



# SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

## Syllabus

### M.Tech (Data Science) I Semester

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Statistics Foundation of Data Science	MTDS -101	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction:** statistics, Theory of probability, random variables. Distributions: frequency distribution, theoretical distributions, chi-square test or  $\chi^2$  distribution, t-distribution, f-distribution.

**Unit-II Statistical Inferences and Estimations:** measure of central tendency, measure of dispersion, Skewness, moments, kurtosis, statistical estimation, point estimates, interval estimates.

**Unit-III Testing of Hypothesis:** types of errors, one-tailed and two tailed tests, test of significance based on  $\chi^2$ , t, f, and z.

**Unit-IV Analysis of Variance and Covariance:** Analysis of Variance (ANOVA), one-way, two-way, three-way, analysis of variance, ANOVA Techniques, Analysis of Co-variance (ANOCOVA), ANOCOVA Techniques.

**Unit-V Correlation and Regression:** Karl Pearson 's coefficient of correlation, regression analysis, simple linear regression, regression using matrices, multiple linear regression, polynomial regression, non-linear regression, correlation analysis.

#### References:

1. Robert M. Bethea, Benjamin S. Duran, Thomas L. Boullion, Statistical Methods for engineers and scientists Marcel dekker, inc. New York .
2. Ray, Sharma, Chaudhary, Mathematical Statistics, Ram Prasad & Sons.
3. Robert v. Hogg, Allen T. Craig, Introduction to mathematical Statistics, Pearson Education.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Advance DBMS	MTDS-102	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I DBMS Concepts:** Data models, Entities and attributes, Relationships, E-R diagram. Relational Data models: Domains, Tuples, Attributes, Keys, Relational database, Schemas, Integrity constraints. Relational algebra and relational calculus, Normalization, Normal forms.

**Unit-II Hash-Based Indexing:** Static hashing, Extendible hashing, Linear hashing, comparisons, Query Processing and Optimization. Distributed databases: client/server database Fragmentation, Replication, Location & Fragment transparency, Distributed Query Processing and Optimization.

**Unit-III Database Protection:** Integrity, Constraints in Query-by-Example, Security, Security in query-by-Example, Security in Statistical Databases. Concurrent Operations on the Database: Basic Concepts, A simple Transaction Model, Model with Read and Write-Locks, Read-only, Write-only Model, Concurrency for Hierarchically Structured Items, Protection against Crashes, Optimistic Concurrency Control.

**Unit-IV Object oriented and object relational databases:** Specialization, Generalization, and Aggregation. Association: Introduction to Image and Multimedia databases and data structures. Data structure- R tree, K d tree, Quad trees, Content based retrieval: Color Histograms.

**Unit-V Web databases:** Accessing databases through web.

### References:

1. R. Elmasri, S. Navathe, “Fundamentals of Database System”, Pearson Education.
2. C.J. Date, “An Introduction to Data base Systems”, Volume I, Pearson Education.
3. SK Singh, “Database Systems”, Pearson Education.
4. H. F. Korth and A. Silberschatz. “Database Concept”, TMH.
5. Narang, “Object Oriented Databases”, Prentice-Hall of India.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Data Structure & Algorithm	MTDS-103	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Fundamentals:** Properties of Big-oh Notation, Conditional Asymptotic Notation, Algorithm Analysis, Amortized Analysis, Introduction to NP-Completeness/NP-Hard, Recurrence Equations, Solving Recurrence Equations, Time-Space Tradeoff.

**Unit-II Search Structures:** Binary Search Trees, AVL Trees, Red-Black trees, Multi-way Search Trees, B-Trees, Splay Trees, Tries.

**Unit-III Heap Structures:** Min/Max heaps, Heaps, Leftist Heaps, Binomial Heaps, Fibonacci Heaps, Skew Heaps, Lazy Binomial Heaps.

**Unit-IV Geometrical Algorithms:** Segment Trees, 1-Dimensional Range Searching, k-d Trees, Line Segment Intersection, Computing the Overlay of Two Subdivisions, Range Trees, Voronoi Diagram.

