Course Title	Mathematical Foundations for Data Science	**Course Number**	CS501T
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Core
Prerequisite	NA	**Effective From**	July 2020
Course Aim	The course will introduce the fundamental concepts of linear algebra, probability and statistics required for a program in data science.		
Course Outcomes	1. Ability to use the mathematical concepts in the field of data science. 2. Employ the techniques and methods related to the area of data science in variety of applications. 3. Apply logical thinking to understand and solve the problem in context.		
Contents of the Course	Module – I 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. Module – II Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes. Module – III Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate 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. Module – IV 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, Wellesley-Cambridge Press, Fifth edition, USA, 2016. 2. Bendat, J. S. and A. G. Piersol. Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA, 2010 3. Montgomery, D. C. and G. C. Runger. Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA, 2011. 4. David G. Luenberger . Optimization by Vector Space Methods, John Wiley & Sons (NY), 1969. 5. Cathy O'Neil and Rachel Schutt . Doing Data Science, O'Reilly Media, 2013.

Course Title	Advanced Data Structures and Algorithms	**Course Number**	CS502T
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Elective
Prerequisite	NA	**Effective From**	July 2020
Course Aim	The aim of the course is to introduce advanced algorithms and practice programming techniques necessary for developing sophisticated computer application.		
Course Outcomes	1. Understand the implementation of symbol table using hashing techniques. 2. Apply advanced abstract data type (ADT) and data structures in solving real world problem. 3. Effectively combine the fundamental data structures and algorithmic techniques in building a solution to a given problem. 4. Develop algorithms for text processing applications.		
Contents of the Course	Module – I Dictionaries: Definition, Dictionary Abstract Data Type, Implementation of Dictionaries, Hashing: Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing. Module – II Skip Lists: Need for Randomizing Data Structures and Algorithms, Search and Update Operations on Skip Lists, Probabilistic Analysis of Skip Lists, Deterministic Skip Lists, Trees: Binary Search Trees (BST), AVL Trees, Red Black Trees: Height of a Red Black Tree, Red Black Trees Bottom-Up Insertion, Top-Down Red Black Trees, Top-Down Deletion in Red Black Trees, Analysis of Operations. Module – III 2-3 Trees: Advantage of 2-3 trees over Binary Search Trees, Search and Update Operations on 2-3 Trees, Analysis of Operations, B-Trees: Advantage of B- trees over BSTs, Height of B-Tree, Search and Update Operations on 2-3 Trees, Analysis of Operations, Splay Trees: Splaying, Search and Update Operations on Splay Trees, Amortized Analysis of Splaying.		

	Module – IV Text Processing: Sting Operations, Brute-Force Pattern Matching, The Boyer-Moore Algorithm, The Knuth-Morris-Pratt Algorithm, Standard Tries, Compressed Tries, Suffix Tries, The Huffman Coding Algorithm, The Longest Common Subsequence Problem (LCS), Applying Dynamic Programming to the LCS Problem, Computational Geometry: One Dimensional Range Searching, Two Dimensional Range Searching, Constructing a Priority Search Tree, Searching a Priority Search Tree, Priority Range Trees, Quadtrees, k-D Trees.
References	1. Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, second Edition, Pearson, 2004. 2. T.H. Cormen, C.E. Leiserson, R.L.Rivest, Introduction to Algorithms, Third Edition Prentice Hall, 2009. 3. Michael T. Goodrich, Roberto Tamassia, Algorithm Design, First Edition, Wiley, 2006.

Course Title	Principles of Data Analytics	**Course Number**	CS503T
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Core
Prerequisite	Algorithms, Linear Algebra, Probability and Statistics	**Effective From**	July 2020
Course Aim	This course will introduce students to the principles of data analytics that underpin key tools and techniques used both to describe and to gain insights into the properties of large and complex datasets. Students will learn the concepts and tools that deal with various facets of data science practice. It includes data collection and integration, exploratory data analysis, predictive modeling, descriptive modeling, data product creation, evaluation, and effective communication.		
Course Outcomes	1. Understand the ideas of statistical approaches to learning. 2. Work with big data platform and explore the data analysis techniques for business applications. 3. Design efficient algorithms for mining the data from large volumes. 4. Perform appropriate statistical tests using R and visualize the outcome. 5. Understand the significance of exploratory data analysis (EDA) in data science and apply basic tools (plots, graphs, summary statistics) to perform EDA. 6. Apply basic machine learning algorithms (Linear Regression, k-Nearest Neighbors (k-NN), k-means, Naive Bayes) for predictive modeling. Explore the merits of Naive Bayes technique. 7. Recognize the characteristics of machine learning techniques that are useful to solve real-world problems. 8. Identify basic approaches used for feature generation and feature selection algorithms (Filters, Wrappers, Decision Trees, and Random Forests) and to apply the techniques in applications. 9. Identify and explain fundamental mathematical and algorithm paradigms that constitute a recommendation engine. Build their own recommendation system using existing components.		
Contents of the Course	Module – I Introduction: What is Data Science? Big Data and Data Science hype		

