AMRITA VISHWA VIDYAPEETHAM

B.Sc Data Science for AHEAD

(2021 Admission onwards)

LEARNING OUTCOMES

The design of the program was developed through discussions between the computer science and mathematics faculty. Each student's plan of study should address a set of learning outcomes developed from these discussions. The learning outcomes answer the question: *"What should a graduate of our data science program be able to do?"*

- Build mathematical / statistical models and understand their power and limitations
- Design an experiment
- Use machine learning and optimization to make decisions
- Acquire, clean, and manage data
- Visualize data for exploration, analysis, and communication
- Collaborate within teams
- Deliver reproducible data analysis
- Manage and analyze massive data sets
- Assemble computational pipelines to support data science from widely available tools
- Conduct data science activities aware of and according to policy, privacy, security and ethical considerations
- Apply problem-solving strategies to open-ended questions

Syllabus

21CSC101A FOUNDATIONS OF DATA SCIENCE 2-0-2-3

Unit-1

Introduction, Causality and Experiments, Data Preprocessing: Data cleaning, Data reduction, Data transformation, Data discretization. Visualization and Graphing: Visualizing Categorical Distributions, Visualizing Numerical Distributions, Overlaid Graphs, plots, and summary statistics of exploratory data analysis, Randomness, Probability, Introduction to Statistics, Sampling, Sample Means and Sample Sizes.

Unit-2

Descriptive statistics – Central tendency, dispersion, variance, covariance, kurtosis, five point summary, Distributions, Bayes Theorem, Error Probabilities; Permutation Testing, Statistical Inference; Hypothesis Testing, Assessing Models, Decisions and Uncertainty, Comparing Samples, A/B Testing, P-Values, Causality.

Unit-3

Estimation, Prediction, Confidence Intervals, Inference for Regression, Classification , Graphical Models, Updating Predictions.

TEXT BOOKS

1. Adi Adhikari and John DeNero, "Computational and Inferential Thinking: The Foundations of Data Science", e-book.

REFERENCES:

1. Data Mining for Business Analytics: Concepts, Techniques and Applications in R, by Galit Shmueli, Peter C. Bruce, Inbal Yahav, Nitin R. Patel, Kenneth C. Lichtendahl Jr., Wiley India, 2018.
2. Rachel Schutt & Cathy O'Neil, "Doing Data Science" O' Reilly, First Edition, 2013.

21CSC102A Problem Solving and Computer Programming with Python 3 0 0 3

Unit 1

Conceptual introduction: topics in computer science, algorithms; modern computer systems: hardware architecture, data representation in computers, software and operating system; Installing Python; basic syntax, interactive shell, editing, saving, and running a script. The concept of data types; variables, assignments; immutable variables; numerical types; arithmetic operators and expressions; comments in the program; understanding error messages;

Unit 2

Conditions, boolean logic, logical operators; ranges; Control statements: if-else, loops (for, while); short-circuit (lazy) evaluation. Strings and text files; manipulating files and directories, os and sys modules; text files: reading/writing text and numbers from/to a file; creating and reading a formatted file (csv or tab-separated). String manipulations: subscript operator, indexing, slicing a string; strings and number system: converting strings to numbers and vice

versa. Binary, octal, hexadecimal numbers

Unit 3

Lists, tuples, and dictionaries; basic list operators, replacing, inserting, removing an element; searching and sorting lists; dictionary literals, adding and removing keys, accessing and replacing values; traversing dictionaries. Design with functions: hiding redundancy, complexity; arguments and return values; formal vs actual arguments, named arguments. Recursive functions. Testing, Debugging, Exceptions, Assertions. Classes and OOP: classes, objects, attributes and methods; defining classes; design with classes, data modeling; persistent storage of objects

TextBook

1. Guttag, John. *Introduction to Computation and Programming Using Python: With Application to Understanding Data Second Edition.* MIT Press, 2016. ISBN: 9780262529624.

21CSC181A Problem Solving Computer Programming Lab 0 0 2 1

1. Installing Python environments
2. Using Python Interpreter to do basic operations like arithmetic computations.
3. Working with variables of different datatypes and using them in expressions.
4. Building stand alone Python scripts
5. Implementing logic requiring conditional expressions and looping
6. Working with strings using inbuilt functionalities of the datatype
7. Working with Python inbuilt datatypes like Lists, Tuples and Dictionaries
8. Working with modularity : Implementing functions and designing logic in a modular fashion
9. Implement unit testing measures assertions and exception handling
10. Use Python to model object oriented programming principles using various use cases.

TextBook

1. Guttag, John. *Introduction to Computation and Programming Using Python: With Application to Understanding Data Second Edition.* MIT Press, 2016. ISBN: 9780262529624.

21CSC111A Advanced Computer Programming with Python 3 0 0 3

Unit 1

Working with packages: How to install/import and use an external Python package. Popular Python packages for applied data science: Exercises to understand usage of libraries like Numpy, SciPy, Pandas in interpreted and script modes.

Unit2

Applied Plotting, Charting & Data Representation in Python: Fundamentals of data reading, streams etc and using Pandas, Basic Charting using Matplotlib, Advanced plots, interactive plots and animated plots, Plotting with Pandas, Seaborn.

Unit 3

Python packages for accessing the Web Data: Regex, urlib, BeautifulSoup, Json, Retrieving and parsing webpages (Json, XML), REST API, Facebook and Twitter API. Connecting DB with Python: Reading and Writing, possible simple SQL queries.

TextBook

1. William McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython*, Second edition (27 October 2017), Shroff/O'Reilly, ISBN-10: 9789352136414, ISBN-13: 978-9352136414

21CSC112A Data Structures 3 1 0 4

Unit 1

Abstraction - Abstract data types; Data Representation; Elementary data types; Basic concepts of data Structures; Mathematical preliminaries - big-Oh notation; efficiency of algorithms; notion of time and space complexity; performance measures for data structures. ADT array - Computations on arrays - sorting and searching algorithms.

Unit 2

ADT Stack, Queue, list - array, linked list, cursor based implementations of linear structures. ADT Tree - tree representation, properties traversal of trees; ADT- Binary Trees – properties and algorithms, ADT Priority Queue - Heaps; heap-based implementations; applications of heaps - sorting; Search Tree - Binary search tree; balanced binary search trees - AVL tree; Applications of Search Trees - TRIE; 2-3-4 tree; concept of B-Tree. ADT Dictionary - array based and tree based implementations; hashing - definition and application .

Unit 3

Graphs: ADT- Data structure for graphs - Graph traversal- Transitive Closure- Directed Acyclic graphs - Weighted graphs – Shortest Paths - Minimum spanning tree – Greedy Methods for MST.

TEXTBOOKS:

1. Goodrich M T, Tamassia R and Michael H. Goldwasser, “Data Structures and Algorithms in Python++”, Wiley publication, 2013.

REFERENCES:

1. Goodrich M T and Tamassia R, “Data Structures and Algorithms in Java”, Fifth edition, Wiley publication, 2010.

2. Tremblay J P and Sorenson P G, “An Introduction to Data Structures with Applications”, Second Edition, Tata McGraw-Hill, 2002.
3. Clifford A. Shaffer, “Data Structures and Algorithm Analysis”, Third Edition, Dover Publications, 2012.

