Course Structure and Syllabus
(From the admitted batch of 2017 – 2018 under CBCS Scheme)

I MTech II Semester
Computer Science and Technology

Sub Code	Subject	Hrs/Week Theory	Hrs/Week Lab	Max Marks Internal	Max Marks External	Total Marks	Credits
MTCST201	BIG DATA ANALYSIS	4	--	25	75	100	4
MTCST202	CLOUD COMPUTING	4	--	25	75	100	4
MTCST203	DATA SCIENCE WITH R	4	--	25	75	100	4
MTCST204	INFORMATION SECURITY AND CRYPTOGRAPHY	4	--	25	75	100	4
MTCST205	ARTIFICIAL INTELLIGENCE/INTERNET OF THINGS/ WEB TECHNOLOGIES	4	--	25	75	100	4
MTCST206	DIGITAL FORENSICS /CYBER SECURITY/ / IMAGE PROCESSING	4	--	25	75	100	4
MTCST207	DATA SCIENCE WITH R LAB	-	3	50	50	100	2
MTCST208	INFORMATION SECURITY AND CRYPTOGRAPHY LAB	--	3	50	50	100	2
	TOTAL	24	6	250	550	800	28

I MTech II Semester
Computer Science and Technology with Big Data Analytics

Sub Code	Subject	Hrs/Week Theory	Hrs/Week Lab	Max Marks Internal	Max Marks External	Total Marks	Credits
MTBDA201	BIG DATA ANALYSIS	4	--	25	75	100	4
MTBDA202	CLOUD COMPUTING	4	--	25	75	100	4
MTBDA203	DATA SCIENCE WITH R	4	--	25	75	100	4
MTBDA204	DATA MINING AND SOCIAL NETWORKING	4	--	25	75	100	4
MTBDA205	MACHINE LEARNING/INTERNET OF THINGS/ WEB TECHNOLOGIES	4	--	25	75	100	4
MTBDA206	STATISTICAL COMPUTING FOR DATA ANALYTICS /PARALLEL AND DISTRIBUTED DATABASES/NATURAL LANGUAGE PROCESSING	4	--	25	75	100	4
MTBDA207	DATA SCIENCE WITH R LAB	-	3	50	50	100	2
MTBDA208	HADOOP LAB	-	3	50	50	100	2
	TOTAL	24	6	250	550	800	28

MTCST201,MTBDA201:BIG DATA ANALYSIS
(M.Tech CST with Big Data Analytics)

Theory	: 4 Hrs/week	**Credits**	: 4
Int Marks	: 25	**Ext Marks**	: 75

UNIT - I

Data structures in Java: Java concepts required for developing Map Reduce Programs: Linked List, Stacks, Queues, Sets, Maps; Generics: Generic classes and Type parameters, Implementing Generic Types, Generic Methods, Wrapper Classes, Concept of Serialization. [Big Java 4th Edition – 15th and 16th chapters]

Introduction to Big Data: Big Data-definition, Characteristics of Big Data 5Vs (Volume, Velocity, Variety, Veracity, Value), Data in the Warehouse and Data in Hadoop, Importance of Big Data, Patterns for Big Data Development [Dirk deRoss et al – 1st and 2nd chapters]

UNIT - II

Introduction to Hadoop: Hadoop- definition, Understanding distributed systems and Hadoop, Comparing SQL databases and Hadoop, Understanding MapReduce, Counting words with Hadoop—running your first program, History of Hadoop.

Starting Hadoop - The building blocks of Hadoop (NameNode, DataNode, Secondary NameNode, JobTracker and Task Tracker)

HDFS: Components of Hadoop -Working with files in HDFS, Anatomy of a MapReduceprogram, Reading and writing [LAM TextBook – 1st, 2nd and 3rd Chapters]

The Hadoop Distributed File system - The Design of HDFS, HDFS Concepts, The Command-Line Interface, Hadoop File system, The Java Interface, Data Flow, Parallel Copying with distcp, Hadoop Archives.[Hadoop the definitive guide – 3rd Chapter]

UNIT - III

MapReduce Programming: Writing basic Map Reduce programs - Getting the patent data set, constructing the basic template of a Map Reduce program, Counting things, Adapting for Hadoop's API changes, Streaming in Hadoop.

