

National Institute of Technology

Department of Computer Science and Engineering

Course Structure of M. Tech Computer Science and Engineering (Analytics)

Distribution of Credits

Programme Core (PC)	Programme Elective (PE)	Total Credits
57	15	72

Semester-wise Distribution of Credits

Semester I	Semester II	Semester III	Semester IV	Total Credits
18	22	16	16	72

Semester-wise Scheduling of Courses

SEMESTER – I

Course No.	Course Name	L-T-P-C
CS-5xx	Core-1	3-0-0-3
CS-5xx	Core-2	3-0-0-3
CS-5xx	Core-3	3-0-0-3
CS-5xx	Elective-1	3-0-0-3
CS-5xx	Elective-2	3-0-0-3
CS-5xx	Core Lab-I	0-0-6-3
	Total	15-0-6-18

SEMESTER – II

Course No.	Course Name	L-T-P-C
CS-5xx	Core-4	3-0-0-3
CS-5xx	Core-5	3-0-0-3
CS-5xx	Elective-3	3-0-0-3
CS-5xx	Elective-4	3-0-0-3
CS-5xx	Elective-5	3-0-0-3
CS-5xx	Core Lab-II	0-0-6-3
CS-516	Minor Project	0-0-8-4
	Total	15-0-14-22

SEMESTER –III

Course No.	Course Name	L-T-P-C
CS-600A	Dissertation-I	0-0-0-16
	Total	0-0-0-16

SEMESTER – IV

Course No.	Course Name	L-T-P-C
CS-600B	Dissertation-II	0-0-0-16
	Total	0-0-0-16

List of Core Subjects and Labs**L-T-P-C**

Course No.	Course Name	L-T-P-C
CS-511	Advanced Data Mining and Databases	3-0-0-3
CS-513	Understanding and Preparing Data for Analytics	3-0-0-3
CS-515	Understanding Various Modeling Techniques	3-0-0-3
CS-517	Applications to Business Problems	3-0-0-3
CS-519	Quantitative Techniques	3-0-0-3
CS-531	Data Mining Lab	0-0-6-3
CS-533	Quantitative Techniques Lab	0-0-6-3
CS-516	Minor Project	0-0-8-4
CS-600A	Dissertation-I	0-0-0-16
CS-600B	Dissertation-II	0-0-0-16

List of Elective Subjects**L-T-P-C**

Course No.	Course Name	L-T-P-C
CS-521	Structuring Problems for Analytics	3-0-0-3
CS-523	Statistical Methods for Data Mining	3-0-0-3
CS-525	Soft Computing	3-0-0-3
CS-527	Simulation and Modeling	3-0-0-3
CS-529	Information Retrieval Systems	3-0-0-3
CS-541	Deployment Strategies	3-0-0-3
CS-543	Natural Language Processing	3-0-0-3
CS-545	Machine Learning	3-0-0-3
CS-547	Operations Research	3-0-0-3
CS-549	Probability and Queuing Theory	3-0-0-3
CS-551	Neural Networks	3-0-0-3

Core Subjects

CS-511 **Advanced Data Mining and Databases**

3-0-0-3

- **Introduction:** Data Mining Concepts, Input, Instances, Attributes and Output, Knowledge Representation & Review of Graph Theory, Lattices, Probability & Statistics
- **Machine learning concepts and approaches:** Supervised Learning Framework, concepts & hypothesis, Training & Learning, Boolean functions and formulae, Monomials, Disjunctive Normal Form & Conjunctive Normal Form, A learning algorithm for monomials
- **Data Preparation:** Data Cleaning, Data Integration & Transformation, Data Reduction
- **Mining Association Rules:** Associations, Maximal Frequent & Closed Frequent item sets, Covering Algorithms & Association Rules, Linear Models & Instance-Based Learning, Mining Association Rules from Transactional databases, Mining Association Rules from Relational databases & Warehouses, Correlation analysis & Constraint-based Association Mining.
- **Classification and Prediction:** Issues regarding Classification & Prediction, Classification by Decision Tree induction, Bayesian classification, Classification by Back Propagation, k-Nearest Neighbor Classifiers, Genetic algorithms, Rough Set & Fuzzy Set approaches
- **Cluster Analysis:** Types of data in Clustering Analysis, Categorization of Major Clustering methods, Hierarchical methods, Density-based methods, Grid-based methods, Model-based Clustering methods
- **Mining Complex Types of Data:** Multidimensional analysis & Descriptive mining of Complex data objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Time-series & Sequence data, Mining Text databases, Mining World -Wide Web
- **Data Mining Applications and Trends in Data Mining:** Massive Datasets/Text mining, Agent-Based Mining