21CSC183A Advanced Computer Programming Lab 0 0 2 1

1. Installing external packages and using them in Python scripts
2. Work with NumPy, SciPy on solving simple mathematical problems
3. Implementing functionalities in Pandas to work with tabular data and do simple database operations on them
4. Implement various plotting and charting methods using packages like Matplotlib and its abstractions like Seaborn
5. Develop Python scripts that can retrieve data from the Web and do operations like parsing, searching, and formatting using packages like BeautifulSoup, urllib, Regex
6. Implement direct database access/manipulations by using Python scripts.

TextBook

1. William McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython*, Second edition (27 October 2017), Shroff/O'Reilly, ISBN-10: 9789352136414, ISBN-13: 978-9352136414

21CSC184A Data Structures Lab 0 0 2 1

Implementing Sample ADT, Templates - Stacks and Queues: Array implementation, Applications - Vector, Lists, using these STLs for other implementations -Linked list: Singly and Doubly Linked Lists Implementation, Linked Stacks, D-Queue, Circular Queue - Implementing STL: Sequences, Iterators - Trees: Binary search tree, Priority Queue, Heaps - Graphs: Graph Representations, Traversals (BFS, DFS) - Hashing: Hash Table creation, creating hash functions, dynamically resizing hash tables.

21CSC201A Database Management Systems 3 1 0 4

Unit I

Introduction to DBMS:Database System Vs File system, Database systems applications, Purpose of database systems - Data models. Relational models: Structure of relational databases – database schema keys – schema diagrams. Relational Query Languages – fundamental relational algebra operations – additional relational algebra operations. Introduction to SQL – Background – SQL data definition –structure of SQL queries – set operations – null values - aggregate functions – modifications to the database.

Unit II

Database design - overview of the design process – the entity-relationship model – constraints – entity-relationship diagrams – reduction to relation schemas - Entity-relationship design issues – weak entity sets – extended E-R features. Intermediate SQL: Nested subqueries - Join expression – Views – Transactions – integrity constraints – authorization. Advanced SQL – Accessing SQL from a program – functions and procedures – triggers.

Unit III

Relational database design – features of good relational designs – atomic domains and normal forms - 1NF, 2NF, 3NF, 4NF and BCNF – decomposition using functional dependencies - functional dependency theory – algorithm for decomposition -decomposition using multi-values dependencies – PJNF and DKNF. Over view of Transaction Management and Concurrency control

Text Book:

1) Silberschatz. A., Korth, H. F. and Sudharshan, S. "Database System Concepts", 6th Edition, TMH, 2010

Reference Books

1) Elmasri, R. and Navathe, S. B. "Fundamentals of Database Systems", 5th Edition, Addison Wesley, 2006

2) Date, C. J. "An Introduction to Database Systems", 8th Edition, Addison Wesley, 2003.

3) Ramakrishnan, R. and Gehrke, J. "Database Management Systems", 3rd Edition, McGrawHill, 2003.

21CSC202A Design and Analysis of Algorithms 3 1 0 4

Unit 1

Introduction: Problem solving -- adding 2 n-bit numbers, multiplication as repeated addition. Running time analysis -- recall of asymptotic notation, big-oh, theta, big-omega, and introduce little-oh and little-omega. Worst case and average case

Basic design paradigms with illustrative examples -- incremental design (e.g., incremental sorting, interpolating polynomials), decremental design (e.g., GCD with discussion on input size, factorial), and pruning (e.g., order statistics). Divide and Conquer: Integer multiplication revisited with an efficient algorithm that motivates and leads into recurrences. Solving recurrences using recurrence trees, repeated substitution, statement of master theorem. Brief recall of merge sort and its recurrence. Median in worst case linear time.

Unit 2

Greedy Algorithms: Greedy choice, optimal substructure property, minimum spanning trees -- Prims and Kruskals, Dijkstras shortest path using arrays and heaps, fractional knapsack, and Huffman coding (use of priority queue). Dynamic Programming: Integral knapsack (contrasted with the fractional variant), longest increasing subsequence, edit distance, matrix chain multiplication, and independent sets in trees.

Unit 3

Graph Algorithms – Graph Traversal: Applications of BFS: distance, connectivity and connected components and cycles in undirected graphs. Applications of DFS: Topological sort, cycles in

directed graphs, Biconnected Components and Strong Connectivity. Path algorithms: Shortest path algorithms (along with analysis) SSSP: Bellman Ford. APSP: Floyd Warshall's. Minimum Spanning Tree (with analysis and applications).

String Matching: Boyer Moore – KMP – Rabin Karp. NP-completeness: reduction amongst problems, classes NP, P, NP-complete, and polynomial time reductions.

Textbooks

1. Introduction to Algorithms, by Cormen, Leiserson, Rivest, and Stein, MIT Press, Third Edition, 2009.

References

1. Algorithms, by Dasgupta, Papadimitrou and Vazirani, McGraw-Hill Education, 2006.
2. Algorithm Design, by Kleinberg and Tardos, Pearson, 2005.
3. Algorithm Design, by Goodrich and Tamassia, Wiley, 2001.

21CSC281A Design and Analysis of Algorithms Lab 0 0 2 1

Implementation of common sorting algorithms – insertion sort, selection sort, quick sort, merge sort, bucket sort, radix sort. Greedy – task scheduling, fractional knapsack and other applications. Divide and Conquer – Closest Pair, Integer multiplication, other applications. Dynamic Programming – matrix chain multiplication, 0-1 knapsack, longest common subsequence, maximum contiguous subarray, edit distance. Graphs- minimum spanning tree algorithms, shortest path algorithms. String matching – KMP, Boyer Moore.

21CSC211A Database Design 3-1-0-4

Unit 1

Overview of DBMS – Database design – Record Oriented File Systems – File Structures, Indexing and Hashing – Disk Storage, Basic File Structures and Hashing – Indexing Structures – Single and Multi-level indexes. Query Processing Optimization and Database Tuning: - Algorithms for Query Processing and Optimization- Physical Database Design and Tuning

Unit 2

Transactions Processing and Concurrency Control : Transaction Concept, Transaction model, Storage Structure, Transaction Atomicity and Durability, Transaction Isolation, Serializability

Concurrency control : Lock-based protocols – Timestamp Ordering based control – Multiversion concurrency control – Locks, Database Recovery Techniques

Unit 3

Advanced Topics: Object Oriented, Object Relational Databases, XML Databases – Concepts, Models and Standards. Parallel and Distributed Databases, NoSQL Databases, Database Security

–Introduction, Attacks and Techniques for Mitigation, Spatio-temporal and Multimedia Databases

TEXT BOOKS

1. Ramesh Elmasri and Shamkant B Navathe, “Fundamentals of Database Systems”, Fifth Edition, Pearson Educaton India, 2008.

REFERENCES

1. Silberschatz A, Korth H F and Sudharshan S, “Database System Concepts”, Sixth Edition, Tata McGraw-Hill Publishing Company Limited, 2010.
2. Niall O’Higgins, “MongoDB and Python”, O’reilly, 2011.
3. Hector Garcia-Molina, Jeff Ullman and Jennifer Widom, “Database Systems: The Complete Book”, Pearson, 2011.
4. Raghu Ramakrishnan and Johannas Gehrke, “Database Management Systems”, Third Edition, McGraw-Hill, 2003.