MapReduce Advanced Programming: Advanced MapReduce - Chaining MapReduce jobs, joining data from different sources, Bloom Filters. [Lam – 4th and 5th Chapters]

UNIT – IV

Utilizing data structures and algorithms: Modelling data and solving problems with graphs, Shortest Path Algorithm, Friends-of-Friends Algorithm, PageRank Algorithm. [Hadoop in Practice by Alex Holmes – 7th Chapter]

Text Books:
1. Big Java Fourth Edition Cay Horstmann Wiley, John Wiley & Sons
2. Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data by Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, 1st Edition, TMH,2012.
3. Hadoop in Action by Chuck Lam, MANNING Publishers.
4. Hadoop in Practice by Alex Holmes, MANNING Publishers
5. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly

Reference Books:
1. Mining of massive datasets, AnandRajaraman, Jeffrey D Ullman, Wiley Publications.

MTCST202,MTBDA202: CLOUD COMPUTING
(Common for M.Tech CST and M.Tech CST with Big Data Analytics)

Theory : 4 Hrs/week   **Credits** : 4
Int Marks : 25   **Ext Marks** : 75

UNIT - I

Introduction: Overview, Peer-to-Peer Systems, Cloud Computing-Delivery Models, services, Ethical Issues and Vulnerabilities, Overview of Parallel and Distributed Systems.

Cloud Infrastructure: Cloud Computing (At Amazon, At Google, Microsoft Windows Azure), Private Clouds, Cloud Storage, Cloud Computing Interoperability (The Intercloud), Responsibility Sharing Between User and Cloud Service Provider.

Cloud Computing (Applications and Paradigms): Challenges, Architectural Styles, Workflows, TheZooKeeper, The MapReduce Programming Model, High Performance Computing on Cloud.

UNIT - II

Cloud Resource Virtualization: Layering and Virtualization, Virtual Machines, Full Virtualization and Para virtualization, Hardware Support for Virtualization.

Cloud Resource Management and Scheduling: Policies and Mechanisms for Resource Management, Applications of Control Theory to Task Scheduling on a Cloud, Stability of a Two-Level Resource Allocation Architecture, Resource Bundling(Combinatorial Auctions for Cloud Resources), Scheduling Algorithms for Computing Clouds, Fair Queuing, Start-Time Fair Queuing, Cloud Scheduling Subject to Deadlines, Scheduling MapReduce Applications Subject to Deadlines.

UNIT - III

Storage System: The Evolution, Storage Models, File Systems, and Databases, Distributed File Systems, General Parallel File Systems, Google File Systems, Apache Hadoop, Locks and Chubby, BigTable, Megastore.

Cloud Security: Security Risks, Security (The Top Concern for Cloud Users), Privacy and Privacy Impact Assessment, Trust, OS Security, VM Security, Security of Virtualization, Xoar: Breaking the Monolithic Design of the TCB.

UNIT -IV

Cloud Application Development: Amazon Web Services: EC2 Cloud Application, How to install Hadoop in Eclipse on a Windows System, Cloud-Based Simulation of a Distributed Trust Algorithm.

Case Study: The GrepTheWeb Application, Xen, a VMM Based on Para virtualization, Migrating an enterprise IT System to IaaS.

Text Books:
1. Cloud Computing: Theory and Practice, Dan C. Marinescu, Morgan Kaufmann, Elsevier.

Reference Books:
1. Distributed and Cloud Computing, Kai Hwang, Geoffrey C.Fox, Jack J.Dongarra, Morgan Kaufmann, 1st Edition, 2011.
2. Cloud Computing-A Practical Approach, Anthony T. Velte, Toby J. Velte, Robert Elsenpeter. McGrawHill.

MTCST203,MTBDA203: DATA SCIENCE WITH R
(Common for M.Tech CST and M.Tech CST with Big Data Analytics)

Theory	: 4 Hrs/week	**Credits**	: 4
Int Marks	: 25	**Ext Marks**	: 75

UNIT - I

Introduction to Data Science- The Data Science process- The roles in a data science project, Stages of a data science project: Defining the goal , Data collection and management Modelling , Model evaluation and critique, Presentation and documentation, Model deployment and maintenance , Setting expectations: Determining lower and upper bounds on model performance

Exploring Data - Using summary statistics to spot problems: Typical problems revealed by data summaries, Spotting problems using graphics and visualization: Visually checking distributions for a single variable, visually checking relationships between two variables

Managing data -Cleaning data, Treating missing values (NAs), Data transformations, Sampling for modelling and validation: Test and training splits, Creating a sample group column, Record grouping, Data provenance

UNIT - II

Modelling Methods - Choosing and evaluating models -Mapping problems to machine learning tasks: Solving classification problems, Solving scoring problems ,Working without known targets, Problem-to-method mapping . Evaluating models: Evaluating classification models, Evaluating scoring models, Evaluating probability models, Evaluating ranking models, Evaluating clustering models. Validating models: Identifying common model problems, Quantifying model soundness, Ensuring model quality

Memorization methods - Building single-variable models: Using categorical features, Using numeric features, Using cross-validation to estimate effects of overfitting .

