

**M.Tech – Data Science Syllabus**  
**(1<sup>st</sup> Semester)**

**CS 5301**

**Mathematics for Data Science**

**3-0-0-3**

**UNIT I: THEORY OF PROBABILITY**

Axioms of Probability, Conditional Probability, Baye's Rule, Random variables: Discrete and Continuous random variables, Probability function and Distribution function, Mathematical Expectation, Variance, Standard Deviation, Moments, Moment generating function, Binomial, Poisson and Normal Distributions.

**UNIT II: DESCRIPTIVE STATISTICS**

Statistical Methods: Definition and scope of Statistics, concepts of statistical population and sample. Data: quantitative and qualitative, attributes, variables, scales of measurement nominal, ordinal, interval and ratio. Presentation: tabular and graphical, including histogram and ogives, consistency and independence of data with special reference to attributes; Measures of Central Tendency: mathematical and positional. Measures of Dispersion: range, quartile deviation, mean deviation, coefficient of variation, Moments, absolute moments, factorial moments, skewness and kurtosis, Sheppard's corrections; Bivariate data: Definition, scatter diagram, simple, partial and multiple correlation (3 variables only), rank correlation. Simple linear regression, principle of least squares and fitting of polynomials and exponential curves, Gradient descent methods.

**UNIT III: MATRICES**

Linear and Orthogonal Transformations, Linear dependence of vectors, Characteristics equation, Eigen values and Eigen vectors, Statement and Verification of Cayley-Hamilton Theorem [without proof], Reduction to Diagonal form, Reduction of Quadratic form to Canonical form by Orthogonal Transformation, Sylvester's theorem[without proof], Solution of Second Order Linear Differential Equations with Constant Coefficients by Matrix method. Largest Eigen value and Eigen vector by Iteration method. principal component analysis and linear discriminant analysis.

**TEXT BOOKS:**

1. Higher Engineering Mathematics by B.S. Grewal, 40th Edition, Khanna Publication
2. Advanced Engineering Mathematics by Erwin Kreyszig, 8th Edition, Wiley India
3. Applied Mathematics for Engineers & Physicist by L.R. Pipes and Harville
4. Theory & Problems of Probability and Statistics by M.R. Spiegel ,Schaum Series, McGraw Hills.
5. M. P. Deisenroth, A. A. Faisal, C. S. Ong, Mathematics for Machine Learning, Cambridge University Press (1st edition)

**Course Outcomes:** At the end of the course students are expected to be:

CO1: Able to use the basic concepts of probability in the field of data science.

CO2: Able to use the basic concept of descriptive statistics in the field of data science to design an intelligent system.

CO3: Able to use the basic concept of matrix algebra to design an intelligent system.

**CS 5302**

**Foundation of Data Science**

**3-0-0-3**

### **UNIT 1: INTRODUCTION TO PYTHON**

Variable creation, Python identifiers, keywords, code blocks, Basic object types, Basic operators, Data types and associated operations, Lists, Tuples, Dictionary, Set Operations, User defined functions, Setting working Directory, Creating and saving a script file, File execution, clearing console, removing variables from environment, learning environment, Modules and Packages: Using and Creating, Commenting script files

### **UNIT 2: INTRODUCTION TO DATA SCIENCE USING MACHINE LEARNING**

Data Science Perspective of Data, Data Science Python Packages, Data Analysis Packages, Machine Learning Core Libraries, Feature Engineering, Exploratory Data Analysis, Data Visualization, Supervised Learning– Regression, Supervised Learning – Classification, Unsupervised Learning Process Flow

### **UNIT 3: TEXT MINING AND RECOMMENDER SYSTEMS**

Text Mining Process Overview, Data Assemble, Data Preprocessing (Text), Model Building, Text Similarity, Text Clustering, Topic Modeling, Text Classification, Sentiment Analysis, Deep Natural Language Processing (DNLP), Recommender Systems

### **Text Books/References:**

1. Introduction to linear algebra - by Gilbert Strang.
2. Applied statistics and probability for engineers – by Douglas Montgomery.
3. Mastering python for data science, Samir Madhavan.
4. Mastering python for machine Learning, Manohar Swamynathan.

### **Course Outcomes (COs):**

CO1: To introduce basic concepts of Data Science with its working principles.

CO2: To understand how Machine Learning can be used to build a DS model with different kinds of data preparation techniques, such as feature engineering, exploratory data analysis, data visualization, etc.

CO3: To introduce basic concepts of text mining.

CO4: To design a DS model and recommender system using machine learning to solve different Problems.

**Unit-1** Introduction: Machine Learning (ML), examples of ML application, ML cycle, Introduction of Supervised Learning, Un-supervised learning, Semi-supervised Learning, and Reinforced Learning. Brief introduction of Feature extraction, Feature Selection technique.

**Unit-2** Bayes decision theory: simplifying Bayes classification, Estimation of parameters: maximum likelihood and Bayesian, Minimization Risk, Naïve Bayes with binary attributes, Naïve Bayes for continuous attribute values

**Unit-3** Decision Tree: Introduction, building a decision tree (ID3), classifying by using a decision tree, ID3 decision tree with missing attribute values. CART Decision Tree, C4.5 Decision Tree, problem solving with Decision Tree.

**Unit-4** Non-parametric Decision making: introduction, histograms, Nearest Neighbor Techniques: the single nearest neighbor technique, the k- nearest neighbor technique, modified k-nearest neighbor Technique; Adaptive decision boundaries, Unsupervised Learning: clustering: agglomerative hierarchical clustering, k-means clustering, K-medoids clustering.

**Unit-5** Linear Discriminant Function: Training a linear classifier, two-class case, higher dimensional attribute space, Support vector machine: Linearly Separable Case, Non-Linearly separable case, Semi-supervised Learning: Introduction, Self-Training, Generative models, Reinforced Learning: Introduction, Markov Decision Process (MDP), Handling Multiclass problems, cross validation, Applications and Implementation Discussions

**Books:**

1. Introduction to Machine Learning – Ethem Alpaydin (PHI)
2. Machine Learning – Mitchell T. M. (Mcgraw Hill)
3. Pattern Classification – Duda R.O., Hart P. E., Strock D. G. (Wiley Interscience)
4. Pattern Recognition An Introduction—V. Susheela Devi, M. Narasimha Murty (University Press)

**Unit-1: Introduction to Algorithm :** Introduction to Data Structures, Algorithms, Time and Space Complexity. **Fundamentals of Data Structures:** Arrays and Lists: Storage, access, and manipulation, 2D arrays: Matrix manipulations. Stacks and Queues: FIFO and LIFO principles, applications in data processing. Hash Tables: Key-value storage, hashing functions, collision resolution, Trees: Binary trees, binary search trees, and their use in hierarchical data, Tries: String storage and retrieval for text processing, Segment Trees: Efficient range queries and updates

**Unit-2: Searching and Sorting:**

Binary Search and linear search, Selection sort, bubble sort, insertion sort, heap sort, merge sort and quick sort.

**Unit-3: 4. Graph Algorithms:**

Depth-First Search (DFS) and Breadth-First Search (BFS), Shortest path algorithms (e.g., Dijkstra's, Bellman-Ford), Minimum spanning trees (e.g., Kruskal's, Prim's)

**Unit-4: Pattern Searching:**

Naive string matching, Advanced algorithms (e.g., KMP, Rabin-Karp)

**Unit-5: Dynamic Programming:**

Memoization and tabulation, Applications in optimization problems

**Unit-6: Greedy Algorithms:**

Concepts and applications (e.g., activity selection, Huffman coding)

**Books/References**

1. Cormen, Thomas H., Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2022.
2. Robert Sedgewick and Kevin Wayne. *Algorithms*. Addison Wesley, 2011
3. Karumanchi, Narasimha. *Data Structures and Algorithms Made Easy*. CareerMonk, 2016.