21CSC212A Machine Learning 3 0 2 4

Supervised Learning (Regression/Classification) : Basic methods: Distance-based methods, Nearest-Neighbors, Decision Trees, Naïve Bayes. Linear models: Linear Regression, Logistic Regression, Generalized Linear Models. Support Vector Machines, Nonlinearity and Kernel Methods. Beyond Binary Classification: Multi-class/Structured Outputs, Ranking

Unsupervised Learning: Clustering: K-means/Kernel K-means. Dimensionality Reduction: PCA and kernel PCA. Matrix Factorization and Matrix Completion. Generative Models (mixture models and latent factor models)

Assorted Topics: Evaluating Machine Learning algorithms and Model Selection. Introduction to Statistical Learning Theory. Ensemble Methods (Boosting, Bagging, Random Forests). Sparse Modeling and Estimation. Modeling Sequence/Time-Series Data. Deep Learning and Feature Representation Learning. Scalable Machine Learning (Online and Distributed Learning). A selection from some other advanced topics, e.g., Semi-supervised Learning, Active Learning, Reinforcement Learning, Inference in Graphical Models, Introduction to Bayesian Learning and Inference.

Text books/ Reference books.

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer 2009 (freely available online)
3. Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007.
4. Hal Daumé III, A Course in Machine Learning, 2015 (freely available online).

21CSC213A Number Theory and Information Security 3 1 0 4

Algorithms for integer arithmetic: Divisibility, GCD, modular arithmetic, modular exponentiation, Montgomery arithmetic, congruence, Chinese remainder theorem, orders and primitive roots, quadratic residues, integer and modular square roots, prime number theorem, continued fractions and rational approximations.

Representation of finite fields: Prime and extension fields, representation of extension fields, polynomial basis, primitive elements, normal basis, optimal normal basis, irreducible polynomials, Root-finding and factorization algorithm, Lenstra-Lenstra-Lovasz algorithm.

Elliptic curves: The elliptic curve group, elliptic curves over finite fields, Schoof's point counting algorithm.

Primality testing algorithms: Fermat Basic Tests , Miller–Rabin Test , AKS Test.

Integer factoring algorithms: Trial division, Pollard rho method, p-1 method, CFRAC method, quadratic sieve method, elliptic curve method.

Computing discrete logarithms over finite fields: Baby-step-giant-step method, Pollard rho method, Pohlig-Hellman method, index calculus methods, linear sieve method, Coppersmith's algorithm.

Quantum Computational Number Theory : Grover's algorithm, Shor's algorithm

Applications in Algebraic coding theory and cryptography.

TEXT BOOKS/REFERENCES:

1. Yan, Song Y. *Computational Number Theory and Modern Cryptography*. John Wiley & Sons, 2012.
2. Meijer, Alko R. *Algebra for Cryptologists*. Springer, 2016
3. Lidl, Rudolf, and Harald Niederreiter. *Introduction to finite fields and their applications*. Cambridge university press, 1994.
4. Apostol, Tom M. *Introduction to analytic number theory*. Springer Science & Business Media, 2013.

21CSC214A OPERATING SYSTEMS 3 0 2 4

Unit 1

Introduction to Operating Systems: Overview - Types of systems - Computer system operations - Hardware Protection - Operating systems services - System calls - System structure - Virtual machines. Process Management: Process concepts- Process scheduling - Operations on Process - Cooperating process - Interprocess communication - Multithreading models - Threading issues - Thread types - CPU scheduling –scheduling algorithms.

Unit 2

Process Synchronization: Critical section problem - synchronization hardware – Semaphores - Classical problems of synchronization - Critical regions – Monitors- Deadlocks - Deadlock

characterization - Methods of handling deadlocks - Deadlock prevention – Avoidance - Detection and recovery.

Unit 3

Storage Management: Memory management – Swapping - Contiguous memory allocation. Paging – Segmentation - Segmentation with Paging - Virtual memory - Demand paging - Process creation – page replacement - Thrashing. File Systems: Directory structure - Directory implementation - Disk scheduling. Case study: Threading concepts in Operating systems, Kernel structures.

TEXTBOOK:

1. Silberschatz and Galvin, “Operating System Concepts”, Ninth Edition, John Wiley and Sons, 2012.

REFERENCES:

1. Deitel. Deitel and Choffnes, “Operating System”, Third edition, Prentice Hall, 2003.
2. Tannenbaum A S, “Modern Operating Systems”, Third edition, Prentice Hall, 2007.
3. Stevens W R and Rago S A, “Advanced Programming in the Unix Environment”, Second Edition, Addison-Wesley, 2013.
4. Gary Nutt, “Operating Systems”, Third Edition, Pearson Education, 2009.

21CSC282A Database Management Systems Lab 0 0 2 1

1) Working with objects using SQL for the following
   i. Data definition language: create, alter, grant, revoke, drop, truncate.
   ii. Data manipulation language: select, insert, update, delete.
   iii. Transaction control statements: commit, rollback, savepoint.

2) Constraints – Queries: Simple selection, projection and selection with conditions.
3) Functions: aggregate functions, group by, order by, date and conversion functions.

4) Set operators, joins, sub query: simple, nested, correlated, existence test, membership test, DDL and sub queries and DML and sub queries.

5) Working with other schema objects: view, sequence, index, synonym, cluster, lock, BLOB, CLOB, nested table, type.

6) PL/SQL programs, cursors, functions, procedures, packages, triggers, exception handling.

7) Front end tool: form creation, validation, trigger and report generation.

8) Mini Project.

21CSC301A DATA VISUALIZATION 3 0 0 3

Unit 1

Introduction to Data Visualization – Classification of Visualization techniques – Structure and representation – Selection of a Visualization – Visualizations for high dimensional data – Graphics and computing.

Unit 2

Principles of Data Visualization : Multivariate data – Linked data – Visualizing trees and forests – Large Datasets – Plots and their variates – Visualizing cluster analysis – contingency tables – finite mixture models.

Unit 3

Methodologies: Visualization in Bayesian data analysis – Matrix visualization – Data visualization by kernel machines .Applications : Visualization for genetic network reconstruction, medical images, financial dataset and Insurance risk processes.

TEXTBOOK

1. Usama Fayyad, Georges G. Grinstein and Andreas Wierse, "Information visualization in Data Mining and Knowledge discovery", Morgan kaufmann publishers, 2002

2. Chun-houh Chen, Wolfgang Hardle and Antony Unwin,"Handbook of Data Visualization", Springer, 2008

21CSC302A Deep Learning 3 0 0 3

- Basics: Biological Neuron, Idea of computational units, McCulloch–Pitts unit and Thresholding logic, Linear Perceptron, Perceptron Learning Algorithm, Linear separability. Convergence theorem for Perceptron Learning Algorithm.
- Feedforward Networks: Multilayer Perceptron, Gradient Descent, Backpropagation, Empirical Risk Minimization, regularization, autoencoders.
- Deep Neural Networks: Difficulty of training deep neural networks, Greedy layerwise training.
- Better Training of Neural Networks: Newer optimization methods for neural networks (Adagrad, adadelta, rmsprop, adam, NAG), second order methods for training, Saddle point problem in neural networks, Regularization methods (dropout, drop connect, batch normalization).
- Recurrent Neural Networks: Back propagation through time, Long Short Term Memory, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs
- Convolutional Neural Networks: LeNet, AlexNet.
- Generative models: Restrictive Boltzmann Machines (RBMs), Introduction to MCMC and Gibbs Sampling, gradient computations in RBMs, Deep Boltzmann Machines.