UNIT - III

Linear and logistic regression -Using linear regression: Understanding linear regression , Building a linear regression model ,Making predictions, Finding relations and extracting advice , Reading the model summary and characterizing coefficient quality , Linear regression takeaways .Using logistic regression: Understanding logistic regression , Building a logistic regression model , Making predictions , Finding relations and extracting advice from logistic models , Reading the model summary and characterizing coefficients.

Unsupervised methods - Cluster analysis: Distances,Preparing the data, Hierarchical clustering with hclust(), The k-means algorithm, Assigning new points to clusters.

UNIT - IV

Delivering Results - Documentation and deployment -The buzz dataset, Using knitr to produce milestone documentation , What is knitr? , knitr technical details. Using comments and version control for running documentation : Writing effective comments , Using version control to record history , Using version control to explore your project , Using version control to share work , Deploying models.

Text Books:
1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.

Reference Books:
1. Jure Leskovec, AnandRajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2014.
2. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.
3. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.
4. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, "Practical Data Science Cookbook", Packt Publishing Ltd., 2014.

MTCST204: INFORMATION SECURITY AND CRYPTOGRAPHY
(MTech CST)

Theory : 4 Hrs/week	**Credits**	: 4
Int Marks : 25	**Ext Marks**	: 75

UNIT – I

Introduction: The need for security, Security approaches, Principles of security, Types of Security attacks, Security Services & Mechanisms, A model for Network Security.

Concepts and Techniques: Cryptology-Cryptography & Cryptanalysis, Plain Text and Cipher Text, Substitution and Transposition Techniques, Encryption and Decryption, Stenography, key range and key size, types of attacks.

Intruders, virus and Firewalls: Intruders, Intrusion detection, password management, virus and related threats, Countermeasures, Firewall design principles, types of firewalls.

UNIT – II

Symmetric Key Cryptography: Modern Block Ciphers, Block vs Stream Cipher, Block Cipher Principles, Feistel Cipher Structure, Algorithm types and Modes, Overview of symmetric key cryptography, DES, Multiple Encryptions and DES- Double & Triple DES, AES, BLOWFISH, Differential and Linear Cryptanalysis.

Asymmetric Key Cryptography: Principles of Asymmetric Key Cryptography- RSA Algorithm, Diffie-Hellman Key Exchange Algorithm, Digital Signatures and Knapsack Algorithm.

UNIT – III

Public Key Infrastructure: Digital certificates, Private Key management, Public Key Cryptography Standards, XML, PKI and Security.

Message Authentication Algorithms and Hash Functions: Authentication requirements, Functions, Message authentication codes, Hash Functions, Secure hash algorithm, HMAC, and CMAC.

User Authentication Mechanisms: Authentication basics, Passwords, Authentication tokens, Certificate Based Authentication, Biometric Authentication and Kerberos.

UNIT – IV

E-Mail Security: Pretty Good Privacy, S/MIME.

IP Security: IP security overview, IP Security architecture, Authentication Header, Encapsulating Security Payload.

Web Security: Web security considerations, Secure Socket Layer and Transport Layer Security, Secure electronic transaction.

Case Studies on Cryptography and security: Secure Inter-branch Payment Transactions, Cross site Scripting Vulnerability, Virtual E lections.

Text Books:
1. Cryptography and Network security, Atul Kahate, Tata McGraw-Hill Pub Company Ltd New Delhi
2. Cryptography and Network Security: William Stallings, Pearson Education, 4th Edition.

Reference Books:
1. Network Security Private Communication in a public world, Charlie Kaufman, Radia Perlman & Mike Speciner, Prentice Hall of India Private Ltd., New Delhi.
2. Network Security: The Complete Reference by Roberta Bragg, Mark Phodes- Ousley, Keith Strassberg Tata McGraw-Hill .

