SCHOOL OF DATA SCIENCE AND FORECASTING

PROGRAM CODE: DS5B BATCH: 2019-21

PROGRAM TITLE: MASTER OF SCIENCE (M.Sc.)
- DATA SCIENCE AND ANALYTICS

PROGRAM OUTCOMES:

- Understanding of the key technologies in data science and analytics: data mining, data visualization techniques, machine learning,statistics, and NLP.
- Getting knowledge on various theoretical and practical aspects of data science.
- Demonstrate use of team work, leadership skills, and decision making.
- Getting opportunities of higher studies in the area of data science.

PROGRAM SPECIFIC OUTCOMES:

- Apply data analysis techniques to the solution of real world business problems, communicate findings, and effectively present results.
- Recognize and analyze ethical issues in business related to intellectual property, data security, integrity, and privacy.
- Demonstrate knowledge of statistical data analysis techniques utilized in business decision making.
- Apply algorithms to build machine intelligence.
- Work with messy data, applying models, and understanding the business context.
- Work with unstructured data from various sources like video and social media.
- Use Data Visualization techniques.
- Write the programming codes in R and Python.

PROGRAM STRUCTURE:

First Semester:

Code	Title	Credits (L T P)
CORE COURSES		
DS5B-501	Database Management	3 (2-0-2)
DS5B-503	Forecasting Methods-I	3 (2-1-0)
DS5B-505	Operations Research	4 (3-1-0)
DS5B-507	Probability and Statistics	3 (2-1-0)
DS5B-509	Statistical Programming in R	3 (2-0-2)
DS5B-511	Advanced Excel	2 (0-0-4)
ELECTIVE COURSES-DISCIPLINE CENTRIC (Any One)		
DS5B-521	Fundamentals of Algorithms	3 (2-0-2)
DS5B-523	Decision Analysis	3 (2-1-0)
ELECTIVE GENERIC: The students can choose following course or any other PG level generic		

course being run in this campus.		
DS5B-541	Communication Skills	3 (2-1-0)
DS5B-551	Comprehensive Viva Voce	4

Second Semester:

Code	Title	Credits (L T P)
CORE COURSES		
DS5B-502	Data Mining and Data Warehousing	3 (2-0-2)
DS5B-504	Machine Learning	3 (2-0-2)
DS5B-506	Linear Algebra and Advanced Calculus	3 (2-1-0)
DS5B-508	Forecasting Methods-II	3 (2-0-2)
DS5B-510	Big Data Technologies	3 (2-0-2)
DS5B-512	Python for Analytics	3 (2-0-2)
ELECTIVE COURSES-DISCIPLINE CENTRIC (Any One)		
DS5B-522	Web Mining	3 (2-0-2)
DS5B-524	Scientific Computing	3 (2-1-0)
ELECTIVE GENERIC: The students can choose following course or any other PG level generic course being run in this campus.		
DS5B-542	Technical Communication	3 (2-1-0)
DS5B-552	Comprehensive Viva Voce	4

Third Semester:

Code	Title	Credits (L T P)
CORE COURSES		
DS5B-601	Cloud Computing	3 (2-1-0)
DS5B-603	Data Visualization	3 (2-0-2)
DS5B-605	Deep Learning	3 (2-1-0)
DS5B-607	Natural Language Processing	3 (2-0-2)
ELECTIVE COURSES-DISCIPLINE CENTRIC (Any One)		
DS5B-621	Cluster Analysis	4 (2-1-0)
DS5B-623	Multivariate Analysis	4 (2-0-2)
ELECTIVE GENERIC: The students can choose following course or any other PG level generic course being run in this campus.		
DS5B-641	Minor Project	4 (0-0-8)
DS5B-651	Comprehensive Viva Voce	4

Fourth Semester:

DS5B-652	Project Dissertation/ Internship	12 (0-0-24)

Note: The above course contents can be modified as per requirement from time to time in accordance with University Ordinance No. 14.

Note: The above course contents can be modified as per requirement from time to time in accordance with University Ordinance No. 14.

DETAILED SYLLABUS:

First Semester:

DS5B-501: Database Management Credits: 3 (2-0-2)

Objective: The purpose of this course is to provide fundamental knowledge of database management system and understanding of how to use and design a DBMS.

Unit I: DBMS Concepts, Comparison between Database approach and Traditional file accessing approach, Advantages of database systems, Schemas and instances, Data Dependency, Data Dictionary, and Meta Data. Data models, Types of Data models (Object Oriented, Record Based and Physical data models), E-R diagram, Relational Data models: Domains, Tuples, Attributes, Keys, Relational database, Schemas, Integrity constraints, Relational algebra and relational calculus.

