# MAHARSHI DAYANAND UNIVERSITY, ROHTAK

## Department of Computer Science & Applications

## M.Sc. in Computer Science (Data Science & Machine Learning) Program w.e.f. 2022-23

## Eligibility for Admission

***BCA/ B. Voc. (Software Development /IT / Web Development)/ B.Sc. (Hons.)/ B.Sc./B.E./ B. Tech. with atleast 50% marks (47.5% marks for SC/ST/ Blind/ Visually and Differently Abled candidates of Haryana only) in aggregate or any other examination recognized as equivalent thereto by M.D. University, Rohtak.***

# 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/<br>23MDM23DA2/<br>23MDM23DA3 | Natural Language Processing<br>Or<br>Stream Processing and Analytics<br>Or<br>Time Series Analysis | 80 | 20 | 100 | 4:0:0 |
| 23MDM23DB1/<br>23MDM23DB2/<br>23MDM23DB3 | Reinforcement Learning<br>Or<br>Information Security and Privacy<br>Or<br>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<br>**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/<br>23MDM24DA2/<br>23MDM24DA3 | AR/VR Systems and Wearable Computing<br>Or<br>Data Wrangling<br>Or<br>Graph Algorithms and Mining | 80 | 20 | 100 | 4:0:0 |
| 23MDM24DB1/<br>23MDM24DB2/<br>23MDM24DB3 | Social Network Analysis<br>Or<br>High Dimensional Data<br>Or<br>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.


**No**te: 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.


**No**te: 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.


**No**te: 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.


**No**te: 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.


**No**te: 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.


**No**te: 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 theuse of data structure and loop functions
CO3: Develop programs in R and analysing data and generate reports based on the dataCO4: Attain exposure of other Programming Languages for Big Data Analytics
CO5: Develop computer image processing programs/applications using OpenCV-Python Libraries and togain exposure of several other relevant case studies/applications.

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

## THIRD SEMESTER

## SUBJECT: DEEP LEARNING
## PAPER CODE: 23MDM23C1

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

CO1: To understand the fundamental principles, theory and approaches for learning with deep neural networks.
CO2: To learn the main variants of deep learning (such recurrent and convolutional architectures), and their typical applications.
CO3: To understand the key concepts, issues and practices when training and modeling with deep architectures.
CO4: To implement programming assignments related to neural network's topics.
CO5: To understand the concept of Deep reinforcement learning.

**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-1

**Deep Learning:** Fundamentals about Deep Learning, Deep learning vs Machine learning, Historical context and motivation for deep learning, Scalars, Vectors, Matrixes, Higher Dimensional Tensors, Manipulating Tensors, Vector data, Time Series data, Image and Video data.

**Deep Neural Networks:** Common architectural Principles of Deep Networks, Building blocks of Deep networks, Role and importance of activation function in deep network, Types of activation functions. Loss Functions, Empirical Risk Minimization, Feature Engineering, Over-fitting and Under-fitting, Regularizing a deep network, model exploration and hyper parameter tuning.

### UNIT-II

**Convolutional Neural Networks:** Basics of convolutional neural networks, Image processing filtering, Building a CNN: Input layers, Convolutional layers, Pooling layers, Dense layers; Backpropagation through Convolutional layer, Filters and feature maps, Backpropagation through Pooling layers, Dropout layers and regularization; Activation functions and optimizers, Applications such as image and text classification, LeNet, AlexNet, VGG16, ResNet, Transfer learning with image data; R-CNN, Fast R-CNN, Faster R-CNN, Mask-RCNN, YOLO

## UNIT-III

**Recurrent Neural Network:** Need of Recurrent Neural Network, Bidirectional RNNs, Encoder-Decoder sequence to sequence architecture, Backpropagation through time, Long Short-Term Memory (LSTM), Bidirectional LSTMs, Gated Recurrent Units (GRU).

**Generative Models:** Restrictive Boltzmann Machines (RBMs), Introduction to MCMC and Gibbs Sampling, Gradient computations in RBMs, Deep Boltzmann Machines.

**Recent Trends:** Varitional Autoencoders (Undercomplete autoencoders, regularized autoencoders, sparse autoencoders, denoising autoencoders), Representational power, layer, size and depth of autoencoders, Stochastic encoders and decoders, Generative Adversarial Networks.

## UNIT-IV

Deep Reinforcement Learning: Basic concepts of Reinforcement learning and Deep Reinforcement Learning (DRL), DRL process and RL approaches, Algorithms of DRL (Value Learning, Policy Learning), Q-Learning algorithm and its implementation, Digging deeper into Q function, Deep Q-Network. Policy optimization: Introduction to policy-based methods, Policy Gradient; Model based RL.
Applications: Large-Scale Deep Learning, Computer Vision, Speech Recognition, Natural Language Processing.

**Suggested Readings:**


1. Josh Patterson, Adam Gibson, Deep Learning-A Practitioners Approach, O'Reilly Media.
2. Nikhil Buduma, Nicholas Locascio, Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms, O'Reilly Media.
3. Li Deng and Dong Yu, Deep Learning Methods and Applications, Foundations and Trends in Signal Processing.
4. Ian Goodfellow, YoshuaBengio, Aaron Courville, Deep Learning (Adaptive Computation and Machine Learning series, MIT Press.
5. SandroSkansi,Introduction to Deep Learning From Logical Calculus to Artificial Intelligence,Springer.
6. Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction, Second Edition, MIT Press.
7. Wiering, Marco, and Martijn Van Otterlo, Reinforcement learning - Adaptation, Learning, and Optimization.
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: DATA VISUALIZATION AND INTERPRETATION
# PAPER CODE: 23MDM23C2

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

CO1: To understand the concept Why and when visualization can be applied in the data science pipeline.
CO2: To understand the pitfalls of visualization design, importance of perception and cognition envelop semantic web related applications.
CO3: To learn Optimization techniques for effective comparison and interpretability of patterns.
CO4: To learn how to evaluate visualization techniques with respect to their quality and the insights.
CO5: To learn about data interpretability.

**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: Visualization as a Discovery tool, Visualization skills for the masses, The Visualization methodology, Visualization design objectives, Understanding the context for data presentations, How visualization affects data interpretation, Role of visualization in data science Two flavors of data visualization: exploratory and communicative.

## UNIT-II

Visualization design principles: Data and task abstraction, Best practices for encoding, Marks and channels, Effectiveness and expressiveness of data visualization, How to critique visualizations, Design problems and consequences, How not to cause misinterpretation

## UNIT-III

Exploratory Visualization techniques: Handling high-dimensional data, Comparison techniques, Handling uncertainty, Depicting time
Interactive visualization: Why interactivity is needed, Handling multiple views, Brushing and Linking,

## UNIT-IV

Data Interpretability: Causes of misinterpretation, Role of communicative visualization in sciences, Graphical interpretability: metrics and approaches.