MTCST205:ARTIFICIAL INTELLIGENCE
(MTech CST)

Theory	**: 4 Hrs/week**	**Credits**	**: 4**
Int Marks	**: 25**	**Ext Marks**	**: 75**

UNIT – I

Introduction to Artificial Intelligence: Artificial Intelligence, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics , Production Systems, , Production System Characteristics
Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And- Test, Hill Climbing, Best-First Search, A* Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis.

UNIT – II

Knowledge Representation: Procedural Vs Declarative Knowledge, Representations and Mappings, Approaches to Knowledge Representation, Issues in Knowledge Representation, Logic Programming Forward Vs Backward Reasoning,
Symbolic Logic: Propositional Logic, First Order Predicate Logic: Representing Instance and is-a Relationships, Computable Functions and Predicates, Syntax & Semantics of FOPL, Normal Forms, Unification &Resolution, Representation Using Rules, Natural Deduction.

UNIT – III

Structured Representations of Knowledge: Semantic Nets, Partitioned Semantic Nets, Frames, Conceptual Dependency, Conceptual Graphs, Scripts, CYC; Matching Techniques, Partial Matching, Fuzzy Matching Algorithms and RETE Matching Algorithms.
Reasoning under Uncertainty: Introduction to Non-Monotonic Reasoning, Truth Maintenance Systems, Logics for Non-Monotonic Reasoning, Model and Temporal Logics; Statistical Reasoning: Bayes Theorem, Certainty Factors and Rule-Based Systems, Bayesian Probabilistic Inference, Bayesian Networks, Dempster-Shafer Theory, Fuzzy Logic & Fuzzy Systems.

UNIT – IV

Experts Systems: Overview of an Expert System, Structure of an Expert Systems, Different Types of Expert Systems- Rule Based, Model Based, Case Based and Hybrid Expert Systems, Knowledge Acquisition and Validation Techniques, Black Board Architecture, Knowledge Building System Tools, Expert System Shells,
Natural Language Processing: Role of Knowledge in Language Understanding, Approaches Natural Language Understanding, Steps in The Natural Language Processing, Syntactic Processing and Augmented Transition Nets, Semantic Analysis, NLP Understanding Systems; Planning, Components of a Planning System, Goal Stack Planning, Hierarchical Planning, Reactive Systems

Text Book:
> 1.Artificial Intelligence, Elaine Rich, Mcgraw-Hill Publications

Reference Books:
> 1. Introduction To Artificial Intelligence & Expert Systems, Patterson, PHI
> 2. Artificial Intelligence, George F Luger, Pearson Education Publications
> 3. Artificial Intelligence, Robert Schalkoff, Mcgraw-Hill Publications
> 4. Artificial Intelligence And Intelligent Systems, N.P. Padhy, Oxford Publications

MTCST206:DIGITAL FORENSICS
(MTech CST)

Theory : 4 Hrs/week	**Credits** : 4
Int Marks : 25	**Ext Marks** : 75

UNIT - I

Computer Forensics and Investigations as a Profession: Understanding Computer Forensics- Preparing for Computer Investigations.

Understanding Computer Investigations: Preparing a Computer Investigation-Taking a Systematic Approach-Understanding Data Recovery Workstations and Software-Conducting an Investigation.

UNIT-II

Data Acquisition: Understanding Storage Formats for Digital Evidence-Determining the Best Acquisition Method-Contingency Planning for Image Acquisitions-Using Acquisition Tools-Validating Data Acquisitions-Using Remote Network Acquisition Tools.

Processing Crime and Incident Scenes: Identifying Digital Evidence-Collecting Evidence in Private-Sector Incident Scenes-Processing Law Enforcement Crime Scenes-Preparing for a Search- Securing a computer Incident or Crime Scene-Seizing Digital Evidence at the Scene-Storing Digital Evidence-Obtaining a Digital Hash-Reviewing a Case.

UNIT-III

Working with Windows and DOS Systems: Understanding File Systems-Exploring Microsoft File Structures-Examining NTFS Disks-Understanding Whole Disk Encryption-Understanding the Windows Registry-Understanding Microsoft startup Tasks-Understanding MS-DOS startup Tasks-Understanding Virtual Machine.

Current Computer Forensics Tools: Evaluating Computer Forensics Tool Needs-Computer Forensics Software Tools- Computer Forensics Hardware Tools- Validating and Testing Forensics Software.