Unit II: Database Design: Introduction to normalization, Normal forms (INF, 2NF, 3NF, BCNF), Functional dependency, Decomposition, Dependency preservation and lossless join, multi-valued dependencies. **Structured Query Language**: DDL, DML, DCL, TCL, SQL Functions, integrity constraints, various joins, sub-query, index, View, Sequence, and Clusters. PL/SQL: manipulating data using PL/SQL, Iteration, Exceptions, Cursors, Trigger. (Experiment on Oracle 11g)

Unit III: Transaction Processing and Concurrency Control: Transaction System, Serializability of schedules, conflict & view serializable schedule, Recovery from transaction failures, Log based recovery. Checkpoints dead lock handling, Concurrency Control, locking Techniques for concurrency control, time stamping protocols for concurrency control, validation based protocol, multiple granularity.

Unit IV: Advance Concepts: Introduction to Distributed databases, data mining, data warehousing, Basic Concepts of Object Oriented Database System, Comparative study of OODBMS V/s RDBMS. Introduction to Image and Multimedia databases and data structures, Web and mobile database, Spatial and Geographic Database, Accessing Database from front-end Application. Case Study: Oracle, MySql, DB2.

Reference Books:

1. A Silberschatz, H Korth, S Sudarshan, “Database System and Concepts”, fifth Edition McGraw-Hill.
2. ElmasriRamez and NovatheShamkant, “Fundamentals of Database Systems”, Benjamin Cummings Publishing. Company.
3. Rob, Coronel, “Database Systems”, Seventh Edition, Cengage Learning**.**

4. Ramakrishnan: Database Management System , McGraw-Hill
5. Fred R.McFadden,JeffreyA.Hoffer& Marry B.Prescott.?Modern Database Management, Fifth Edition,Pearson Education Asia,2001
6. Gray Jim and Reuter Address, "Transaction Processing: Concepts and Techniques", Morgan Kauffman Publishers.

Course Outcomes:

- The student will learn the basics of database management.
- The student will be able to design database using ER diagram.
- The student will be able to optimize database using normalization.
- The student will be able to work on database software MYSQL/Oracle.
- The student will learn how to write SQL queries.

DS5B-503: Forecasting Methods-I Credits: 3 (2-1-0)

Objective: This subject is designed in such a way to provide the basic concepts of quantitative forecasting methods, risk and uncertainty in forecasting.

COURSE DESCRIPTION:

Unit I: Introduction:
Forecasting perspective, an overview of forecasting methods, basic steps in forecasting. Common timeseries pattern, graphical and numerical summaries. Measuring forecasting accuracy: Mean, Median, Mode, Error Standard Deviation, MAD, ME, MSE, MAE prediction intervals, transformations and adjustments.

Unit II: Time Series Decomposition:
Principles of decomposition, moving averages, local regression smoothing, classical decomposition, Census Bureau methods, STL decomposition, forecasting and decomposition. Correlation measurement, Auto correlation, ACF, Durbin-Watson statistics.

Unit III: Exponential Smoothing Methods:
Forecasting scenarios, Averaging methods: mean, moving averages; Exponential smoothing methods-single exponential smoothing, ARRSES, Holt's linear method, Holt-Winter's trend and seasonality method, comparison of methods, general aspects of smoothing methods.

Unit IV: Regression:
Simple regression, forecasting with simple regression, non-linear relationships. Regression with time series, regression and forecasting, econometric models for forecasting. Univariate ARIMA model: examining stationary, ARIMA models, forecasting with ARIMA models, ARIMA application.

Text Books:

1. Spyros Makridakis: Forecasting Method's and application Wiley
2. N.P. Nagpal: Forecasting Techniques,RBSA

3. Stephen A. Delurgio: Forecasting Principals and Application, McGraw Hill

Course Outcomes:

- Discuss the key factors which affect the success of forecasting procedures.
- Use Basic Statistical Techniques and statistical Graphics to forecast values.
- Find different sets of Smoothed or Average values to be used when forecasting.
- Understand the key concepts needed to use the Linear Regression model when forecasting.
- Model and Forecast the Seasonal component of a set of values.
- Model the different types of cyclical behaviour observed in different sets of values.
- Understand and use the Box-Jenkins or ARMA Procedure.

DS5B-505: Operations Research Credits: 4 (3-1-0)

COURSE OBJECTIVE:

This course exposes the students in mathematical modelling, solving and analysing business and industrial problems using operations research methods.

COURSE DESCRIPTION:

Unit -I:

Introduction, History, Development of Operations Research, Characteristics of Operations Research, Models in Operations Research, Principles of Modelling. Pre-modelling-Need Recognition, Problem Formulation. Modelling-Model development, Data collection, Model solution, Model validation, and sensitivity analysis. Post-modelling-Interpretation of results and implications, Decision making, Implementation, and Control.

Unit-II:

Linear Programming, Formulation of Linear programming Problems. Solution Methods-Graphical, Simplex, M-Technique, Two-Phase. Special cases of LP problems. Duality, Primal-dual relationships, Dual simplex method. Sensitivity analysis. Solving LP problems using computer software.

