, Normal Distribution, Distributions of Several Random Variables.

UNIT-III

Statistical Inference: Types of Statistical Inference, Descriptive Statistics, Inferential Statistics, Importance of Statistical Inference in Machine Learning, Descriptive Statistics, Measures of Central Tendency: Mean, Median, Mode, Midrange, Measures of Dispersion: Range, Variance, Mean Deviation, Standard Deviation. Coefficient of variation: Moments, Skewness, Kurtosis, One sample hypothesis testing, hypothesis, Testing of Hypothesis, Binomial distribution and normal distribution, Chi-Square Tests, t-test, ANOVA. Pearson Correlation.

Measure of Relationship: Covariance, Karl Pearson's Coefficient of Correlation, Measures of Position: Percentile, Z-score, Quartiles, Bayes' Theorem, Bayes Classifier, Bayesian network, Probabilistic models with hidden variables

UNIT-IV

Linear Algebra: Matrix and vector algebra, systems of linear equations using matrices, linear independence, Matrix factorization concept/LU decomposition, Eigen values and eigenvectors. **Understanding of calculus:** concept of function and derivative, Multivariate calculus: concept, Partial Derivatives, chain rule, the Jacobian and the Hessian.

Regression Model: Introduction, types of regression. Simple regression- Types, Making predictions, Cost function, Gradient descent, Training, Model evaluation. Multivariable regression: Growing complexity, Normalization, making predictions, initialize weights, Cost function, Simplifying with matrices, Bias term, Model evaluation.

Suggested Readings:

1. Bruce, Peter, Andrew Bruce, and Peter Gedeck: Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, O'Reilly Media.
2. Liu, Chung Laung: Elements of Discrete Mathematics, Tata McGraw-Hill Education.
3. Heumann, Christian, Schomaker, Michael, Shalabh: Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R, Springer.
4. Douglas C. Montgomery, George C. Runger: Applied Statistics and Probability for Engineers, Wiley (Low price edition)
5. Robert V. Hogg. Allen T. Craig: Introduction to Mathematics and Statistics, Pearson Education.
6. Richard A. Johnson, Irwin Miller, John Freund: Probability and Statistics for Engineers.
7. Irwin Miller, Marylees Miller: Mathematical Statistics with Applications, Pearson Education.
8. Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoit Liquet: The R Software-Fundamentals of Programming and Statistical Analysis, Springer.
9. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: INFORMATION RETRIEVAL AND DATA MINING

PAPERCODE: 22MDM21C2

Course Outcomes:
By the end of the course the students will be able to:

CO1: Understand the basic concepts of the information retrieval.
CO2: Analyse the involvement of the information retrieval in modern life style & social media.
CO3: Apply data pre-processing, indexing, retrieval methods and concepts.
CO4: Evaluate the effectiveness and efficiency of different information retrieval systems
CO5: Understand the basic concepts of Data Warehouse and analysis and mining of data.

Maximum Marks: 100(External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Introduction to Information Retrieval Systems: Definition and Objectives of Information Retrieval Systems, Functional Overview, Relationship to Database Management Systems, Digital Libraries and Data Warehouses.

Information Retrieval System Capabilities: Search Capabilities, Browse Capabilities, Miscellaneous Capabilities

Cataloging and Indexing: History and Objectives of Indexing, Indexing Process, Automatic Indexing, Information Extraction.

Data Structure: Introduction to Data Structure, Stemming Algorithms, Inverted File Structure, N-Gram Data Structures, PAT Data Structure, Signature File Structure, Hypertext and XML Data Structures, Hidden Markov Models.

UNIT-II

Automatic Indexing: Classes of Automatic Indexing, Statistical Indexing, Natural Language, Concept Indexing, Hypertext Linkages.

Document and Term Clustering: Introduction to Clustering, Thesaurus Generation, Item Clustering, Hierarchy of Clusters.

User Search Techniques: Search Statements and Binding, Similarity Measures and Ranking, Relevance Feedback, Selective Dissemination of Information Search, Weighted Searches of Boolean Systems, Searching the INTERNET and Hypertext.

UNIT-III

Information Visualization: Introduction to Information Visualization, Cognition and Perception, Information Visualization Technologies.

Text Search Algorithms: Introduction to Text Search Techniques, Software Text Search Algorithms, Hardware Text Search Systems.