Visualization in data science: methods and examples, Explaining machine learning models, Interpretability challenges and solutions, Transparency and human-machine trust

**Suggested Readings:**


1. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications.
2. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications.
3. Ritchie S. King, Visual story telling with D3, Pearson.
4. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly.
5. A Julie Steele and Noah Iliinsky, Designing Data Visualizations: Representing Informational Relationships, O'Relly
6. Andy Kirk, Data Visualization: A Successful Design Process, PAKT 6. Scott Murray, Interactive Data Visualization for Web, O'Relly
7. Nathan Yau, "Data Points: Visualization that means something", Wiley.
8. Tamara Munzner, Visualization Analysis and Design, AK Peters Visualization Series, CRC Press.
9. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: NATURAL LANGUAGE PROCESSING
# PAPER CODE: 23MDM23DA1

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

CO1: Understand key concepts from NLP those are used to describe and analyze language.
CO2: Realize semantics and pragmatics of English language for text processing.
CO3: Analyze the syntax, semantics and pragmatics of a statement written in a natural language.
CO4: Demonstrate the state-of-art algorithms and techniques for text-based processing of natural language with respect to morphology.
CO5: To understand computational phonology.

**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

**Overview of Natural Language Processing:** Origins of NLP, Challenges of NLP, Stages of NLP, Applications such as information extraction, question answering, and machine translation. The problem of ambiguity, Why NLP is difficult.

**Word Level Analysis:** Regular Expressions, Finite-State Automata, Morphology Parsing, Spelling Error Detection and Correction, Words and Word Classes, Part-of-Speech Tagging.

**Regular Expressions:** Regular Expressions, Automata, Similarity Computation: Regular Expressions, patterns, FA, Formal Language, NFSA, Regular Language and FSAs, Raw Text Extraction and Tokenization, Extracting Terms from Tokens, Vector Space Representation and Normalization, Similarity Computation in Text.

## UNIT-II

**Morphology and Finite-State Transducers:** Inflection, Derivational Morphology, Finite- State Morphological Parsing, The Lexicon and Morphotactics, Morphological Parsing with Finite State Transducers, Combining FST Lexicon and Rules, Lexicon-free FSTs: The Porter Stemmer, Human Morphological Processing.

**Matrix Factorization and Topic Modeling:** Introduction, Singular Value Decomposition, Nonnegative Matrix Factorization, Probabilistic Latent Semantic Analysis, Latent Dirichlet Allocation

## UNIT-III

**Computational Phonology and Text-to-Speech**: Speech Sounds and Phonetic Transcription, The Phoneme and Phonological Rules, Phonological Rules and Transducers, Advanced Issues in Computational Phonology, Machine Learning of Phonological Rules, Mapping Text to Phones for TTS, Prosody in TTS .

**Probabilistic Models of Pronunciation and Spelling**: Dealing with Spelling Errors, Spelling Error Patterns, Detecting NonWord Errors, Probabilistic Models, Applying the Bayesian method to spelling, Minimum Edit Distance, English Pronunciation Variation, The Bayesian method, Pronunciation in Humans.

**UNIT-IV**

**N-gram Language Models**: The role of language models. Simple N-gram models. Estimating parameters and smoothing. Evaluating language models. Smoothing, Backoff, Deleted Interpolation, N-grams for Spelling and Pronunciation, Entropy.

**Markov Model and POS Tagging:** Overview of Hidden Markov Models, Parameter estimation, Information sources in tagging: Markov model taggers, The Viterbi Algorithm Revisited,
Word Classes and Part-of-Speech Tagging: Tagsets for English, Part of Speech Tagging, Rule-based Part-of-speech Tagging, Stochastic Part-of-speech Tagging, Transformation- Based Taggin.

**Suggested Readings:**


1. Christopher D.Manning and Hinrich Schuetze: Foundations of Statistical NaturalLanguage Processing MIT press.
2. Steven Bird,Ewan Klein and Edward Loper: Natural Language Processing with Python, O'Reilly Media.
3. Allen J., Natural Language understanding, Benjamin/Cummings Publishing Company Inc.
4. Jensen K., Heidorn G.E., Richardson S.D., Natural Language Processing: The PLNLP Approach, Springer.
5. Siddiqui and Tiwari U.S., Natural Language Processing and Information Retrieval, Oxford University Press
6. Roach P., Phonetics, Oxford University Press.
7. NitinIndurkhya, Fred J. Damerau "Handbook of Natural Language Processing", Second Edition, CRC Press.
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: STREAM PROCESSING AND ANALYTICS
# PAPER CODE: 23MDM23DA2

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

CO1: Concept of streaming data and challenges in mining data streams.
CO2: Recognize the characteristics of data streams that make it useful to solve real-world problems.
CO3: Identify and apply appropriate algorithms for analyzing the data streams for variety of problems.
CO4: Implement different algorithms for analyzing the data streams.
CO5: Identify the metrics and procedures to evaluate a model.

**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

Streaming Data – Sources – Difference between Streaming Data and Static Data, Characteristics of the data streams, Challenges in mining data streams. Requirements, characteristics and principles for real time processing
Stream Processing overview, stream processing applications and its characteristics, information flow processing technologies and stream processing systems-data, processing, architecture.

## UNIT-II

Overview of Large Scale Stream Processing Engines – Issues in Stream Processing. Components in Streaming Analytics Architecture – Vital Attributes – High Availability – Low Latency – Horizontal Scalability-Fault Tolerance.

Service Configuration and coordination – Motivation, maintaining distributed state-unreliable network connection- clock synchronization- consensus, Apache ZooKeeper

## UNIT-III

Data Flow Management in streaming Analysis- Distributed Data Flows – At Least One Delivery, Apache Kafka – design and implementation, Apache Flume, Processing Streaming Data: Processing Data with Samza.

Overview of Data Storage: NoSQL Storage- Redis and MongoDB, Other Storage Technologies, Choosing a Technology, Warehousing, Visualization- topology, metrics, status and data.

## UNIT-IV

Stream Analytics: data preprocessing and transformation- Overview, mining process, notation, descriptive statistics, sampling, sketches, quantization, dimensionality reduction, transforms.

Modeling and evaluation of stream- Overview, Offline vs Online Evaluation, stream classification, stream clustering, and anomaly detection.