CO1: Understand and explain fundamental algorithmic concepts such as time complexity, space complexity, and asymptotic notation.

CO2: Design efficient algorithms using techniques like divide and conquer, greedy, dynamic programming, and backtracking.

CO3: Analyze the performance of algorithms using best-case, worst-case, and average-case scenarios.

CO4: Solve complex computational problems using graph algorithms, searching, sorting, and optimization methods.

**Unit 1: Introduction to Distributed Database systems**

Distributed Computing, What is/is not a Distributed DBMS, History, Distributed DBMS promises, DDBMS Design issues, Typical application – IoT network database, Distributed DBMS architecture, DDBMS Implementation Alternatives

**Unit 2: Design of Distributed RDBMS**

Relational Model, ER Model, Relational Query languages, Distributed Data fragmentation, Replication and Allocation, Horizontal Fragmentation, Vertical Fragmentation, Hybrid or Mixed Fragmentation, Full and Partial Database replication, Various options for Distributing A Database

**Unit 3: SQL & Multi database connectivity**

Structured Query Language, Multi-Databases Access architectures, Open Database Connectivity (ODBC) Standard, Java Support for SQL, Java Database Connectivity, JDBC Application Architecture

**Unit 4: Functional of Distributed Database design**

Theory of Functional Dependency, Multivalued Dependency, Distributed Database decomposition, Normal Forms, BCNF, 4NF, 5NF, Database Normalization & De-Normalization.

**Unit 5: Distributed Query Processing & Optimization**

Query Processing in Distributed environment, Distributed Query decomposition, Data localization, Query cost calculation, Query optimization principles, Distributed Query optimization strategies

**Unit 6: Distributed Transaction Processing**

Distributed Transaction Management, Read set, Write set & Base set characterization, Serializability in Distributed DBMS, Distributed Concurrency Control algorithms, Multi-version Concurrency Control (MVCC), Distributed recovery and replication servers, Distributed Deadlock management.

**Unit 7: Current trends and developments,**

Recent developments & innovative solutions related to distributed database applications technologies,

**Text/References:**

1. M. Tomer Ozsu and P. Valduriez. Principles of Distributed Database Systems, Pearson Education.
2. S. Ceri and G. Pelagapati. Distributed Database, Principles and Systems, McGraw Hill Publication

CO1: Aware of Distributed Database systems architecture, Fragmentation and Replication methods.  
CO2: Familiar with multi site Relational database & Multi database connectivity and Querying.  
CO3: Use the different techniques of Distributed query processing & Optimization.  
CO4: Set the rules over management of transaction and concurrency control.  
CO5: Apprehend current trends, tools & innovations in DDBMS applications.

**Unit-1: Python Environment Setup:** Setting up Python environment using Anaconda (Conda), Installing and running Jupyter Notebook, Introduction to Google Colab for cloud-based Python coding.

**Unit-2: Scientific Computing with Python:** Installing and exploring features of NumPy, SciPy, and Pandas, Performing numerical operations, linear algebra, and data manipulation.

**Unit-3: Data Visualization Techniques :** Visualizing data using Matplotlib, Creating advanced plots with Seaborn.

**Unit-4: File Handling and Data Serialization :** Working with file input/output operations, Reading and writing files in different formats (CSV, JSON, etc.), Data serialization using pickle and other methods.

**Unit-5: Machine Learning with Python:** Implementing basic machine learning algorithms (e.g., Linear Regression, KNN, Decision Trees), Applying data preprocessing, model training, and evaluation techniques using scikit-learn.

**Unit-6: Deep Learning Fundamentals :** Introduction to Neural Networks and Python libraries: Keras and PyTorch, Building and training a simple neural network, Understanding and implementing the Backpropagation algorithm.

**Unit-7: Advanced Deep Learning Models :** Exploring Convolutional Neural Networks (CNNs) for image classification, Working with sequential models: Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) network.

**Books:**

1. "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython" by William McKinney
2. "Python Data Science Handbook" by Jake VanderPlas

CO1: Applies Python programming and essential libraries (NumPy, pandas, matplotlib) for data manipulation and visualization.

CO2: Implement basic machine learning and deep learning models using scikit-learn, Keras, and PyTorch.

CO3: Analyze and solve data-driven problems using neural networks, including CNNs, RNNs, and language models.

**M.Tech – Data Science Syllabus**  
**(2<sup>nd</sup> Semester)**

**CS 5307**

**Big Data Analytics and Visualization**

**3-0-0-3**

**Unit 1: Introduction**

Big data definition, classification, characteristics. Scalability and parallel processing. Designing data architecture. Data sources: quality, preprocessing and storing. Data storage and analysis – traditional and big data systems. Application of big data analytics in various areas.

**Unit 2: Data Storage for Big Data: NoSQL**

A very brief overview of triggers, views, schedules and joins with reference to their implementation details. ACID Vs BASE properties of transactions and CAP theorem. Schema less models and their conformance to CAP theorem. NoSQL: definition, its importance and relevance in big data processing. NoSQL data architecture patterns: key-value store, document store, tabular data (column family store, bigtable data store etc.), object data store, graph databases. Example databases: MongoDB and CASSANDRA.

**Unit 3: Hadoop**

Introduction and overview: Hadoop components – Hadoop common, HDFS, MapReduce, YARN, SPARK etc. and their layered organization. Hadoop data storage – features, rack and data node organisation in a cluster switch network, HDFS architecture: NameNode and secondary NameNode, DataNode, FSImage and EditLogs, replication management, HDFS read and write architecture. MapReduce architecture: input file structure and distribution, mapper, combiner, partitioner, shuffling and sorting, reducer, RecordWriter. Basic idea of MapReduce programming and optimizing MapReduce tasks. Discussion of HADOOP ecosystem: HDFS, YARN, Pig, Hive, HBase, Mahout, Oozie, Zookeeper.

**Unit 4: Data Preprocessing**

Data types: attributes and measurements, types of data sets. Data preprocessing: aggregation, sampling, dimensionality reduction, feature subset selection, feature creation, discretization and binarization, variable transformation. Proximity measures: similarity and dissimilarity between attributes and data objects, examples of proximity measures, selection of appropriate proximity measures. Summary statistics: frequency and the mode, percentiles, measure of location – mean and median, measure of spread – range and variance, multivariate summary statistics.

**Unit 5: Theory and Methods of Data Modelling**

Association rules, clustering, regression, classification, time series analysis, text analysis. Data stream mining.

**Unit 6: Data Visualization**

Introduction to visualization tools: Visual tools – Orange, KNIME, Geogebra, Tableau, Programming tools – Python Matplotlib+Pyplot, MATLAB/Octave, R. Visualization techniques: bar chart, stacked bar chart, line chart, histogram, pie chart, frequency polygon, box plot, scatter plot, heat map, tree map. Multivariate data visualization: geometric projection techniques, icon-based techniques, pixel-

oriented techniques, hierarchical techniques, scatterplot matrix, parallel coordinates. Data Dashboard (e.g Tableau): taxonomies, user Interaction, organizational functions, dashboard design, worksheets and workbooks. Dashboard creation using visualization tool use cases: finance, marketing, insurance, healthcare etc.