	and getting past the hype, Why now?, Datafication, Current landscape of perspectives, Skill sets, Life cycle of Data Science, Different phases. Module - II Exploratory Data Analysis and the Data Science Process: Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA, The Data Science Process, Case Study: RealDirect (online real estate firm), Three Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbors (k-NN), k-means. Module – III One More Machine Learning Algorithm and Usage in Applications: Motivating application: Filtering Spam, Why Linear Regression and k-NN are poor choices for Filtering Spam, Naive Bayes and why it works for Filtering Spam, Data Wrangling: APIs and other tools for scrapping the Web, Feature Generation and Feature Selection (Extracting Meaning From Data), Motivating application: user (customer) retention, Feature Generation (brainstorming, role of domain expertise, and place for imagination), Feature Selection algorithms: Filters; Wrappers; Decision Trees; Random Forests, Recommendation Systems: Building a User-Facing Data Product: Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis, Exercise: build your own recommendation system. Module – IV Data Visualization: Basic principles, ideas and tools for data visualization, Case study on industry projects, Exercise: create your own visualization of a complex dataset, Data Science and Ethical Issues: Discussions on privacy, security, ethics, A look back at Data Science, Next-generation data scientists.
References	1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly, 2014. 2. Jure Leskovek, Anand Rajaraman and Jerey Ullman. Mining of Massive Datasets, Cambridge University Press, 2014. 3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2013. 4. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic

	Thinking. O′Reilly, 2013. 5. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. Springer, 2009. 6. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science.2018. 7. Mohammed J. Zaki and Wagner Miera Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, 2014. 8. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, Third Edition. Morgan Kaufmann, 2011.

<table>
<tr><td>Course Title</td><td>Statistical Learning</td><td>Course Number</td><td>CS504T</td></tr>
<tr><td>Department</td><td>Computer Science and Engineering</td><td>Structure (I-P-C)</td><td>3-0-3</td></tr>
<tr><td>Offered To</td><td>M. Tech. CSE-DADS</td><td>Status (Core/Elective)</td><td>Core</td></tr>
<tr><td>Prerequisite</td><td>Probability, Statistics and Algorithms</td><td>Effective From</td><td>July 2020</td></tr>
<tr><td>Course Aim</td><td colspan="3">This course will provide an introduction to the theoretical analysis of prediction methods, focusing on statistical and computational aspects. It will provide an introduction to Machine Learning and its core models and algorithms. and probabilistic and game theoretic formulations of prediction problems, and it will focus on tools for the theoretical analysis of the performance of learning algorithms and the inherent difficulty of learning problems.</td></tr>
<tr><td>Course Outcomes</td><td colspan="3">1. Familiarize with a broad range of approaches and techniques in machine learning.
2. Choose effective methods to solve various learning problem.
3. Explore statistical learning methods and their application to modern problems in science, industry, and society.
4. Build analytics pipelines for regression problems and classification problems.
5. Build analytics pipelines for recommendation problems.</td></tr>
<tr><td>Contents of the Course</td><td colspan="3">Module – I

Probabilistic formulations of prediction problems: Plug-in estimators, empirical risk minimization, Linear threshold functions, perceptron algorithm, Risk bounds, Concentration inequalities, Uniform convergence, Rademacher averages; combinatorial dimensions, Convex surrogate losses for classification, Linear regression, Regularization and linear model selection, Feature Selection Methods, Cross Validation methods.

Module - II

Game-theoretic formulations of prediction problems, High Dimensional methods, Lasso, Ridge Regression, Dimensionality Reduction, Minimax strategies for log loss, linear loss, and quadratic loss, Universal portfolios, Online convex optimization.

Module – III</td></tr>
</table>

	Neural networks: Stochastic gradient methods, Combinatorial dimensions and Rademacher averages, Hardness results for learning, Efficient learning algorithms. Module - IV Kernel methods: Reproducing kernel Hilbert spaces, Mercer's theorem, Convex optimization for kernel methods, Representer theorem, Ensemble methods: AdaBoost, AdaBoost as I-projection, Convergence and consistency of AdaBoost.
References	1. James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013. 2. Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer, 2009.

Course Title	Social Media and Network Analysis	**Course Number**	CS515T
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Elective
Prerequisite	NA	**Effective From**	July 2020
Course Aim	This course introduces the basic concepts and analysis techniques in SNA. Students learn how to identify key individuals and groups in social systems, to detect and generate fundamental network structures, and to model growth and diffusion processes in networks. Students will be trained in interpreting the meaning of the aforementioned phenomena and suggesting potential courses of action to reinforce or change the observed trends.		
Course Outcomes	1. Formalize different types of entities and relationships as nodes and edges and represent this information as relational data. 2. Plan and execute network analytical computations. 3. Use advanced network analysis software to generate visualizations and perform empirical investigations of network data. 4. Interpret and synthesize the meaning of the results with respect to a question, goal, or task. 5. Collect network data in different ways and from different sources while adhering to legal standards and ethics standards.		
Contents of the Course	Module – I Class logistics, Overview on Network Analysis, The Network Analysis Process and Methodology, Network Visualization, Sociometric Analysis, Preliminaries on Network Structures, Models and simulation on Network evolution, Sociological direction. Module – II Subgroups and Cliques, Clustering of graphs, Organization of relational data, Mining Social-network graphs, Direct discovery of communities in graph, Neighborhood properties in graphs, Ego Networks, Structural Holes, Cognitive Social Structures. Module – III Introduction: Integration of Text and Network Analysis, Types of Networks Extracted From Texts Across Disciplines, Local Centrality, Global Centrality, Bank centrality in corporate network.		