Unit-III:

Transportation Model-Formulating the model, Initial Feasible Solution-North-West Method, Least Cost Method, Vogel's Approximation Method. Optimum Solution-MODI method, Stepping Stone Method. Special issues of transportation problems. Assignment Model: Formulating the model, Solving the assignment problem using Hungarian method. Special issues of assignment problems.

Integer programming: Types of integer programming problems, Formulating the model, Solution using Branch and bound method. Dynamic Programming.

Solving the problems using computer software.

Unit-IV:

Network models: Minimal spanning tree algorithm, Shortest-route problem, Maximal flow model, Minimum –cost capacitated flow problem. Project Scheduling: CPM and PERT.

Inventory models: Functions of inventory, information requirements for inventory management-demand, lead time, inventory costs, and quantity on hand. Objectives of inventory management. Economic Order Quantity Models-Economic order size, economic production run, quantity discounts. Determining the reorder point. Material Requirement Planning.

Queuing Models: Goals, elements and characteristics of queuing systems, Measures of system performance, Waiting line models-single channel, multiple channel. Cost considerations.

TEXT BOOKS

Hamdy A. Taha: Operations Research: An introduction, Pearson Prentice Hall
David R. Anderson, Dennis J. Sweeney, Thomas A. Williams: An Introduction to Management Science, South-Western College Publishing.
William J. Stevenson: Introduction to Management Science, IRWIN.

Course Outcomes:

- Understand the verbal description of the real system and accordingly identify and development of operational research models.
- Understand the mathematical tools that are needed to solve optimisation problems.
- Use of mathematical software to solve the OR models developed.
- Develop a technical report that describes the model, solving technique, results analysis and recommendations.

DS5B-507: Probability and Statistics Credits: 3 (2-1-0)

Objective: The main objective of this course is to provide students fundamental knowledge of probability and statistics.

COURSE DESCRIPTION:

UNIT I: General Introduction, including the Uses and Applications of Statistics, Types of Data and their Collection Methods, Stages of Statistical Investigation; Descriptive Analysis of Data including Exploratory Data Analysis; Central Tendency: mean, mode, median, percentile, moments, Kurtosis, and Skewness,

UNIT II:

Probability Theory: Set Theory, Random Experiments and Outcomes, Measure of Probability of Events, Independence Events, Conditional Probability, Some Basic Rules/Theorems of Probability; Counting Techniques and Application to Problems

UNIT III:

Random Variables and Probability Distributions: Discrete distributions: degenerate, Bernoulli, Binomial, Poisson, Geometric, Negative binomial, hyper-geometric, uniform,

Multinomial distribution and its marginal and conditional distributions. Continuous distributions: Uniform, Normal, Lognormal, Exponential, Laplace, Pareto, Beta, Gamma, Cauchy.

UNIT IV:

Uses of different software:Uses of different package of R language for analysis and distribution of data in the field of probability function and distribution function.

SPSS: Analysis and interpretation of different data sets.

Text Books:

1. Chung K.L. "Elementary Probability Theory with Stochastic Process", Springer / Narosa
2. Feller W. "An Introduction to Probability Theory & its Applications", John Wiley
3. Goon A.M., Gupta M. K., Dasgupta B., "Fundamentals of Statistics (V-1)",World Press
4. Yule G.U & Kendall M.G., "An Introduction to the Theory of Statistics", C.Griffin
5. Snedecor& Cochran "Statistical Methods" 6th edition, Iowa State Univ. Press.

Course Outcomes:

1. Learn about data driven decision making under probabilistic framework.
2. Translate a business problem into a statistical inference problem.
3. Identify an appropriate statistical method to solve the problem.
4. Perform statistical analysis and draw conclusions from the analysis to solve the business problem.

DS5B-509: Statistical Programming in R Credits: 3 (2-0-2)

Objective: This course is an introduction to R, a powerful and flexible statistical language and environment that also provides more flexible graphics capabilities than other popular statistical packages. After taking this course, students will be able to –

1. Use R for statistical programming, computation, graphics, and modeling,
2. Write functions and use R in an efficient way,
3. Fit some basic types of statistical models,
4. Use R in their own research,
5. Be able to expand their knowledge of R on their own

COURSE DESCRIPTION:

Unit I: Introduction to R programming language: Getting R, Managing R, Arithmetic and Matrix Operations, Introduction to Functions, Control Structures. Working with Objects and Data: Introduction to Objects, Manipulating Objects, Constructing Data Objects, types of Data items, Structure of Data items, Reading and Getting Data, Manipulating Data, Storing Data.

Unit II: Data Distribution and Statistical Testing: Types of Data distribution, Normal distribution, Poisson distribution, Random number generation, Chi-Square Testing, Student's t-test, F-test, Monte Carlo Simulation.