**Textbooks:**

1. **For units 1 & 2:** Big Data Analytics – Introduction to Hadoop, Spark and Machine Learning. Authors: Raj Kamal, Preeti Saxena. Publisher: McGraw Hill India. ISBN-10: 9353164966, ISBN-13: 978-9353164966.
2. **For units 3 & 5:** Big Data Analytics. Authors: G Sudha Sadasivam, R Thirumahal. Publisher: Oxford University Press. ISBN-10: 0-19-949722-2, ISBN-13: 978-0-19-949722
3. **For units 4 & 5:** Introduction to Data Mining. Authors: Pang-Ning Tan, Micheal Steinbach, Vipin Kumar. Publisher: Pearson. ISBN: 978-93-3257-140-2.
4. **For unit 6:**
  - a. Visualization Analysis and Design. Author: Tamara Munzner. Publisher: CRC Press, New York. eBook ISBN: 9781498759717.
  - b. Data Visualization: Exploring and Explaining with Data. Authors: Michael Fry, Jeffrey Ohlmann, Jeffrey Camm, James Cochran. Publisher: South-Western College Publishing. Print ISBN: 9780357631348, eBook ISBN: 9780357631430.
  - c. Data Visualization in R and Python. Author: Marco Cremonini. Publisher: Willey. Print ISBN: 9781394289486, Online ISBN: 9781394289516.
  - d. For Orange, KNIME, Geogebra and Tableau: Online materials should be sufficient.

**CO1:** Appreciate current trends and to have an overview of big data analytics from an application, architectural and research point of view.

**CO2:** To have an overview of NoSQL from an application, architectural and research point of view.

**CO3:** To have an overview of Hadoop and its ecosystem from an application, architectural and research point of view.

**CO4:** To have an overview of data preprocessing and modelling.

**CO5:** To learn and practice various data visualization techniques with special emphasis to big data.



**UNIT I: INTRODUCTION:** Basics of Neural Network, McCulloch–Pitts unit and Thresholding logic, Linear Perceptron, Perceptron Learning Algorithm, Linear separability, Convergence theorem for Perceptron Learning Algorithm, Type of network architecture, Activation functions, Basic Learning rules.

**UNIT II: FEEDFORWARD NETWORKS:** Multilayer Neural Network, Gradient Descent learning, Back propagation, Empirical Risk Minimization, regularization, Radial Basis Neural Network

**UNIT III: RECURRENT NEURAL NETWORKS:** Back propagation through time, Long Short Term Memory, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs.

**UNIT IV: DEEP NEURAL NETWORKS:** Introduction to Deep NN: Convolutional Neural Networks: LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, Visualizing Convolutional Neural Networks, Guided Back propagation, Deep Dream, Deep Art, Fooling Convolutional Neural Networks. Auto Encoders, Deep Reinforcement Learning

**UNIT V: PARAMETER TUNING:** Model-Optimization Algorithms (Adagrad, adadelata, rmsprop, adam, NAG), second order methods for training, Saddle point problem in neural networks, Regularization methods (dropout, drop connect, batch normalization).

#### Text Books

1. Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, 2016..
2. Bishop, C., M., Pattern Recognition and Machine Learning, Springer, 2006.
3. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
4. Golub, G., H., and Van Loan, C., F., Matrix Computations, JHU Press, 2013.
5. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.

#### UNIT I: OPTIMIZATION

Need for unconstrained methods in solving constrained problems, necessary conditions of unconstrained optimization, structure methods, quadratic models, methods of line search, steepest descent method; conjugate-direction methods: methods for sums of squares and nonlinear equations; linear programming: simplex methods, duality in linear programming, transportation problem, max/min Network flow.

## **UNIT II: UNCONSTRAINED OPTIMIZATION**

Line search method: Wolf condition, Goldstein condition, sufficient decrease and backtracking, Newtons method and Quasi Newton method; trust region method: the Cauchy point, algorithm based on Cauchy point, improving on the Cauchy point, the Dog-leg method, two-dimensional subspace reduction; nonlinear conjugate gradient method: the Fletcher Reeves method.

## **UNIT III: CONSTRAINED OPTIMIZATION**

Penalty method, quadratic penalty method, convergence, non-smooth penalty function, L1 penalty method, augmented Lagrangian method; quadratic programming, Schur complementary, null space method, active set method for convex QP; sequential quadratic programming, convex programming.

**UNIT IV:** Gradient based techniques such as Adam, AdaGrad, AdaDelta, Gradient Descent (GD), Stochastic Gradient Descent (SGD) etc. Metaheuristic techniques such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Differential Evolution (DE), Simulated annealing, TLBO, Greywolf optimization, etc.

### **Text book:**

1. Chong, E. K. and Zak, S. H., An Introduction to Optimization, 2nd Ed., Wiley India (2001).
2. Luenberger, D. G. and Ye, Y., Linear and Nonlinear Programming, 3rd Ed., Springer (2008).
3. Kambo, N. S., Mathematical Programming Techniques, East-West Press (1997).
4. Boyd, S. and Vandenberghe, L., Convex Optimization, Cambridge Univ. Press (2004).
5. Nocedal, J. and Wright, S. Numerical Optimization, Springer (2006).

CO1: To understanding different types of optimization problems like, linear, non linear, direction search etc.

CO2: To explain the working principle of optimization techniques, convergence, iteration.

CO3: To use optimization techniques in various problems on Constraint or unconstraint, soft optimization.

**UNIT I: Neural Network:** Perceptron, Back propagation NN, Radial Basis Neural Network

**UNIT II: Convolutional Neural Networks:** Customized, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet.

**UNIT III: Recurrent Neural Networks:** Back propagation through time, Long Short Term Memory, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs.

**UNIT IV: Model-Optimization Algorithm:** Gradient Descent, Stochastic Gradient Descent (SGD), Mini-batch Gradient Descent, AdaGrad (Adaptive Gradient Algorithm), RMSprop (Root Mean Square Propagation), AdaDelta, Adam (Adaptive Moment Estimation)

**Text Books:**

1. Seth Weidman, Deep Learning from Scratch: Building with Python from First Principles (Greyscale Indian Edition) Paperback –2019
2. Deepak Gowda : Practical Deep Learning with PyTorch: PyTorch implementation for computer vision, NLP, audio, and language translation (English Edition) Paperback- 2025
3. Ronald T. Kneusel : Practical Deep Learning: A Python-Based Introduction Paperback –2021

**Unit-1:** Introduction: Problem framing, feature selection, dimensionality reduction using PCA and other methods;

**Unit-2:** Linear discriminant functions: Gradient descent procedures, Perceptron, Support vector machines - a brief introduction; Artificial neural networks: Multilayer perceptron – feed forward neural network.

**Unit-3:** A brief introduction to deep neural networks, convolutional neural networks, recurrent neural networks; Non-metric methods for pattern classification: Non-numeric data or nominal data;

**Unit-4:** Decision trees: Classification and Regression Trees (CART); Unsupervised learning: Clustering, Vector Quantization, Kohonen Map, EM Algorithm;

**Unit-5:** Generative models: Definition and characteristics, probabilistic graphical models, density estimation in learning; Combining classifiers: Advantages, boosting, hierarchical classifiers, and issues;

**Unit-6:** Application(s): Face recognition - preprocessing, face detection algorithms, selection of representative patterns, classification algorithms, results and discussion.

Texts:

1. S. Marsland, *Machine Learning: An Algorithmic Perspective*, Chapman & Hall/CRC, 2009.
2. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2nd Edn., Wiley India, 2007.
3. S.Theodoridis and K.Koutroumbas, *Pattern Recognition*, 4th Ed., Academic Press, 2009
4. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2006.
5. P.A Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, Englewood Cliffs, NJ, 1980
6. I. H. Witten, *Data Mining: Practical Machine Learning Tools And Techniques*, 2nd Edn., Elsevier India, 2008.

**CS 5322**

**Speech Processing**

**3-0-0-3**

**Unit-1:** Overview of Speech Technology; What is Speech Technology? Why is it important? Its applications and issues.

**Unit-2:** Speech Production; Mechanism of speech production; Categories of sounds; Sound units in indian languages.

**Unit-3:** Nature of Speech Signal; Source-system characteristics; Segmental and suprasegmental features; Temporal and spectral parameters for sound units in indian languages.

