Curriculum and Syllabus for M. Tech in Data Science and Engineering
Department of Computer Science and Engineering
National Institute of Technology Silchar Assam India

Semester	Type of Course	Code	Course Name
DSE 1st Sem	Core	CS 5301	Mathematics for Data Science
	Core	CS 5302	Foundation of Data Science
	Core	CS 5303	Introduction of Machine Learning
	Core	CS 5304	Algorithms for Data Science
	Core	CS 5305	Distributed Database Management Systems
	Core Lab	CS 5306	Python Programming Laboratory
DSE 2st Sem	Core	CS 5307	Big Data Analytics and Visualization
	Core	CS 5308	Artificial Neural Network and Deep Learning
	Core	CS 5309	Introduction to Optimization Techniques
	Core Lab	CS 5310	Deep Learning and Optimization Laboratory
	Elective I	CS 5321	Pattern Recognition for Machine learning
		CS 5322	Speech Processing
		CS 5323	Data Mining
		CS 5324	Image Processing and Computer Vision
		CS 5325	Fundamentals of Information Retrieval
		CS 5326	Computaions System Biology
		CS 5327	Data Analysis for Machine Learning
		CS 5328	Quantum Computing
		CS 5329	Bioinformatics
	Elective II	CS 5341	Optimization Methods
		CS 5342	Machine Learning Using Cloud Computing
		CS 5343	Topics and Tools in Social Media Data Mining
		CS 5344	Multimedia systems
		CS 5345	Advanced Topics in Artificial Intelligence
		CS 5346	High Performance Computing and Data Science
		CS 5347	Health Informatics
		CS 5348	Natural Image Processing
		CS 5349	Data Communications and Networks
DSE 3rd Sem	Core Lab	CS 6301	Data Science Laboratory
	Project	CS 6310	Project
DSE 4th Sem	Project	CS 6311	Project

CS5301	**Mathematics for Data Science**	**L**	**T**	**P**	**C**
M.Tech. (DSE), First Semester (Core)		3	0	0	3

Course objective: The course will introduce the fundamental concepts of probability, statistics, and linear algebra, required for a program in data science.

Course outcome (CO): At the end of the course students are expected to,

1. be able to use the basic concepts of probability in the field of data science
2. be able to use the basic concept of descriptive statistics in the field of data science to design an intelligent system
3. be able to use the basic concept of matrix algebra to design an intelligent system

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 Kreysizig, 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. Spiegal ,Schaum Series, McGraw Hills.
5. M. P. Deisenroth, A. A. Faisal, C. S. Ong, Mathematics for Machine Learning, Cambridge University Press (1st edition)

CS5243 Foundation of Data Science **Credit (3-0-0-3)**

UNIT I: 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 II: 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 III: 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

Course Outcomes (COs):

1. To introduce basic concepts of Data Science with its working principle.
2. To understand how Machine Learning can be used to build DS model with different kinds of data preparation techniques such as feature engineering, exploratory data analysis, data visualization etc.
3. To introduce basic concepts of text mining.
4. To design DS model and recommender system using machine learning to solve different Problems.

Text Books and 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.

Relevance of the subject:

The subject “Foundation of Data Science” is fundamental for students aspiring to build a career in data-driven technologies. It provides a comprehensive introduction to the core principles, methodologies, and tools used in data science. By covering topics such as data types, data wrangling, basic statistics, and introductory machine learning, this course lays the groundwork for understanding how to extract meaningful insights from raw data. Students learn how to approach real-world problems with a data-centric mindset, enabling them to formulate questions, gather and clean data, perform exploratory analysis, and make data-informed decisions. The course fosters analytical thinking and equips students with essential skills that are applicable across diverse domains such as business, healthcare, finance, and technology, making it a highly relevant and valuable component of modern education.

M.Tech (Data Science and Engineering) **1st Semester**

CS- 5303 **Introduction of Machine Learning** **3-0-0**

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)

National Institute of Technology Silchar
Department of Computer Science & Engineering

Syllabus:- CS 5305 Distributed Database Management systems

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.