**Unit-V Additional Topics:** Approximation Algorithms, Vertex Cover & Euclidean, Travelling Salesperson Problem, Randomized Algorithms: Closest Pair Problem & Minimum Spanning Trees, Online Algorithm: Euclidean Spanning Tree.

### References:

1. Horowitz, S. Sahni, "Fundamentals of Computer Algorithms", Galgotia Publishers.
2. D.E. Knuth, "The Art of Computer Programming", Vols. 1 and 3, Pearson Education.
3. V Aho, J E Hopcroft, J D Ullman, "Design & Analysis of Algorithms", Pearson Education.
4. K. Mehlhorn, "Data Structures and Algorithms", Vols. 1 and 2, Springer Verlag.
5. Purdom, Jr. and C.A. Brown, "Analyses of Algorithms", Holt Rinechart and Winston.
6. Parag H Dave, Himanshu B Dave, "Design and Analysis of Algorithms", Pearson Education.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	AI Techniques & Soft Computing	MTDS-104	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction to AI:** AI Techniques, Applications, Evolution of Computing, Soft Computing Constituents From Conventional AI to Computational Intelligence. Machine Learning Basics, Artificial Intelligence Planning: Classical Planning, Partial Order Planning, Graph Plan and SAT Plan, Hierarchical Planning, Planning and Acting in Nondeterministic Domains, Multiagent Planning.

**Unit-II Reasoning with Lower Order Logics:** Logical Agent, Proposition Logic, Syntax and Semantics Theorem Proving, Model Checking, Inference in First Order Logic, Learning Techniques: Logical Formulation of Learning, Knowledge in Learning, Explanation Based Learning, Learning using Relevance Information, Inductive Logic Programming, Statistical Learning, Learning with Complete Data, Learning with Hidden Data, Applications.

**Unit-III Search Techniques:** Problem Solving by Search, Uninformed Search, Searching with Costs, Informed State Space Search, Heuristic Search: DFS, BFS, Heuristic Search, Problem Reduction Search, Game Search, Constraint Satisfaction Problems.

**Unit-IV Intelligent Agents and Knowledge Representation:** Agents and Environments, Good Behavior: The concepts of Rationality, The Nature of Environments, The Structure of Agents, Knowledge Representation, Object Oriented Approach, Semantic Nets, Frames, Semantic Web, Ontology.

**Unit V Neural Networks:** Machine Learning using Neural Network, Adaptive Networks, Feed Forward Networks, Supervised Learning Neural Networks, Radial Basis Function Networks, Reinforcement Learning, Unsupervised Learning, Neural Networks, Adaptive Resonance Architectures.

### References:

1. Stuart J. Russell, Peter Norvig, “Artificial Intelligence –A Modern Approach”, Third Edition, Pearson Education.
2. Elaine Rich, Kevin Knight, Shivashankar B. Nair, “Artificial Intelligence”, Third Edition, Tata McGraw-Hill.
3. Dheepak Khemani, “A First Course in Artificial Intelligence”, McGraw-Hill.
4. NPTEL Artificial Intelligence Course by Prof. Dasgupta <http://nptel.ac.in/courses/106105079/25>. Sebastian Thrun, Peter Norvig, Udacity: Introduction to Artificial Intelligence, <https://in.udacity.com/course/intro-to-artificial-intelligence—cs271>
5. KwangH.Lee, “First course on Fuzzy Theory and Applications”, Springer.
6. George J. Klir and Bo Yuan, “Fuzzy Sets and Fuzzy Logic-Theory and Applications”, Prentice Hall.
7. James A. Freeman and David M. Skapura, “Neural Networks Algorithms, Applications, and Programming Techniques”, Addison Wesley.
8. David E.Goldberg, “Genetic Algorithms in Search, Optimization and Machine Learning”, Addison Wesley.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Data Visualization	MTDS-105	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction:** Context of data visualization –Definition, Methodology, Visualization design objectives. Key Factors, Purpose, visualization function and tone, Visualization design options, Data representation, Data Presentation, Seven stages of data visualization, widgets, and Data visualization tools.