Unit III: Graphical Analysis using R: Basic Plotting, Manipulating the plotting window, Box-Whisker Plots, Scatter Plots, Pair Plots, Pie Charts, Bar Charts.

Unit IV: Advanced R: Statistical models in R, Correlation and regression analysis, Analysis of Variance (ANOVA), creating data for complex analysis, Summarizing data, and case studies.

Practical:

1. Practical based on matrix operation and retrieving matrix elements
2. Adding and deleting columns/ rows from existing data-frame
3. Coding for simple functions and nested functions
4. Coding for structures (i.e. loops and controls)
5. Calculate probabilities for Binomial distribution
6. Calculate probabilities for Poisson distribution
7. Calculate probabilities for Normal distribution
8. Generate random numbers from uniform distribution
9. Monte-Carlo simulation
10. Chi-Square (Goodness of fit, independence and homogeneity) test for the given data
11. Testing of mean and variance for the given data (t-test and F- test)
12. Generating statistical plots for the given data
13. Fitting a linear model for the given data and interpretation of the results
14. Fitting a logistic regression model for the given data and interpretation of the results
15. Fitting a Poisson regression model for the given data and interpretation of the results
16. Comparing the mean of multiple groups using ANOVA.

Text Books:

1. Mark Gardener: Beginning R: The Statistical Programming Language, Willey publications
2. Norman Matloff: The Art of R Programming: A Tour of Statistical Software Design, OREILLY & Associates Inc.

Course Outcomes:

- Data manipulation - acquiring skills in flexible matrix manipulation
- Access online resources for R and import new function packages into the R workspace
- Scripting in such a way that the script can be used with minimal effort for similar datasets and analyses and for especially large datasets
- Explore data-sets to create testable hypotheses and identify appropriate statistical tests
- Perform appropriate statistical tests using R
- Learn how to create high-quality figures, especially associated with more complex analyses (e.g. three dimensional scatter plots, animated chart Trellis displays, etc.).

DS5B-511: Advanced Excel Credits: 2 (0-0-4)

Objective: The main objective of this course is to learn analysis of data using MS Excel, resulting in less time and better understands what the data means.

COURSE DESCRIPTION:

Unit I: Introduction to Excel User Interface**,** Application, Workbook, Worksheets & its Components, Named Ranges; Formatting: Cell Color, Font Color, Indents, Alignments, Number Formats, Custom Formats, Editing commands; Data Sorting: Built-in Sort, Sorting Levels, Custom Sort; Data Filtering: Auto Filter – Filter By Color, Filter by Icono Advanced Filter, Remove Duplicates; Data Subtotal – Built-In Subtotal (Nested Subtotal).

Unit II: Data Validation: Based on cell values (text length, whole no Based on Formulas, List Dropdown, Circle Invalid Data, Input & Error Messages; Data Grouping: Grouping Rows, Grouping Columns. Data Tables: Conditional Formatting, Formatting based on Cell values, Formatting based on Formulas, Icon Sets (bars, scales, icons), Freezing Panes, Text-to-Columns, Delimited, Fixed Length; Data Consolidation (from multiple files), Getting External Data into Excel, From MS Access, From Text files, From Web, Other Data Sources.

Unit III: Formulas, TEXT Functions, IF, ERROR Functions, LOGICAL Functions, VLOOKUP, HLOOKUP, COUNTIF, SUMIF, SUMPRODUCT, DATE & TIME FUNCTIONS, FORMULA TEXT, Information Functions (ISNA, ISEVEN, ISERR…).

Unit IV: Charts: Chart Types, Chart Components, Primary Vs Secondary Axis, Chart Formatting, Sparkline (2010 and above); Pivot Tables: Introduction & Creation, Slicer, TimeLine, Pivot Charts, Calculated Fields, Calculated Items, Grouping, Formatting – Number/Conditional, PowerPivot, PowerView.

Assignment: Dashboard design

Text Books:

1. John Walkenbac, “Excel 2016 Bible”, John Willey & sons.
2. Jordan Goldmeier , “Advanced Excel Essentials”, Apress Publisher.
3. Conrad George Carlberg , “Business Analysis with Microsoft Excel”, Que Publishers.
4. Bernd Held, ”Microsoft Excel Functions & Formulas”, Wordware publishing, Inc.
5. Steven Roman ,”Writing Excel Macros with VBA” O’Reilly Media.

Course Outcomes:

- The student will be able to perform basic operations in Excel.
- The student will be able to summarize data using Grouping and pivot table.
- The student will be able to write conditional statements and perform LOOKUP operations.
- The student will be able to create charts in Excel.

- The student will be able to create a dashboard in Excel.

DS5B-521: Fundamentals of Algorithms Credits: 3 (2-0-2)

Objective: The main object of this course is to provide an introduction to create analytical skills, to enable the students to design algorithms for various applications, and to analyse the algorithms.

