Syllabus

M.Tech Data Science and Engineering

Semester I

DSE6101: MATHEMATICAL FOUNDATIONS OF DATA SCIENCE [3 1 0 4]

Basics of Data Science: Introduction, typology of problems, importance of linear algebra, statistics and optimization from a data science perspective, structured thinking for solving data science problems; **Linear Algebra:** Matrices and their properties (determinants, traces, rank, nullity, etc.), eigenvalues and eigenvectors, Matrix factorizations, inner products, distance measures, projections, notion of hyper planes, half-planes; **Probability, Statistics and Random Processes:** Probability theory and axioms, random variables, probability distributions and density functions (Uni-variate and multivariate), expectations and moments, covariance and correlation, statistics and sampling distributions, hypothesis testing of means, proportions, variances and correlations, confidence (statistical) intervals, correlation functions, white-noise process; **Optimization:** Unconstrained optimization, necessary and sufficiency conditions for optima, gradient descent methods, constrained optimization, KKT conditions, introduction to non-gradient techniques, introduction to least squares optimization, optimization view of machine learning; **Introduction to Data Science Methods:** Linear regression as an exemplar function approximation problem, linear classification problems.

References:

1. G. Strang *Introduction to Linear Algebra*, (5e), Wellesley-Cambridge Press,2016.
2. Bendat, J. S. and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, (4e), John Wiley & Sons, 2010.
3. Montgomery, D. C. and G. C. Runger, *Applied Statistics and Probability for Engineers*, (5e), John Wiley & Sons, 2011. 4. Cathy O'Neil and Rachel Schutt, *Doing Data Science*, (4e), O'Reilly Media, Fourth Edition, 2016.

DR6001: Research Methodology [3 0 0 3]

Introduction to Research, Meaning and Concepts, Types of Research, Process of Research, Classification of variables, Parts of Research, Formulating Research Problem, Determining research Objective, Ethics in research, Limitations in research, Research repositories, **Research Design**: Meaning and concept of research design, Types of research design, Need for research design,

Features of a good research design, Developing a research plan, **Data Collection**: Types of Data, Primary and secondary data - benefits and drawbacks, Various methods of data collection, **Measurement scales** - classification of scales, Measurement error, Criteria for good measurement, **Interpretation and report writing**: Meaning of Interpretation, Technique of interpretation, Precaution in interpretation, Types of reports, Significance of report writing, Different steps in writing report, Layout of the research report, Oral Presentation, Thesis writing, Research paper writing, Preparing of synopsis and summary of research thesis work, **Reference writing**: Footnote, Various electronics tool for citation and referencing, in-text citation, Bibliography, Citation styles, Bibliometrics, **Sampling techniques**: Basic terms, Importance of sampling in research, Essentials of a good sample, Sampling error, Standard error of the mean (standard deviation), Estimation of parameters, Accuracy and precision of estimation, Sampling procedure, Types/methods of sampling, Central limit theorem, Sample size determination, Confidence interval and confidence level, Measurements and scaling techniques, Types of data primary and secondary, **Types of scales**: Ratio, Interval, Ordinal nominal, Mapping rules, Characteristic of a good measurements, Sources of error in measurements.

References :

1. C. R. Kothari and G. Garg, *Research Methodology: methods and techniques*, New Age International (P) Ltd, 2019.
2. K.S. Bordens and B. B. Abbitt, *Research Design & Methods, A process approach*, (10e), McGraw Hill Education, 2018.
3. D. Chawla and N. Sondhi, *Research Methodology-Concepts and Cases*, Vikas Publishing House, 2011.
4. D. R. Cooper, P. S. Schindler, *Business Research Methods*, (12e), McGraw Hill, 2014.
5. A. Bryman and E. Bell, *Business Research Methods*, (5e), Oxford University Press, 2018.
6. W. G. Zikmund, *Business Research Methods*, (8e), Cenagage Learning, 2010.
7. W. Guddard and S. Melville, *Research Methodology: An Introduction for Science & Engineering Students*, (2e), Kenwyn, South Africa: Juta & Co., 2011.
8. R. Ganesen, *Research Methodology for Engineers*, MJP Publishers, Chennai, 2011.