With respect to data sources:

- Identify the assumptions and subjectivity inherent in data and how it influences analytic insights.
- Describe and contrast the features and characteristics of data types (textual, numeric, visual, etc.).
- Define and describe the data life cycle; associate different reporting techniques with the life cycle of data.
- Proactively assess the quality of data using data profiling techniques.
- Understand the importance of data management and governance and to understand the different roles associated with management of data.
 - ✓ Understand the legal, ethical, organizational, structural and procedural imperatives related to data collection, storage, and use.
 - ✓ Understand the legal as well as cost/benefit implications associated with managing the life cycle of the data.

With respect to data preparation:

- Define the dimensions and characteristics of data quality in each stage of the analytical process.
 - ✓ Prescribe and implement strategies for dealing with missing (null) values including techniques for imputing values.
 - ✓ Prescribe and implement strategies for dealing with outliers, extreme values, and complex distributions of values.
- Integrate data from multiple sources and understand ETL tools and concepts.
- Determine whether or not sampling and partitioning of data sets is needed and to specify appropriate sampling methods and sample sizes.
- Apply data cleansing techniques, including correcting and filtering bad data from the data set, validating codes against lists of acceptable values.
- Balance and weight data as needed to increase representativeness.
- Determine and manage sparsely populated, high dimensional data sets.
 - ✓ Prescribe and implement appropriate data reduction techniques for both continuous and categorical variables.
- Understand the relationships between business problems and data requirements.
 - ✓ Specify appropriate atomicity of data collection, storage, and use.
 - ✓ Be able to articulate the level of data quality appropriate and sufficient to the business problem.

- Understand the relationships between data requirements and analytic models.
 - ✓ Distinguish data structures (one row per subject, multiple rows per subject, etc.) appropriate for specific analytic techniques.
 - ✓ Understand how to analyze hierarchical variable structures.
 - ✓ Understand the implications of null values, outliers, measurement levels, for various analytic techniques.
 - ✓ Understand the characteristics and idiosyncrasies associated with data from different specific sources (e.g., the web, historical time series, purchase patterns, text data, recommendation data, etc.).

With respect to querying:

- Perform query-based data extracts from relational databases to flat files.
- Design and execute high-level queries for direct access to databases.
 - ✓ Specify SQL queries with different types of joins.
 - ✓ Become well versed in tools such as Oracle 11g, Microsoft SQL Server, etc.
 - ✓ Perform slice and dice, drill down and roll up, and pivot operations.
 - ✓ Become well versed with analytic query languages such as MDX.
- Specify techniques for obtaining data from web sources, footfall analysis, social media, email, etc.

Predictive Analytics and Data Mining:

- Construct, validate, and interpret data mining and predictive analytics models using large multivariate data sets.
- Apply data mining and predictive analytic techniques to problems in fund raising, retailing, direct marketing, market segmentation, bankruptcy prediction, credit scoring, and fraud detection.
- Perform data exploration to evaluate variables for data mining and to suggest and implement approaches to handle data problems such as missing values, outliers, and skewed distributions.
- Compute and interpret key predictive accuracy measures and methods including lift charts, gain charts, and ROC curves (with both symmetric and asymmetric costs of outcomes).
- Understand and utilized techniques for problems with continuous dependent variables
 - ✓ Understand and apply ordinary multiple regression, Poisson regression, and negative binomial regression.
 - ✓ Understand and apply neural networks.
 - ✓ Understand and apply regression trees (CART).
 - ✓ Understand and apply survival analysis (Cox regression).
 - ✓ Understand the assumptions, limitations, advantages and disadvantages of each of these techniques.
- Understand and utilized techniques for problems with binary and categorical dependent variables
 - ✓ Understand and apply multiple logistic regression
 - ✓ Understand and apply discriminant analysis
 - ✓ Understand and apply artificial neural networks and perceptrons.
 - ✓ Understand and apply decision tree models (CHAID, C4.5, and others)
 - ✓ Understand and apply support vector machine algorithms.
 - ✓ Understand and apply naïve Bayesian classifiers.
 - ✓ Understand and apply k-nearest neighbor models.
 - ✓ Be able to describe the assumptions, limitations, advantages and disadvantages of each of these techniques.
- Understand and utilize techniques for unsupervised modeling situations
 - ✓ Understand and apply principal components analysis.
 - ✓ Understand and apply hierarchical and non-hierarchical (k-means) cluster analysis.
 - ✓ Understand and apply self-organizing (Kohonen networks).