**Unit-II Visualizing Data Methods:** Mapping, Time series, Connections and correlations, Indicator, Area chart, Pivot table, Scatter charts, Scatter maps, Tree maps, Space filling and non-space filling methods-Hierarchies and Recursion, Networks and Graphs, Displaying Arbitrary Graphs, Node link graph, Matrix representation for graphs, Info graphics

**Unit-III Visualizing Data Process:** Acquiring data, Where to Find Data, Tools for Acquiring Data from the Internet, Locating Files for Use with Processing, Loading Text Data, Dealing with Files and Folders, Listing Files in a Folder, Asynchronous Image Downloads, Advanced Web Techniques, Using a Database, Dealing with a Large Number of Files. Parsing data, Levels of Effort, Tools for Gathering Clues, Text Is Best, Text Markup Languages, Regular Expressions (regexps), Grammars and BNF Notation, Compressed Data, Vectors and Geometry, Binary Data Formats, Advanced Detective Work.

**Unit-IV Interactive Data Visualization:** Drawing with data, Scales, Axes, Updates, Transition and Motion, Interactivity, Layouts, Geo mapping, Exporting, Framework-T3, .js, tableau.

**Unit-V Security Data Visualization:** Port scan visualization, Vulnerability assessment and exploitation, Firewall log visualization, Intrusion detection log visualization, Attacking and defending visualization systems, Creating security visualization system.

### References:

1. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly.
2. Edward R. Tufte, "The visual display of quantitative information", Second Edition, Graphics Press.
3. Evan Stubbs, "The value of business analytics: Identifying the path to profitability", Wiley.
4. Gert H. N. Laursen and Jesper Thorlund, "Business Analytics for Managers: Taking business intelligence beyond reporting", Wiley.
5. Nathan Yau, "Data Points: Visualization that means something", Wiley.
6. Stephen Few, "Information dashboard design: Displaying data for at-a-glance monitoring", second edition, Analytics Press.
7. Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press.
8. Tamara Munzner, "Visualization Analysis and Design", AK Peters Visualization Series, CRC Press.



# SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

## Syllabus

### M.Tech (Data Science) II Semester

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Data Mining & Warehousing	MTDS-201	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction:** Data Mining: Definitions, KDD v/s Data Mining, DBMS v/s Data Mining, DM techniques, Mining problems, Issues and Challenges in DM, DM Application areas.

**Unit-II Association Rules & Clustering Techniques:** Introduction, Various association algorithms like A Priori, Partition, Pincer search etc., Generalized association rules. Clustering paradigms; Partitioning algorithms like K-Medoid, CLARA, CLARANS; Hierarchical clustering, DBSCAN, BIRCH, CURE; categorical clustering algorithms, STIRR, ROCK, CACTUS.

**Unit-III Other DM techniques & Web Mining:** Application of Neural Network, AI, Fuzzy logic and Genetic algorithm, Decision tree in DM. Web Mining, Web content mining, Web structure Mining, Web Usage Mining. Temporal and spatial DM: Temporal association rules, Sequence Mining, GSP, SPADE, SPIRIT, and WUM algorithms, Episode Discovery, Event prediction, Time series analysis.

**Unit-IV Spatial Mining:** Spatial Mining tasks, Spatial clustering, Spatial Trends. Data Mining of Image and Video: A case study. Image and Video representation techniques, feature extraction, motion analysis, content based image and video retrieval, clustering and association paradigm, knowledge discovery.

**Unit-V The vicious cycle of Data mining:** Data mining methodology, Measuring the effectiveness of data mining, Data mining techniques. Market baskets analysis, Memory based reasoning, Automatic cluster detection, Link analysis, Artificial Neural Networks, Generic algorithms, Data mining and corporate data warehouse, OLA.

#### References:

1. Arun K. Pujari, "Data Mining Techniques", University Press.
2. Adriaans & Zantinge, "Data Mining", Pearson education.
3. Berry Linoff, "Mastering Data Mining", Wiley.
4. Dunham, "Data Mining", Pearson Education.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Machine Learning	MTDS-202	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction:** Machine Learning, Types of Machine Learning, Machine Learning process-preliminaries, Testing Machine Learning algorithms, Turning data into Probabilities, and Statistics for Machine Learning, Probability theory, Probability Distributions, Decision Theory.