DSE6102: ADVANCED DATA STRUCTURES AND ALGORITHMS [3 1 0 4]

Advanced Search Trees: Review of Binary Search Trees, AVL Tree, R-B Trees and Splay Trees, Advanced Search Data Structures Like- Treaps, Skip Lists, Finger Search Trees, Biased Search Trees; **Data Structures For External Storage:** Review of 2-3-4 Trees and 2-3 Trees, B-Tree, B+ Trees, Priority Queues and Concatenable Queues Using 2-3 Trees; **Advanced Heaps:** Review of Heaps, Binomial Trees, Implementing Binomial Heaps and its Operations, Structure of Fibonacci Heaps, Mergeable Heap Operations, Decreasing Key and Deleting a Node, Bounding the Maximum Degree, Amortized Analysis of Fibonacci Heaps; **Dictionaries and Hashing:** Review of Dictionaries and Implementation, Review of Hashing- The Bucket Approach, Index File Approach, Universal Hashing, Perfect Hashing,

LocalitySensitive Hashing, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing, Synopses, Fingerprints, Fault Tolerant Data Structures; **Graph Theory And Sorting Network**: Review Of Graph Representation and Basic Algorithms, Algorithms for Connectedness, Finding all Spanning Trees in a Weighted Graph, Bipartite Graphs: Maximum Matching, the Hungarian Algorithm, Maximum Flow in a Transport Network - the Ford– Fulkerson Algorithm; **Sorting Network:** Comparison Network, Zero-One Principle, Bitonic Sorting and Merging Network Sorter.

References:

1. Cormen, T. H., Leiserson, C. E., Rivest, R.L., and Stein, C. *Introduction to Algorithms*, (3e), MIT Press, 2009.
2. Aho, V., Hopcroft, J. E. and Ullman, J. D. *The Design and Analysis of Computer Algorithms*, (1e), Fourth Impression, Pearson Education, 2009.
3. Horowitz, E., Sahni, S. and Rajasekaran, S, *Computer Algorithms*, (2e), University Press, 2007.
4. Weiss, M. A, *Data Structures and Algorithm Analysis in C++*, (2e), Pearson Education India, 2004.
5. Goodrich, M. T., Tamassia, R., *Algorithm Design*, (1e), John Wiley, 2002.

DSE6103: MACHINE LEARNING TECHNIQUES [3 1 0 4]

Introduction: learning problems, perspectives and issues, concept learning, version spaces and candidate eliminations, inductive bias, decision tree learning, representation, algorithm, heuristic space search. Linear and logistic regression; **Neural Networks and Genetic Algorithms:** Neural network representation, problems, perceptron's, multilayer networks and back propagation algorithms, advanced topics, hyper parameter optimization, genetic algorithms, hypothesis space search, genetic programming, models of evaluation and learning; **Bayesian and Computational Learning:** Bayes theorem, concept learning, maximum likelihood – minimum description length principle, Bayes optimal classifier, Gibbs algorithm, naïve Bayes classifier, Bayesian belief network, EM algorithm, probability learning, sample complexity, finite and infinite hypothesis spaces, mistake bound model; **Instant Based Learning**: K- nearest neighbor learning, locally weighted regression, radial basis functions, case based learning; **Advanced Learning:** Learning sets of rules, sequential covering algorithm, learning rule set, first order rules, sets of first order rules, induction on inverted deduction, inverting resolution, analytical learning, perfect domain theories, explanation base learning, FOCL algorithm, reinforcement learning task, Q-learning, temporal difference learning.

References:

1. Tom M. Mitchell, *Machine Learning*, (2e), McGraw-Hill Education India Private Limited, 2017.
2. Ethem Alpaydin, *Introduction to Machine Learning*, (3e), The MIT Press, 2015.
3. Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, (2e), CRC Press, 2014.

DSE6141: DATA PREPARATION AND ANALYSIS [3 0 0 3]

Data Gathering and Preparation: Data Formats, Parsing and Transformation, Scalability and Real-time Issues; **Data Cleaning:** Consistency Checking, Heterogeneous and Missing Data, Data Transformation and Segmentation; **Exploratory Analysis:** Descriptive and comparative statistics, Clustering and association, Hypothesis Generation; **Visualization:** Designing Visualizations, Time Series, Geolocated Data, Correlations and Connections, Hierarchies and Networks, Interactivity.

References:

1. Glenn J. Myatt., *Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining*, (2e), Wiley Press, 2006.
2. E. Tufte. *The Visual Display of Quantitative Information*, (2e), Graphics Press, 2007.
3. Jules J., Berman D., *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, (2e), 2013.

DSE6143: KNOWLEDGE DISCOVERY [3 0 0 3]

Introduction to Knowledge Management: Data, Information and Knowledge, Types of Knowledge, Knowledge Management Models, Organizational Memory, Knowledge Creation and Organizational Learning, Knowledge Codification, Knowledge Sharing and Transfer, Business and Competitive Intelligence, Ethical Issues in Knowledge Management, The Role of Culture in Knowledge Management, Business Process and Knowledge Management, Alignment of Business and Knowledge Management Strategies, Intellectual Capital and Knowledge Management, Measurement of Impact of Knowledge management programs; **Technologies for Knowledge Management:** Artificial Intelligence, Digital Libraries, Repositories, ECM, Knowledge-Based Systems, Information/Knowledge Audit; **An Architecture for Knowledge Discovery:** KM Cycle its Vision and Search, Generation, Acquisition, Capture, Transformation, Transfer, Application; **Knowledge Capture Systems:** Systems that Preserve and Formalize Knowledge; Concept Maps, Process Modeling, RSS, Wikis, Delphi Method; **Knowledge Sharing Systems:** Systems that Organize and Distribute Knowledge; Ontology Development Systems, Categorization and Classification Tools, XML-Based Tools; **Knowledge Discovery Methods:** Correlation, Class, Novelty, Association, Preprocessing Methods, Supervised Methods, Unsupervised Methods, Soft Computing Methods, Supporting Methods , Advanced Methods.