- ✓ Understand and apply market basket analysis.
- ✓ Understand and apply methods for anomaly detection.
- ✓ Understand and apply methods for recommendation systems.
- ✓ Understand the assumptions, limitations, advantages and disadvantages of each of these techniques.
- Understand and apply natural language techniques (e.g., text mining).
- Understand and apply techniques for model enhancement including boosting, bagging, and ensemble methods.
- Perform complete start to finish analyses including problem definition, data cleaning, variable selection, model selection, model building, testing and validation, and interpretation of results.

Simulation:

- Describe and explain the role of simulation in identifying, quantifying, and estimating the benefits of decision models.
- Design and parameterize discrete event Monte Carlo and Markov Chain Monte Carlo computer simulations of business problems by
 - ✓ Specifying and evaluating data sources;
 - ✓ Fitting probability distributions to input variables;
 - ✓ Creating and running simulations using commercial software packages.
- Design formal experiments in the context of simulation models.
- Analyze simulation outputs using statistical tools and methods

Forecasting:

- Select the appropriate forecasting technique/method based upon key criteria such as:
 - ✓ Forecast time horizon
 - ✓ Amount and relevance of historical data
 - ✓ Data patterns (seasonality, trends)
 - ✓ Accuracy requirements and purpose of the forecast
 - ✓ Assessment of the potential for forecasting – can the variable be forecasted?
- Understand the organizational and decision making context of forecasting
- Understand the basic principles for each of the following and be able to apply each using appropriate software:
 - ✓ Judgmental models such as Delphi, using expert opinion, consumer intentions and expectations, judgmental bootstrapping, analogies, prediction markets
 - ✓ Extrapolation and growth models (e.g., the Bass model)

- ✓ Time series models such as smoothing (Holt-Winters, multiplicative and additive), ARIMA,
- ✓ Causal models (e.g., ordinary regression, econometric models)
- ✓ Crowd sourcing and prediction markets
- Evaluate and improve forecast accuracy through consideration of:
 - ✓ Uncertainty assessment and estimation of forecast error
 - ✓ Assessment and control of bias
 - ✓ Multi-method approaches and triangulation
 - ✓ Techniques such decomposition, segmentation, and causal chains
 - ✓ The incorporation of subjective judgments and adjustments
 - ✓ Systems for tracking, monitoring and learning

Optimization models:

- Understand and apply basic linear optimization models, including:
 - ✓ Multi-Period Models
 - ✓ Blending Models
 - ✓ Perform and interpret sensitivity analysis
 - ✓ Define and explain the role of hill climbing algorithms, derivatives, and gradients
 - ✓ Define and construct constraints on models, including:
 - Global versus local constraints
 - Linear versus non-linear constraints
 - ✓ Define and explain the role of Lagrange multipliers
 - ✓ Define and explain the Efficient Frontier
 - ✓ Understand practical issues such as:
 - Solution times
 - Dealing with hard problems (how large is too large?)
- Understand and apply network and integer optimization models, including the following:
 - ✓ The special structure of network models
 - ✓ Applications to facility location models
 - ✓ Applications to set covering models
- Understand and apply non-linear optimization models, including:
 - ✓ Convex and concave functions
 - ✓ Global optimal solution search techniques
 - ✓ Portfolio optimization
- Understand the nature and application of genetic search algorithms in artificial intelligence, including the following:

- ✓ Encoding schemes
- ✓ Convergence criteria
- ✓ Application to location problems

Web Analytics:

- Understand managerial issues related to web analytics:
 - ✓ Leveraging benchmarks and goals for web analytics to create executive dashboards
 - ✓ Being cognizant of legal, ethical and privacy issues in the use of web trails
- Understand and employ different types of data used in web analytics, e.g., server logs, visitor's data, search engine data.
- Understand and explain the issues related to clickstream data quality and the implications for applications.
- Understand key terms and terminology as well as different types of web-data based metrics that can be tracked, e.g., visitors, session, page views, hits, session summary, referrals, most popular search engine, search terms by engine, keyword searches, top entrance pages, top exit pages, bounce rate, length and depth of session.
- Perform both internal site search analytics and search engine optimization (SEO).
- Employ web analytics for:
 - ✓ Personalization and recommender systems.
 - ✓ Competitive intelligence.
- Become familiar with the capabilities and limitations of currently available web analytics tools.

Social Media and Online Marketing Analytics:

- Understand how to use web analytics to optimize paid search campaigns and affiliate programs
- Learn how to achieve greater return on investment for online marketing initiatives and evaluate the impact of campaign results
- Optimize display ad campaigns and measure offline drive to websites.
- Understand web-based advertising pricing (auction) models.
- Identify the critical metrics to assess and improve campaigns including methodologies for tracking the impact of online channels such as e-mail, paid search (pay-per-click), organic search (higher page ranking in search engines), and display/banner ads.
- With respect to brand tracking methods, understand what can be tracked and measured for online campaigns.
- Analyze display advertising, rich media, brand and buzz.
- Examine email campaign effectiveness from delivery to website conversion.

Supply Chain Analytics:

- Understand how to improve plan replenishment by being able to:
 - ✓ Set the basis for rationalizing the supplier set.
 - ✓ Define forecast procurement performance indicators and cycle times.
 - ✓ Predict and better control expense processes and costs.
 - ✓ Employ different techniques for forecasting supply and demand.
 - ✓ Interactively develop and evaluate alternative plans to mitigate overall supply chain risk.
- Understand how to design, manage, and control product and service quality:
 - ✓ Predict and manage quality issues.
 - ✓ Identify true cost of warranty programs.
 - ✓ Understand the impact of adding or removing plants and warehouses on the distribution and manufacturing network.
- Understand how to manage materials (both sourcing and delivery) by being able to:
 - ✓ Predict historical inventory fluctuations, stock-outs and allocation situations.
 - ✓ Better isolate obsolete and slow-moving inventories.

Healthcare Analytics:

- Understand patient analytics

- ✓ Optimize every phase of a health care product's life cycle by building and refining marketing messages and promotions to incorporate customer segmentation, targeting, and predictive models built on patient-level healthcare data with expanded demographic and lifestyle behavior data.
- ✓ Design and perform hospital inpatient and outpatient analyses
- Understand clinical analytics
 - ✓ Improve clinical trial site targeting
 - ✓ Reduce time, cost and uncertainty of clinical trial execution
- Understand retail pharmacy operations
 - ✓ Maximize revenue by better understanding the dynamics between cash pricing and third-party reimbursements in a pharmacy's local markets
 - ✓ Better create opportunities for the pharmacy chain to partner with physicians on health initiatives
- Understand disease profiling
 - ✓ Understand and forecast critical flu/respiratory illness status levels for a regional market area, reflecting seasonal severity and potential spread rate of illness throughout the season
- Understand managed care analytics
- Understand hospital operations analytics
 - ✓ Optimize staff management

IT Analytics

- Understand different types of metrics related to IT operations, that is, IT application performance metrics and how they relate to customer experiences as well as the overall business performance.
 - ✓ Traditional IT infrastructure metrics, e.g., CPU load, memory utilization, performance of operating systems, database and networks.
 - ✓ End-user experience metrics, e.g., average web page response time, transaction success rate.
 - ✓ Business performance metrics, e.g., transaction of trades per minute, peak throughput.
- Employ application performance monitoring (for example, transactions, processes, networks, servers) to:
 - ✓ Understand and optimize utilization of existing IT resources (capacity analysis)
 - ✓ Forecast performance latencies before customers are impacted
 - ✓ Understand application performance in the context of its underlying infrastructure, e.g., Correlate behavior and isolate root cause across IT silos (automated diagnostics)
 - ✓ Reduce false alerts

- Understand configuration management challenges associated with moving operations from a physical data center to virtual to a cloud environment:
 - ✓ Proactively manage virtualization for data centers.
 - ✓ Proactively manage applications running in complex multi-vendor cloud-based environments that depend on many moving parts generating performance data.