**Unit-4:** Basics of Digital Signal Processing; Signals and systems; Discrete fourier transform; Digital filtering; Stochastic processes.

**Unit-5:** Speech Signal Processing Methods: Short-time spectrum analysis; Spectrograms; Linear prediction analysis; Cepstrum analysis.

**Unit-6:** Speech Recognition; Isolated word recognition; Connected word recognition Continuous Speech Recognition; Speech recognition problem; Hidden markov models.

**Unit-7:** Other Applications: Word spotting; Speaker recognition; Speech enhancement; Speech synthesis; Practical issues in speech technology.

**Text/Reference Books:**

1. L R Rabiner and R W Schafer, "Theory and Application of Digital Speech Processing," PH, Pearson, 2011.
2. L R Rabiner, B-H Juang and B Yegnanarayana, "Fundamentals of Speech Recognition," Pearson, 2009 (Indian subcontinent adaptation).
3. Xuedong Huang, Alex Acero, Hsiao-wuen Hon, "Spoken Language Processing:A guide to Theory, Algorithm, and System Development," Prentice Hall PTR, 2001.
4. Oppenheim and Schafer, "Discrete-Time Signal Processing," PHI, 2001
5. T W Parsons, "Voice and Speech Processing," McGraw Hill, 1986.
6. Thomas Quatieri, "Discrete-time Speech Processing: Principles and Practice," PH, 2001.
7. Rabiner and Schafer, "Digital Processing of Speech Signals," Pearson Education, 1993.
8. Douglas O' Shaughnessy, "Speech Communications," University Press, 2001.

**UNIT 1: Introduction to Data Mining:**

Definition, functionalities, classification of data mining systems, data mining tasks, and the relationship with Knowledge Discovery in Databases (KDD).

**UNIT 2: Data Warehouse Fundamentals:**

OLTP vs. OLAP, characteristics of data warehouses, and data warehouse development methodologies.

**UNIT 3: Data Preprocessing:**

Data cleaning, integration, transformation, reduction, discretization, and concept hierarchy generation.

**UNIT 4: Association Rule Mining:**

Frequent itemset mining, Apriori algorithm, FP-growth algorithm, and measuring rule quality.

**UNIT 5: Classification:**

Basic concepts, decision tree induction, Bayesian classifiers, support vector machines, and model evaluation techniques.

**UNIT 6: Clustering:**

Basic concepts, different clustering methods (K-means, hierarchical clustering, DBSCAN), and cluster evaluation.

**UNIT 7: Prediction:**

Linear and non-linear regression, model evaluation, and prediction using various techniques.

**Text / Reference Books**

1. Jiawei Han, Micheline Kamber and Jian Pei "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2011.
2. Alex Berson and Stephen J. Smith "Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint 2007.
3. K.P. Soman, Shyam Diwakar and V. Ajay "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006.
4. G. K. Gupta "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar "Introduction to Data Mining", Pearson Education, 2007.

CO1: At the end of the module the students will be able to understand the role of data mining in knowledge discovery process.

CO2: Familiarize with various data mining functionalities and how it can be applied to various real-world problems.

CO3: Learn about finding data characteristics and evaluating the outcome of data mining process.

CO4: Familiarize with various machine learning algorithms used in data mining.

**UNIT 1:** Introduction to Computer Vision and Image processing (CVIP): Basics of CVIP, History of CVIP, Evolution of CVIP, CV Models, Image Filtering, Image Representations, Image Statistics, Recognition Methodology: Conditioning, Labeling, Grouping, Extracting, and Matching,

**UNIT 2:** Morphological Image Processing: Introduction, Dilation, Erosion, Opening, Closing, Hit-or-Miss transformation, Morphological algorithm operations on binary images, Morphological algorithm operations on gray-scale images, Thinning, Thickening. Region growing, region shrinking.

**UNIT 3:** Image Representation and Description: Representation schemes, Boundary descriptors, and Region descriptors, Binary Machine Vision: Thresholding, Segmentation, Connected component labeling, Hierarchical segmentation, spatial clustering, Split& merge, Rule-based Segmentation, Motion-based segmentation. Edge, Line-Linking, Hough transform, Line fitting, Curve fitting (Least-square fitting).

**UNIT 4:** Region Analysis: Region properties, External points, Spatial moments, Mixed spatial gray-level moments, Boundary analysis: Signature properties, Shape numbers. General Frame Works For Matching: Distance relational approach, Ordered structural matching, View class matching, Models database organization

Books:

1. Digital Image Processing: R.C. Gonzalez & R. E. Woods, Addison Wesley, Pearson Education, 2nd Ed, 2004.
2. Robert Haralick and Linda Shapiro, "Computer and Robot Vision", Vol I, II, Addison Wesley, 1993.
3. David A. Forsyth, Jean Ponce, "Computer Vision: A Modern Approach" Pearson
4. Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis, and Machine Vision" Thomson Learning.

**UNIT 1:** Introduction: concepts and terminology of information retrieval systems, Information Retrieval Vs Information Extraction; Indexing: inverted files, encoding, Zipf's Law, compression, boolean queries;

**UNIT 2:** Fundamental IR models: Boolean, Vector Space, probabilistic, TFIDF, Okapi, language modeling, latent semantic indexing, query processing and refinement techniques;

**UNIT 3:** Performance Evaluation: precision, recall, F-measure;

**UNIT 4:** Classification: Rocchio, Naive Bayes, k-nearest neighbors, support vector machine; Clustering: partitioning methods, k-means clustering, hierarchical;

**UNIT 5:** Introduction to advanced topics: search, relevance feedback, ranking, query expansion.

**Texts/References:**

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schtze, Introduction to Information Retrieval, Cambridge University Press. 2008.
2. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval, Addison Wesley, 1st edition, 1999.
3. Soumen Chakrabarti, Mining the Web, Morgan-Kaufmann Publishers, 2002.
4. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, Corr. 2nd printing edition, 2009.
5. David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 2nd edition, 2004.
6. William B. Frakes, Ricardo Baeza-Yates, Information Retrieval Data Structures and Algorithms, Prentice Hall, 1992.
7. G. Salton, M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1986.
8. C. J. Van Rijsbergen, Information Retrieval, Butterworth-Heinemann; 2nd edition, 1979.

**Unit 1: Introduction to Systems Biology**

Overview of molecular and cellular biology, Introduction to omics data (genomics, transcriptomics, proteomics, metabolomics), Introduction to Modelling, Fundamentals of mathematical modelling, Some example models, Representation of Biological Networks, MATLAB Basics

**Unit 2: Introduction to Networks**

Introduction to Network Biology, Network models, Biological networks, network perturbations, community detection, network motifs, Cytoscape, Gene regulatory networks, Protein networks, Signalling networks

**Unit 3: Dynamic Modelling and Genetic Algorithms**

Introduction to Dynamic modelling, Solving ODEs in MATLAB, Example Biological Model, parameter estimation, Genetic Algorithms, Other evolutionary algorithms, Drug development, modelling in drug development, Quantitative Systems

**Unit 4: Omics Data Integration and Analysis**

Integration of multi-omics datasets, dimensionality reduction and clustering techniques (PCA, t-SNE, hierarchical clustering), machine learning applications in systems biology

**Textbooks/references:**

1. Raman K (2021) An Introduction to Computational Systems Biology: Systems-Level Modelling of Cellular Networks. 1/e ISBN 9781138597327 (Chapman and Hall/CRC)
2. Voit E (2012) A First Course in Systems Biology. Garland Science, 1/e. ISBN 081534467
3. Klipp E (2009) Systems biology: a textbook. Wiley-VCH, 1/e. ISBN 9783527318742
4. Newman MEJ (2011) Networks: an introduction. Oxford Univ. Press. ISBN 9780199206650

**Course outcomes (CO):**

- CO1: Understand biological networks and their computational representations
- CO2: Model and simulate the dynamic behavior of biological systems
- CO3: Analyze and integrate omics data for systems-level insights
- CO4: Apply computational tools to investigate real biological problems



**UNIT 1: INTRODUCTION**

Introduction to Data, Data types, Introduction to Stages of Data Processing: Data PreProcessing; Data Imputation; Data Cleaning; Data Transformation; Data Visualization; Inferential Statistics: Estimation, test of hypothesis, analysis of variance (ANOVA)

**UNIT 2: EXPLORATORY DATA ANALYSIS AND VISUALIZATION**

Introduction to the Chicago Train Ridership data, Visualizations for Numeric Data: Exploring Train Ridership Data, Visualizations for Categorical Data: Exploring the OkCupid Data, Visualizing Relationships between Outcomes and Predictors, Exploring Relationships between Categorical Predictors, Post Modelling Exploratory Visualizations.