	Module – IV Introduction: Multi-Agent Models for Representing Networks, Social Network Based Multi-Agent systems, Trends on social networks, Case Studies.
References	1. Scott, J. Social network analysis: A handbook, Second Edition, Newbury Park. Sage, 2008 2. Knoke. Social Network Analysis, Second Edition, Sage, 2008. 3. Hanneman, Robert A. and Mark Riddle, Introduction to social network methods, Riverside, CA, 2005.

Course Title	Decision Sciences	**Course Number**	CS505I
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Elective
Prerequisite	NA	**Effective From**	July 2020
Course Aim	This course provides an introduction to the concepts and methods of Decision Science, which involves the application of mathematical modeling and analysis to management problems. It also provides a foundation in modeling with spreadsheets. The primary goal of the course is to help you become a more skilled builder and consumer of models and model-based analyses.		
Course Outcomes	1. Familiarize with the usage of Microsoft Excel for business analysis. 2. Understand the role of quantitative techniques for managerial decision making. 3. Ability to structure problems and to perform logical analyses. 4. Explore the functionalities from various models and to use those insights to communicate, persuade and motivate change.		
Contents of the Course	Module – I Decision making process, Introduction to spreadsheet modeling and problem solving, Spreadsheet engineering, Spreadsheet analysis, Modeling and Prototyping, Module – II Measures of Central Tendency and Measures of Dispersion, Data analysis, Modeling in Practice, Enterprise Resource Planning, Introduction to Optimization and Solver, Case study. Module – III Creating optimization models, Optimization models and sensitivity analysis, Production planning, Revenue management, Sales force sizing and allocation, Case Study. Module – IV Introduction to simulation, Simulation techniques and examples, Simulation modeling and analysis, Managing risk with insurance, Optimization in simulation, Assessing acquisition value with simulation.		

References	1. Francis J. Clauss, Applied Management Science and Spreadsheet Modeling, Duxbury. 2. Jeffrey D. Camm and James R. Evans, Management Science: Modeling, Analysis, and Interpretation, South-Western. 3. Cliff Ragsdale, Spreadsheet Modeling and Decision Analysis, South-Western. 4. Wayne L. Winston and S. Christian Albright, Practical Management Science: Spreadsheet Modeling and Applications, Duxbury. 5. Jeffrey Moore et al., Introductory Management Science, Prentice-Hall.

Course Title	Decision Sciences Practice	**Course Number**	CS506I
Department	Computer Science and Engineering	**Structure (I-P-C)**	3-0-3
Offered To	M. Tech. CSE-DADS	**Status (Core/Elective)**	Elective
Prerequisite	NA	**Effective From**	July 2020
Course Aim	This course provides an introduction to the concepts and methods of Decision Science, which involves the application of mathematical modeling and analysis to management problems. It also provides a foundation in modeling with spreadsheets. The primary goal of the course is to help you become a more skilled builder and consumer of models and model-based analyses.		
Course Outcomes	1. Familiarize with various features of R language. 2. Understand the role of quantitative techniques for managerial decision making. 3. Ability to perform data analysis with popular datasets. 4. Explore and familiarize with experiments in linear model selection, cross validation and other statistical learning techniques.		
Contents of the Course	1. Designing a spreadsheet that will allow an analyst to predict the monthly expenses in a company. 2. Importing and exporting spreadsheet data. 3. Regression: linear regression, test of significance, residual analysis, polynomial regression using R tool. 4. Experiments on cross validation, and linear model selection. 5. Experiments on advanced linear regression methods such as lasso and ridge regression. 6. Exploratory data analysis: Charts and Plots, Data Visualization, combining visualization and data transformation that allows to efficiently explore the data. 7. Information Extraction from given Textual data. 8. Apply the techniques of Linear Algebra using R tool on given dataset. 9. Optimization in R: Common R Packages for Linear, Quadratic and Non-linear optimization and sensitivity analysis. 10. Linear Programming in R: lpsolve. 11. Capstone Project.		
References	1. Francis J. Clauss, Applied Management Science and Spreadsheet Modeling, Duxbury. 2. Jeffrey D. Camm and James R. Evans, Management Science: Modeling, Analysis, andInterpretation, South-Western. 3. Cliff Ragsdale, Spreadsheet Modeling and Decision Analysis,		

	South-Western. 4. Wayne L. Winston and S. Christian Albright, Practical Management Science: SpreadsheetModeling and Applications, Duxbury. 5. Jeffrey Moore et al., Introductory Management Science, Prentice-Hall.