**Unit-II Supervised Learning:** Linear Models for Regression, Linear Models for Classification, Discriminant Functions, Probabilistic Generative Models, Probabilistic Discriminative Models, Decision Tree Learning, Bayesian Learning, Naïve Bayes, Ensemble Methods, Bagging, Boosting, Neural Networks , Multi-layer Perceptron, Feed-forward Network, Error Back propagation, Support Vector Machines.

**Unit-III Unsupervised Learning:** Clustering-K-means, EM Algorithm, Mixtures of Gaussians, Dimensionality Reduction, Linear Discriminant Analysis, Factor Analysis, Principal Components Analysis, Independent Components Analysis.

**Unit-IV Probabilistic Graphical Models:** Graphical Models, Undirected Graphical Models, Markov Random Fields, Directed Graphical Models, Bayesian Networks, Conditional Independence properties, Markov Random Fields, Hidden Markov Models, Conditional Random Fields (CRFs).

**Unit-V Advanced Learning:** Sampling-Basic Sampling methods, Monte Carlo, Gibbs Sampling, Computational Learning Theory, Mistake Bound Analysis, Reinforcement learning, Markov Decision processes, Deterministic and Non-deterministic Rewards and Actions, Temporal Difference Learning Exploration.

### References:

1. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer.
2. Stephen Marshland, "Machine Learning –An Algorithmic Perspective", Chapman and Hall, CRC Press, Second Edition.
3. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press.
4. Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, Third Edition.
5. Tom Mitchell, "Machine Learning", McGraw-Hill.





## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Information Retrieval	MTDS-203	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction to Information Retrieval:** The nature of unstructured and semi-structured text, Inverted index and Boolean queries.

**Unit-II Text Indexing, Storage and Compression:** Text encoding: tokenization, stemming, stop words, phrases, index optimization. Index compression: lexicon compression and postings, lists compression, Gap encoding, gamma codes, Zipf's Law, Index construction, Postings size estimation, merge sort, dynamic indexing, positional indexes, n-gram indexes, real-world issues.

**Unit-III Retrieval Models:** Boolean, vector space, TFIDF, Okapi, probabilistic, language modeling, latent semantic indexing. Vector space scoring, The cosine measure, Efficiency considerations, Document length normalization, Relevance feedback and query expansion, Rocchio.

**Unit-IV Text Clustering:** Clustering versus classification, Partitioning methods, k-means clustering, Mixture of Gaussian model, Hierarchical agglomerative clustering, Clustering terms using documents

**Unit-V Web Information Retrieval:** Hypertext, web crawling, search engine, ranking, link analysis, Page Rank, HITS, XML and Semantic web.

### References:

1. Manning, Raghavan and Schutze, "Introduction to Information Retrieval", Cambridge University Press.
2. Baze-Yates and Ribeiro-Neto, "Modern Information Retrieval", Addison-Wesley.
3. Soumen Charabharti, "Mining the Web-Discovering Knowledge from Hypertext Data", Morgan-Kaufman Publishers.
4. Pierre Baldi, Paolo Frasconi and Padhraic Smythe, "Modelling the internet and the web: Probabilistic methods and Algorithms", John Wiley.
5. Gerald J.Kowalski, Mark T.Maybury, "Information storage and Retrieval systems: theory and implementation", Kluwer Academic Publishers.





## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Data Analytics	MTDS-204	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Introduction:** Limitations of predictive analytics, Opportunities for predictive analytics, Key roles for successful analytic project, Main phases of life cycle, Developing core deliverables for stakeholders.

**Unit-II Predictive Modeling Techniques:** Predictive analytics in general, supervised learning, Data validation and cleaning, Data recording, Missing data, the issue of over fitting. Linear regression: Interpreting the coefficients of a linear model.