**Suggested Readings:**


1. Byron Ellis: "Real-Time Analytics: Techniques to Analyze and Visualize Streaming Data", Wiley, 1st edition.
2. Andrade, Henrique CM: "Fundamentals of stream processing: application design, systems,and analytics", Cambridge University Press.
3. Joao Gama: "Knowledge Discovery From Data Streams", Chapman & Hall/ CRC.
4. SherifSakr: "Large Scale and Big Data: Processing and Management", CRC Press.
5. Sean Allen, Mathew Jankowski, Peter Pathirana: " Storm Applied", Manning Publications.
6. 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: TIME SERIES ANALYSIS

# PAPER CODE: 23MDM23DA3

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

CO1: Understand basics and importance Time Series Data.
CO2: Learn about stationary models used for Time Series Analysis.
CO3: Choose an appropriate ARMA/ARIMA model for a given set of data and to fit the model.
CO4: Compute forecasts for different Linear methods and models.
CO5: Implement Non Stationary Time Series and use of deep learning in Time series forecasting.

**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

**Time Series Forecasting:** Stochastic process and its main characteristics: Basics of Stochastic process, Time series as a discrete stochastic process. Different types of data: Cross sectional data, Time Series data, Panel data; Internal structures of time series, General trends, Seasonality, Cyclical changes, Unexpected variations;

**Get Started with Time Series Data Analysis:** Common data preparation operations for Time Series, Data cleaning of missing values in Time series, Time Series data normalization and Standardization, Time Series Feature Engineering, Date Time features, Lag features and Window features, Rolling window statistics, Expanding Window Statistics.

## UNIT-II

**Linear Stationary Models:** Stochastic Processes and Stationary, Wold's decomposition and Autocorrelation, Autoregressive (AR): First Order Autoregressive Processes, Second Order Autoregressive Processes; Moving Average (MA): First Order Moving Average processes, Second Order Moving Average processes, Estimation of Partial Autocorreation function, Autoregressive-Moving Average Models (ARMA and ARIMA models) Model Building and Estimation.

**Stationary processes in the frequency domain:** The spectral density function, The periodogram, spectral analysis. State-space models: Dynamic linear models and the Kalman filter.

## UNIT-III

**Box-Jenkins' approach:** Box-Jenkins methodology to identification of stationary time series models. Forecasting, trend and seasonality in Box-Jenkinsmodel.

**Non-stationary time series:** Basics of Non-stationary time series. Time series with non-stationary variance. Non-stationary mean. ARIMA (p,d,q) models. Box-Jenkins methodology to determination of order of integration.

**Unit root problem:** Spurious trends and regressions. Unit root tests: Dickey-Fuller and ADF test. Regressive dynamic models: Regressive dynamic models. Auto regressive models with distributed lags(ADL). Basic concepts of Time series co-integration.

**UNIT-IV**

**Deep learning for Time Series Forecasting:** Automatically Learning and Extracting features from Raw and Imperfect data; Deep learning supports Multiple inputs and outputs, Deep RNN, Bi-directional RNN, Training RNN, Solving Long range dependency problem, LSTM, GRU, RNN for Time Series forecasting, 1-D convolutions, 2-D convolutions, 1-D convolution for Time Series forecasting.

**Suggested Readings:**


1. Brockwell ,P.J. and Davis,A.Time Series,Theory and Methods, second edition. Springer.
2. Brockwell,P.J. and Davis,A. Introduction to Time Series and Forecasting,3rd edition. Springer.
3. Chatfield, C. and Xing,H.The Analysis of Time Series. An Introduction with R, 17th edition. Chapman & Hall/CRC.
4. Shumway, R. H. and Stoffer, D. S. Time Series Analysis and Its Applications: With R Examples, fourth edition. Springer.
5. James D. Hamilton , Time Series Analysis, Princeton
6. Box,G.E.P.,Jenkins,G.M.andReinsel,G.C.Time Series Analysis: Forecasting and Control, 3rd Edition, Prentice Hall, NewJersey.
7. Chatfield,C.The Analysis of Time Series, 5th edition, Chapman and Hall, NewYork.
8. Shumway,R.H., Stoffer, D.S.Time Series Analysis and Its Applications (with R examples).Springer-Verlag, NewYork.
9. James D.Hamilton.TimeSeriesAnalysis,1stEdition,Princeton University Press,
10. Galit Shmueli and Kenneth C. Lichtendahl Jr. Practical Time Series Forecasting with R:A Hands-On Guide,2nd Edition, Axelrod Schnall Publishers.
11. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: REINFORCEMENT LEARNING
# PAPER CODE: 23MDM23DB1

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

CO1: Understand basics of Reinforcement Learning.
CO2: Understand RL Framework and Markov Decision Process.
CO3: To understand and apply basic RL algorithms for simple sequential decision making problems in uncertain conditions
CO4: To understand Hierarchical reinforcement learning.
CO5: To understand applications of reinforcement learning in different fields.

**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:** Basics of RL, RL task formulation (action space, state space, environment definition), Defining RL framework, The Reinforcement Learning problem: evaluative feedback, nonassociative learning, Rewards and returns, Markov Decision Processes, Value functions, optimality and approximation.

**Bandit Problems:** Explore-exploit dilemma, Binary Bandits, Learning automata, exploration schemes Dynamic programming: value iteration, policy iteration, asynchronous DP, generalized policy iteration.

## UNIT-II

**Monte-Carlo methods and Temporal Difference Learning:** Monte Carlo: Prediction, Estimation of Action values, Control and Control without Exploring Starts, Policy evaluation, Roll outs, On Policy and Off Policy learning, Temporal Difference Prediction: TD(0), Optimality of TD(0), SARSA: On Policy TD control, Q-learning: Off Policy TD control, R-learning, Games and after states.

**Eligibility traces:** n-step TD prediction, TD (lambda), forward and backward views, Q(lambda), SARSA(lambda), replacing traces and accumulating traces.

## UNIT-III

**Function Approximation:** Value prediction, gradient descent methods, linear function approximation, Control algorithms, Fitted Iterative Methods Policy Gradient methods: non-associative learning - REINFORCE algorithm, exact gradient methods, estimating gradients, approximate policy gradient algorithms, actor-critic methods.

**Deep Reinforcement Learning:** Deep Q-Networks, Double Deep-Q Networks(DQN, DDQN, Dueling DQN, Prioritized Experience Replay)

**UNIT-IV**

**Hierarchical RL:** MAXQ framework, Options framework, HAM framework, Inverse reinforcement learning, Maximum Entropy Deep Inverse Reinforcement Learning, Generative Adversarial Imitation Learning, Recent Trends in RL Architectures, Applications of Reinforcement learning: NLP, healthcare, finance, education, robotics, games computer vision.