COURSE DESCRIPTION:

Unit I: Algorithm, Pseudo code for expressing algorithms, Performance Analysis-Space complexity, Time complexity, Asymptotic Notation- Big oh notation, Omega notation, Theta notation and Little oh notation

Unit II: Divide and conquer: General method, applications-Binary search, Quick sort, Merge sort, Strassen's matrix multiplication.

Unit III: Greedy method: General method, applications-Job sequencing with deadlines, 0/1 knapsack problem, Minimum cost spanning trees, Single source shortest path problem. Dynamic Programming: General method, applications-Matrix chain multiplication, Optimal binary search trees, 0/1 knapsack problem, All pairs shortest path problem, Travelling sales person problem.

Unit IV: Backtracking: General method, applications-n-queen problem, sum of subsets problem, graph colouring, Hamiltonian cycles. Branch and Bound, NP-Hard and NP-Complete problems: Basic concepts, non-deterministic algorithms, NP - Hard and NP Complete classes, Cook's theorem.

TEXT BOOKS

1. Anany Levitin, "Introduction to the Design and Analysis of Algorithm", Pearson Education Asia, 2003.
2. T.H.Cormen,C.E.Leiserson, R.L.Rivest,andC.Stein "Introduction to Algorithms", second edition, ,PHI Pvt. Ltd./ Pearson Education
3. R.C.T.Lee, S.S.Tseng, R.C.Chang and T.Tsai, "Introduction to Design and Analysis of Algorithms A strategic approach", Mc Graw Hill.
4. Ellis Horowitz,SatrajSahni and Rajasekharam," Fundamentals of Computer Algorithms",Galgotia publications pvt. Ltd.
5. Parag Himanshu Dave, HimanshuBhalchandra, " Design and Analysis Algorithms", Dave Publisher: Pearson

Course Outcomes:

- The student will be able to understand the basics of algorithms.
- The student will be able to work with the divide and conquer method.
- The student will be able to work with Greedy methods.
- The student will be able to work with backtracking algorithms.
- The student will be able to understand NP-Hard and NP-Complete Problem.

DS5B-523: Decision Analysis Credits: 3 (2-1-0)

OBJECTIVE:

This course introduces and applies advanced modelling techniques to decision problems with the objective of enhancing the decision-making skills as well as the spreadsheet knowledge base.

COURSE DESCRIPTION:

Unit-I: Game Theory: Introduction, definitions, two-person zero sum game. Game with pure strategies, saddle point, game value. Game with mixed strategies, solution methods-algebraic method, graphical method, dominance, linear programming method.

Unit-II: Decision Theory: Structuring the decision problem-payoff tables, decision trees. Decision making under certainty. Decision making under uncertainty: Optimistic, Conservative, Minimax regret.Decision making under risk: Expected value criterion, expected value of perfect information. Sensitivity analysis. Decision making with sample information, expected value of sample information, efficiency of sample information. Computing branch probabilities. Utility and decision making: meaning of utility, developing utilities and payoffs, the expected utility approach.

Unit-III: Multi-criteria Decision Making: Multi-criteria decision making: Goal programming, Scoring models, Analytical Hierarchy Process.

Unit-IV: Markov Analysis: Transition probabilities, system behaviour. Methods of Analysis: Tree diagram, Matrix multiplication, Algebraic solution. Cyclical and absorbing states, Market share analysis, Accounts receivable analysis.

TEXTBOOKS

1. Hamdy A. Taha: Operations Research: An introduction, Pearson Prentice Hall
2. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams: An Introduction to Management Science, South-Western College Publishing.
3. William J. Stevenson: Introduction to Management Science, IRWIN.

Course Outcomes:

1. Learn basics of decision making and uncertainty analysis under a risky situation.
2. To understand concepts in Strategic Game Theory.
3. To learn how to model and solve real life cases using Game Theory.

DS5B-541: Communication Skills Credits: 3 (2-1-0)

OBJECTIVE

To make use of the opportunities and meet the challenges of the modern world, students need to develop effective communication skills. This is a skill that is most often quoted as lacking in engineers and IT professionals. Proper communication skills are also extremely important in interpersonal relationships. This course will help students in becoming confident and effective communicators, who can project themselves positively to others.

COURSE DESCRIPTION:

Unit I: Introduction to Communication: Need for Effective Communication. The Process of Communication: Levels of communication; Flow of communication; Use of language in communication; Communication networks; Significance of technical communication. Barriers to Communication: Types of barriers; miscommunication; noise; overcoming measures.

Unit II: Listening Skills: Listening as an active skill; Types of Listeners; Listening for general content; Listening to fill up information; Intensive Listening; Listening for specific information; Developing effective listening skills; Barriers to effective listening skills.