Network Forensics: Network Forensics Overview-Securing a Network.

UNIT-IV

Computer Forensics Analysis and Validation: Determining What Data to Collect and Analyze-Validating Forensic Data-Addressing Data-Hiding Techniques- Performing Remote Acquisitions.

E-mail Investigations: Exploring the Role of E-mail in Investigations- Exploring the Roles of the Client and Server in E-mail- Investigating E-mail Crimes and Violations-Understanding E-mail Servers-Using Specialized E-mail Forensics Tools.

Cell Phone and Mobile Device Forensics: Understanding Mobile Device Forensics-Understanding Acquisition Procedures for Cell Phones and Mobile Devices.

Text Books:

1.Bill Nelson, Amelia Philips and Christopher Steuart, "Guide to computer forensics and investigation ", 3rd edition

2.Kevin Mandia, Chris Prosise, "Incident Response and computer forensics", Tata McGrawHill, 2006

Reference Books:

1. "Computer Forensics: Computer Crime Scene Investigation", JOHN R. VACCA, Firewall Media.

2. Peter Stephenson, "Investigating Computer Crime: A Handbook for Corporate Investigations", Sept 1999

3. Eoghan Casey, "Handbook Computer Crime Investigation's Forensic Tools and Technology", Academic Press, 1st Edition, 2001

4. Computer Forensics: Investigating Network Intrusions and Cyber Crime (Ec-Council Press Series: Computer Forensics), 2010

MTBDA204: DATA MINING AND SOCIAL NETWORKING
(MTech CST with Big Data Analytics)

Theory : 4 Hrs/week	**Credits**	: 4
Int Marks : 25	**Ext Marks**	: 75

UNIT - I
Introduction to Data Mining: Motivation and importance, What is Data Mining, Data Mining on What Kind of Data, Data Mining Functionalities, Major Issues in Data Mining.
Data Preprocessing: Why Preprocess The Data, Descriptive Data Summarization, Data Cleaning,Data Integration And Transformation, Data Reduction, Data Discretization And Concept Hierarchy Generation.

UNIT- II
Concept Description: Characterization and comparison What is Concept Description, Data Generalization by Attribute-Oriented Induction(AOI), AOI for Data Characterization, Efficient Implementation of AOI, AOI for Class comparisons.
Mining Frequent Patterns: Frequent Itemsets, Closed Itemsets And Association Rules, Frequent Itemset Mining Methods: Apriori Method, Generating Association Rules, Improving The Efficiency of Apriori, Mining Frequent Itemsets Using Vertical Data Format, Mining Closed Frequent Itemsets .
Introduction to Supervised and Unsupervised Classification: Support Vector Machines, Cluster Analysis, types of data in cluster analysis, clustering methods.

UNIT - III
Social Networks and Web Mining: Web Community and Social Network Analysis, Characteristics of Web Data ,Web Community , Social Networking.
Extracting and Analyzing Web Social Networks: Extracting Evolution of Web Community from a Series of Web Archive: Types of Changes, Evolution Metrics, Web Archives and Graphs, Evolution of Web Community Charts. Temporal Analysis on Semantic Graph using Three-Way Tensor Decomposition: Background, Algorithms, Examples of Formed Community . Analysis of Communities and Their Evolutions in Dynamic Networks: Motivation, Problem Formulation, Community Discovery Examples.

UNIT IV
Recommendation Systems: User-based and Item-based Collaborative Filtering Recommender Systems: User-based Collaborative Filtering, Item-based Collaborative Filtering Algorithm , Performance Evaluation.
Hybrid User-based and Item-based Web Recommendation System: Problem Domain, Hybrid User and Item-based Approach, Experimental Observations, User Profiling for Web Recommendation Algorithm Based on PLSA model, Recommendation Algorithm Based on LDA Model.

Text books:
1. Data Mining Concepts and Techniques Micheline Han and Kamber(Unit I and II)
2. Web Mining and Social Networking, G.Xu, Y.Zhang, L.Li(Unit III, IV)

Reference Books:
1. John Scott , "Social Network Analysis: A Handbook" , SAGE Publications Ltd; 2nd edition, 2000.
2. Borgatti, S. P., Everett, M. G., & Johnson, J. C., "Analyzing social networks", SAGE Publications Ltd; 1 edition, 2013.