Multimedia Information Retrieval: Spoken Language Audio Retrieval, Non-Speech Audio Retrieval, Graph Retrieval, Imagery Retrieval, Video Retrieval.

UNIT-IV

Data Warehousing: OLAP, Dimensional Modelling (facts, dimensions), cube, Schema, defining schema's star schema, snow-flakes schema and fact constellation, ETL process.

Overview of Data Mining: Concept of Data Mining, Association rules, Knowledge Discovery from Databases, Classification, and Clustering.

Classification methods: Decision tree(ID3,C4.5,CART), Bayesian Classification, Rule based, Neural Network, Lazy and Eager Learners, Parameters for measuring Accuracy.

Data Mining Prediction methods: Linear and nonlinear regression, Logistic Regression Use of open source data mining tool – WEKA, XLMiner, MOA.

Suggested Readings:

1. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze: Introduction to information retrieval, Cambridge University Press.
2. J. Han, M Kamber: Data Mining Concepts and Techniques, Morgan Kaufmann.
3. M. Dunham: Data Mining - Introductory and Advance Topics, Pearson Education.
4. F. Wilfrid Lancaster: Information Retrieval Systems: Characteristics, Testing and Evaluation, New York: Wiley.
5. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

PAPER CODE: 22MDM21C3

Course Outcomes:
By the end of the course the students will be able to:

CO1: Identify the need of Intelligent Agents in problem solving and learn the concept of artificial intelligence, problem solving and searching process.
CO2: Understand the concepts of knowledge and handling of uncertain and probabilistic knowledge
CO3: Design and analyze a learning technique for a given system in different AI application domains like marketing, healthcare, banking, finance, education, etc.
CO4: Learn the concepts of computational intelligence evolutionary computation and neural networks.
CO5: Handle the uncertainty in knowledge using fuzzy logic and understand concepts of fuzzy logic.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Introduction and Intelligent Agents: What is AI? Foundations History of Artificial Intelligence; State of the Art Intelligent Agents: Agents and Environments; Good Behavior: Concept of Rationality, Nature of Environments, and Structure of Agents.

Case Studies (if any): Intelligent agents in autonomous systems.

Problem-solving: Solving Problems by Searching: Problem-Solving Agents, Uninformed Search Strategies, Informed (Heuristic) Search Strategies, Heuristic Functions, Beyond Classical Search Local Search Algorithms and Optimization Problems, Local Search in Continuous Spaces, Searching with Nondeterministic Actions, Searching with Partial Observations, Online Search Agents and Unknown Environments.

Case Studies (if any): Search techniques for a sliding tile problem.

UNIT-II

Knowledge, reasoning, and planning: Knowledge based Agents, Types of knowledge, Knowledge acquisition and its techniques; Knowledge representation: Level of representation; First-Order Logic and Its Inference;

Reasoning and Uncertain knowledge: What is reasoning? Types of reasoning, Quantifying Uncertainty, Probabilistic Reasoning, Probabilistic Reasoning over Time, Bayes Theorem in reasoning, Bayesian Belief Network, Making Simple Decisions, Making Complex Decisions.

Planning: Components of a Planning system, Classical Planning, Planning and Acting in the Real World.

Case Studies (if any): Application of planning to a production system

UNIT-III

Computational Intelligence: Introduction to Computational Intelligence, Biological and Artificial Neural **N**etwork (ANN), artificial neural network models; learning in artificial neural networks; neural network and its applications.

Evolutionary Computing & Optimization: Fundamentals of evolutionary computation, Design and Analysis of Genetic Algorithms, Evolutionary Strategies, comparison of GA and traditional search methods. Genetic Operators and Parameters, Genetic Algorithms in Problem Solving; Optimization: Particle Swarm Optimization, Ant Colony Optimization, Artificial Immune Systems; Other Algorithms: Harmony Search, Honey-Bee Optimization, Memetic Algorithms, Co-Evolution, Multi-Objective Optimization, Tabu Search, Constraint Handling.

UNIT-IV

Fuzzy Systems: Crisp sets, Fuzzy sets: Basic types and concepts, characteristics and significance of paradigm shift, Representation of fuzzy sets, Operations, membership functions, Classical relations and fuzzy relations, fuzzyfication, defuzzyfication, fuzzy reasoning, fuzzy inference systems, fuzzy control system, fuzzy clustering, applications of fuzzy systems. Neuro-fuzzy systems, neuro-fuzzy modeling; neuro-fuzzy control.