**Suggested Readings:**


1. Sutton, Richard S., and Andrew G. Barto,"Reinforcement learning: An introduction," First Edition, MIT press.
2. Sugiyama, Masashi, "Statistical reinforcement learning: modern machine learning approaches," First Edition, CRC Press.
3. "Bandit algorithms," First Edition, Cambridge University Press.
4. Lattimore, T. and C. Szepesvári, "Reinforcement Learning Algorithms: Analysis and Applications," First Edition, Springer.
5. Alexander Zai and Brandon Brown "Deep Reinforcement Learning in Action," First Edition, Manning Publications.
6. Saba Szepesvari. Algorithms for Reinforcement learning. Morgan & Claypool Publishers.
7. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: INFORMATION SECURITY AND PRIVACY
# PAPER CODE: 23MDM23DB2

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

CO1: Understand the most important classes of information security/privacy risks in today environment
CO2: To become able to explain various information security threats and controls for it.
CO3: To become able to explain the usages of cryptography.
CO4: To become able to analyze privacy and attacks.
CO5: Understand the privacy control and to learn about Online Privacy.

**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-1

**Security Concepts:** Security Objectives, The Challenges of Information Security, Security Attacks -passive Attacks, Active Attacks. Security Services- Authentication, Access Control, Data Confidentiality, Data Integrity, Nonrepudiation, Availability Service, Cyber security, Information Security , Network Security, Communications Security, Device Security, Security Mechanisms.

**Cryptography:** Public-Key Infrastructure, Certificates and PKI Architecture, Cryptographic Algorithms-Keyless, Single-Key and Two-key. Symmetric and Asymmetric Encryption, Cryptographic Hash Functions, Digital Signatures.

## UNIT-II

**Information Privacy Threats and Vulnerabilities:** Evolving Threat Environment, Overall Impact of Advances in Technology, Repurposing Collected Data, Means of Collection of PII, Privacy Threat Taxonomy-Information collection, Processing, Dissemination, Invasions, NIST Threat Model, Threat Sources, Identifying Threats, Privacy Vulnerabilities- Vulnerabilities Categories, Location of Privacy Vulnerabilities, National Vulnerabilities Database and Common Vulnerability.

**Malicious Software and Intruders:** Malware Protection Activities, Types and Threat of Malware, Capabilities and Managing of Malware Protection Software. Firewalls: Characteristics,Types of Firewalls, Next Generation Firewalls, DMZ Networks. Intrusion Detection: Basic Principles, Approaches to Intrusion Detection, Host –Based Intrusion Detection Techniques, Network-Based Intrusion Detection Systems.

## UNIT-III

**Privacy Concepts:** Key Privacy Terminology, Privacy by Design- Privacy by design Principles, Requirements and policy Development, Privacy Risk Assessment, Privacy and Security Control Selection,

Privacy Program and Integration Plan. Privacy Engineering: Privacy Implementation, System Integration, Privacy Testing and Evaluation, Privacy Auditing and Incident Response.

**Technical Security Controls for Privacy:** System Access Concepts: Privileges, functions and Privacy Considerations for System access, Privacy Authorization. User Authentication: Means and Multifactor Authentication, Model for Electronic User Authentication. Access Control: Subject, Object, Access Rights, Access Control Policies, Discretionary Access Control, Role-based Access Control. Identity and Access Management: IAM Architecture Federated Identity Management.

**UNIT-IV**

**Privacy Enhancing Technologies:** Privacy in databases: Personal Data Attributes, Types of Data Files, Re-Identification Attacks:Types of Attacks, Potential Attackers, Disclosure Risks, and Applicability to Privacy Threats. De-Identification of Direct Identifiers: Anonymization, Pseudonymization, De- Identification of Quasi-Identifiers in Microdata, Files: Privacy –Preserving Data Publishing, Disclosure Risk versus Data Utility, PPDP Techniques. K-Anonymity, L-Diversity nad T-Closeness.Privacy in Queryable databases, Privacy Threats, Protecting Queryable Databases

**Online Privacy:** Online Ecosystem for Personal Data, Web Security and Privacy: Web server, Web Browser and Web Application Security and Privacy, Mobile App Security, Online Privacy Threats: Web Application Privacy, Mobile App Privacy, Online Privacy Requirements, Principals and Framework, Tracking, Technologies. IoT Security and Privacy .Cloud Computing and Privacy

**Suggested Readings:**


1. William Stallings , "Cryptography and Network Security: Principles and Practice"
2. Mark Stanislav and Andrew Hoffman."Introduction to Information Security"
3. Mehdi KhosrowPour, "Information Security: Concepts, Methodologies, Tools, and Applications"
4. Bruce Schneier, "Applied Cryptography: Protocols, Algorithms, and Source Code in C"
5. William Stallings. “Information Privacy Engineering and Privacy by Design”
6. Ross Anderson, "Security Engineering: A Guide to Building Dependable Distributed Systems"
7. 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: PREDICTIVE ANALYTICS FOR IoT

# PAPER CODE: 23MDM23DB3

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

CO1: Acquire fundamental knowledge of IoT Data Types, Types of IoT Data Analytics.
CO2: Understand and overcome various Challenges in IoT Data Analytics.
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 IoT data.
CO5: Designing applications of 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

**Introduction to IoT:** Definitions, frameworks and key technologies. Challenges to solve in IoT-Key hardware and software elements. Real-World Data representation and visualization, Introduction to Data Analytics for IoT. IoT Applications**,** Structured vs Unstructured IoT Data.

**IoT Data Analytics:** Definition, Challenges, Devices, Connectivity protocols, data messaging protocols-MQTT, HTTP, CoAP, Data Distribution Services (DDS), IoT Data Analytics – Elastics Analytics Concepts, Scaling.

**Introduction to Predictive Analytics:** Types of Analysis- Descriptive, Diagnostic, Predictive, and Prescriptive Analysis. Process of Predictive Analytics: Data Gathering, Cluster Analysis, Association Rule Mining, Outlier Analysis, Conclusive Statement Gathering.

## UNIT-II

**Data Visualization in IoT:** Data Visualization, Need for Data Visualization, Role of Data Visualization in IoT, Visualization and Dashboard – Designing visual analysis for IoT data- creating dashboard – creating and visualizing alerts – basics of geospatial analytics- vector-based methods-raster based methods- storage of geospatial data-processing of geospatial data- Anomaly detection forecasting. Case studies of various problems like pollution reporting problems etc.

## UNIT-III

**Introduction to Machine Learning for IoT:** Applications of Machine Learning for IoT, Introduction to Azure Machine Learning, Code-First Machine Learning with Python.

**Data Preparation for Predictive Maintenance Modeling:** Exploring IoT Data with Python, Cleaning and Standardization of IoT Data, and Applying Advanced IoT Data Exploration Techniques.