- Recent trends: Variational Autoencoders, Generative Adversarial Networks, Multi-task Deep Learning, Multi-view Deep Learning
- Applications: Vision, NLP, Speech (just an overview of different applications in 2-3 lectures)

Textbook:

1. Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, MIT Press, 2016.

References:

1. Ian Goodfellow, YoshuaBengio and Aaron Courville, Deep Learning, MIT press 2016
2. Neural Networks: A Systematic Introduction, Raúl Rojas, 1996
3. Pattern Recognition and Machine Learning, Christopher Bishop, 2007

21CSC304A Software Engineering 3 1 0 4

Software process and lifecycle: Software Product, Software Processes, Study of different process models, Project Management Concepts, Planning and Scheduling, Team organization and people management.
Software requirement engineering: Software requirements, extraction and specification, Feasibility Studies, Requirements Modeling, object oriented analysis.
Design Concepts: Object oriented design, Architectural design. Component level Design, User Interface Design, Distributed Systems Architecture, Real Time Software Design, User Interface Design, Pattern Based Design.
Risk Management: Metrics and Measurement, Estimation for software projects, software configuration management, Maintenance and Reengineering.
Software Testing: Unit testing, integration testing, black box and white box testing, regression testing, performance testing, object oriented testing. Verification and validation of Software:Software Inspections and Audit, Automated Analysis, Critical systems validation.
Software Quality Assurance, Quality Standards, Quality Planning and Control, Various Quality models. Overview of recent trends in Software Engineering, Security Engineering, Agile Methods, Service Oriented Software Engineering, Aspect Oriented Software Development.Self-Study:

Text Books: 1.Ian Sommerville, Software Engineering, Addison – Wesley
References:

1.Roger Pressman, Software Engineering A Practitioners Approach, McGraw Hill Publication
2.Rajib Mall, Fundamentals of Software Engineering, Prentice Hall of India
3.Ivar Jacobson, Object Oriented Software Engineering A use case Approach, Pearson

21CSC331A ADVANCED BIG DATA ANALYTICS 3-0-0-3

Unit - I

How MapReduce Works - Anatomy of a MapReduce Job Run, Failures, Shuffle and Sort, Task Execution

Unit -II

MapReduce Types and Formats - MapReduce Types, Input Formats, output formats,

Unit- III

MapReduce Features- Counters, Sorting, Joins, Side Data Distribution

Unit -IV

Simple analytics using MapReduce, Calculating frequency distributions and sorting using MapReduce, Calculating histograms using MapReduce, Calculating scatter plots using MapReduce

Unit – V

Hierarchical clustering, Clustering algorithm to large dataset, classification using Navie bayes classifier, other applications

Text Books/References:

1. Tom White , Hadoop: The Definitive Guide, Fourth Edition , O'Reilly Media ,2009
2. Srinath Perera and Thilina Gunarathne , Hadoop MapReduce Cookbook : Recipes for analyzing large and complex datasets with Hadoop MapReduce, Packt PublishingLtd,2013.

21CSC334A Data Analytics in Computational Biology 3 0 0 3

Introduction to Bioinformatics - applications of Bioinformatics - challenges and opportunities - introduction to NCBI data model- Various file formats for biological sequences.

Bioinformatics resources – Importance of databases - Biological databases- Primary & Secondary databases (Genbank, EMBL, DDBJ, Swiss Prot , PDB, NDB, BLOCKS, Pfam, ProSITE, etc.).

Sequence alignment methods: Sequence analysis of biological data-Significance of sequence alignment- pairwise sequence alignment methods- Use of scoring matrices and gap penalties in sequence alignments- PAM and BLOSUM Scoring Matrices. Introduction to Dynamic Programming, Global alignments: Needleman Wunsch Algorithm, Local Alignments: Smith Waterman Algorithm, Gap Penalties.

Multiple sequence alignment methods – Tools and application of multiple sequence alignment. Sequence alignment tools (BLAST, FASTA, CLUSTAL-W/X, MUSCLE, TCOFFEE), Variants of BLAST (BLASTn, BLASTp, PSIBLAST, PHI-BLA

Phylogenetic analysis algorithms: Maximum Parsimony, UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining, jackknife, Probabilistic models and associated algorithms

such as Probabilistic models of evolution and maximum likelihood algorithm, Bootstrapping methods, use of tools such as PHYLIP, MEGA, PAUP.

References/ Textbooks

1 Higgins,Des and Taylor Williw: Bioinformatics: Sequence , Structure and databanks, Oxford , University Press,2000.
2. Baxenvants, AD., Bioinformatics: A practical guide to the analysis of genes and proteins", Third edition, John wiley & Sons ,2005
3. Teresa Attwood, Introduction To Bioinformatics ,Pearson Education Singapore Pte Ltd, 2007
4. Mount, DW, Bioinformatics: Sequence and Genome analysis", Second edition, Cold Spring Harbor Laboratory Press. Baxevanis 5. A.D., Davison D.B., Page R. D. M. & Petsko G.A. Current Protocols in Bioinformatics. New York, John Wiley & Sons Inc., 2004. ISBN: 0555015254
6. S.C. Rastogi et al, Bioinformatics: Methods and Applications: (Genomics, Proteomics and Drug Discovery) Kindle Edition.

21CSC335A Data Compression 3 0 0 3

Unit 1

Information Theory Foundation: Entropy, its properties, conditional entropy, mutual information, Types of codes, Krafts McMillan Inequality theorem, Source coding theorem.Introduction to Compression Techniques: Introduction, Types of compression - Lossy, lossless. Performance measures, Modeling, Coding. Text Compression: Huffmann - static and dynamic, application in text compression, Shannon Fano Elias Coding, Arithmetic coding, Dictionary based coding- static, adaptive, UNIX compress.

Unit 2

Scalar and Vector Quantization: Scalar Quantization – Introduction, Uniform and Adaptive quantization. Vector Quantization- Introduction, Advantages, LBG, Tree vector quantization, Trellis coded quantization
Audio Compression: Distortion criteria- Auditory perception, PCM, DPCM, ADPCM, Predictive coding- basic algorithm, Basic sub-band coding, MPEG Audio Coding

Unit 3

Image Compression: Distortion criteria- The human visual system, Transform coding- DCT, JPEG, JBIG II, GIF, Wavelet based compression- wavelets, the scaling function, Haar Transforms, JPEG-2000. Video Compression: Motion Estimation and Compensation- Full search and Fast search algorithms, H.261, MPEG-1, MPEG-2, MPEG-4, MPEG -7.

TEXTBOOKS:

1. Sayood and Khalid, "Introduction to Data Compression", Third Edition, Morgan Kaufmann, 2006.

REFERENCES:

1. Richardson I E G, "Video Codec Design: Developing Image and Video Compression Techniques", John Wiley & Sons, 2002.

2. Salomon D, “Data Compression: The Complete Reference”, Fourth Edition, Springer, 2007.
3. Gersho A and Kluwer R M G, “Vector Quantization and Signal Compression”, Academic Press, 1992.

21CSC336A DEEP LEARNING FOR IMAGE PROCESSING 3 0 0 3

Unit1

Mathematical Background for Image Processing: Review of Vectors and Matrices - Review of Probability and statistics. Digital Image Fundamentals: Elements of Visual Perception- Image Sensing and Acquisition – Image Sampling and Quantization – Basic Relationships between Pixels- Image interpolation. Intensity Transformations and Spatial Filtering: Basic Intensity transformation Functions – Histogram Processing – Fundamentals of Spatial Filtering – Smoothing and Sharpening Spatial Filters.