MTBDA205:MACHINE LEARNING
(MTech CST with Big Data Analytics)

Theory : 4 Hrs/week	**Credits** : 4
Int Marks : 25	**Ext Marks** : 75

UNIT - I

Introduction - Well-posed learning problems, Designing a learning system, Perspectives and issues in machine learning Concept learning and the general to specific ordering – Introduction, A concept learning task, Concept earning as search, Find-S: finding a maximally specific hypothesis, Version spaces and the candidate elimination algorithm, Remarks on version spaces and candidate elimination, Inductive bias

UNIT - II

Decision Tree learning – Introduction, Decision tree representation, Appropriate problems for decision tree learning, The basic decision tree learning algorithm, Hypothesis space search in decision tree learning, Inductive bias in decision tree learning, Issues in decision tree learning

Artificial Neural Networks – Introduction, Neural network representation, Appropriate problems for neural network learning, Perceptions, Multilayer networks and the back propagation algorithm, Remarks on the back propagation algorithm, An illustrative example face recognition Advanced topics in artificial neural networks

UNIT - III

Bayesian learning – Introduction, Bayes theorem, Bayes theorem and concept learning, Maximum likelihood and least squared error hypotheses, Maximum likelihood hypotheses for predicting probabilities, Minimum description length principle, Bayes optimal classifier, Gibs algorithm, Naïve Bayes classifier, An example learning to classify text, Bayesian belief networks The EM algorithm

Instance-Based Learning- Introduction, k -Nearest Neighbour Learning, Locally Weighted Regression, Radial Basis Functions, Case-Based Reasoning, Remarks on Lazy and Eager Learning

UNIT - IV

Genetic Algorithms – Motivation, Genetic Algorithms, An illustrative Example, Hypothesis SpaceSearch, Genetic Programming, Models of Evolution and Learning, Parallelizing Genetic Algorithms

Learning Sets of Rules – Introduction, Sequential Covering Algorithms, Learning Rule Sets: Summary, Learning First Order Rules, Learning Sets of First Order Rules: FOIL, Induction as Inverted Deduction, Inverting Resolution

Text Books:
1. Machine Learning – Tom Mitchell, McGraw Hill Education; First edition (1 July 2017) **ISBN-10:**1259096955
2. Machine Learning: An Algorithmic Perspective, Second Edition, Stephen Marsland, Taylor & Francis (CRC) 2014. ISBN-13: 978-1-4665-8333-7 (eBook - PDF)

Reference Books:
1. Machine Learning Methods in the Environmental Sciences, Neural Networks, William WHsieh, Cambridge Univ Press. 2009. ISBN-13 978-0-511-59557-8
2. R ichard o. Duda, Peter E. Hart and David G. Stork, pattern classification, John Wiley & SonsInc., 2001. ISBN: 978-0-471-05669-0
3. Chris Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.ISBN-10: 0-387-31073-8
4. Machine Learning by Peter Flach, Cambridge. 2012. ISBN 978-1-107-09639-4

MTBDA206: STATISTICAL COMPUTING FOR DATA ANALYTICS
(MTech CST with Big Data Analytics)

Theory	: 4 Hrs/week	Credits	: 4
Int Marks	: 25	Ext Marks	: 75

UNIT- I

Data Analytics Life Cycle: Data Analytics Life Cycle Overview-Key Roles of a successful Analytics Project, Background and Overview of Data Analytics Life Cycle, Discovery, Data Preparation, Model Planning, Model Building, Communicate Results, Operationalize.

UNIT- II

Introduction to Statistics: Introduction, Data Collection and descriptive statistics, Inferential statistics and Probabilty models, Descriptive Statistics: Describing Data Sets, Summarising Data Sets, Quartiles and Deciles. Measures of Variations-Range, IQR, Quartile deviation, Mean deviation, Standard Deviation, Coefficient variance, Skewness, Moments and kurtosis.

UNIT- III

Probability, Variables and Special Random Variables: Axiomas of Probability, Conditional Probability, Bayes Formula, Independent Events, Random Variables, Types, Jointly Distributed RV,Variance, Covariance, Special RVs, Bernouli and Binomial RVs, Poison, Hypergeometric, Uniform, Normal, Exponential. Distributions-Normal, Gamma, Chi-Square, t, F and Logistics. Sampling Statistics-Central Limit Theorem, Sample Variance

UNIT- IV

Hypothesis Testing and Predictive Analytics: Testing for Attributes -Mean of Normal Population – One-tailed and two-tailed tests, F-test and Chi-Square test. Predictive Modeling and Analysis-Regression Analysis, Linear Regression, Polynomial Regression, Multiple Linear Regression, Multivariate Regression, Multicollinearity, ANOVA. Correlation Analysis-Rank Correlation coefficient, Multiple correlation, Least Square, Curve fitting and goodness of fit.