Marketing Analytics:

- Understand customer classification and segmentation using different types of cluster analysis.
- Apply analytics-based product positioning models.
- Apply models to problems related to customer acquisition/retention/recapture.
- Apply and interpret analytics for customer relationship management (CRM).
- Apply and interpret analytics customer lifetime value analysis.
- Apply and interpret analytics to media allocation decisions
- Understand upselling and cross-selling using market basket analysis.
- Understand the use of social media in churn analysis.
- Design, deploy, analyze and interpret product and service choice models.
- Select and apply sales forecasting techniques appropriate to the market type and maturity.
- Apply analytic models to price setting and optimization.
- Apply optimization models to category management and product portfolios.
- Develop and apply market response models.

Financial Analytics:

- Understand portfolio optimization with data uncertainty using quadratic programming
- Understand option pricing and structuring asset backed securities using dynamic pricing
- Understand asset/liability management using stochastic programming
- Understand asset pricing using linear programming

- **Introduction:** OR modelling approach and various real life situations, Linear programming problems & Applications, Various components of LP problem formulation, Solving Linear Programming problem using simultaneous equations, Graphical Method.
- **Simplex Method & Extensions:** Sensitivity analysis, Duality theory, Revised Simplex, Dual Simplex, Transportation and Assignment Problems.
- **Network Analysis including PERT-CPM:** Concepts of network, The shortest path, minimum spanning tree problem, Maximum flow problem, Minimum cost flow problems, The network simplex method, Project planning & control with PERT & CPM.
- **Integer Programming Concepts:** Formulation solution and applications
- **Dynamic Programming Concepts:** Formulation, solution and application, Game Theory
- **Queuing Theory & Applications:** Linear Goal Programming methods and applications, Simulation.

CS-531 Data Mining Lab**0-0-6-3**

Students are required to perform practicals in Oracle/MS SQL Server, Weka Tool and STATISTICA *Data Miner*

- Building a Database Design using ER Modeling and Normalization Techniques
- Implementation of functions ,Procedures, Triggers and Cursors
- Load Data from heterogeneous sources including text files into a predefined warehouse schema.
- Design a data mart for a bank to store the credit history of customers in a bank. Use this credit profiling to process future loan applications.
- Feature Selection and Variable Filtering (for very large data sets)
- Association Mining in large data sets
- Interactive Drill-Down, Roll up, Slice and Dice operations
- Generalized EM & *k*-Means Cluster Analysis
- Generalized Additive Models (GAM)
- General Classification and Regression Trees (GTrees)
- General CHAID (Chi-square Automatic Interaction Detection) Models
- Interactive Classification and Regression Trees
- Goodness of Fit Computations

This is only the suggested list of practicals. Instructor may frame additional practicals relevant to the course contents.

CS-533 Quantitative Techniques Lab**0-0-6-3**

Implementation of the followings in High Level Programming Language:

- Transportation Problem.
- Assignment Problem.
- Simplex method.
- Integer Programming.
- 2 x n Problem.
- PERT and CPM.
- Queuing Model.
- Simulation.
- Dynamic Programming.
- Goal Programming.

This is only the suggested list of practicals. Instructor may frame additional practicals relevant to the course contents.

Elective Subjects

CS-521 Structuring Problems for Analytics

3-0-0-3

- Apply an “analytics lens” to business problems, identify applications areas for analytics, define business analytics and distinguish business analytics from other related frameworks.
- Define and explain the business analytics process (problem definition; data preparation; technical analysis and modeling; evaluation of results; implementation and deployment) (CRISP, etc.).
- Apply frameworks, techniques, and structured and unstructured processes for framing analytics problems and to develop quantifiable objectives and metrics for success.
- Identify sources and opportunities for applications of syndicated (external) data, assess the quality, validity and potential weaknesses of external, internal, and experimental data, and create databases which combine external, internal, and experimental data.
- Understand and describe the functionality and role of analytic techniques in statistics, data mining, forecasting, simulation, and optimization.
- Describe and provide brief evaluations of the capabilities of the most important commercial and open-source analytics tools and to be able to perform basic analyses with at least one commonly available program.
- Propose and apply effective reporting and visualization tools to effectively communicate (in writing and orally) the results and insights from analytics projects.
- Be able to identify and articulate challenges associated with implementation and deployment of the results of analytics projects; to propose structural and business process changes needed for effective deployment; to specify metrics for monitoring and controlling applications of analytics.
- Prepare a project proposal in which a business problem is identified, data requirements are determined, and appropriate analytics tools and processes are specified.
- Design experiments and perform basic statistical analysis of data from experiments (both field and laboratory) to investigate business issues.
- Propose and evaluate analytics teams that are properly aligned with the analytics problem, organizational capabilities, and resources.
- Review and appraise the ethical, legal, and privacy issues associated with applications of analytics.