References:

1. Wesley, W Chu, *Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities*, (1e), Springer, 2013.
2. Dalkir, K, *Knowledge Management in Theory and Practice*, (2e), MIT Press, 2011.
3. Oded, M and Lior, R, *The Data Mining and Knowledge Discovery Handbook*, (2e), Springer, 2010.

DSE6121 Machine Learning Application for Music Data Processing [3 0 0 3]

Introduction: Basics of Sound engineering and Sound and Waveforms, Types of Music patterns and various sounds, Basic Terminology (Amplitude, Frequency, Intensity, Loudness, Timber), Audio Feature Extraction, Filtering and Windowing, Time-Domain Features of Audio Signals (Amplitude envelope, Root-mean-square energy, Zero-crossing rate), Frequency-Domain Analysis of Audio Signals. Fourier transform: Intuition, Discrete Fourier transform, Fourier transform properties, Short-time Fourier transform. Spectrogram: Spectrogram Analysis, Mel Spectrogram, Mel-Frequency Cepstral Coefficients (MFCC). Machine Learning in Audio Processing: Overview of machine learning concepts and applications in audio processing, Introduction to neural networks and deep learning, Convolutional Neural Networks (CNNs) for audio classification, Recurrent Neural Networks (RNNs) for sequential audio data, Python libraries for audio data processing.

CASE STUDY: Music Genre Classification related case study.

Books:

Knees P. & Schedl M (2016) Music Similarity and Retrieval: An Introduction to Audio and Web based Strategies

DSE6130: ADVANCED DATA STRUCTURES LAB [0 0 4 2]

Application of Search Trees: Search, insert and delete operations and their comparison for binary search tree, R-B Trees and 2-3 trees. Analysis of operations on structures studied like skip-lists; **Application of Heaps:** Converting a set of data records in to Binomial heaps and Fibonacci heaps. Comparing the search, add and delete operations based on time and space requirements; **Application of Graph:** Maximum flow problem should be solved using Ford Fulkerson method. A mapping solved using Hungarian Algorithm and Ford Fulkerson algorithm; **Lab Project:** A generalized case study where student will choose appropriate data-structure(s) and apply based on their utility for specific problem.

References:

1. Cormen, T. H., Leiserson, C. E., Rivest, R.L., and Stein, C. *Introduction to Algorithms*, (3e), MIT Press, 2009.
2. Aho, V., Hopcroft, J. E. and Ullman, J. D, *The Design and Analysis of Computer Algorithms*, (1e), Fourth Impression, Pearson Education India, 2009.
3. Horowitz, E., Sahni, S. and Rajasekaran, S, *Computer Algorithms*, (2e), University Press, 2007.
4. Weiss, M. A, *Data Structures and Algorithm Analysis in C++*, (2e), Pearson, 2004.
5. Goodrich, M. T., Tamassia, R., *Algorithm Design*, (1e), John Wiley, 2002.

DSE6131: MACHINE LEARNING TECHNIQUES LAB [0 0 4 2]

Introduction to Machine Learning and Python Environment: Setting up the development environment (Python, Jupyter Notebooks, libraries), Vectors, Lists, Matrices, Arrays, Factors, Data Frames. Essential libraries: NumPy, Pandas, Matplotlib. **Data Preprocessing:** Data cleaning techniques and handling missing values. Data transformation (scaling, encoding categorical variables). **Learning Techniques:** Use various Supervised and Unsupervised learning techniques like Linear Regression and Nonlinear Regression, Model evaluation and metrics. Apply Statistical approaches for advance learning techniques. Construct models for Classification. Build neural network models. **Lab Project:** Data preprocessing and regression analysis, Classification and neural network model construction

REFERENCES:

1. Johannes Ernesti & Peter Kaiser,Python 3: The Comprehensive Guide to Hands-On Python Programming (1st Edition), Rheinwerk Computing Publication,1st edition,2022.

2. Vijayvargia, Abhishek, Machine Learning with Python: An Approach to Applied Machine Learning, BPB Publications, 1st edition,2018.

2. Aurelien Geron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, Oreilly, March 2017.

3. Dr. M Gopal, Applied Machine Learning, 1st Edition, McGraw-Hill,2018