**Unit-III Classification Techniques:** Naïve Bayesian Classifier, K-means Clustering, Association Rules, Decision Trees, Linear and Logistic Regression, Time Series Analysis, Text Analytics; ,Classification with a rare class, Thresholds and lift, ranking examples, Training to overcome imbalance, a case study in classification with a rare class.

**Unit-IV Rapid Miner:** Standardization of features, Example of a Rapid miner process, other notes on Rapid miner. Support vector machines: Loss functions, Regularization, Linear soft-margin SVMs, Dual formulation, Nonlinear kernels, Radial basis function kernels, Selecting the best SVM settings, Doing valid experiments: Cross-validation, Cross-validation for model selection,

**Unit-V Technology and Tools:** Introduction to “R”, analyzing and exploring data with “R”, statistics for model building and evaluation. Map Reduce /Hadoop, In- database Analytics, MAD lib and advanced SQL Tools.

### References:

1. Vignesh Prajapati, “Big data analytics with R and Hadoop”, Packt Publication.
2. Noreen Burlingame, “The little book on Big Data”, New Street publisher.
3. Norman Matloff , “The Art of R Programming: A Tour of Statistical Software Desig”, No Starch Press.
4. [http://www.johndcook.com/R\\_language\\_for\\_programmers.html](http://www.johndcook.com/R_language_for_programmers.html)
5. <http://bigdatauniversity.com/>
6. <http://home.ubalt.edu/ntsbarsh/stat-data/topics.htm#rintroduction>.



## SCHOOL OF INFORMATION TECHNOLOGY, RGPV BHOPAL

Category of Course	Course Title	Course Code	Credits			Theory Papers
			L	T	P	
M.Tech	Cyber Law & IPR	MTDS-205	3	1	-	Max. Marks – 70 Min Marks – 28 Duration: 3 Hrs.

**Unit-I Information Technology:** Understanding the Basic concepts, Introduction about the cyberspace, Cyber world: an overview, internet and online resources, regulation of cyber space, Historical background of Cyber law, UNCITRAL Model law.

**Unit-II IT Act:** Evolution of the IT Act 2000, Genesis and Necessity, Nature, Scope and Importance of IT Act, cyber law in India with special reference to Information Technology Act, 2000, Introduction to computer and cyber-crimes, Cyber-crimes and related concepts, distinction between cyber-crimes and conventional crimes, Cyber criminals and their objectives. Kinds of cyber-crimes, cyber stalking; cyber pornography, forgery and fraud, crime related to IPRs, cyber terrorism; computer vandalism etc, Scope of cyber laws: e-commerce; online contracts etc. Salient features of the IT Act 2000, Electronic records and Digital Signature, Regulation of Certifying Authorities.

**Unit-III Cyber Crime and Issues:** Regulation of cyber-crimes, Issues relating to investigation, issues relating to jurisdiction, issues relating to evidence, relevant provisions under Information Technology Act 2000, Indian penal code, IPRs (copyright, trademarks and software patenting), e-taxation, e-governance and cyber-crimes, security of information, digital signature, intellectual property (IP).

**Unit-IV Introduction to IPR:** Historical background of IP, IPR governance, National patent offices, the world intellectual property organization (WIPO), International Conventions on Copyright, Berne Convention, WIPO Treaty, Scope of Copyright protection in the digital environment, Concept of Trademarks in Internet Era, Jurisdiction in Trademark Disputes, Protecting Trademarks in Digital Environment, International Conventions on Patents, Provisions of Patent Act 1970 in relation to cyber world, Procedure of Patenting relating to digital technology, Types of IPR and its relevant acts.

**Unit-V IPR Issues:** Challenges, Settlement of Disputes, Uniform Dispute Resolution Policy, Legal framework: National & International level, ICANN dispute resolution policy, legal position on database protection, protection of multimedia works in cyberspace, copyright infringement, study of landmark cases.

### Reference:

1. Harish chander, “Cyber laws & IT Protection”, PHI.
2. Vinod V. Sople, “Managing Intellectual Property” PHI Learning Private Limited.
3. Duggal, Pavan (2014) Legal Framework on Electronic Commerce and Intellectual Property Rights in Cyberspace, Universal Law Publishing - An imprint of Lexis Nexis.