Reading Skills: Previewing techniques; Skimming; Scanning; Understanding the gist of an argument;
Identifying the topic sentence; Inferring lexical and contextual meaning; recognizing coherence and sequencing of sentences; Improving comprehension skills.

Writing Skills: Sentence formation; Use of appropriate diction; Paragraph and Essay Writing; Coherence and Cohesion.

Unit III: Letter Writing: Formal, informal and demi-official letters; business letters. Job Application: Cover letter, Differences between bio-data, CV and Resume. Report Writing: Basics of Report Writing; Structure of a report; Types of reports. Non-verbal Communication and Body Language: Forms of non-verbal communication; Interpreting body-language cues; Kinesics; Proxemics; Chronemics; Effective use of body language.

Unit IV: Interview Skills: Types of Interviews; Ensuring success in job interviews; Appropriate use of non-verbal communication. Group Discussion: Differences between group discussion and debate; Ensuring success in group discussions. Presentation Skills: Oral presentation and public speaking skills; business presentations. Technology-based Communication: Netiquettes: effective e-mail messages; power-point presentation; enhancing editing skills using computer software.

Text Books:

Spoken English & Effective Communication (With 2 CDs): Mind Power Spoken English Institute by by Prof. Sharad Srivastava and Dr AmitaSinghvi; Franklin International; New edition (2015)
Communication Skills by Sanjay Kumar and PushpLata; Oxford University Press India; (2015)
El Dorado: A Textbook of Communication Skills by R. Pushkala and P.A.Sarada; Orient Blackswan; (2013)

Course Outcomes:

At the end of the course the students will be able-

1. To make communication with the parties concerned.
2. To write memorandum, circulars, notices, business letters, and business reports.

3. To write resume and job application.
4. To participate in group discussion and interviews.

Second Semester:

DS5B-502: Data Mining and Data Warehousing Credits: 3 (2-0-2)

Course Objective The main objective of this course is to provide understanding of data warehouse fundamentals and data mining techniques for business applications.

COURSE DESCRIPTION:

UNIT I: Data Warehousing: Introduction data warehousing, Data Mart, Data Warehouse Architecture; Star, Snowflake and Galaxy Schemas for Multidimensional databases, Fact and dimension data, Partitioning Strategy-Horizontal and Vertical Partitioning. ETL Concepts.OLAP technology: Multidimensional data models and different OLAP Operations, OLAP Server: ROLAP, MOLAP, Data Warehouse implementation, Efficient Computation of Data Cubes, Processing of OLAP queries, indexing data.

UNIT II: Data Mining: Basics of data mining, Data mining techniques, KDP (Knowledge Discovery Process), Application and Challenges of Data Mining; Introduction to Web Mining, Text Mining.Data Processing: Data Cleaning, Data Integration and Transformation; Data Reduction: Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Data Discretization and Concept hierarchy generation for numerical and categorical data.

UNIT III: Mining Association Rules in Large Databases: Association Rule Mining, Single-Dimensional Boolean Association Rules, Multi-Level Association Rule, Apriori Algorithm, FP-Growth Algorithm, latest trends in association rules mining.

UNIT IV: Classification methods: Decision tree, Bayesian Classification, Rule based; clustering methods: Partitioning methods(K-Means, K-Mediods) and Hierarchical Clustering (Agglomerative and Divisive Clustering, Multi-phase method) Prediction: Linear and non-linear regression.

Text Books:

1. P.Ponnian, “Data Warehousing Fundamentals”, John Wiley.
2. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann.
3. P. N. Tan, M. Steinbach, Vipin Kumar, “Introduction to Data Mining”, Pearson Education.
4. G. Shmueli, N.R. Patel, P.C. Bruce, “Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner”, Wiley India.
5. Michael Berry and Gordon Linoff “Data Mining Techniques”,Wiley Publications.
6. M.H.Dunham, “Data Mining Introductory & Advanced Topics”, Pearson Education.

Course Outcomes:

- The student will learn basics of Data Warehouse.
- The student will learn basics of Data Mining.
- The student will be able to Preprocess the data.
- The student will be able to perform Market Basket analysis.
- The student will be able to work with Classification Algorithms.
- The student will be able to cluster the data.

DS5B-504: Machine Learning Credits: 3 (2-0-2)

Objective: The main objective is to help students to understand the fundamental concepts of machine learning.

COURSE DESCRIPTION:

UNIT I: Introduction to Machine Learning, History and Overview of machine learning, Applications, Types of Machine Learning, Basic Concepts. Concept Learning and candidate elimination learning Algorithm.

UNIT II: Artificial Neural Network: biological neural network, evolution of artificial neural network, McCulloch-Pitts neuron models, Learning (Supervise & Unsupervised) and activation function. Supervised Learning: Perceptron learning, Single layer/multilayer, linear Separability, Adaline, Madaline, Back propagation network, RBFN.