Case Studies: Credit card Fraud Analysis, Sentiment Analysis, Recommendation Systems and Collaborative filtering, Uber Alternative Routing.

Suggested Readings:

1. Russell S. and Norvig P.: Artificial Intelligence: A Modern Approach, Prentice-Hall.
2. Elaine Rich, Kevin Knight and Nair: Artificial Intelligence, TMH.
3. Luger G. F. and Stubblefield W. A.: Artificial Intelligence: Structures and strategies for Complex Problem Solving, Addison Wesley.
4. Nilsson Nils J.: Artificial Intelligence: A New Synthesis, Morgan Kaufmann Publishers Inc.
5. Patrick Henry Winston: Artificial Intelligence, Addison-Wesley Publishing Company.
6. M. Mitchell: An Introduction to Genetic Algorithms, Prentice-Hall.
7. J.S.R. Jang, C.T. Sun and E. Mizutani: Neuro-Fuzzy and Soft Computing, PHI, Pearson Education.
8. Davis E. Goldberg: Genetic Algorithms: Search, Optimization and Machine Learning, Addison Wesley.
9. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: MACHINE LEARNING AND PYTHON PROGRAMMING

PAPER CODE: 22MDM21C4

Course Outcomes:
By the end of the course the students will be able to:

CO1: Acquire fundamental knowledge of learning theory
CO2: Design and evaluate various machine learning algorithms
CO3: Use machine learning methods for multivariate data analysis in various scientific fields
CO4: Choose and apply appropriate Machine Learning Techniques for analysis, forecasting, categorization and clustering of the data
CO5: Using Python programming for various machine learning problems.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Machine Learning Concepts: 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 non-parametric models, Predictive and descriptive learning, Classification concepts, Binary and multi-class classification.

Learning Theory: 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 Chhernoff/Hoeffding bounds, VC dimension, Probably Approximately Correct (PAC) learning, Concept learning, the hypothesis space, Least general generalization, Internal disjunction, Paths through the hypothesis space, model Evaluation and selection

UNIT-II

Geometric Models: Regression, Logistic regression , Assessing performance of regression - Error measures, Overfitting, Least square method, Multivariate Linear regression, Regression for Classification, Perceptron, Multi-layer perceptron, Simple neural network, Kernel based methods, Support vector machines(SVM), Soft margin SVM, Support Vector Machines as a linear and non-linear classifier, Limitations of SVM, Concept of Relevance Vector, K-nearest neighbor algorithm

Logical, Grouping And Grading Models: 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-III

Probabilistic Models: Uncertainty, Normal distribution and its geometric interpretations, Baye's theorem, Naïve Bayes Classifier, Bayesian network, Discriminative learning with maximum likelihood, Probabilistic models with hidden variables, Hidden Markov model, Expectation Maximization methods, Gaussian Mixtures and compression based models.

Case Studies on Advanced Machine Learning Techniques: Diagnosis of human disease, Diagnosis of crop disease, Text mining tasks like semantic analysis, author profiling, author identification, language identification, summarization etc., Prediction & forecasting, Fraud detection, Learning to rate vulnerabilities and predict exploits.

UNIT-IV

Programming with Python: Introduction to Python and Computer Programming; Data Types, Variables, Basic Input-Output Operations, Basic Operators; Boolean Values, Conditional Execution, Loops, Lists and List Processing, Logical and Bitwise Operations; Functions, Tuples, Dictionaries, and Data Processing; Modules, Packages, String and List Methods, and Exceptions; The Object-Oriented Approach: Classes, Methods, Objects, and the Standard Objective Features; Exception Handling, and Working with Files.

Suggested Readings:

1. C.M. Bishop: Pattern Recognition and Machine learning, Springer.
2. Hastie, Tibshirani, Friedman: Introduction to statistical machine learning with applications in R, Springer.
3. Tom Mitchell: Machine Learning, McGraw Hill.
4. Parag Kulkarni: Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEEPress.
5. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SOFTWARE LAB-I
PAPER CODE: 22MDM21CL1

(Based on 22MDM21C2, 22MDM21C3 and 22MDM21C4)

Course Outcomes:
By the end of the course the students will be able to:

CO1: Acquire programming skills in core Python.
CO2: Acquire Object Oriented Skills in Python.
CO3: Develop the skill of designing Graphical user Interfaces in Python.
CO4: Develop the ability to write database applications in Python.
CO5: Gain exposure of several relevant case studies/applications.