## UNIT-IV

**Feature Engineering for Predictive Maintenance Modeling**: Explore Core Concepts of Feature Engineering, Feature Selection Techniques.

Cloud Analytics and Security, AWS / Azure, the design of data processing for analytics, the application of big data technology to storage, exploring and visualizing data, and solutions for industry-specific analysis problems.

**Suggested Readings:**


1. Andrew Minteer , Analytics for Internet of Things — Packt Publications Mumbai.
2. Anasse Bari, Mohamed Chaouchi, and Tommy Jung, Predictive Analytics For Dummies
3. Alvaro Fuentes , Hands-On Predictive Analytics with Python: Master the Complete Predictive Analytics Process, from Problem Definition to Model Deployment- Packt Publications Ltd.
4. Kai Hwang, Min Chen,Big–Data Analytics for Cloud, IoT and Cognitive Computing Hardcover
5. C.M. Bishop: Pattern Recognition and Machine learning Springer.
6. Hastie, Tibshirani, Friedman: Introduction to statistical machine learning with applications in R,Springer.
7. Tom Mitchell: Machine Learning, McGraw Hill.
8. Parag Kulkarni: Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEE Press.
9. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

## Software Lab-III

## PAPER CODE: 23MDM23CL1

## (Based on 23MDM23C1, 23MDM23C2, 23MDM23DA1/2/3, DB1/2/3)

**Course Outcomes:**

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

CO1: Acquire programming skills in core Python.
CO2: Develop and evaluate various Neural network problems.
CO3: To implement Visualization and optimization techniques.
CO4: To implement programming assignments related to neural network.
CO5: Gain exposure of several relevant case studies.

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

## FOURTH SEMESTER

## SUBJECT: BLOCKCHAIN TECHNOLOGY
## PAPER CODE: 23MDM24C1

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

CO1: Contentedly discuss the history, types and applications of blockchain.
CO2: Explain the fundamental characteristics of blockchain using bitcoin.
CO3: Explain the modern concepts of block chain technology systematically.
CO4: Understand the modern currencies and its market usages.
CO5: Design blockchain based applications.

**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 Blockchain:** History, Definition, Types of Blockchain, Hash Functions, Properties of Hash Function, Digital Signature,Working of Blockchain, Issues and needs of Blockchain, Benefits and Challenges of Blockchain, features of Blockchain, Blockchain Network and Nodes, Peer-to-Peer Network.

**Blockchain Architecture:** Mining Mechanism, Life cycle of Blockchain, Merkle Patricia Tree, Gas Limit, Transaction Fees, Anonymity, Reward, Chain policy, Applications of Blockchain, Fork and its Types, Generic elements of Blockchain, Cryptography in Blockchain, Nash Equilibrium, Prisner's Dilemma, Zero-Sum Games.

### UNIT-II

**Cryptocurrency:** History, Distributed Ledger, Creation of coins ,Bitcoin blockchain, Bitcoin Transactions, Bitcoin Mining, Block reward, Challenges and solutions, Proof of work, Proof of Stake, Alternatives to Bitcoin consensus, Bitcoin scripting Languages and their use.

**Ethereum blockchain:** Ethereum network work, Decentralized Autonomous Organization (DAO), Smart Contracts, Objectives and Principles for the design of Ethereum, Blockchain systems, GHOST, Ethereum Attacks, Sidechain, Namecoin (NMC), Consensua Mechanism, Proof of Work, Proof of Stake, Delegated Proof of Stake.

## UNIT-III

**Smart Contract with Solidity:** What is Smart Contract? Life Cycle of a Smart Contract, Solidity, The Ethereum Contract ABI, Smart Contract templates; Oracles: Types of Blockchain oracles; Deploying smart Contracts, Statements and Expressions in Solidity, Data types of Solidity, Mining Ether, Truffle Suite, Ganache, Deploying using Ganache. Private Ethereum Blockchain with Geth; Smart Contract Security.

**Cryptography Regulation:** Stakeholders, Roots and Bitcoin, Legal Aspects, Crypto currency exchange, Black market and Global economy. Security and safeguards: Protecting blockchain from attackers, Blockchain security: Key management in Bitcoin

## UNIT-IV

**Block Chain for Government:** Digital identity, land records and other kinds of record keeping between government entities, public distribution system/social welfare systems.

**Enterprise Applications of Blockchain:** Internet of Things, Smart Health Care, Transportation Smart city, Supply chain Management. Government Identify management , Auto executing contracts , Three signature escrow, Triple entry accounting, Elections and voting, Property records, titles Micropayments, Notary, challenges and research Issue in Blockchain.

**Suggested Readings:**


1. Paul Laurence, "Blockchain: Step-By-Step Guide to Understand.
2. Frank walrtin, " Blockchain: The comprehensive beginner?
3. Daniel Drescher, "Block chain basics A non-technical introduction in 25 steps", Apress.
4. Mark Gates, "Block Chain: Ultimate guide to understanding block chain, bitcoin, crypto currencies, smart contracts and the future of money", Wise Fox Publishing and Mark Gates.
5. Andreas Antonopoulos, "Mastering Bitcoin: Unlocking Digital Crypto currencies", O'Reilly Media Inc.
6. Paul Vigna and MichealJ.Casey. The Age of Cryptography.
7. 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: AR/VR SYSTEMS AND WEARABLE COMPUTING

# PAPER CODE: 23MDM24DA1

**Course Outcomes:**

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

CO1: Understand the components of the virtual reality system, working and applications of VR/AR systems in designing wearable devices.

CO2: Describe and understand the implementation of hardware for designing virtual reality devices

CO3: Apply the different modeling concepts to visual virtualization

CO4: Analyze the performance of given simple applications related to virtual reality

CO5: Understand the concepts of the augmented reality system

**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

**Overview and Introduction to Immersive Technologies (AR/VR/MR):** Augmented-Virtual and Mixed Reality, wearable technologies, History and Taxonomy, technology and features of augmented reality, the difference between AR , VR and MR, Challenges with AR, AR systems and functionality, Augmented reality methods, and visualization techniques for augmented reality.

**VR systems**: VR as a discipline, Basic features of VR systems, Architecture of VR systems.

**VR hardware:** VR input hardware: tracking systems, motion capture systems, data gloves, VR output hardware: visual displays.

## UNIT-II

**Fundamentals of Computer Graphics in AR/VR**: Graphics, Software and Hardware Technology on Stereoscopic Display, Advanced Techniques in CG- Management of Large-Scale Environments & Real-Time Rendering, Development Tools and Frameworks in Virtual Reality.

**Stereoscopic vision & Haptic rendering:** Fundamentals of the human visual system, Depth cues, Stereopsis, Retinal disparity, Haptic sense, Haptic devices, Algorithms for haptic rendering and parallax, Synthesis of stereo pairs, and Pipeline for stereo images.