Unit2

Filtering in Frequency Domain: 2D Discrete Fourier Transforms - Basics of filtering - Image Smoothing and Image Sharpening Using Frequency Domain Filters- Selective Filtering, Image Restoration: Noise Models – Restoration using Spatial Filters – Periodic Noise Reduction by Frequency Domain Filters.

Unit 3

Morphological Image Processing: Erosion – Dilation – Opening – Closing – Hit-or-Miss Transform- Extraction of Connected Components. Image Segmentation: Fundamentals – Point, Line and Edge Detection – Thresholding- Region Based Segmentation – Region Growing – Region Splitting and Merging. Color image processing.

TEXTBOOK:

Gonzalez R C and Woods R E, “Digital Image Processing”, Third Edition, Pearson Education, 2009.

REFERENCES:

1. *Pratt W K, “Digital Image Processing”, Fourth Edition, John Wiley & Sons, 2007.*
2. *Castleman K R, “Digital Image Processing”, Prentice Hall, 1996.*
3. *Gonzalez, Woods and Eddins, “Digital Image Processing Using MATLAB”, Prentice Hall, 2004.*
4. *Russ J C, “The Image Processing Handbook”, CRC Press, 2007.*

21CSC338A Mining of Massive Datasets 3 0 0 3

Basics of Data Mining - Computational Approaches - Statistical Limits on Data Mining - Bonferroni’s Principle - MapReduce - Distributed File Systems . MapReduce . Algorithms Using MapReduce . Extensions to MapReduce. Finding Similar Items - Applications of Near-Neighbor Search - Shingling of Documents - Similarity-Preserving Summaries of Sets - Locality-Sensitive Hashing for Documents - Distance Measures

Mining Data Streams: The Stream Data Model - Sampling Data in a Stream - Filtering Streams. Link Analysis: PageRank - Efficient Computation of PageRank - Topic-Sensitive PageRank - Link Spam. Frequent Itemsets : The Market-Basket Model - Market Baskets and the A-Priori Algorithm - Handling Larger Datasets in Main Memory. Clustering: Introduction to Clustering Techniques - Hierarchical Clustering - K-means Algorithms – CURE algorithm.
Recommendation Systems: A Model for Recommendation Systems - Content-Based Recommendations - Collaborative Filtering - Dimensionality Reduction. Mining Social-Network Graphs: Social Networks as Graphs - Clustering of Social-Network Graphs - Direct Discovery of Communities - Partitioning of Graphs - Finding Overlapping Communities – Simrank. Dimensionality Reduction: Eigenvalues and Eigenvectors of Symmetric Matrices- Principal-Component Analysis - Singular-Value Decomposition . Large-Scale Machine Learning - Machine-Learning Model - Perceptrons - Support-Vector Machines .

Text Book

Jure Leskovec , Anand Rajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2014.

References

Tom White, Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale , O'Reilly Media; 4 edition , 2015.

21CSC340A **Social Network Analytics** **3 0 0 3**

Unit 1 : Online Social Networks (OSNs)

Introduction - Types of social networks (e.g., Twitter, Facebook), Measurement and Collection of Social Network Data. Techniques to study different aspects of OSNs -- Follower-followee dynamics, link farming, spam detection, hashtag popularity and prediction, linguistic styles of tweets. Case Study: An Analysis of Demographic and Behaviour Trends using Social Media: Facebook, Twitter and Instagram

Unit 2: Fundamentals of Social Data Analytics

Introduction - Working with Social Media Data, Topic Models, Modelling social interactions on the Web – Agent Based Simulations, Random Walks and variants, Case Study: Social Network Influence on Mode Choice and Carpooling during Special Events: The Case of Purdue Game Day

Unit 3 : Applied Social Data Analytics

Application of Topic models, Information Diffusion, Opinions and Sentiments - Mining, Analysis and Summarization, Case Study: Sentiment Analysis on a set of Movie Reviews using Deep Learning techniques, Recommendation Systems, Language dynamics and influence in online communities, Community identification, link prediction and topical search in social networks, Case Study: The Interplay of Identity and Social Network: A Methodological and Empirical Study

Text and Reference Literature

1. Cioffi-Revilla, Claudio. *Introduction to Computational Social Science*, Springer, 2014.
2. Matthew A. Russell. *Mining the Social Web: Data Mining Facebook, Twitter, Linkedin, Google+, Github, and More*, 2nd Edition, O'Reilly Media, 2013.
3. Robert Hanneman and Mark Riddle. *Introduction to social network methods.* Online Text Book, 2005.
4. Jennifer Golbeck, *Analyzing the social web*, Morgan Kaufmann, 2013.
5. Claudio Castellano, Santo Fortunato, and Vittorio Loreto, *Statistical physics of social dynamics*, Rev. Mod. Phys. 81, 591, 11 May 2009.
6. S. Fortunato and C. Castellano, *Word of mouth and universal voting behaviour in proportional elections*, Phys. Rev. Lett. 99, (2007).
7. Douglas D. Heckathorn, *The Dynamics and Dilemmas of Collective Action*, American Sociological Review (1996).
8. Michael W. Macy and Robert Willer, *From factors to actors: Computational Sociology and Agent-Based Modeling*, Annual Review of Sociology Vol. 28: 143-166 (2002).
9. Nilanjan Dey Samarjeet Borah Rosalina Babo Amira Ashour, *Social Network Analytics - Computational Research Methods and Techniques, First Edition*, eBook ISBN: 9780128156414, Paperback ISBN: 9780128154588, Imprint: Academic Press, Published Date: 23rd November 2018

21CSC341A **Soft Computing** **3 0 0 3**

Unit I

Artificial Intelligence (*AI*): A Brief review – Pitfalls of Traditional *AI*–Why computational intelligence (*CI*) ? – Concepts of *CI* – Importance of tolerance of imprecision and uncertainty– Constituent techniques of *CI*– overview of Artificial Neural Networks, Fuzzy Logic, Evolutionary Computation.

Unit II

Fuzzy Logic:Introduction – the case of imprecision, the utility and limitation of fuzzy systems. Classical sets and Fuzzy sets: operations, properties and mapping.

Unit III

Classical relations and fuzzy relations: cardinality, operations, properties and composition – tolerance and equivalence relations. Properties of membership function, fuzzification and defuzzification. Logic and fuzzysystems. Fuzzy control systems – Aircraft landing control problems.

Unit IV

Evolutionary computation: Introduction – Constituent algorithms - Using Genetic Algorithm for solving simple optimization problems. Swarm intelligence algorithms – Overview of other bio-inspired algorithms – Over view of Hybrid approaches (neural networks, fuzzy logics, genetic algorithm etc).

Text Books:

1) Kumar S. ‘Neural Networks – A classroom approach’, TMH, 20014.
2) Ross T J ‘Fuzzy Logic with Engineering Applications’, TMH, 2002.
3) Eiben A E and Smith J E, ‘ Introduction to Evolutionary Computing’, Second Edition, Springer, Natural Computing Series, 20017.