CS-523 Statistical Methods for Data Mining**3-0-0-3**

- **Foundations of Statistics:** Basic concepts of Statistics, Collection, Sampling, Classification and Graphical Representation of Data, Measures of Central Tendency, Measures of Dispersion, Moments, Skewness and Kurtosis.
- **Theoretical Distributions:** Binomial, Poisson and Normal Distributions.
- **Hypothesis Testing:** Procedure of Testing Hypothesis, Standard Error and Sampling Distribution, Tests of Significance, Student's t-Test, Z-test.
- χ^2 Test: χ^2 defined, Yate's Corrections, Properties of χ^2 , Analysis of χ^2
- **Variance Analysis:** F-test, Techniques of Analysis of Variance, Analysis of Variance in two-way Classification Model.
- **Miscellaneous Topics:** Correlation Analysis, Regression Analysis, Curve Fitting using Least Square Method.

CS-525 Soft Computing**3-0-0-3**

- **Artificial Neural Networks:** Basic concepts, Single layer perception, Multilayer Perception, Supervised and Unsupervised learning, Back propagation networks, Kohonen's self organizing networks, Hopfield network.
- **Fuzzy Systems:** Fuzzy sets and Fuzzy reasoning, Fuzzy matrices, Fuzzy functions, Decomposition, Fuzzy automata and languages, Fuzzy control methods, Fuzzy decision making.
- **Neuro - Fuzzy Modeling:** Adaptive networks based Fuzzy interface systems, Classification and Regression Trees, Data clustering algorithms, Rule based structure identification, Neuro-Fuzzy controls, Simulated annealing, Evolutionary computation.
- **Genetic Algorithms:** Survival of the Fittest, Fitness Computations, Cross over, Mutation, Reproduction, Rank method, Rank space method.
- **Softcomputing And Conventional AI:** AI search algorithm, Predicate calculus, Rules of inference, Semantic networks, Frames, Objects, Hybrid models, Applications.

CS-527 Simulation and Modeling**3-0-0-3**

- **Introduction to Simulation:** System & System Environment, Components of a System, Discrete and Continuous Systems, Model of a System and Types of Models, Discrete Event System Simulation, Advantages and Disadvantages of Simulation, Areas of Application
- **Techniques of Simulation:** Monte Carlo Method, Types of System Simulations, Real Time Simulation, Stochastic Variables, Discrete Probability Functions
- **General Principles:** Concepts in Discrete Event Simulation, Event Scheduling /Time Advance Algorithm, List Processing, Using Dynamic Allocation & Linked List
- **Simulation Software:** History of Simulation Software, Selection of Simulation Software, Simulation in C++, GPSS, Simulations Packages, Trends in simulation Software.
- **Statistical Models in Simulation:** Useful Statistical Models, Discrete Distributions, Continuous Distributions, Poisson Process, Empirical Distributions

- **Queuing Models:** Characteristics of Queuing systems, Queuing Notation, Long Run Measures of performance of Queuing Systems, Steady State Behavior of infinite Population Markovian Models, Steady State Behavior of finite Population Models, Networks of Queues
- **Random Number Generation:** Properties of Random Numbers, Generation of Pseudo-Random Numbers, Techniques for Generating Random Numbers, Tests for Random Numbers, Inverse transform Techniques, Convolution Methods, and Acceptance –Rejection Techniques
- **Input Modeling:** Data Collection, Identifying the Distribution with Data, Parameter Estimation, Chi – Square Test, Selecting Input Models with Data
- Verification & Validation of simulation Modeling: Model Building, Verification & Validation, Verification of simulation Models, Calibration & Validation of Models.