## UNIT-III

**VR/AR software development:** Challenges in VR/AR software development, Master/slave and Client/server architectures, Cluster rendering, Game Engines, and available SDK to develop VR applications for different hardware (HTC VIVE, Oculus, Google VR), AR Software, AR Toolkit, Marker Based augmented reality, AR Toolkit.

**3D interaction techniques:** 3D Manipulation tasks, Manipulation Techniques, Input Devices, and Interaction Techniques for 3D Manipulation.

**UNIT-IV**

**Wearable Devices:** Definitions, goals, dreams, constraints, managing cognitive bandwidth, UMass Wearables, Role of Wearables, Attributes of Wearables, The Meta Wearables – Textiles and clothing, Social Aspects: Interpretation of Aesthetics, Adoption of Innovation, On-Body Interaction, Overview of Human-Computer Interaction (HCI), Principles of interaction and user interface design.

**Wearable Computing:** Wearable computers (Body Borne Computers), history of wearable technologies, Evolution, and Operating Systems for Wearable Computers: Free RTOS, LiteOS, Tizen OS, Watch OS, and Wear OS. Applications of Wearable Computing, Wearable computing Algorithms, Architecture Standardization.

**Case Study:** Google Glass, health monitoring.

**Suggested Readings:**


1. Steven M. LaValle ,Virtual Reality, , Cambridge University Press.
2. William R Sherman and Alan B Craig, (The Morgan Kaufmann Series in Computer Graphics), Understanding Virtual Reality: Interface, Application and Design,". Morgan Kaufmann Publishers, San Francisco, CA.
3. Alan B Craig, William R Sherman, and Jeffrey D Will, Morgan Kaufmann. Developing Virtual Reality Applications: Foundations of Effective Design,
4. Edward Sazonov, Michael R Neuman, "Wearable Sensors: Fundamentals, Implementation and Applications" Elsevier.
5. Gerard Jonghyun Kim, "Designing Virtual Systems: The Structured Approach".
6. Doug A Bowman, Ernest Kuijff, Joseph J LaViola, Jr and Ivan Poupyrev, "3D User Interfaces, Theory and Practice", Addison Wesley, USA.
7. Oliver Bimber and Ramesh Raskar, "Spatial Augmented Reality: Merging Real and Virtual Worlds".
8. Burdea, Grigore C and Philippe Coiffet, "Virtual Reality Technology", Wiley Interscience, India.
9. Subhas C. Mukhopadhyay, "Wearable Electronics Sensors-For Safe and Healthy Living", Springer International Publishing.
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.

# SUBJECT: DATA WRANGLING
# PAPER CODE: 23MDM24DA2

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

CO1: Perform data wrangling
CO2: Explain principles of visual perception
CO3: Apply core skills for visual analysis
CO4: Apply visualization techniques for various data analysis tasks
CO5: Evaluate visualization techniques.

**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

**Data Wrangling:** What is Data Wrangling, The Role of Data Wrangling, How is Data Wrangling Performed?, Tasks of Data Wrangling, Data Wrangling Tools-Data meant to be read by machines-CSV data, JSON data, XML data.

**Data Cleanup:** Need of data cleanup, data clean up basics – formatting, outliers, duplicates, Normalizing and standardizing data, Saving the data determining suitable data cleanup, Scripting the Cleanup, Testing with New Data.

## UNIT – II

**Data Exploration and Analysis:** Exploring data, importing data, exploring table functions, joining numerous datasets, Identifying Correlations, Identifying outliers, Creating Grouping, Analyzing data, Separating and Focusing the data, Data presentation

**Data Visualization:** Introduction of visual perception, visual representation of data, Gestalt principles, information overloads. Creating visual representations, visualization reference model, visual mapping, visual analytics, Design of visualization applications.

## UNIT – III

**Data Visualization:** Classification of visualization systems, Interaction and visualization techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents. Visualization of groups, trees, graphs, clusters, networks, software, Metaphorical visualization; Visualization of volumetric data, Vector fields, Processes and Simulations, Visualization of maps, Geographic Information, GIS, Collaborative Visualizations, Evaluating visualizations.

## UNIT – IV

**Web Scraping:** What to Scrape and How, Analyzing a Web Page, Network/Timeline, Interacting with JavaScript, In-depth Analysis of a page, Getting Pages; Web Scraping: Browser based parsing, Screen reading with Selenium, Screen reading with Ghost, PySpidering the Web, Building a Spider with Scrapy, Crawling whole websites with Scrapy

**Suggested Readings:**


1. Jacqueline Kazil and Katharine Jarmul, Data Wrangling with Python: Tips and Tools to Make Your Life Easier, O'Reilly.
2. Tirthajyoti Sarkar, Shubhadeep: Data Wrangling with Python: Creating actionable data from raw sources, Packet publishing Ltd., 2019
3. Ward, Grinstein Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. Natick A K Peters, Ltd.
4. E. Tufte, The Visual Display of Quantitative Information, Graphics Press
5. Allan Visochek: Practical Data Wrangling, Packet Publishing Ltd.
6. Tye Rattenbury: Principles of Data Wrangling: Practical Techniques for Data Preparation, O'Reilly Media, Inc.
7. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: GRAPH ALGORITHMS AND MINING

# PAPER CODE: 23MDM24DA3

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

CO1: Understand the graph theory and graph mining foundations.
CO2: Analyze Graph Mining methods.
CO3: Formulate and solve graph-related problems.
CO4: Apply graph mining algorithms to analyze large-scale datasets on various domains.
CO5: to learn how Graph mining is implemented in various domains.

**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 Graphs**: Introduction to Graphs and basic terminology, Representations of Graph, Types of Graphs, Basic algorithms for decomposing graphs into parts, Connectivity of graphs, Matching on graphs.
**Data Mining**: Overview of Data Mining, Basic Mining algorithms for Item Sets and Sequences. Evaluation of Mining approaches.

## UNIT-II

**Graph Algorithms:** Graph Coloring, Graphs on surface, Directed graphs, Shortest path algorithms, Algorithms to discover minimum spanning tree, Flows in Networks and Flow algorithms, Searching Graphs and Related algorithms.

## UNIT-III

**Graph Mining:** Motivation for Graph Mining, Applications of Graph Mining, Mining Frequent Sub graphs –Transactions, BFS/Apriori Approach (FSG and others), DFS Approach (GSPAN and others), Diagonal and Greedy Approaches, Constraint-based mining and New algorithms, Mining Frequent Sub graphs, Graph visualizations.