Reference Books:

1) Konar A, 'Computational Intelligence : Principles, Techniques and Applications", Springer Verlat, 2005.
2) Engelbercht AP, 'Fundamentals of Computational Swarm Intelligence', John Wiley and Sons , 2005.
3) Jang J S R and Sun C T ,Mizutani E, 'Neuron – Fuzzy and Soft Computing' , PHI, 2002.
4) Rajasjekaran S and VijayalakshmiPai G A 'Neural Networks, Fuzzy Logic and Genetic Algorithm', PHI, 2003.

21CSC342A Statistical Pattern Recognition 3-0-2-4

UNIT I

Introduction and Bayesian Decision Theory– Pattern recognition systems – the design cycle – learning and adaptation – Bayesian decision theory – continuous features – Minimum error rate classification – discriminant functions and decision surfaces – the normal density based discriminant functions.

UNIT II

Maximum likelihood estimation – Bayesian estimation - Bayesian parameter estimation – Gaussian case and general theory – problems of dimensionality – components analysis and discriminants – hidden Markov models.

UNIT III

Nonparametric techniques and linear discriminant functions- density estimation – Parzen windows – nearest neighbourhood estimation – rules and metrics – linear discriminant functions and decision surfaces – generalized linear discriminant functions – two-category linearly separable case – minimizing the perception criterion function.

UNIT IV

Nonmetric methods and algorithm-independent machine learning- decision trees – CART methods – algorithm-independent machine learning – lack of inherent superiority of any classifier – bias and variance for regression and classification – resampling or estimating statistics – estimating and comparing classifiers.

UNIT V

Unsupervised learning and clustering – mixture densities and identifiability – maximum likelihood estimates – application to normal mixtures – unsupervised Bayesian learning – data description and clustering – criterion functions for clustering – hierarchical clustering – component analysis – low-dimensional representations and multi-dimensional scaling.

References:

1. *Richard O. Duda, Peter E. Hart and David G. Stork, "Pattern Classification", Second Edition, 2003, John wily & sons.*
2. *Earl Gose, Richard Johnsonbaugh and Steve Jost, "Pattern Recognition and Image Analysis, 2002, Prentice Hall of India.*
3. *Nilsson N J, "The Quest for Artificial Intelligence", Cambridge University Press, 2009.*

21CSC343A **Time series** **3-0-0-3**

Multivariate Data: Random Vector: Probability mass and density functions, Distribution function, Mean vector & Dispersion matrix. Multivariate and Bivariate normal distributions.

Time series as a discrete parameter stochastic process, Auto - covariance, Auto-correlation functions and their properties, moving average models , autoregressive models, Autoregressive Moving Average models.

Text Books:

1. Johnson, R and Wichern(1992): Applied Multivariate Statistical Analysis, Prentice Hall, India, 6thedition.
2. Anderson, T.W. The Statistical Analysis of Time Series, John Wiley, New York, 1971.

References:

1. Anderson, T. W. (1983): An Introduction to Multivariate Statistical Analysis. 3rdEd. Wiley.
2. Box, G.E.P. and Jenkins, G.M. Time Series Analysis- Forecasting and Control, Holden-day, San Francisco,1976.

21MAT107A **Linear Algebra** **3 1 0 4**

Vector Spaces: Vector spaces - Sub spaces - Linear independence - Basis – Dimension.
Inner Product Spaces: Inner products - Orthogonality - Orthogonal basis - Gram Schmidt Process - Change of basis - Orthogonal complements - Projection on subspace - Least Square Principle.

Linear Transformations: Positive definite matrices - Matrix norm and condition number - QR-Decomposition - Linear transformation - Relation between matrices and linear transformations - Kernel and range of a linear transformation - Change of basis - Nilpotent transformations - Trace and Transpose, Determinants, Symmetric and Skew Symmetric Matrices, Adjoint and Hermitian Adjoint of a Matrix, Hermitian, Unitary and Normal Transformations, Self Adjoint and Normal Transformations, Real Quadratic Forms.

Eigen values and Eigen vectors: Problems in Eigen Values and Eigen Vectors, Diagonalization, Orthogonal Diagonalization, Quadratic Forms, Diagonalizing Quadratic Forms, Conic Sections. Similarity of linear transformations - Diagonalisation and its applications - Jordan form and rational canonical form.

TEXT BOOKS

1. Howard Anton and Chris Rorres, "*Elementary Linear Algebra*", Tenth Edition, John Wiley & Sons, 2010.

REFERENCES:

1. Nabil Nassif, Jocelyne Erhel, Bernard Philippe, Introduction to Computational Linear Algebra, CRC press, 2015.
2. Gilbert Strang, "*Linear Algebra and Its Applications*", Fourth Edition, Cengage, 2006.
3. Kenneth Hoffmann and Ray Kunze, *Linear Algebra*, Second Edition, Prentice Hall, 1971.
4. I. N. Herstein, 'Topics in Algebra', Second Edition, John Wiley and Sons, 2000.

21MAT109A CALCULUS 3 1 0 4

Unit 1

Differentiation: The Derivative as a Function – Differentiation Rules – The Derivative as a Rate of Change – Derivatives of Trigonometric Functions – The Chain Rule and Parametric Equations – Implicit Differentiation – Linearization and Differentials.

Chapter 2- Sec: 2.1 to 2.7 and Chapter 3- Sec: 3.1 to 3.6, 3.7, Self Study - Sec: 3.7.

Unit 2

Application of Derivatives: Extreme values of Functions – The Mean Value Theorem – Monotonic Functions and the First Derivative Test – Concavity and Curve Sketching – Intermediate Forms and L' Hospital's Rule – Anti Derivatives.

Chapter 4- Sec: 4.1 to 4.4, 4.6 to 4.8, Self Study - Sec: 4.5

Unit 3

The Definite Integral – The Fundamental Theorem of Calculus – Indefinite Integrals and the Substitution Rule – Substitution and Area between Curves.

Chapter 5- Sec: 5.1 to 5.6

Unit 4

Techniques of Integration: Basic Integration Formulas – Integration by Parts – Integration of Rational Functions by Partial Fractions – Trigonometric Integrals – Trigonometric Substitutions – Numerical Integration – Improper Integrals.

Chapter 8: 8.1 to 8.5, 8.7,8.8, Self Study - Sec: 8.6

Unit 5

Application of Definite Integrals: Volumes by Slicing and Rotation about an Axis – Volumes by Cylindrical Shells – Lengths of Plane Curves – Moments and Centre of Mass – Areas of Surface of Revolution and the Theorems of Pappus – Work – Fluid Pressure and Forces.

Chapter 6 – Sec: 6.1 to 6.7

TEXTBOOK:

1. Finney and Thomas, "Calculus", Pearson, Eleventh Edition, 2008.

REFERENCE BOOKS:

1. Howard Anton, IRl Bivens, Stephens Davis, "Calculus" Wiley, 10th Edition, 2016 Reprint.
2. M. J. Strauss, G. L. Bradley and K. J. Smith, "Calculus", 3rd Edition, Dorling Kindersley (India) Pvt. Ltd. (Pearson Education), 2007.
3. James Stewart, "Calculus: Early Transcendentals", Cengage (India), 8th Edition, 2016.