M.Sc. in Computer Science (Data Science & Machine Learning) First Year Second Semester

SUBJECT: INTRODUCTION TO DATA SCIENCE

PAPER CODE: 22MDM22C1

Course Outcomes:
By the end of the course the students will be able to:

CO1: Apply data science processes to an e-commerce data and demonstrate the use of estimation methods for analyzing this data.
CO2: Compare and apply appropriate machine learning algorithms for classification.
CO3: Compare and choose one data visualization method for effective visualization of data.
CO4: Design a model of recommendation system based on the content of the data.
CO5: Apply standard clustering methods to analyze social network graph.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Introduction to Data Science: What is Data Science, importance of data science, Big data and data Science, The current Scenario, Industry Perspective Types of Data: Structured vs. Unstructured Data, Quantitative vs. Categorical Data, Big Data vs. Little Data, Data science process, Role of Data Scientist.

Case Studies (if any): Ecommerce Marketplace.

UNIT-II

Statistical Interference and Exploratory Data Analysis: Introduction-Population and samples, Data Preparation, Exploratory Data Analysis-Summarizing Data, Data Distribution, Outlier Treatment, Measuring Symmetry, Continuous Distribution, Kernel Density, Estimation: Sample and Estimated Mean, Variance and Standard Scores, Covariance, and Pearson's and Spearman's Rank Correlation.

Machine Learning Algorithms: Linear Regression, K-nearest Neighbors (K-NN), K-mean, Spam Filters, Naïve Bayes, and Wrangling: Naive Bayes, Comparing Naive Bayes to k-NN, Scraping the Web: APIs and Other Tools.

UNIT-III

Data Visualisation: Introduction, Types of data visualisation, Data for visualisation: Data types, Data encodings, Retinal variables, Mapping variables to encodings, Visual encodings.

Recommendation Systems: A Model for Recommendation Systems - The Utility Matrix, The Long Tail, Applications of Recommendation Systems, Populating the Utility Matrix; **Content-Based**

Recommendations: Item Profiles, Discovering Features of Documents, Obtaining Item Features From Tags, Representing Item Profiles, User Profiles, Recommending Items to Users Based on Content; **Collaborative Filtering:** Measuring Similarity, The Duality of Similarity, Clustering Users and Items, Evaluation of Recommendation System

Case Studies (if any): Movie Lens Case Study.

UNIT-IV

Social Network Analysis: Social Networks as Graphs, Varieties of Social Networks, Graphs With Several Node Types, Clustering of Social-Network Graphs: Distance Measures for Social-Network Graphs, Applying Standard Clustering Methods, Betweenness, The Girvan-Newman Algorithm, Using Betweenness to Find Communities

Case Studies (if any): Community detection in social network.

Suggested Readings:

1. Russell S. and Norvig P.: Artificial Intelligence: A Modern Approach, Prentice-Hall.
2. Cathy O'Neil and Rachel Schutt: Doing Data Science, Straight Talk From The Frontline, O'Reilly.
3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman: Mining of Massive Datasets, Cambridge University Press.
4. Laura Igual and Santi Segui: Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications, Springer.
5. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: BIG DATA ANALYTICS AND R PROGRAMMING

PAPER CODE: 22MDM22C2

Course Outcomes:
By the end of the course the students will be able to:

CO1: Design the data analytics life cycle for selected problem statement.
CO2: Develop insights into the big data and present results for selected problem statement through visualization techniques.
CO3: Demonstrate the use of Hadoop and its ecosystem elements to analyze big data.
CO4: Demonstrate use of advanced FOSS computing environments for big health care data.
CO5: Exhibit R programming skills.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Basics of Big data: characteristics, types, sources, architectures, Data analysis process, Data analytics lifecycle, Pre-processing data, Market and Business Drivers for Big Data Analytics, Business Problems Suited to Big Data Analytics.

Case Studies (if any): Case study on data analytics lifecycle.

UNIT-II

Technologies for big data analytics: 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.

Case Studies (if any): Using MapReduce to scale algorithms for Big Data analytics.