**UNIT 3: FEATURE SELECTION AND ENGINEERING**

Feature Selection, Classes of Feature Selection Methodologies: intrinsic (or implicit) methods, filter methods, and wrapper methods, Feature Engineering, Feature Engineering techniques: Binning, Feature Hashing, Log Transforms, n-grams, Binarisation, Bag-of-words.

**UNIT 4: MACHINE LEARNING MODELS AND EVALUATION**

Review of mathematical concepts in machine learning, linear and logistic regression, regularization, gradient descent, bayesian modeling, expectation-maximization algorithm, gaussian mixture models, markov models, clustering, k-nearest neighbor, support vector machines, different types of artificial neural networks, principal component analysis, deep learning, model evaluation approaches: Cross Validation, Confusion Matrix, Gain and Lift cgart, Kolmogorov-Smirnov Chart, Chi Square, ROC curve, Gini Coefficient,  $L^1$  version of RSME.

**Textbooks/References:**

1. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013.
2. Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann, 2011.
3. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. Springer, 2009.
4. Murphy, K. Machine Learning: A Probabilistic Perspective. - MIT Press, 2012.
5. Zumel, N. , Mount, J. Practical Data Science with R". Manning, 2014.
6. G. Strang (2016). Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA.
7. Bendat, J. S. and A. G. Piersol (2010). Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA:
8. Montgomery, D. C. and G. C. Runger (2011). Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA.
9. David G. Luenberger (1969). Optimization by Vector Space Methods, John Wiley & Sons (NY)

CO1: Understand the types of data and apply basic data processing techniques.

CO2: Perform exploratory data analysis (EDA) using visual and statistical methods to uncover patterns, relationships, and trends in real-world datasets.

CO3: Implement feature selection and feature engineering techniques to improve the performance of machine learning models.

CO4: Apply machine learning algorithms and evaluate their performance using appropriate metrics and visualizations.

**CS 5328**

**Quantum Computing**

**3-0-0-3**

**Unit 1:** Introduction and overview:

Basics of quantum computing, Quantum Bits or Qubit, representation of qubits, reversible Computation, basics of Quantum mechanics and linear algebra, Quantum Algorithms, Parallelism, Postulates of Quantum Mechanisms, bra-ket notation, single and multiple qubit, Entanglement, EPR pair.

**Unit 2:** Quantum Circuits:

Characteristics, Design, Single and multi qubit operations, Operators, Quantum States, Control Operations, Measurement, Universal Quantum Gates, design of common quantum circuits, List of merits in quantum circuits, synthesis of reversible logic-MDM method etc, Simulation of Quantum Systems.

**Unit 3:** Common quantum algorithms:

Deutsch's algorithm, Deutsch-Jozsa algorithm, Quantum Fourier transform, Simon's algorithm, Shor algorithm, Phase estimation, Quantum search algorithms, Oracle, Grover search, Hamiltonian simulation etc.

**Unit 4:** Quantum noise and Quantum Operations:

Classical Noise and Markov Processes, Noise and Environment, Bit flip and phase flip channels, Trace and Partial trace etc. Quantum error correction: Introduction, Three qubit flip code, Three qubit phase code, Shor code for error correction, Stabilizers codes etc.

**Unit 5:** Basics of Entropy and information

Quantum information theory, Quantum Cryptography, Quantum annealing, Different Physical realization of quantum computers--Harmonic Oscillator Quantum Computer, Optical Photon Quantum Computer, etc.

**Books:**

1. Micheal A. Nielsen. & Issac L. Chiang, "Quantum Computation and Quantum Information", Cambridge University Press, 2010, ISBN: 1139495488, 9781139495486.
2. Colin P. Williams, "Explorations in Quantum Computing", Springer Science & Business Media, 2010, ISBN: 1846288878, 9781846288876.

3. Eleanor G. Rieffel, Wolfgang H. Polak, "Quantum Computing: A Gentle Introduction", MIT Press, 2011, ISBN: 0262015064, 9780262015066.
4. Mika Hirvensalo, "Quantum Computing", Springer Science & Business Media, 2013, ISBN: 3662096366, 9783662096369.
5. Sahni, "Quantum Computing", Tata McGraw-Hill Education, 2007, ISBN: 0070657009, 9780070657007.
6. Robert Wille, Rolf Drechsler, "Towards a Design Flow for Reversible Logic", Edition illustrated, Publisher Springer Science & Business Media, 2010, ISBN9048195799, 9789048195794.
7. Authors Nabila Abdessaied, Rolf Drechsler, "Reversible and Quantum Circuits: Optimization and Complexity Analysis", Edition illustrated, Publisher Springer, 2016, ISBN 331931937X, 9783319319377.
8. Bhunia C T., "Quantum Computing", New Age International Publishers, ISBN-10: 8122430759, ISBN-13: 978-8122430752.
9. Phillip Kaye, Raymond Laflamme, Michele Mosca, "An Introduction to Quantum Computing", Oxford University Press, 2007, ISBN: 019857049X, 9780198570493.
10. Giuliano Benenti, Giulio Casati, Giuliano Strini, "Principles of Quantum Computation and Information", World Scientific, 2004, ISBN: 9812388583, 9789812388582.
11. Noson S. Yanofsky, Mirco A. Mannucci, "Quantum Computing for Computer Scientists", Cambridge University Press, 2008, ISBN: 1139643908, 9781139643900.
12. N. David Mermin, "Quantum Computer Science: An Introduction", Cambridge University Press, 2007, ISBN: 1139466801, 9781139466806.
13. Jiannis K. Pachos, "Introduction to Topological Quantum Computation", Cambridge University Press, ISBN: 1107005043, 9781107005044.
14. Salvador Elías Venegas-Andraca, "Quantum Walks for Computer Scientists", Morgan & Claypool, ISBN: 1598296566, 9781598296563.

CO1: Introduction of quantum computation, classical bit and qubit, basics of linear algebra, Postulates, notations and representation.

CO2: Basic notions of quantum computing- qubits, mathematical representation, quantum gates, operators, measurement operators, quantum evolution, quantum circuits, Entanglement etc.

CO3: Quantum algorithms, parallelism, search algorithm etc.

CO4: Physical realisation, quantum noise, error correction. quantum cryptography etc.