CS 5306	**Python Programming Laboratory**	**L T P C**
	M. Tech (DSE) First Year	0 0 2 2

Prerequisite: Basic knowledge of machine learning and Python programming.

Course Objectives:
- To develop proficiency in using Python and its libraries for data analysis, visualization, and machine learning.
- To provide hands-on experience in building and training deep learning models using frameworks like Keras and PyTorch.
- To explore the application of Python in real-world data science problems, including image classification and language modeling.

Course Outcomes (COs):
- 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.

Syllabus Outline:

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

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

Data Visualization Techniques : Visualizing data using Matplotlib, Creating advanced plots with Seaborn,
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

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

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

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:

- **"Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython"** by William McKinney
- **“Python Data Science Handbook”** by Jake VanderPlas

CS5307 Big Data Analytics and Visualization
Elective for M Tech Data Science, 2nd Semester
3-0-0

Course Objectives (CO)

1. **[CO1]** Appreciate current trends and to have an overview of big data analytics from an application, architectural and research point of view.
2. **[CO2]** To have an overview of NoSQL from an application, architectural and research point of view.
3. **[CO3]** To have an overview of Hadoop and its ecosystem from an application, architectural and research point of view.
4. **[CO4]** To have an overview of data preprocessing and modelling.
5. **[CO5]** To learn and practice various data visualization techniques with special emphasis to big data.

Syllabus

1. Introduction [Contributes to CO1]

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.

2. Data Storage for Big Data: NoSQL [Contributes to CO1 and CO2]

A very brief overview of triggers, views, schedules and joins with reference to their implementation details. ACID Vs BASE properties of transections 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.

3. Hadoop [Contributes to CO1 and CO3]

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. MapRedude 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, Zookeepeer.

4. Data Preprocessing [Contributes to CO4]

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.

5. **Theory and Methods of Data Modelling [Contributes to CO4]**

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

6. **Data Visualization [Contributes to CO5]**

Introduction to visualization tools: Visual tools – Orange, KNIME, Geogebra, Tableou, 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 Tableou): 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 modules 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 modules 3 & 5:** *Big Data Analytics*. Authors: *G Sudha Sadasivam, R Thirumahal*. Publisher: *Oxford University Press*. ISBN-10: 0-19-949722-2, ISBN-13: 978-0-19-949722-5.
3. **For modules 4 & 5:** *Introduction to Data Mining*. Authors: *Pang-Ning Tan, Micheal Steinbach, Vipin Kumar*. Publisher: Pearson. ISBN: 978-93-3257-140-2.
4. **For module 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 Tableou*: Online materials should be sufficient.

Artificial Neural Network and Deep Learning (CS 5308)

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, adadelta, rmsprop, adam, NAG), second order methods for training, Saddle point problem in neural networks, Regularization methods (dropout, drop connect, batch normalization).

Text Books

Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016..

Bishop, C., M., Pattern Recognition and Machine Learning, Springer, 2006.

Reference Books

Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.

Golub, G., H., and Van Loan, C., F., Matrix Computations, JHU Press,2013.

Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.

Name of the module: Introduction to Optimization Techniques

Code: CS 5309

Credit: 3 0 0 [L T P]

Course: Core, M.Tech. (DSE), Second Semester

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. Nocedel, J. and Wright, S. Numerical Optimization, Springer (2006).

Course Outcomes (COs):

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

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

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

Deep Learning and Optimization Laboratory (CS 5310)

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

CS5321- Pattern Recognition for Machine Learning L – T – P – C: 3 – 0 – 0 - 0

Introduction: Problem framing, feature selection, dimensionality reduction using PCA and other methods; Linear discriminant functions: Gradient descent procedures, Perceptron, Support vector machines - a brief introduction; Artificial neural networks: Multilayer perceptron – feed forward neural network. 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; Decision trees: Classification and Regression Trees (CART); Unsupervised learning: Clustering, Vector Quantization, Kohonen Map, EM Algorithm; Generative models: Definition and characteristics, probabilistic graphical models, density estimation in learning; Combining classifiers: Advantages, boosting, hierarchical classifiers, and issues; 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

References:

1.

1. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2006.
2. P.A Devijver and J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall International, Englewood Cliffs, NJ, 1980
3. I. H. Witten, *Data Mining: Practical Machine Learning Tools And Techniques*, 2nd Edn., Elsevier India, 2008.

CS 5322 **Speech Processing**

Course content:

• Overview of Speech Technology; What is Speech Technology? Why is it important? Its applications and issues.

• Speech Production; Mechanism of speech production; Categories of sounds; Sound units in indian languages.

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

• Basics of Digital Signal Processing; Signals and systems; Discrete fourier transform; Digital filtering; Stochastic processes.

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

• Speech Recognition; Isolated word recognition; Connected word recognition Continuous Speech Recognition; Speech recognition problem; Hidden markov models.

• Other Applications: Word spotting; Speaker recognition; Speech enhancement; Speech synthesis; Practical issues in speech technology.

Text Book:

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.

References:

1. Oppenheim and Schafer, "Discrete-Time Signal Processing," PHI, 2001
2. T W Parsons, "Voice and Speech Processing," McGraw Hill, 1986.
3. Thomas Quatieri, "Discrete-time Speech Processing: Principles and Practice," PH, 2001.
4. Rabiner and Schafer, "Digital Processing of Speech Signals," Pearson Education, 1993.
5. Douglas O' Shaughnessy, "Speech Communications," University Press, 2001.

CS 5323 | **Data Mining** | **L-T-P-C**

2nd Semester M.Tech (DSE) **3-0-0-3**

Introduction to Data Mining:

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

Data Preprocessing:

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

Data Warehouse Fundamentals:

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

Association Rule Mining:

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

Classification:

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

Clustering:

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

Prediction:

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

Text Book

1. Jiawei Han, Micheline Kamber and Jian Pei“Data Mining Concepts and Techniques”, Third Edition, Elsevier, 2011.

Reference Books

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

Course Objectives (CO):

At the end of the module the students will be able to

1. Understand the role of data mining in knowledge discovery process.
2. Familiarize with various data mining functionalities and how it can be applied to various real-world problems.
3. Learn about finding data characteristics and evaluating the outcome of data mining process.
4. Familiarize with various machine learning algorithms used in data mining.

M.Tech (Data Science and Engineering) 2nd Semester

CS- 5324 Image Processing and Computer Vision (Elective-I) 3-0-0

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, 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.

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).

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

Commented [H1]:

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,VaclavHlavac, Roger Boyle, "Image Processing, Analysis, and Machine Vision" Thomson Learning.

CS5325- Fundamentals of Information Retrieval L – T – P – C: 3 – 0 – 0 - 0

Introduction: concepts and terminology of information retrieval systems, Information Retrieval Vs Information Extraction; Indexing: inverted files, encoding, Zipf's Law, compression, boolean queries; Fundamental IR models: Boolean, Vector Space, probabilistic, TFIDF, Okapi, language modeling, latent semantic indexing, query processing and refinement techniques; Performance Evaluation: precision, recall, F-measure; Classification: Rocchio, Naive Bayes, k-nearest neighbors, support vector machine; Clustering: partitioning methods, k-means clustering, hierarchical; Introduction to advanced topics: search, relevance feedback, ranking, query expansion.

Texts:

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.

References:

1. Soumen Chakrabarti, Mining the Web, Morgan-Kaufmann Publishers, 2002.
2. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer, Corr. 2nd printing edition, 2009.
3. David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 2nd edition, 2004.
4. William B. Frakes, Ricardo Baeza-Yates, Information Retrieval Data Structures and Algorithms, Prentice Hall, 1992.
5. G. Salton, M. J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1986.
6. C. J. Van Rijsbergen, Information Retrieval, Butterworth-Heinemann; 2nd edition, 1979.