21MAT118A **Discrete Mathematics** **3 0 2 4**

Phase I

Logic, Mathematical Reasoning and Counting: Logic, Prepositional Equivalence, Predicate and Quantifiers, Theorem Proving, Functions, Mathematical Induction. Recursive Definitions, Recursive Algorithms, Basics of Counting, Pigeonhole Principle, Permutation and Combinations. (Sections: 1.1 -1.3, 1.5 -1.7, 2.3, 4.1 - 4.4, 5.1 - 5.3 and 5.5)

Phase II

Relations and Their Properties: Representing Relations, Closure of Relations, Partial Ordering, Equivalence Relations and partitions. (Sections: 7.1, 7.3 - 7.6)

Advanced Counting Techniques and Relations: Recurrence Relations, Solving Recurrence Relations, Generating Functions, Solutions of Homogeneous Recurrence Relations, Divide and Conquer Relations, Inclusion-Exclusion. (Sections: 6.1 - 6.6)

Phase III

Graph Theory: Introduction to Graphs, Graph Operations, Graph and Matrices, Graph Isomorphism, Connectivity, Euler and Hamilton Paths, Shortest Path Problem, Planar Graph, Graph Colorings and Chromatic Polynomials. (Sections: 8.1 - 8.8)

TEXTBOOKS:

1. Kenneth H. Rosen, “Discrete Mathematics and its Applications”, Tata McGraw- Hill Publishing Company Limited, New Delhi, Sixth Edition, 2007.

REFERENCES:

1. R.P. Grimaldi, “Discrete and Combinatorial Mathematics”, Pearson Education, Fifth Edition, 2007.

2. Thomas Koshy, “Discrete Mathematics with Applications”, Academic Press, 2005.
3. Liu, “Elements of Discrete Mathematics”, Tata McGraw- Hill Publishing Company Limited , 2004.

21MAT119A **Probability and Statistics** **3 1 0 4**

Unit – I

Sample Space and Events, Interpretations and Axioms of Probability, Addition rules, Conditional Probability, Multiplication and Total Probability rules, Independence, Bayes theorem.

Unit – II

Random variables, Probability Distributions and Probability mass functions, Cumulative Distribution functions, mathematical expectation, variance, moments and moment generating function. Standard discrete distributions – Binomial and Poisson Distributions -Standard continuous distributions - Exponential and Normal distributions.

Unit – III

Two dimensional random variables-Joint, marginal and conditional probability distributions for discrete and continuous cases, independence, expectation of two dimensional random variables - conditional mean, conditional variance, covariance and correlation.

Unit-IV

Point Estimation, Sampling Distributions and Central limit theorem, Point estimation, Maximum likelihood Estimation, - Confidence Interval on the mean of a Normal Distribution with Variance known and unknown, -Confidence interval on the variance and ratio of variances.

Unit – V

Hypothesis Testing, Tests on the mean of normal distributions with variance known and unknown. Tests on the variance single and two samples, Test of goodness of fit.

Textbooks:

1. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers, John Wiley and Sons Inc., 2005
2. Amir D Azcel, Jayavel Sounderpandian, Palanisamy Saravanan and Rohit Joshi, Complete Business Statistics, 7th edition McGrawHill education 2012.
3. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, Probability and Statistics for Engineers and Scientists, 8th Edition, Pearson Education Asia, 2007.

Reference books:

1. Ross S.M., *Introduction to Probability and Statistics for Engineers and Scientists*, 3rd edition, Elsevier Academic Press.
2. Ravichandran, J. Probability and Statistics for engineers, First Reprint Edition, Wiley India, 2012.

21MAT208A **CONVEX OPTIMIZATION** **3 0 0 3**

Unit 1. Introduction: Mathematical optimization, Least-squares and linear programming, Convex optimization, Nonlinear optimization.

Chapter 1.

Unit 2

Convex sets: Affine and convex sets. Some important examples. Operations that preserve convexity. Generalized inequalities. Separating and supporting hyperplanes. Dual cones and generalized inequalities.

Chapter-2

Unit 3

Convex functions: Basic properties and examples. Operations that preserve convexity. The conjugate function. Quasiconvex functions. Log-concave and log-convex functions. Convexity with respect to generalized inequalities.

Chapter-3.

Unit 4

Convex optimization problems. Optimization problems. Convex optimization. Linear optimization problems. Quadratic optimization problems. Geometric programming. Generalized inequality constraints. Vector optimization.

Chapter-4.

Unit 5:

Duality: The Lagrange dual function. The Lagrange dual problem. Geometric interpretation. Saddle-point interpretation. Optimality conditions. Perturbation and sensitivity analysis. Theorems of alternatives. Generalized inequalities.

TEXT BOOKS:

1. Stephen Boyd and Lieven Vandenberghe, Convex Optimization, Cambridge University Press, 2009.

REFERENCES:

1. Dimitri P. Bertsekas, Convex Optimization Theory, University Press, 2016.
2. Hamdy A. Taha, “Operations Research-An Introduction”, Prentice Hall, 9th Edition, 2010.
3. Edwin K.P. Chong and Stanislaw H. Zak, “An Introduction to Optimization”, Second Edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, 2004.

21MAT209A **Regression Analysis** **3 1 0 4**

Unit I

Simple Linear Regression: Linear Regression Model, Least square estimation of the parameters, Hypothesis Testing on the slope and intercept, Interval estimation in Simple linear Regression, Prediction of New Observations and Coefficient of Determination.

Unit II

Multiple Linear Regression: Multiple Linear Regression Models, Estimation of the Model Parameters, Hypothesis testing in Multiple Linear Regression, Confidence Interval on the Regression and Prediction of New observations.

Unit III

Generalized linear models - Logistic regression Models, Poisson regression - hypothesis testing on model parameter. Model Adequacy Checking: Introduction, Residual Analysis, Detection, treatment of Outliers and Lack of fit of the Regression Model.

Unit IV

Polynomial regression models – polynomial models in one variable – Polynomial models in two or more variables – variable selection and model building – computational techniques for variable selection.

Unit V

Introduction to analysis of variance- one way and two way ANOVA – Analysis of variance in Regression: Response surface designs – Introduction to response surface methodology, Method of steepest accent, Analysis of second order response surface, experimental design for fitting response surfaces.

Text Books/References:

1. Douglas C. Montgomery and Elizabeth A.Peck and G.Geoffrey Vining, Introduction to Linear Regression Analysis",3rd Edition ,John Wiley& Sons, Inc
2. Douglas C. Montgomery,Design and analysis of Experiments, 8th edition, John Wiley& Sons, Inc
3. Ravichandran, J. Probability and Statistics for engineers, First Reprint Edition, Wiley India, 2012

21ENG102A **Communicative English** **2-0-2-3**

Objectives:

To help students obtain an ability to communicate fluently in English; to enable and enhance the students skills in reading, writing, listening and speaking; to impart an aesthetic sense and enhance creativity

Course Contents:

Unit I

Kinds of sentences, usage of preposition, use of adjectives, adverbs for description, Tenses, Determiners- Agreement (Subject – Verb, Pronoun- Antecedent) collocation, Phrasal Verbs, Modifiers, Linkers/ Discourse Markers, Question Tags

Unit II

Paragraph writing – Cohesion - Development: definition, comparison, classification, contrast, cause and effect - Essay writing: Descriptive and Narrative

Unit III

Letter Writing - Personal (congratulation, invitation, felicitation, gratitude, condolence etc.) Official (Principal / Head of the department/ College authorities, Bank Manager, Editors of newspapers and magazines)

Unit IV

Reading Comprehension – Skimming and scanning- inference and deduction – Reading different kinds of material –Speaking**:** Narration of incidents / stories/ anecdotes- Current News Awareness