**Unit 1: Introduction to bioinformatics and data generation**

What is bioinformatics and its relation with molecular biology. Examples of related tools (FASTA, BLAST, BLAT, RASMOL), databases (GENBANK, Pubmed, PDB) and software (RASMOL, Ligand Explorer). Data generation; Generation of large scale molecular biology data through Genome sequencing, Protein sequencing, Gel electrophoresis, NMR Spectroscopy, X-Ray Diffraction, and microarray. Applications of Bioinformatics

**Unit 2: Sequence Alignments and Visualization**

Introduction to Sequences, alignments and Dynamic Programming; Local alignment and Global alignment (algorithm and example), Pairwise alignment (BLAST and FASTA Algorithm) and multiple sequence alignment (Clustal W algorithm). Methods for presenting large quantities of biological data: sequence viewers (Artemis, SeqVISTA), 3D structure viewers (Rasmol, SPDBv, Chime, Cn3D, PyMol), Anatomical visualization.

**Unit 3: Gene Expression and and Representation of patterns and relationship**

General introduction to Gene expression in prokaryotes and eukaryotes, transcription factors binding sites. SNP, EST, STS. Introduction to Regular Expression, Hierarchies, and Graphical models (including Markov chain and Bayes notes). Genetic variability and connections to clinical data.

**Unit 4: Protein structure and function**

Protein secondary structure classification databases: HSSP, FSSP, CATH, SCOP. Protein secondary structure prediction methods: GOR, Chou-Fasman, PHD, PSI-PRED, J-Pred. Protein Tertiary structure prediction methods: Homology Modeling, Fold Recognition, Ab-initio Method. Protein folding, Molecular Dynamics of Protein, Molecular Docking of Protein, Small molecule and Nucleotide, Concepts of Force Field

**Textbooks/References:**

1. Bioinformatics- a Practical Guide to the Analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellette, B.F., Wiley India Pvt Ltd. 2009
2. Essential Bioinformatics by Jin xiong., Cambridge University press, New York. 2006
3. Discovering Genomics, Proteomics and Bioinformatics 2nd edition - by A. Malcolm Campbell and Laurie J. Heyer. by Cold Spring Harbor Laboratory Press 2006.
4. Bioinformatics: Sequence and Genome Analysis by Mount D., Cold Spring Harbor Laboratory Press, New York. 2004
5. Introduction to bioinformatics by Teresa K. Attwood, David J. Parry-Smith. Pearson Education. 1999 Old editions
6. Bioinformatics in the Post-Genomic Era by Jeffrey Augen, Addison-Wesley Publisher, 2004.

CO1: Ability to access, clean, and manipulate large biological datasets (e.g., genomic sequences, protein structures).

CO2: Apply appropriate statistical methods to analyze and visualize biological data.

CO3: Implement algorithms for sequence alignment, gene expression, and protein structure prediction.

**UNIT 1: Introduction and Preliminaries:**

**Mathematical Modeling:** Formulating optimization problems, defining objective functions, decision variables, and constraints. **Types of Optimization Problems:** Classifying problems based on linearity, convexity, continuity, and constraints. **Classical Optimization Techniques:** Understanding optimality criteria, calculus-based methods (single and multi-variable), and the Kuhn-Tucker conditions.

**UNIT 2: Linear Programming:**

**Fundamentals:** Standard form, graphical methods, simplex method, duality theory, sensitivity analysis. **Applications:** Transportation and assignment problems, structural and water resources problems.

**UNIT 3: Nonlinear Programming:**

**Nonlinear Optimization:** One-dimensional minimization methods, unconstrained and constrained optimization, transformation methods. **Constrained Optimization:** Penalty function methods, Lagrange multipliers.

**UNIT 4: Numerical Optimization Methods:**

**Line Search Methods:** Bracketing methods, region elimination methods, point estimation methods. **Gradient-based Methods:** Steepest descent, Newton's method, quasi-Newton methods. **Conjugate Gradient Methods:** Conjugate direction methods.

**UNIT 5: Dynamic Programming:**

**Introduction:** Recursive equations, computational procedure, and applications. **Applications:** Design of continuous beams, water allocation, and other sequential decision-making problems.

**UNIT 6: Constrained Optimization:**

**Penalty and Barrier Methods:** Interior and exterior penalty functions, barrier methods. **Lagrange Multipliers:** Kuhn-Tucker conditions.

**UNIT 7: Metaheuristic Optimization Methods:**

**Genetic Algorithms:** Principles of natural selection, operators like crossover and mutation. **Other Metaheuristics:** Simulated annealing, particle swarm optimization, ant colony optimization.

**Text / Reference Books**

1. S.S. Rao, "Engineering Optimization: Theory and Practice", New Age International P) Ltd., New Delhi, 2000.
2. G. Hadley, "Linear programming", Narosa Publishing House, New Delhi, 1990.
3. H.A. Taha, "Operations Research: An Introduction", 5th Edition, Macmillan, New York, 1992.
4. K. Deb, "Optimization for Engineering Design Algorithms and Examples", Prentice-Hall of India Pvt. Ltd., New Delhi, 1995.

CO1: At the end of the module the students will be able to understand the need and origin of the optimization methods.

CO2: Get a broader picture of the various applications of optimization methods used in

- engineering.
- CO3: Define an optimization problem and its various components.
- CO4: Formulate optimization problems as mathematical programming problems.
- CO5: Classify optimization problems to suitably choose the method needed to solve the particular type of problem.
- CO6: Briefly learn about classical and advanced techniques in optimizations.

**CS 5342**

**Machine Learning using Cloud Computing**

**3-0-0-3**

**UNIT 1: Introduction:**

Overview of Machine Learning (ML) and its types, Overview of Cloud Computing (IaaS, PaaS, SaaS), Importance of cloud in ML (scalability, deployment, cost-efficiency), ML lifecycle and where cloud fits, What is Machine Learning? Types (Supervised, Unsupervised, Reinforcement), Traditional Programming vs. Machine Learning, Applications of ML in various domains, Basics of Python and ML libraries (NumPy, pandas, scikit-learn)

**UNIT 2: Data Preprocessing and Feature Engineering:**

Data cleaning (handling missing, noisy, and inconsistent data), Encoding categorical variables, scaling, and normalization, Feature selection and dimensionality reduction, Train-test split, cross-validation

**UNIT 3: Supervised Learning Algorithms**

Regression: Linear, Polynomial; Classification: Logistic, K-Nearest Neighbour, Decision Tree and Random Forests, Support Vector Machines, Model evaluation: Confusion Matrix, Accuracy, Precision, Recall, F1-score, ROC-AUC

**UNIT 4: Unsupervised Learning Algorithms:**

Clustering: K-Means, DBSCAN, Hierarchical Clustering, Dimensionality Reduction: PCA, t-SNE Association Rule Mining: Apriori, Eclat

**UNIT 5: Ensemble Learning and Model Improvement**

Bagging and Boosting: Random Forest, AdaBoost, Gradient Boosting, Hyperparameter tuning: Grid Search, Random Search, Cross-validation techniques

**UNIT 6: Neural Networks and Deep Learning Basics:**

Perceptron and Multilayer Perceptron (MLP), Introduction to TensorFlow/PyTorch, Backpropagation and gradient descent, Activation functions and loss functions

**UNIT 7: ML Workflow on the Cloud:**

Data collection and storage in cloud, Data preprocessing using cloud tools (e.g., AWS Glue, Azure Data Factory), Model training on cloud (using Jupyter Notebooks, Google Colab, or SageMaker), Using cloud GPU/TPU for large-scale training

## References

1. Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*. Shroff/O'Reilly; Third edition, 2022
2. Christopher Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
3. Sebastian Raschka & Vahid Mirjalili. *Python Machine Learning*. Pckt, 2017.

CO1: Understand the fundamental concepts, types, and applications of Machine Learning.  
CO2: Apply supervised and unsupervised learning algorithms to real-world problems.  
CO3: Preprocess and transform raw data for use in machine learning models.  
CO4: Preprocess and transform raw data for use in machine learning models.  
CO5: Understand cloud computing for machine learning.