CS5326 Computations System Biology **Credit (3-0-0-3)**

Unit I: 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 II: 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 III: Dymanic 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 IV: Omics Data Integration and Analysis

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

Course outcomes (CO):

1. Understand biological networks and their computational representations

2. Model and simulate dynamic behavior of biological systems

3. Analyze and integrate omics data for systems-level insights

4. Apply computational tools to investigate real biological problems

Text books and 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

Relevance of the subject:

Computational Systems Biology is a highly interdisciplinary subject that equips students with the ability to model, simulate, and analyze complex biological systems using computational and mathematical tools. In the era of big data and high-throughput technologies, understanding biological phenomena at the systems level is essential for advancing fields like personalized medicine, drug discovery, synthetic biology, and precision agriculture. This course enables students to bridge the gap between biology and computation, preparing them to tackle real-world challenges such as understanding disease mechanisms, designing gene circuits, and analyzing omics data. By mastering this subject, students gain valuable skills in systems thinking, programming, data analysis, and modeling—making them well-suited for careers in academia, biotech, healthcare, and bioinformatics industries.

CS5327: Data Analysis for Machine Learning **Credit (3-0-0-3)**

UNIT I: 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 II: 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 III: 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 IV: 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.

Course Outcomes (COs):

1. Understand the types of data and apply basic data processing techniques.

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

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

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

Text books:

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.

References

1. G. Strang (2016). Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA.
2. Bendat, J. S. and A. G. Piersol (2010). Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA:
3. Montgomery, D. C. and G. C. Runger (2011). Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA.
4. David G. Luenberger (1969). Optimization by Vector Space Methods, John Wiley & Sons (NY)

Relevance of the subject:

The subject "Data Analysis for Machine Learning" plays a crucial role in equipping students with the foundational and practical skills required in today's data-driven world. As data becomes central to decision-making across industries, the ability to preprocess, analyze, visualize, and draw inferences from data has become indispensable. This course introduces students to essential stages of data handling — from cleaning and transformation to statistical inference — which are critical for developing robust machine learning models. By exploring real-world datasets, students gain hands-on experience in exploratory data analysis, feature engineering, and model evaluation techniques. These competencies not only enhance their understanding of machine learning workflows but also prepare them for careers in data science, artificial intelligence, and related fields where analytical thinking and technical proficiency are highly valued.

Name of the Module: Quantum Computing
Module Category: Elective-I MTech (DSE)
Code: CS 5328
Semester: Second
Credit Value: 3[L=3, T=0, P=0]

Subject matters:

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.

Reading lists:
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, ISBN 9048195799, 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., "Quantium 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.*

Course outcomes (CO):

After successful completion of the course, students will have knowledge of -

1. Introduction of quantum computation, classical bit and qubit, basics of linear algebra, Postulates, notations and representation.
2. Basic notions of quantum computing- qubits, mathematical representation, quantum gates, operators, measurement operators, quantum evolution, quantum circuits, Entanglement etc.
3. Quantum algorithms, parallelism, search algorithm etc.
4. Physical realisation, quantum noise, error correction. quantum cryptography etc.

Bioinformatics subject syllabus (core elective) CS 5329: **Credit (3-0-0-3)**

Unit I: 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 II: 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 III: 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 IV: 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-intio Method. Protein folding, Molecular Dynamics of Protein, Molecular Docking of Protein, Small molecule and Nucleotide, Concepts of Force Field

Course Outcomes (CO):

1. Ability to access, clean, and manipulate large biological datasets (e.g., genomic sequences, protein structures).
2. Apply appropriate statistical methods to analyze and visualize biological data.
3. Implement algorithms for sequence alignment, gene expression, and protein structure prediction.

Text books:
1. Bioinformatics- a Practical Guide to the Analysis of Genes and Proteins by Baxevanis, A.D. and Francis Ouellellette, 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.