Unit V

Prose: John Halt's 'Three Kinds of Discipline' [**Detailed**]

Max Beerbohm's 'The Golden Drugget' [**Detailed**]

Poems: Ogden Nash- 'This is Going to Hurt Just a Little Bit' [**Detailed**]

Robert Kroetsch– 'I am Getting Old Now', Langston Hughes- 'I, Too' [**Detailed**]

Wole Soyinka- 'Telephone Conversation' **[Non- Detailed**]

Kamala Das- 'The Dance of the Eunuchs' [**Non-Detailed**]

Short Stories: Edgar Allan Poe's 'The Black Cat', Ruskin Bond's 'The Time Stops at Shamili' **[Non- Detailed**]

CORE READING:

1. *Ruskin Bond, Time Stops at Shamli and Other Stories, Penguin Books India Pvt Ltd, 1989*
2. *Syamala, V. Speak English in Four Easy Steps, Improve English Foundation Trivandrum: 2006*
3. *Beerbohm, Max, The Prince of Minor Writers: The Selected Essays of Max Beerbohm (NYRB Classics), Phillip Lopate (Introduction, Editor), The New York Review of Book Publishers.*
4. *Edger Allan Poe. The Selected Works of Edger Allan Poe. A Running Press, 2014.*
5. *Online sources*

References:

6. *Ruskin Bond, Time Stops at Shamli and Other Stories, Penguin Books India Pvt Ltd, 1989*
7. *Martinet, Thomson, A Practical English Grammar, IV Ed. OUP, 1986.*
8. *Murphy, Raymond, Murphy's English Grammar, CUP, 2004*
9. *Online sources*

21ENG111A **Professional Communication** **1- 0-2-2**

Objectives:

To convey and document information in a formal environment; to acquire the skill of self projection in professional circles; to inculcate critical and analytical thinking.

Unit I

Vocabulary Building: Prefixes and Suffixes; One word substitutes, Modal auxiliaries,Error Analysis: Position of Adverbs, Redundancy, misplaced modifiers**,** Dangling modifiers – Reported Speech

Unit II

Instruction, Suggestion & Recommendation - Sounds of English: Stress, Intonation

- Essay writing: Analytical and Argumentative

Unit III

Circulars, Memos – Business Letters - e - mails

Unit IV

Reports: Trip report, incident report, event report - Situational Dialogue - Group Discussion

Unit V

Listening and Reading Practice - Book Review

References

1. *FelixaEskey. Tech Talk, University of Michigan. 2005*
2. *Michael Swan. Practical English Usage, Oxford University Press. 2005*
3. *Anderson, Paul. Technical Communication: A Reader Centered Approach, V Edition, Hercourt, 2003.*
4. *Raymond V. Lesikar and Marie E. Flatley. Basic Business Communication, Tata Mc Graw Hill Pub. Co. New Delhi. 2005. Tenth Edition.*
5. *Thampi, G. Balamohan. Meeting the World: Writings on Contemporary Issues. Pearson, 2013.*
6. *Lynch, Tony. Study Listening. New Delhi: CUP, 2008.*
7. *Kenneth, Anderson, Tony Lynch, Joan Mac Lean. Study Speaking. New Delhi: CUP, 2008.*
8. *Marks, Jonathan. English Pronunciation in Use. New Delhi: CUP, 2007.*
9. *Syamala, V. Effective English Communication For You (Functional Grammar, Oral and Written Communication): Emerald, 2002.*

18MAT288 **Data Science Lab-I: Statistics Lab** **0 0 2 1**

1. Data Visualization using plot, pie chart, bar chart, histogram and Box plot
2. Find the central measures for given data, like, mean, mode, median and deviations
3. Root finding
4. Regression models
5. Correlation coefficient
6. Rank correlation.

21CSC283A **Machine Learning Lab** **0 0 2 1**

18CSC401 **Parallel and Distributed Systems** **3 1 0 4**

Unit 1

Introduction – parallelism and goals, parallel computing models – RAM, PRAM , CTA. Reasoning about Performance – Introduction -Basic Concepts - Performance Loss - Parallel Structure - Measuring Performance. Shared memory architecture.

Unit 2

Parallel Programming: Task and Data Parallelism with examples –Comparison Programming with Threads - POSIX Threads- Thread Creation and Destruction. Mutual Exclusion- Synchronization - Safety and Performance Issues – Reduction – threads Inter process communication – internet protocols – multicast communication – MPI. Remote invocation:Remote procedure call – remote method invocation -

Unit 3

System models : physical models, architecture models, operating system support. Distributed file systems – introduction- time and global states – synchronization of physical clocks – coordination and agreements: Mutual exclusion, election, consensus.

Text Books

1. George Coulouris , Jean Dollimore , Tim Kindberg , Gordon BlairDISTRIBUTED SYSTEMS Concepts and Design Fifth Edition , Addison Wiley, 2012.
2. Calvin Lin ,Larry Snyder, Principles of Parallel Programming, Pearson, 2009

References

1. Bertil Schmidt, Jorge Gonzalez-Dominguez, Christian Hundt , Moritz Schlarb, Parallel Programming: Concepts and Practice 1st Edition, Morgan Kaufmann, 2017.
2. Ajay D. Kshemkalyani, MukeshSinghal , Distributed Computing: Principles, Algorithms, and Systems, Cambridge University Press, 1 edition, 2008.

18CSC449 IoT Workshop 3 0 0 3

Unit - 1

Introduction to loT - loT definition - Characteristics - Things in loT - loT Complete Architectural Stack - loT enabling Technologies - loT Challenges - loT Levels - A Case Study to realise the stack.

Sensors and Hardware for loT - Accelerometer, Proximity Sensor, lR sensor, Gas Sensor,Temperature Sensor, Chemical Sensor, Motion Detection Sensor.Hardware Kits - Arduino,Raspberry Pi, Node MCU. A Case study with any one of the boards and data acquisition from sensors (Lab Component)

Unit - 2

Protocols for loT - infrastructure protocol IPV4/V6|RPL), Identification (URLs), Transport (Wi-Fi, Li-Fi, BLE), Discovery, Data Protocols, Device Management Protocols. - A Case Study with MQTT/CoAP usage. (Lab Component)

Cloud and Data analytics- Types of Cloud - loT with cloud challenges - Selection of cloud for IoT applications - Fog computing for loT - Edge computing for loT - Cloud security aspects for loT applications - RFM for Data Analytics - Case study with AWS / AZURE / Adafruit / IBM Bluemix (Lab Component).

Unit - 3

Case studies with architectural analysis:

loT applications - Smart City - Smart Water - Smart Agriculture - Smart Energy - Smart Healthcare - Smart Transportation - Smart Retail -Smart waste management . (Lab Component - As a project)

Text and Reference Books

1. "lnternet of Things: A Hands-on Approach", by ArshdeepBahga and Vijay Madisetti (UniversitiesPress)
2. lnfosys Training E Materials.
3. "The Internet of Things: Enabling Technologies, Platforms, and Use Cases',, by pethuru Rajand Anupama C. Raman (CRC press)

4. Adrian McEwen, Designing the internet of Things, Wiley (B November 2Ot3), ISBN-13:978-.11-L1,8430620,
5. NPTEL Reference :https://onlinecourses.nptel.ac.in/noc17_cs22/preview