Text Books:
1. EMC Education Services, "Data Science and Big Data Analytics", Wiley Publications.(I Unit)
2. Sheldon M. Ross,"Introduction to Probability and Statistics for Engineers and Scientists", 3rd edition, Academic Press. (II, III, IV Units).
3. Alberto Cordoba, "Understanding the Predictive Analytics Lifecycle", Wiley, 2014. (IV Unit)
4. Eric Siegel, Thomas H. Davenport, "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die", Wiley, 2013

Reference Books:
1. Dalgaard, Peter, "Introductory statistics with R", Springer Science& Business Media.(IV Unit)
2. R. N. Prasad, Seema Acharya, "Fundamentals of Business Analytics", Wiley, 2015.
3. Brain S. Everitt, "A Handbook of Statistical Analysis Using R", Second Edition,

MTCST207, MTBDA207: DATA SCIENCE WITH R LAB
(Common for MTech CST and MTech CST with Big Data Analytics)

Lab	**: 3 Hrs/week**	**Credits**	**: 2**
Int Marks	**: 50**	**Ext Marks**	**: 50**

1. Learn Basics of R by creating Vectors, Lists, Arrays, Matrices, Data frames.

2. Write Programs using Control Structures in R – Sequence, Selection, Iteration

3. Load Data into R by various methods.

4. Demonstrate the visualization and graphics using visualization packages.

5. Explore data using Single Variables: Unimodal, Bimodal, Histograms, Density Plots, Bar charts

6. Explore data using two Variables: Line plots, Scatter Plots, smoothing cures, Bar charts

7. Generate Decision Tree using any Data Set and name the branches of the tree by showing the splits

8. Cluster Random data in a pane and calculate the distance between items in every pair of clusters

9. Write R Program to Implement Linear and Multiple Regression on 'mt cars' data set, to estimate the value of mpg variable

10. Generate a Markdown file using Knit – R with any example

References:
1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.
3. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.

MTCST208: INFORMATION SECURITY AND CRYPTOGRAPHY LAB
(MTech CST)

Lab	**: 3 Hrs/week**	**Credits**	**: 2**
Int Marks	**: 50**	**Ext Marks**	**: 50**

Note:
   i.   *Use C/C++/Java or equivalent tool to implement the following experiments.*
   ii. *A minimum of 7 experiments should be completed.*

1. Consider a file with composite data substitute the content and transpose the ciphers.
2. Implement encryption and decryption substitution technique using Modified Caesar-Cipher.
3. Implement One Time Pad Cipher algorithm.
4. Implement Rail Fence transposition technique.
5. Consider an alphanumeric data, Encrypt and Decrypt the data using Advance Encryption Standards and verify for the correctness.
6. Apply the RSA algorithm on a text file to produce cipher text file.
7. Develop a mechanism to setup a security channel using Diffie-Hellman Key Exchange between client and server.
8. Implementation of Message Authentication Code using cryptography VMAC function.
9. Calculate the message digest of a text using the MD5/SHA-1 algorithm .
10. Generate digital signature using RSA & MD5/SHA-1.

References:
1. Cryptography and Network security, Atul Kahate, Tata McGraw-Hill Pub company Ltd., New Delhi

Lab	: 3 Hrs/week	Credits	: 2
Int Marks	: 50	Ext Marks	: 50

1. Write a Java Program to implement Linked Lists, Stacks and Queues.
2. Write Java Program that implements Generic Types which collects pair of elements of different types.
3. Write a Java Program that uses object serialization and deserialization.
4. Know about setting up and Installing Hadoop in its three operating modes and implement in Standalone.
5. Implement the following file management tasks in Hadoop: Adding, Retrieving and deleting files.
   Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere
and copies them into HDFS using one of the above command line utilities.
6. Write a Map-Reduce Program to find average of numbers.
7. Implement Matrix Multiplication with Hadoop Map Reduce
8. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

References:

1. Big Java Fourth Edition Cay Horstmann Wiley, John Wiley & Sons
2. www.hadoop.apache.org
3. www.gist.github.com