## UNIT-IV

**Applications of Graph Mining**: Web Mining, Centrality analysis, Link analysis algorithms, Graph clustering and Community Detection, Node classification and Link prediction, Influential spreaders, Influence maximization, Geo-social and location based networks.

**Suggested Readings:**


1. Diestel,R.: Graph Theory, 4th ed. Springer-Verlag, Heidelberg
2. J. Han and M. Kamber: Data Mining–Concepts and Techniques, 2nd Edition, Morgan Kaufman Publishers.
3. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer publishing.
4. Jure Leskovec, Anand Rajaraman, Jeff Ullman. Mining of Massive Datasets. Book 2nd edition. Cambridge University Press
5. David Easley and Jon Kleinberg. Networks, Crowds, and Markets. Cambridge University Press.
6. Deepayan Chakrabarti and Christos Faloutsos: Graph Mining: Laws, Tools, and Case Studies. Synthesis Lectureson Data Mining and Knowledge Discovery, Morgan & Claypool Publishers.
7. Albert-László Barabási. Network Science. Cambridge University Press.
8. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: SOCIAL NETWORK ANALYSIS
# PAPER CODE: 23MDM24DB1

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

CO1: To understand the concept of semantic web and related applications.
CO2: To develop semantic web related applications.
CO3: To learn knowledge representation using ontology.
CO4: To understand human behaviour in social web and related communities.
CO5: To learn visualization of social networks.

**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:**

Introduction to Semantic Web: Limitations of current Web - Development of Semantic Web - Emergence of the Social Web - Social Network analysis: Development of Social Network Analysis - Key concepts and measures in network analysis - Electronic sources for network analysis: Electronic discussion networks, Blogs and online communities - Web-based networks - Applications of Social Network Analysis.

**Modelling, Aggregating And Knowledge Representation:**

Ontology and their role in the Semantic Web: Ontology-based knowledge Representation - Ontology languages for the Semantic Web: Resource Description Framework - Web Ontology Language - Modelling and aggregating social network data: State-of-the-art in network data representation - Ontological representation of social individuals - Ontological representation of social relationships - Aggregating and reasoning with social network data - Advanced representations.

## UNIT-II

**Extraction And Mining Communities In Web Social Networks:**

Extracting evolution of Web Community from a Series of Web Archive - Detecting communities in social networks - Definition of community - Evaluating communities - Methods for community detection and mining - Applications of community mining algorithms - Tools for detecting communities social network infrastructures and communities - Decentralized online social networks - Multi-Relational characterization of dynamic social network communities.

## UNIT-III

**Predicting Human Behaviour And Privacy Issues:**

Understanding and predicting human behaviour for social communities - User data management - Inference and Distribution - Enabling new human experiences - Reality mining - Context - Awareness - Privacy in online social networks - Trust in online environment - Trust models based on subjective logic - Trust network analysis - Trust transitivity analysis - Combining trust and reputation - Trust derivation based on trust comparisons - Attack spectrum and countermeasures.

## UNIT-IV

**Visualization And Applications Of Social Networks:**

Graph theory - Centrality - Clustering - Node-Edge Diagrams - Matrix representation - Visualizing online social networks, Visualizing social networks with matrix-based representations - Matrix and Node-Link Diagrams - Hybrid representations - Applications - Cover networks - Community welfare - Collaboration networks - Co-Citation networks.

**Suggested Readings:**


1. Peter Mika, ―Social Networks and the Semantic Web‖, First Edition, Springer.
2. BorkoFurht, ―Handbook of Social Network Technologies and Applications‖, 1st Edition, Springer.
3. Guandong Xu ,Yanchun Zhang and Lin Li, ―Web Mining and Social Networking – Techniques and applications, First Edition, Springer.
4. Dion Goh and Schubert Foo, ―Social information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively‖, IGI Global Snippet.
5. Max Chevalier, Christine Julien and Chantal Soulé-Dupuy, ―Collaborative and Social Information Retrieval and Access: Techniques for Improved user Modelling, IGI Global Snippet.
6. John G. Breslin, Alexander Passant and Stefan Decker, ―The Social Semantic Web, Springer.
7. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: HIGH DIMENSIONAL DATA

# PAPER CODE: 23MDM24DB2

**Course Outcomes:**

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

CO1: To understand the key concepts of high-dimensional data and their significance in data science.
CO2: To become familiar with various feature selection, regularization, and dimensionality reduction techniques for high-dimensional data analysis.
CO3: To understand the algorithms used for high-dimensional data analysis, including random forests, support vector machines, and deep learning architectures.
CO4: To gain practical experience in working with high-dimensional data using Python and relevant libraries such as NumPy, pandas, scikit-learn, and TensorFlow.
CO5: To explore the Applications of high dimensional data analysis.

**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 High-Dimensional Data:** Overview of high-dimensional data, Characteristics and Examples of high-dimensional data; Challenges in analyzing high-dimensional data, Opportunities presented by high-dimensional data.

**Feature selection techniques:** Filter, Wrapper, Embedded, and Regularization methods. **Filter methods:** Chi-squared, Mutual Information, ReliefF, Correlation-based Feature Selection, and others. **Wrapper methods:** Recursive Feature Elimination, Sequential Feature Selection, and others. **Embedded methods:** Lasso, Ridge Regression, Elastic Net, and others. Regularization methods: L1 and L2 Regularization, Elastic Net, and Dropout Regularization.

## UNIT-II

**Dimensionality Reduction Techniques:** PCA, t-SNE, UMAP.

**Exploratory Data Analysis for High-Dimensional Data:** Visualization techniques for high-dimensional data (e.g. scatterplots, heatmaps, parallel coordinates, treemaps); Clustering and outlier detection in high-dimensional data (e.g. k-means, hierarchical clustering, DBSCAN, LOF); Feature selection and feature engineering for high-dimensional data (e.g. recursive feature elimination, feature importance, feature scaling).

## UNIT-III

**Machine Learning for High-Dimensional Data:** Supervised learning algorithms for high-dimensional data (e.g. linear regression, logistic regression, SVMs, decision trees, random forests); Unsupervised learning algorithms for high-dimensional data (e.g. PCA, t-SNE, UMAP, autoencoders, k-means, hierarchical clustering).

**Deep learning and neural networks for high-dimensional data** (e.g. convolutional neural networks, recurrent neural networks, attention mechanisms).

## UNIT-IV

**Applications of High-Dimensional Data Analysis:** Computer Vision - Overview, Image classification, Object detection, Semantic segmentation, Instance segmentation, and Image generation; Natural Language Processing: Overview of NLP, Text classification, Sentiment analysis, Named Entity Recognition, Machine Translation, and Language Generation; Other relevant domains: Overview of related domains - Bioinformatics, Neuroscience, Finance. Applications of high-dimensional data analysis in Bioinformatics, Neuroscience and Finance; Ethical considerations and limitations of high-dimensional data analysis (e.g. privacy, bias, reproducibility).