UNIT-III

Hadoop ecosystem: Sqoop, Impala, Apache Flume, Pig, Hive, Data transformation and analysis using Pig, Data analysis using Hive and Impala, Mahout, Oozie, Zookeeper etc.

Case Studies (if any): Sentiment analysis.

Big data analytics with Apache Spark: Apache Spark, Spark core, Interactive data analysis with spark shell, Writing a spark application, Spark RDD Optimization Techniques, Spark Algorithm, Spark SQL.

UNIT-IV

R Programming: R interpreter, Introduction to major R data structures like vectors, matrices, arrays, list

and data frames, Control Structures, vectorized if and multiple selection, functions.

Installing, loading and using packages: Read/write data from/in files, extracting data from web-sites, Clean data, Transform data by sorting, adding/removing new/existing columns, centering, scaling and normalizing the data values, converting types of values, using string in-built functions.

Designing GUI: Building interactive application and connecting it with database.

Programming languages for Big data analytics: 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.

Suggested Readings:

1. DT Editorial Services: Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization.
2. David Dietrich, Barry Hiller: Data Science and Big Data Analytics, EMC education services, Wiley Publications.
3. Mohammed Guller: Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large Scale Data Analysis.
4. David Loshin: Big Data Analytics From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, Morgan Kaufmann.
5. Venkat Ankam, "Big Data Analytics", Packt Publishing.
6. Jenny Kim, Benjamin Bengfort: Data Analytics with Hadoop, O'Reilly Media, Inc.
7. Glenn J. Myatt: Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining.
8. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: COMPUTER VISION AND IMAGE PROCESSING

PAPER CODE: 22MDM22C3

Course Outcomes:
By the end of the course the students will be able to:

CO1: Understand the fundamental of Computer vision and Machine Learning for Computer vision.
CO2: Understand the foundation of Image Processing and concepts of Image enhancement
CO3: Understand the concepts of Image restoration and feature extraction.
CO4: Know about segmentation and compression of image.
CO5: Develop computer image processing programs and applications using OpenCV-Python Libraries.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT - I

Computer Vision: Concepts of Computer vision, Computer vision vs Image Processing, Geometric techniques in Computer vision- Image transformations, Camera projections, Camera calibration, Depth from Stereo, Two view structure from motion, Object tracking.

Machine Learning for Computer Vision: Introduction to Machine learning, types of Machine Learning; Image Classification: Supervised and Unsupervised Image classification, Algorithms for Image classification; Object detection: Concepts of Object detection, Methods of Object detection; Image Segmentation: Classes of segmentation, Types of segmentation, Methods of segmentation

UNIT - II

Image Processing Foundations: Review of image processing techniques – classical filtering operations – thresholding techniques – edge detection techniques – corner and interest point detection – mathematical morphology – texture.

Shapes and Regions: Binary shape analysis – connectedness – object labelling and counting – size filtering – distance functions – skeletons and thinning – deformable shape analysis – boundary tracking procedures – active contours – shape models and shape recognition – centroidal profiles – handling occlusion – boundary length measures – boundary descriptors – chain codes – Fourier descriptors – region descriptors – moments.

Image Enhancement: Spatial and Frequency domain-Histogram processing-Spatial filtering-Smoothening spatial filters Sharpening spatial filters- Discrete Fourier Transform-Discrete Cosine Transform-Haar Transform -Hough Transform-Frequency filtering-Smoothening frequency filters-Sharpening frequency filters-Selective filtering.

UNIT – III

Image Restoration & Image Registration: Noise models - Degradation models-Methods to estimate the degradation-Image de-blurring Restoration in the presence of noise only spatial filtering-Periodic noise reduction by frequency domain filtering-Inverse filtering-Wiener Filtering. Geometrical transformation-Point based methods- Surface based methods-Intensity based methods

Feature Extraction: Region of interest (ROI) selection - Feature extraction: Histogram based features -

Intensity features-Color, Shape features-Contour extraction and representation-Homogenous region extraction and representation-Texture descriptors - Feature Selection: Principal Component Analysis (PCA).

Image Segmentation: Discontinuity detection-Edge linking and boundary detection. Thresholding-Region oriented segmentation- Histogram based segmentation. Object recognition based on shape descriptors. Dilation and Erosion-Opening and Closing-Medial axis transforms-Objects skeletons-Thinning boundaries.