CS-529 Information Retrieval Systems

3-0-0-3

- **Introduction:** Definition, Objectives, Functional Overview, Relationship to DBMS, Digital libraries and Data Warehouses.
- Information Retrieval System Capabilities: Search, Browse, Miscellaneous
- **Cataloging and Indexing:** Objectives, Indexing Process, Automatic Indexing, Information Extraction.
- **Data Structures:** Introduction, Stemming Algorithms, Inverted file structures, N-gram data structure, PAT data structure, Signature file structure, Hypertext data structure.
- **Automatic Indexing:** Classes of automatic indexing, Statistical indexing, Natural language, Concept indexing, Hypertext linkages
- **Document and Term Clustering:** Introduction, Thesaurus generation, Item clustering, Hierarchy of clusters.
- **User Search Techniques:** Search statements and binding, Similarity measures and ranking, Relevance feedback, Selective dissemination of information search, Weighted searches of Boolean systems, Searching the Internet and hypertext. Information Visualization: Introduction, Cognition and perception, Information visualization technologies.
- **Text Search Algorithms:** Introduction, Software text search algorithms, Hardware text search systems.
- **Information System Evaluation:** Introduction, Measures used in system evaluation, Measurement example – TREC results.

CS-541 Deployment Strategies

3-0-0-3

- Understand various cognitive issues related to visually rendering insights using graphics
 - ✓ Recognize the significance of cognitive processes that influence the use of graphical rendering of insights for decision making
 - ✓ Understand different forms of external representations and how they influence search, recognition and inferences made by end users

- ✓ Realize the different categories of pre-attentive attributes (e.g., color, shape) and how they can be employed in the development of a graphic
- Understand the properties of information and those of graphics, and how to create a “correct” mapping between information and graphics
- Realize the importance of task (e.g., part-whole analysis) and individuals (e.g., dichromatic) in graphical rendering of insights
- Understand the implications of business analytics on:
 - ✓ Processes that are used to create data
 - ✓ Information technology that is used to automate the processes
 - ✓ People that employ IT to accomplish a task as defined by the process.

CS-543 Natural Language Processing

3-0-0-3

- **Introduction to Natural Language Understanding:** The study of Language, Applications of NLP, Evaluating Language Understanding Systems, Different levels of Language Analysis, Representations and Understanding, Organization of Natural language Understanding Systems, Linguistic Background: An outline of English syntax. Introduction to semantics and knowledge representation, Some applications like machine translation, database interface.
- **Grammars and Parsing:** Grammars and sentence Structure, Top-Down and Bottom-Up Parsers, Transition Network Grammars, Top-Down Chart Parsing. Feature Systems and Augmented Grammars: Basic Feature system for English, Morphological Analysis and the Lexicon, Parsing with Features, Augmented Transition Networks.
- **Grammars for Natural Language:** Auxiliary Verbs and Verb Phrases, Movement Phenomenon in Language, Handling questions in Context-Free Grammars. Human preferences in Parsing, Encoding uncertainty, Deterministic Parser.
- **Computational morphology:** Lemmatization, Part-of-Speech Tagging, Finite-State Analysis.
- **Ambiguity Resolution:** Statistical Methods, Probabilistic Language Processing, Estimating Probabilities, Part-of-Speech tagging, Obtaining Lexical Probabilities, Probabilistic Context-Free Grammars, Best First Parsing. Semantics and Logical Form, Word senses and Ambiguity, Encoding Ambiguity in Logical Form.
- **Application of NLP:** Intelligent Work Processors: Machine Translation; User Interfaces; Man-Machine Interfaces: Natural language Querying Tutoring and Authoring Systems. Speech Recognition Commercial use of NLP.

CS-545 Machine Learning

3-0-0-3

- **Introduction:** Basic concepts.
- **Supervised learning:** Supervised learning setup, LMS, Logistic regression, Perceptron, Exponential family, Generative learning algorithms, Gaussian discriminant analysis, Naive Bayes, Support vector machines,

Model selection and feature selection, Ensemble methods: Bagging, boosting. Evaluating and debugging learning algorithms.