Reference Books:

1. Bioinformatics: Sequence and Genome Analysis by Mount D., Cold Spring Harbor Laboratory Press, New York. 2004
2. Introduction to bioinformatics by Teresa K. Attwood, David J. Parry-Smith. Pearson Education. 1999 Old editions
3. Bioinformatics in the Post-Genomic Era by Jeffrey Augen, Addison-Wesley Publisher, 2004.

Relevance of the course for students:

Bioinformatics combines biology, computer science, and data analysis, making it a natural fit for data science students. It provides exposure to handling complex biological data, which enhances their ability to work with diverse datasets. This subject involves working with algorithms, databases, and tools for sequence analysis, structural biology, and systems biology. These skills are transferable to other domains in data science, such as natural language processing, image analysis, and predictive modeling. Bioinformatics is at the forefront of personalized medicine, drug discovery, and genomics. Data science students can leverage their expertise to contribute to cutting-edge research and innovation in these areas.

CS 5341 **Optimization Methods** **L-T-P-C**

2nd Semester M.Tech (DSE) **3-0-0-3**

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.

2. Linear Programming:

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

3. Nonlinear Programming:

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

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.

5. Dynamic Programming:

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

6. Constrained Optimization:

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

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.

Course Objectives (CO):

At the end of the module the students will be able to

1. Understand the need and origin of the optimization methods.
2. Get a broader picture of the various applications of optimization methods used in engineering.
3. Define an optimization problem and its various components.
4. Formulate optimization problems as mathematical programming problems.
5. Classify optimization problems to suitably choose the method needed to solve the particular type of problem.
6. Briefly learn about classical and advanced techniques in optimizations.

CS 5343 Topics and Tools in Social Media Data Mining

Course contents:

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.

Link mining: Topics on social media link analysis including centrality, community, link prediction, influence analysis and any other relevant topics.

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

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

CS5344- Multimedia Systems L – T – P – C: 3 – 0 – 0 - 0

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

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

References:

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

CS 5345	**Advanced Topics in Artificial Intelligence**	**L T P C**
	M. Tech (DSE) First Year (Elective)	3 0 0 3

Prerequisite: Basic knowledge of Machine Learning, AI concepts, and programming in Python

Course Objectives:

- To introduce learners to the cutting-edge developments and research trends in Artificial Intelligence.
- To explore advanced techniques in deep learning, reinforcement learning, probabilistic reasoning, and AI planning.
- To analyze and apply AI in complex domains such as multi-agent systems, explainable AI, and ethical AI.
- To provide a foundation for pursuing research or advanced projects in AI.

Course Outcomes (COs):

- 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.

Syllabus Outline:

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:

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

CS 5346 High Performance Computing and Data Science 3-0-0-3

M.Tech. (DSE), Second Semester (Elective-II)

Prerequisites, if any:

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

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

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

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

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 5349	Data Communications and Networks	L	T	P	C
M.Tech Data Science, 2nd Semester (Elective II)		3	0	0	3

Course Outcome

1. To understand fundamental and advanced concepts of data communication and network protocols.
2. 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).
3. To identify and understand security threats, cryptographic techniques and authentication methods.
4. To implement IPsec and learn emerging technologies like Hadoop, Spark for scalable data processing.

Content	CO's
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).	CO1
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.	CO1, CO2
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.	CO1, CO2, CO4
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.	CO1, CO2, CO4
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),	CO1, CO2, CO3, CO4

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.

M.Tech (Data Science and Engineering) **2^{nd} Semester**

CS- 5348 **Natural Image Processing (Elective-II)** **3-0-0**

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.

REFERENCES:

1. Fundamentals of Digital Image Processing: A. K. Jain , PHI.
2. Digital Image Processing using MAT LAB: Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins: Pearson Education India, 2004.
3. Digital Image Processing: William K. Pratt, John Wilely, 3rd Edition, 2004.

Course Outcome:

After completion of this subject students will be able to:

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.

CS6301	CS6301 – Data Science Laboratory	L	T	P	C
	M.Tech (DSE), Third Semester (Core)	0	0	3	2

Prerequisites, if any:

CS 5302 - Data Science

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