UNIT - IV

Image Coding & Compression: Lossless compression versus lossy compression-Measures of the compression efficiency- Hufmann coding-Bitplane coding-Shift codes-Block Truncation coding-Arithmetic coding-Predictive coding techniques-Lossy compression algorithm using the 2-D. DCT transform-The JPEG 2000 standard Baseline lossy JPEG, based on DWT.

OpenCV: Installation of OpenCV-Python, OpenCVStandard Images and Data Sets, Python for IPCV, Python for Image Processing, Contrast Stretching, Linear Filtering, Histogram Equalization, Gaussian Convolution, Separable Gaussian Convolution, Gaussian Derivatives, Comparison of theory and practice, Canny Edge Detector, Histogram of Oriented Gradients, Preprocessing, Calculate the Gradient Images, Calculate HOG in 8x8 Cells, Block Normalization, Calculate the HOG feature vector, Visualizing the HOG.

Case study: Human Iris location, hole detection, Generalized Hough Transform (GHT), spatial matched filtering, GHT for ellipse detection, object location, GHT for feature collation.

Suggested Readings:

1. D. L. Baggio et al.: Mastering OpenCV with Practical Computer Vision Projects‖, Packt Publishing.
2. Jan Erik Solem: Programming Computer Vision with Python: Tools and algorithms for analyzing images, O'Reilly Media.
3. Mark Nixon and Alberto S. Aquado: Feature Extraction & Image Processing for Computer Vision‖, Third Edition, Academic Press.
4. R. Szeliski: Computer Vision: Algorithms and Applications‖, Springer.
5. Simon J. D. Prince: Computer Vision: Models, Learning, and Inference‖, Cambridge University Press.
6. Richard Szeliski, Computer Vision: Algorithms and Applications (Texts in Computer Science) 2nd ed. 2022 Edition
7. E. R. Davies: Computer and Machine Vision: Theory, Algorithms and Practicalities.
8. Valliappa Lakshmanan , Martin Görner: Practical Machine Learning for Computer Vision: End-to-End Machine Learning for Images, O'reilly
9. Forsyth Ponce, Computer Vision: A Modern Approach (2nd Edition), Pearson Publication.
10. Alberto FernándezVillán: Mastering OpenCV 4 with Python, Packt Publication.
11. David MillánEscrivá , Robert Laganiere: OpenCV 4 Computer Vision Application Programming Cookbook - Fourth Edition, Packt Publication
12. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time.

SUBJECT: CLOUD COMPUTING AND IoT

PAPER CODE: 22MDM22C4

Course Outcomes:
By the end of the course the students will be able to:

CO1: Understand the various cloud computing service models and use of various cloud services.
CO2: Perform service management in cloud computing and to know various security concepts in cloud computing.
CO3: Understand cloud functionality on the basis of various case-studies.
CO4: To understand the concepts of IoT and its applications & the OSI Model for the IoT/M2M Systems.
CO5: Understand the architecture and design principles for IoT.

Maximum Marks: 100 (External: 80, Internal: 20) **Time: 3 hours**

Note: Examiner will be required to set nine (09) questions in all and each carrying 16 marks. Question Number 1 will be compulsory and will consist of total 8 parts (short-answer type questions) covering the entire syllabus. In addition to the compulsory question, examiner will set two questions from each Unit of the syllabus. Student will be required to attempt five (05) questions in all including Question Number 1, which is compulsory and four (04) more questions by selecting one question from each Unit.

UNIT-I

Cloud Computing Fundamentals: Definition & Evolution of Cloud Computing, Cloud types-NIST model, cloud cube model, Deployment models, Service models, Cloud Reference model, Characteristics & Computing Benefits and Limitations of Cloud, Cloud Architecture, Communication protocols; Cloud computing vs. Cluster computing vs. Grid computing; Applications: Technologies and Process required when deploying Web services; Deploying a web service from inside and Outside of a Cloud. Services and Applications by Types: IaaS, PaaS, SaaS, IDaaS, and CaaS.

Virtualization: Objectives, Benefits of Virtualization, Emulation, Virtualization for Enterprise, VMware, Server Virtualization, Data Storage Virtualization, Load balancing and Virtualization, Improving Performance through Load Balancing, Hypervisors, Machine Imaging, Porting of applications in the cloud. Concept of Software- Defined Networking (***SDN***), Network-Function Virtualization (***NFV***) *and* Virtual Network Functions (***VNF***).