UNIT III: Bayesian Learning, Bayes Theorem, Naïve Bayesian classifier, Bayesian belief, EM Algorithm. Dimensionality Reduction: Factor Analysis, Principal Component Analysis, Linear Discriminant Analysis.

UNIT IV: Markov and Hidden Markov Models, PAC Learning, Support Vector Machine, Evolutionary Learning: Genetic Algorithm, generating offspring, applications and genetic programming.

Text Books:

1. Tom Mitchell, "Machine Learning", McGraw-Hill, 1997.
2. Stephen Marsland, “Machine Learning –An Algorithmic Perspective”, CRC Press, 2009.
3. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012
4. EthemAlpaydin, “Introduction to Machine Learning”, Prentice Hall of India, 2005.
5. Christopher Bishop, “Pattern Recognition and Machine Learning” Springer, 2006.
6. Sanjeev Kulkarni, Gilbert Harman, “An Elementary Introduction to Statistical Learning Theory”, 2011.
7. N. Shivnandam, “Principle of soft computing”, Wiley.

Course Outcomes:

- The student will be able to understand the basics of machine learning.
- The student will be able to understand Regression analysis.
- The student will be able to work on classification problems.
- The student will be able to work with unsupervised learning approaches.
- The student will be able to perform machine learning operations in scikit-learn.

DS5B-506: Linear Algebra and Advanced Calculus Credits: 3 (2-1-0)

Objective: The main objective is to students will learn to solve many types of data science problems using Linear Algebra and Calculus.

Course Description:

Unit I: Vector spaces over fields, subspaces, bases and dimension. Systems of linear equations, matrices, rank, Gaussian elimination. Linear transformations, representation of linear transformations by matrices, rank-nullity theorem, duality and transpose.

Unit II: Determinants, cofactors, adjoint, Cramer's Rule. Eigenvalues and Eigenvectors, characteristic polynomials, minimal polynomials, Cayley-Hamilton Theorem, triangulation, diagonalization, rational canonical form, Jordan canonical form.

Unit III: Inner product spaces, Gram-Schmidt orthonormalization, orthogonal projections, linear functionals and adjoints, Hermitian, self-adjoint, unitary and normal operators, Spectral Theorem for normal operators, Rayleigh quotient, Min-Max Principle.

Unit IV: Limits and Continuity – Partial derivatives – Total derivative – Differentiation of implicit functions – Jacobian and properties – Taylor's series for functions of two variables – Maxima and minima of functions of two variables – Lagrange's method of undetermined multipliers. Double integrals in cartesian and polar coordinates – Change of order of integration – Area enclosed by plane curves – Change of variables in double integrals – Area of a curved surface - Triple integrals – Volume of Solids.

Text Books:

1. K. Hoffman and R. Kunze, Linear Algebra, Pearson Education (India), 2003. Prentice-Hall of India, 1991.
2. S. Lang, Linear Algebra, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1989.
3. P. Lax, Linear Algebra, John Wiley & Sons, New York,. Indian Ed. 1997
4. H.E. Rose, Linear Algebra, Birkhauser, 2002.
5. Hildebrand, Francis. Advanced Calculus for Applications. 2nd ed. Englewood Cliffs: Prentice Hall, March 31, 1976.
6. Kaplan W., "Advanced Calculus", Addison Wesley (Pearson Education, Inc.), 5 th Edition, 2003
7. Grewal. B.S, "Higher Engineering Mathematics", 41st Edition, Khanna Publications, Delhi, (2011).

Course Outcomes:

After successful completion of this course students will be able to:

1) demonstrate competence with the basic ideas of linear algebra including concepts of vector spaces, linear systems, independence, theory of matrices, linear transformations, bases and dimension, eigenvalues, eigenvectors and diagonalization;

2) describe and apply the key concepts advance calculus;
3) communicate and understand mathematical statements, ideas and results, both verbally and in writing.

DS5B-508: Forecasting Method-II Credits: 3 (2-0-2)

Objective: This subject is designed in such a way to provide the advance concepts of forecasting models based on quantitative and qualitative analysis. Risk and uncertainty in forecasting and also see the cause effect relationship in given scenario and forecast the best possible outcomes.

Unit I:

Technology forecasting: Its meaning and need, alternatives to forecasting, stages of innovation. Delphi: advantages and disadvantages of committees, Delphi procedure, conducting a Delphi based study. Forecasting by analogy. Growth curves-substitution curves, Pearl curve, Gompertz curve, Fisher-Pry curve, selection of proper growth curve, estimation of upper limit.

Unit II:

Trend extrapolation- exponential trends, non-exponential growth, qualitative trends etc. Measures of technology-scoring models, Causal models-technology only models, techno-economic model. Normative methods-relevance trees, morphological models, mission flow diagrams.

Unit III:

Basic concept of model,regression model,regression with ARIMA error, dynamic regression model intervention analysis, multivariate autoregressive models, state space model.