- **Learning theory:** Bias/variance tradeoff, Union and Chernoff/Hoeffding bounds, VC dimension, Worst case (online) learning, Practical advice on how to use learning algorithms.
- **Unsupervised learning:** Clustering, K-means, EM, Mixture of Gaussians, Factor analysis, PCA (Principal components analysis), ICA (Independent components analysis).
- **Reinforcement learning and control:** MDPs, Bellman equations, Value iteration and policy iteration, Linear quadratic regulation (LQR), LQG, Q-learning, Value function approximation, Policy search, Reinforce, POMDPs.

CS-547 Operations Research

3-0-0-3

- Introduction to Operations Research (OR), Introduction to Foundation mathematics and statistics.
- Linear Programming (LP), LP and allocation of resources, LP definition, Linearity requirement.
- Maximization Then Minimization problems, Graphical LP Minimization solution. Introduction, Simplex method definition, formulating the Simplex model.
- Linear Programming: Simplex Method for Maximizing, Simplex maximizing example for similar limitations, Mixed limitations, Example containing mixed constraints, Minimization example for similar limitations.
- Sensitivity Analysis: Changes in Objective Function, Changes in RHS,
- The Transportation Model: Basic Assumptions, Solution Methods: Feasible Solution: The Northwest Method, The Lowest Cost Method; Optimal Solution: The Stepping Stone Method, Modified; Distribution (MODI) Method.
- The Assignment Model: Basic Assumptions, Solution Methods: Different Combinations Method, Short-Cut Method (Hungarian Method)
- MSPT: The Dijkstra algorithm, and Floyd's Algorithm, Shortest Route Algorithm.

CS-549 Probability and Queuing Theory

3-0-0-3

- **Probability:** The concept of probability, The axioms of probability, Some important theorems on Probability, Assignment of Probabilities, Conditional Probability, Theorems on conditional probability, Independent Event's, Bayes' Theorem.
- **Random Variables and Probability Distributions:** Random variables, Discrete probability distributions, Distribution functions for discrete random variables, Continuous probability distribution, Distributions for continuous random variables, joint distributions, Independent random variables.
- **Mathematical Expectation:** Definition, Functions of random variables, some theorems on Expectation, The variance and Standard Deviation, Moments, Moment Generating Functions, Covariance, Correlation Coefficient.
- **Special Probability Distributions:** The Binomial Distribution, The Normal Distribution, The Poisson Distribution, Relations between different distribution,

Central limit theorem, Uniform distribution, Chi-square Distribution, Exponential distribution.

- **Sampling Theory:** Population and Sample, Sampling with and without replacement, the sample mean, Sampling distribution of means, proportions, differences and sums, the sample variance, the sample distribution of variances.
- **Tests of Hypotheses and Significance:** Statistical Decisions, Statistical hypotheses, Null Hypotheses, Tests of hypotheses and significance, Type I and Type II errors, level of significance, Tests involving the Normal distribution, One-Tailed and Two-tailed tests, Special tests of significance for large and small samples, The Chi-square test for goodness of fit.
- **Curve Fitting Regression and Correlation:** Curve Fitting, The method of least squares, The least squares line, multiple regression, the linear correlation coefficient, Rank correlation, Probability interpretation of regression and correlation.
- **Discrete-Parameter Markov Chains:** Introduction, Computation of n-step Transition Probabilities, State Classification and Limiting Distributions, Distribution of times between state changes, Irreducible finite chains with aperiodic states, The M/G/1 Queuing System, Discrete-parameter, Birth-Death processes, Finite Markov chains with absorbing states.

CS-551 Neural Networks

3-0-0-3

- **Introduction:** Historical perspective, Basic Neurobiology, Why Artificial Networks? Network architectures, The tasks neural networks can perform, Characteristics of Neural Networks.
- **Basic neuron models:** McCulloch-Pitts model, radial basis function model, etc, learning algorithms.
- **Basic neural network models:** The Hebbian Hypothesis, Single-layered neural networks, multilayer perceptron, nearest neighbor based multilayer perceptron, Training of artificial Neural Networks.
- **Basic learning algorithms:** Supervised learning, constructive algorithms, Single-hidden layer algorithms, The Upstart algorithm, The Cascade Correlation algorithm, Neural networks and temporal sequences, Sequence recognition, Sequence generation, Unsupervised learning, Competitive Learning, the back propagation algorithm, self-organization learning, winner-take-all competitive learning, evolutionary learning, etc.
- **Applications:** character recognition, signal restoration, etc.