**Unit 1: Content mining**

Topics on social media content mining including retrieval, ranking, trends detection, event detection, event forecasting, opinion mining, sentiment analysis, fake news detection and any other relevant topics.

**Unit 2: Link mining**

Topics on social media link analysis including centrality measures, community detection, link prediction, source detection, influence analysis and any other relevant topics.

**Unit 3: Log analysis**

Topics related to user's behavioral analysis, personalization, recommendation, and any other relevant topics.

**Unit 4: Various Tools**

Various state-of-the-art big data analytics tools for mining social media data.

**Reference Books:**

1. C.D. Manning, P.Raghavan and H.Schutze, Introduction to Information Retrieval, Cambridge University Press. 2008.
2. M.A. Russell, Mining the Social Web, 2nd Edn., O'Reilly Media, 2013.
3. R.Zafarani, M.A.Abbasi and H.Liu, Social Media Mining: An Introduction, Cambridge University Press, 2014

**Course Outcome (CO):**

CO1: Aware of various topics and different tools of social media data mining.

CO2: Understand the working principles of different algorithms related to social media data mining.

CO3: Analysis and evaluation of performance of algorithms related to social media data mining.



**UNIT 1:** Introduction:

Components of multimedia systems, Multimedia applications, description and architecture of multimedia system, system specifications;

**UNIT 2:** Multimedia content and formats:

Text, digital audio concepts, Sound MIDI, audio file formats, audio and video capture, Sampling, Quantization, Image, audio and Video formats, Animation;

**UNIT 3:** Multimedia content coding and compression:

Lossy and Lossless Compression, Image, Audio and Video Compression, Huffman Coding, Shannon Fano Algorithm, Adaptive Coding, Arithmetic Coding, Higher Order Modeling, Finite Context Modeling, Dictionary based Compression, Transform Coding,

**UNIT 4:** Case study: LZ77, LZW, JPEG, DCT, MPEG; Multimedia Server and access: Multimedia protocols, services, Interfaces, Multimedia server and clients, Multimedia streaming, Multimedia on demand, Security issues; Contemporary research topics.

## Texts/References:

1. Kamisetty Rao, Zoran Bojkovic, Dragorad Milovanovic, "Introduction to Multimedia Communications: Applications, Middleware, Networking", Wiley-Interscience, 2006
2. Nikil Jayant, "Signal Compression: Coding of Speech, Audio, Text, Image and Video, World Scientific", 1997
3. Vasudev Bhaskaran, and Konstantinos Konstantinides, "Image and Video Compression Standards: Algorithms and Architectures", Springer 1997
4. Ralf Steinmetz and Klara Nahrstedt, "Multimedia Systems", Springer, 2004
5. Khalid Sayood, "Lossless Compression Handbook (Communications, Networking and Multimedia)", 1/E, Academic Press; 2002

**Unit 1:** Foundations and State-of-the-Art in AI, Revisiting classical AI: Search, logic, knowledge representation, Modern AI trends: Deep learning, transformers, generative AI, Overview of research directions: Explainable AI, embodied AI, Neuro-symbolic AI

**Unit 2:** Advanced Deep Learning Architectures, Convolutional Neural Networks (CNNs) – ResNet, DenseNet, Recurrent Neural Networks (RNNs) – LSTM, GRU, Attention Mechanism and Transformers (BERT, GPT), Generative Models – GANs, VAEs, Applications: NLP, computer vision, speech

**Unit 3:** Reinforcement Learning and Decision Making, Markov Decision Processes (MDPs), Dynamic Programming, Q-learning, SARSA, DQN, Policy Gradient, Actor-Critic Methods, Deep Reinforcement Learning, Applications: Robotics, games, recommendation systems

**Unit 4:** Probabilistic Reasoning and Graphical Models, Bayesian networks, Hidden Markov Models, Probabilistic inference and learning, Structured prediction and approximate inference, Temporal and spatial models

**Unit 5:** Multi-Agent Systems and Swarm Intelligence, Introduction to multi-agent systems, Game theory and strategic decision making, Coordination, communication, and distributed AI, Swarm intelligence and emergent behavior,

**Unit 6:** Explainable, Ethical and Responsible AI, Explainable AI (XAI): Interpretable models and techniques, Fairness, accountability, transparency (FAT), Bias in datasets and mitigation techniques, Ethical considerations and societal impacts of AI, Regulations and AI governance frameworks

**Unit 7:** AI in Practice and Research Frontiers, Real-world AI case studies (healthcare, finance, security), Open-source AI frameworks (PyTorch, TensorFlow, HuggingFace), AI for social good and sustainable development, Quantum AI and neuromorphic computing (overview)

**Books:**

1. **"Artificial Intelligence: A Modern Approach"** by Stuart Russell & Peter Norvig
2. **"Deep Learning"** by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
3. **"Probabilistic Graphical Models: Principles and Techniques"** by Daphne Koller and Nir Friedman
4. **"Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations"** by Yoav Shoham and Kevin Leyton-Brown

CO1: Evaluate recent advancements and state-of-the-art AI algorithms.

CO2: Apply advanced deep learning models to real-world problems.

CO3: Design intelligent agents using reinforcement learning and planning algorithms.

CO4: Analyze AI systems for interpretability, fairness, and robustness.

CO5: Critically assess ethical and societal implications of AI deployment.

**CS 5346**

**High Performance Computing and Data Science**

**3-0-0-3**

**Unit 1:** Foundations of HPC and Data Science:

Introduction to HPC & Data Science, Cluster, Grid, and Cloud HPC Systems, Von Neumann Architecture and Modern Processors, Memory Hierarchies: Cache, RAM, Disk, Shared vs. Distributed Memory Architectures

**Unit 2:** Parallel Computing Concepts:

Types of Parallelism, Amdahl's Law, Gustafson's Law, Efficiency, Scalability, Load Balancing, Shared Memory Programming Basics (OpenMP Concepts), Threading Models and Synchronization (critical, barriers), MPI: Point-to-Point and Collective Communication, MPI Data Distribution and Scalability Concepts

**Unit 3:** Accelerators, Scheduling & Optimization: CPU vs. GPU, CUDA Programming Model Overview, GPU Memory and Execution Optimization (concepts only), HPC Job Scheduling (e.g., SLURM) – Queues, Nodes, Partitions

**Unit 4:** Scalable Data Science and Machine Learning: Data Types, Storage, Cleaning, Transformation, ML at Scale: Common Bottlenecks, Model Parallelism, Parallel ML Algorithms and Toolkits: XGBoost, joblib, RAPIDS, Distributed Deep Learning: PyTorch Distributed, Horovod, Data vs. Model Parallelism in Deep Learning

**Unit 5:** Advanced Topics & Integration: Bottleneck Analysis and Pipeline Optimization, Reproducibility: Version Control, MLFlow, Metadata Management, Reproducibility: Version Control, MLFlow, Metadata Management, Containers: Docker, Singularity for HPC Environments. Future of HPC: AI Accelerators (TPUs), Quantum HPC

Text Books:

1. Introduction to High Performance Computing for Scientists and Engineers – Georg Hager and Gerhard Wellein (CRC Press)
2. Using OpenMP: Portable Shared Memory Parallel Programming – Barbara Chapman, Gabriele Jost, and Ruud van der Pas (MIT Press)
3. Programming Massively Parallel Processors: A Hands-on Approach – David B. Kirk and Wen-mei W. Hwu (Morgan Kaufmann)
4. Data Science from Scratch: First Principles with Python – Joel Grus (O'Reilly Media)
5. Learning Spark: Lightning-Fast Big Data Analysis – Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee (O'Reilly Media)

**CS 5347**

**Health Informatics**

**3-0-0-3**

**Unit 1:** Health Data, Information, and Knowledge: Describe the data, information, and knowledge environment of health informatics, role of informatics to improve healthcare

**Unit 2:** Machine Learning for Healthcare: Brief history of AI and ML in Healthcare, Examples of Machine learning in health care, Algorithms as medical devices, Dynamic machine learning devices, Challenges with machine learning for healthcare,

**Unit 3:** Electronic Health Records: Describe and demonstrate the electronic health record (EHR) and its derivatives, and the functions for which it is used, including clinical decision support and re-use of clinical data

**Unit 4:** Personal Health Records and Decision Aids: Describe and demonstrate the personal health record (PHR) and decision aids, and the functions for which they are used to inform personal health decision-making and clinical decision support system.