UNIT - II

Cloud and Service Management: Features of Network management system, Monitoring of an entire cloud computing deployment stack, lifecycle management of cloud services (six stages of lifecycle). Service Oriented Architecture (SOA), Event driven SOA, Enterprise Service bus, Service Catalogs, Service Level Agreements (SLAs); Managing Data - Scalability & Cloud Services, Database & Data Stores in Cloud , Large Scale Data Processing.

Cloud Security Concepts: Cloud security challenges, Cloud security approaches: encryption, tokenization/ obfuscation, cloud security alliance standards, cloud security models and related patterns.

Use of Platforms & Case Study in Cloud Computing: Concepts of Platform as a Service, Use of PaaS application frameworks; Use of Google, Amazon and Microsoft Web Services. Cloud vendors and Service

Management: Amazon cloud, AWS Overview, Installation of AWS, Google app engine, azure cloud, salesforce.
Case Study on Open Source & Commercial Clouds: Eucalyptus, Microsoft Azure, Amazon EC2.

UNIT III

IoT Overview: Introduction to Internet of Things (IoT) and IoT Applications, Conceptual Framework & Architectural View of IoT, Technology Behind IoT, Sources of IoT, M2M communication, Modified OSI Model for the IoT/M2M Systems, data enrichment, data consolidation and device management at IoT/M2M Gateway, Web communication protocols used by connected IoT/M2M devices, Message communication protocols (CoAP-SMS, CoAPMQ, MQTT,XMPP) for IoT/M2M devices.

Architecture and Design Principles for IoT: Internet connectivity, Internet-based communication, IPv4, IPv6, 6LoWPAN protocol, IP Addressing in the IoT, Application layer protocols: HTTP, HTTPS, FTP, TELNET and ports.

UNIT-IV

Data Collection, Storage and Computing using a Cloud Platform: Introduction, Cloud computing paradigm for data collection, storage and computing, Cloud service models, IoT Cloud- based data collection, storage and computing services using Nimbits.

Prototyping and Designing Software for IoT Applications: Introduction, Prototyping Embedded device software, Programming Embedded Device Arduino Platform using IDE, Reading data from sensors and devices, Devices, Gateways, Internet and Web/Cloud services software development.

Programming MQTT clients and MQTT server.

IoT Security: Introduction to IoT privacy and security, Vulnerabilities, Security requirements and threat analysis, IoT Security Tomography and layered attacker model.

Suggested Readings:

1. Anthony T. Velte Toby J. Velte, Robert Elsenpeter: Cloud Computing: A Practical Approach, McGraw-Hill.
2. Kris Jamsa: Cloud Computing: SaaS, PaaS, IaaS, Virtualization and more.
3. Tim Mather, Subra Kumaraswamy, Shahed Latif: Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, O'Reilly Media Inc.
4. Ronald L. Krutz, Russell Dean Vines: Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Wiley-India.
5. Raj Kamal: Internet of Things-Architecture and design principles, McGraw Hill Education.
6. Jan Holler, VlasiosTsiatsis, Catherine Mulligan, Stefan Avesand, Stamatis Karnouskos, David Boyle: From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence, Academic Press.
7. Peter Waher, Learning Internet of Things, PACKT publishing, BIRMINGHAM – MUMBAI
8. Bernd Scholz-Reiter, Florian Michahelles: Architecting the Internet of Things, Springer.
9. Daniel Minoli: Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications, Willy Publications
10. Any other book(s) covering the contents of the paper in more depth.

Note: Latest and additional good books may be suggested and added from time to time

SOFTWARE LAB-II
PAPER CODE: 22MDM22CL1

(Based on 22MDM22C2, 22MDM22C3, 22MDM22C4)

Course Outcomes:
By the end of the course the students will be able to:

CO1: Demonstrate how to install and configure RStudio
CO2: Explain critical R programming concepts, apply OOP concepts in R programming and explain the use of data structure and loop functions
CO3: Develop programs in R and analysing data and generate reports based on the data
CO4: Attain exposure of other Programming Languages for Big Data Analytics
CO5: Develop computer image processing programs/applications using OpenCV-Python Libraries and to gain exposure of several other relevant case studies/applications.