**Suggested Readings:**


1. Hastie, T., Tibshirani, R., & Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
2. Goodfellow, I., Bengio, Y., &Courville, A.: Deep Learning. MIT Press.
3. Inge Koch: Analysis of Multivariate and High-Dimensional Data, Cambridge University Press.
4. FatemehEmdad, SeyedZekavat: High Dimensional Data Analysis: Overview, Analysis, and Applications, VDM Verlag.
5. Bishop, C. M.: Pattern Recognition and Machine Learning. Springer.
6. Witten, I. H., Frank, E., & Hall, M. A.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
7. Murphy, K. P.: Machine Learning: A Probabilistic Perspective. MIT Press.
8. Python Machine Learning (2nd Edition) by Sebastian Raschka and Vahid Mirjalili
9. Scikit-learn documentation: https://scikit-learn.org/stable/documentation.html
10. TensorFlow documentation: https://www.tensorflow.org/api_docs
11. NumPy documentation: https://numpy.org/doc/
12. Pandas documentation: https://pandas.pydata.org/docs/
13. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

# SUBJECT: ARTIFICIAL INTELLIGENCE AND DECISION SCIENCES
# PAPER CODE: 23MDM24DB3

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

CO1: Understand the concepts of using data analysis for decision making of linear problems.
CO2: To develop the required reasoning and analysis related to transportation and assignment of work related problems.
CO3: To analyses and control projects with critical path identification.
CO4: To understand and analyses decision related problems with/without probabilities.
CO5: To critically evaluate the process of multi-criteria decision making.

**Max. Marks: 100** (**External: 80, Internal: 20)** **Time: 3 Hrs**

**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

**Representational AI:** Introduction, representations used in AI, constraint satisfaction, continuous and discrete problems, logical representation and inference, Monte Carlo tree search, probabilistic graphical models.
**Inference and Reasoning:** Inference, planning: discrete, continuous, deterministic, probabilistic models, MDPS and POMDPS, the Tiger Problem, Importance of Decision Science

## UNIT II

**Optimizing Techniques:** Concept, Formulation & Graphical Solution, LPP, feasible solutions, other cases.
**Assignment Models**: Concept, Flood's Technique/ Hungarian Method, applications including restricted & multiple assignments.
**Transportation Models**: Types of Problem: Balanced, Unbalanced, Minimization, Maximization Basic initial Solution using North West Corner, Least Cost and VAM, Optimal Solution using MODI.

## UNIT III

**Project Scheduling and Networking:**Concept, Drawing network, identifying critical path, Calculating EST, LST, EFT, LFT, Slack and Probability of Project Completion, PERT, CPM, Critical path identification, time-cost trade-offs, Gantt Charts.
**Sequencing problems:** Introduction, Problems involving n jobs- 2 machines, n jobs- 3 machines & n jobs-m machines; Comparison of priority sequencing rules.

## UNIT IV

**Decision Analysis:** Pay-off tables, decision trees
**Decision making without probabilities**: optimistic approach, conservative approach, min-max regret approach

**Decision making with probabilities:** Optimism (Maximax or Minimin) Criterion, Pessimism (Maximin or Minimax) Criterion, Equal Probabilities (Laplace) Criterion, Coefficient of Optimism (Hurwicz) Criterion, Regret (Savage) Criterion.

**Suggested Readings:**


1. N.D. Vohra, Quantitative Techniques in Management, 4th Edition,Tata McGraw Hill Publications
2. J. K. Sharma, Operations Research-theory-and-applications, 4th Edition, MacMillan Publishers, India.
3. Anderson, Sweeney, Williams, Camm, Cochran, Fry, and Ohlmann, An Introduction to Management Science: Quantitative Approaches to Decision Making, 14th Edition;;Cenage Learning.
4. Billey E. Gilett, Introduction to Operations Research, Tata McGraw Hill Publications
5. R. Pannerselvam, Operations Research, Prentice Hall India, 2nd Edition
6. Hans G. Daellenbach, Donald C McNickle, Management Science: Decision making through systems thinking, Palgrave Macmillan.
7. Hamdy A. Taha, Operations Research, Pearson Publication
8. David R. Andersion, Dennis J Sweeney, Thomas A Williams, Mik Wisniewski, An introduction to Management Science, quantitative approaches to decision making, second edition, Cenage Learning.
9. Raghu Nandan Sengupta, Aparna Gupta, Joydeep Dutta, Decision Sciences theory and practice, CRC Press.
10. Ranjeet H Chitale, Decision Sciences, Nirali Prakashan.
11. Any other book(s) covering the contents of the paper in more depth.


**No**te: Latest and additional good books may be suggested and added from time to time.

## Software Lab-IV

## PAPER CODE: 23MDM24CL

## (Based on 23MDM24C1, 23MDM24DA1/2/3, DB1/2/3)

**Course Outcomes:**

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

CO1: To develop semantic web related applications.
CO2: Visualization & Analysis of social networks
CO3: Understand and implement the concepts of blockchain technology.
CO4: Design blockchain based applications.
CO5: Gain exposure of implementation of different concepts related to IoT and Smart City.

# PROJECT REPORT/DISSERTATION

## PAPER CODE: 23MDM24C3

A student may opt either for a live Industry Project or a Research Project/Problem/Study. Every student will have to identify/select/explore a **live Project/Research problem** during 3rd semester of the program and each student will carry out an Industry Project/Research work during this semester as well as the successive M.Sc. in Computer Science (Data Science & Machine Learning) 4th semester. The students will have freedom to showcase their innovation/creativity/research potential either:

*In developing a software solution/App to meet out live/realistic requirement(s) using any type of software development tools/ languages/ technologies in view of the ongoing Software Industry trends.*

Or

*Undertaking a research project/problem/study which is socially relevant and meaningful.*

Each student will be assigned a Faculty Member as Supervisor/Guide by the Head of the Department/Director/Principal for Project/Dissertation and the supervisor will be guiding/supervising/mentoring/supporting as well as tracking the progress of the student on the assigned project/dissertation task.

Each student will be required to submit required number of Project Report/Dissertation to the Department/Institute/College as stipulated by the University.

The students undertaking the research project/problem/study are required to submit a dissertation as a partial fulfillment of their course curriculum. Dissertation work involves carrying out an original research project/problem/study under the guidance/supervision of a faculty member.

Generally, a dissertation may have five or six chapters: Introduction, Review of Literature, Methodology (Research Design & Methods), Presentation of Research and Summary, Implications, Conclusions (Discussion).