**Unit 5:** Information Retrieval (Search): Discuss the discovery and dissemination of health-related knowledge and demonstrate the ability to retrieve and appraise it

**Unit 6:** Data Science, Analytics, and Visualization: Apply analytical and visualization skills to medical data sets, Perform visualization and simple analysis of a data set to assess difficulty of predicting cardiovascular risk in a synthetic patient dataset

**Unit 7:** Informatics Applications in Public Health: Describe the applications of informatics to public health, Ethical Issues in Health Informatics: Discuss the ethical challenges for the use of data and information in health-related areas.

**Text / Reference Books**

1. Hoyt, RE and Yoshihashi, A, Eds. (2014). Health Informatics: Practical Guide for Healthcare and Information Technology Professionals, Sixth Edition. Pensacola, FL, Lulu.com.
2. Biomedical Informatics, Computer Applications in Health Care and Biomedicine Shortliffe, E.&Cimino, J. (Eds) Springer,4thEdition 2014

CO1: Develop knowledge about problems and challenges that health informatics addresses

CO2: Apply basic knowledge to the research and practice of health informatics

CO3: Demonstrate basic skills and knowledge in health informatics for application in future health-related careers

CO4: Demonstrate ability to ask, search for information, and answer health-related questions

CO5: Demonstrate ability to identify genomic variants associated with a disease phenotype and communicate this association

CO6: Analyze ethical and diversity issues in health informatics

**Unit-I: Image Fundamentals:**

Natural Image and Digital Image through Scanner, Image Model, Concept of Gray Levels, Gray Level to Binary Image Conversion, Sampling and Quantization, Relationship between Pixels, Imaging Geometry, 2D Transformations-FFT, DFT, DCT, KLT and SVD.

**Unit-II: Image Enhancement in Spatial Domain Point Processing:**

Image Enhancement in Spatial Domain Point Processing, Grey level transformation function, Logarithmic transformation function, Power-law function, Histogram Processing, Spatial Filtering, Enhancement in Frequency Domain, Image Smoothing, Image Sharpening.

**Unit-III: Image Restoration Degradation Model:**

Image Restoration Degradation Model, Algebraic Approach to Restoration, Inverse Filtering, Least Mean Square Filters, Constrained Least Squares Restoration, Interactive Restoration.

**Unit-IV: Image Segmentation Detection of Discontinuities:**

Image Segmentation, Edge Detection, Edge operators, Laplacian Operator, Edge Linking and Boundary Detection, Thresholding, Adaptive Thresholding, Optimum Thresholding, Region Oriented Segmentation.

**Unit-V: Image Compression Redundancies and their Removal Methods:**

Image Compression Redundancies and their Removal Methods, Fidelity Criteria, Image Compression Models, Source Encoder and Decoder, Error Free Compression, Lossy Compression.

**TEXT BOOK:**

1. Digital Image Processing: R.C. Gonzalez & R. E. Woods, Addison Wesley, Pearson Education, 2nd Ed, 2004.
2. Fundamentals of Digital Image Processing: A. K. Jain , PHI.
3. Digital Image Processing using MATLAB: Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins: Pearson Education India, 2004.
4. Digital Image Processing: William K. Pratt, John Wiley, 3rd Edition, 2004.

CO-1 Understand the relevant aspects of Natural Image representation and their practical implications.

CO-2 Understand 2-D convolution, the 2-D DFT, and have the ability to design systems using these concepts.

CO-3 Understand and apply Image Restoration & Enhancement techniques in various Image based applications.

CO-4 Apply image processing technique in solving real-time problems.

**Unit -1 Fundamentals of Data Communication**

Overview of data communication and networking, data transmission concept analog and digital, transmission media guided and unguided, bandwidth, latency and throughput, encoding and modulation techniques (ASK, FSK, PSK).

**Unit-2 Network Models and Protocols**

OSI and TCP/IP models, layered architecture and protocols, internetworking principles, network devices, IP addressing and Subnetting (IPv4 and IPv6), Data Link Layer functions; error detection and correction, parity, CRC, hamming code, framing, ARQ, sliding window protocols, HDLC, SDLC, Point to point protocol (PPP), ATM, Medium Access sublayer; MAC protocols; CSMA/CD, CSMA/CA, channelization; FDMA, TDMA, CDMA, IEEE 802, 802.3 standard, Ethernet and VLANs, Routing algorithms, static algorithms, dynamic routing, hierarchical routing, routing for mobile hosts, broadcast / multicast routing, ARP, RARP, ICMP, RIP, OSPF, BGP, IGMP, Mobile IP, Mobile IP route optimization.

**Unit -3 Transport layer, session layer and presentation layers**

Sockets, socket address, difference between IP address and port numbers, connection-oriented service, Internet transport protocols (TCP and UDP), TCP transmission policy, TCP congestion control, wireless TCP and UDP, performance issues.

**Unit- 4 Application Layer services**

Client-Server paradigm, RPC, DNS in Big data application, e-mail architecture and services, MIME, SMTP, delivery protocols, mail server, FTP, Trivial file management protocol (TFTP), SNMP, DHCP, WWW, URL, HTTP, Browser architecture, structure of a web page, dynamic document, server replication, remote login, wireless application protocol.

**Unit-5 Network Security fundamental**

Security requirements and attacks, cryptography components; plaintext, ciphertext, encryption/decryption algorithm, symmetric-key cryptography, asymmetric-key cryptography, hash function, digital signature, traditional cipher, block cipher, overview of IPSec, authentication header format (transport/ tunnel mode), Encapsulating Security Payload (ESP) format, message/user authentication, pseudorandom number generation. key management and distribution techniques, VPNs, firewalls, secure communication in data science application, fundamentals of cloud/edge computing for data storage/collection, software-defined networking (SDN), Hadoop, Spark, AWS.

**Text Books:**

1. Data Communications and Networking, Behrouz A. Forouzan.
2. Computer Networks, Andrew S. Tanenbaum and David J. Wetherall.
3. Data and Computer Communications, William Stallings.
4. Cryptography and Network Security Principles and Practice by William Stalling.
5. Data Science for Network Cybersecurity by Amiya Nayak and Loannis Chochilours
6. Security in Computing by Charles P. Pfleeger and Shari L. Pfleeger.

CO1: To understand fundamental and advanced concepts of data communication and network protocols.

CO2: To explore the TCP/IP model and able to configure the network IP addresses, and apply subnetting techniques to optimize the network traffic flow and reduce latency in distributed data processing (e.g., Hadoop/Spark clusters).

CO3: To identify and understand security threats, cryptographic techniques and authentication methods.

CO4: To implement IPsec and learn emerging technologies like Hadoop, Spark for scalable data processing.

**M.Tech – Data Science Syllabus**  
**(3<sup>rd</sup> Semester)**

**CS 6301**

**Data Science Laboratory**

**0-0-2-2**

Data Collection – Data Preparation – Business and Data Understanding - Exploratory Data Analysis – Data Visualization – Supervised Learning – Regression – Classification – Unsupervised Learning – Case studies – Text Mining