Unit IV:

Uses of forecasting methods in practice, advantages and limitation of forecasting.

Text Books:

Spyros Makridakis ,Forecasting Methods and Application Wiley publication
Joseph P. Martino,Technology forecasting for decision making, McGraw-Hill engineering and technology management series
Stephen A. Delurgio, Forecasting principals and applications,Irwin McGraw-Hill

Course Outcomes:

- Understand the ways of forecasting in the new immerging field of technology.
- Model and Forecast the different possible Trend, growth and seasonal components of a set of values
- Use EViews to perform the key operations needed to obtain Descriptive Statistics and Regression Analysis.
- To understand the different causing variables affecting forecast and model them.
- Combine the Trend, Seasonal and Cyclical components to produce a more accurate forecast of a set of values.

DS5B-510: Big Data Technologies Credits: 3 (2-0-2)

Course Objective: The main objective of this course is to introduce big data technologies such as

Hadoop, spark and analyzing big datasets in spark using python.

COURSE DESCRIPTION:

UNIT I: Introduction
Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting – Modern Data Analytic Tools. Big Data Analytics Process, Big Data Analytics for Business. Identifying problem and solving problem in Big Data environment. Analyzing Unstructured vs. Structured Data, Databases.

UNIT II: Hadoop and MapReduce
Introduction to Hadoop, Hadoop architecture, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystem, Hadoop Releases; Hadoop Distributed File system: Design of HDFS, HDFS Concepts. Introduction to MapReduce: MapReduce Basic Concepts, Understanding the Map Reduce architecture, Writing MapReduce Programs. understanding Map phase, shuffling, sorting, and reducing phase.

UNIT III: Spark
Introduction to Spark, Resilient Distributed Dataset (RDD), RDD Operations: actions and transforamtionfuntions. Spark Dataframes, operations on Dataframes: Join, groupby, aggregate, handling missing data.

UNIT IV:SparkSQL and MLLib
SparkSQL and its basic operations. MLLib: Data types, Basic statistics, Classification(Logistic regression, Decision tree classifier)and linear regression model generation, Model Evaluation, Collaborative filtering, and Clustering.

Practicals:

1. Experiment on Spark RDD operations: actions and transformations
2. Experiment on Spark DataFrame operations: Join, GroupBy, aggregate, handling missing data.
3. Experiment on SparkSQL operations.
4. Develop Logistic Regression Model using Spark MLLib.
5. Develop Decision Tree Classification Model Using MLLib
6. Develop Linear Regression Model using Spark MLLib.
7. Model performance evaluation using Spark MLLib.
8. Develop K-Means Clustering model using SparkMLLib.
9. Develop programs to implement Recommendation Systems using Spark MLLib.
10. Experiment on NoSQL database using MongoDB to create, update and insert.

Text Books:

1. Arvind Sathi, “Big Data Analytics: Disruptive Technologies for Changing theGame”, 1st Edition, IBM Corporation, 2012.
2. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, 1st Edition, Wiley and SAS Business Series, 2012.
3. Tom White,"Hadoop: The Definitive Guide", 3rd Edition, O'Reilly Media, 2012.
4. Donald Miner, Adam Shook, Eric Sammer, “Hadoop Operation”, O’Reilly 2012.
5. Donald Miner, Adam Shook “MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems”, O’Reilly 2012.
6. Chuck Lam, "Hadoop in Action", Manning Publications, 2010.

7. https://spark.apache.org/docs/2.0.0/programming-guide.html

Course Outcomes:
- The student will be able to understand core concepts of Hadoop.
- The student will know the HDFS and MapReduce.
- The student will be able to work with Big Data tool such as Spark.
- The student will be able to analyze big data sets.
- The student will be able to create machine learning models for Big data.

DS5B-512: Python for Analytics Credits: 3 (2-0-2)

Objective: The main objective is to help students to understand the fundamentals of python. Student will learn how to analysis data using Python.

COURSE DESCRIPTION:

UNIT I: Introduction to Python: Python versus Java, Python Interpreter and it's Environment, Python installation, Python basics: variables, operators, Strings, Conditional and Control Statements, loops; Data structures: lists and dictionaries; functions: global functions, local functions, lambda functions and methods.

UNIT II: Object Oriented Programming Concepts: Class, object, constructor, destructor and inheritance; Modules & Packages, File Input and Output, Catching exceptions to deal with bad data, Multithreading, Database Connectivity.

UNIT III: Numpy: Creating Arrays, Arrays Operations, Multidimensional ArraysArrays transformation, Array Concatenation, Array Math Operations, Multidimensional Array and its Operations, Vector and Matrix. **Visualization:** Visualization with matplotlib, Figures and subplots, Labeling and arranging figures, Outputting graphics.

UNIT IV: Pandas: Manipulating data from CSV, Excel, HDF5, and SQL databases, Data analysis and modelling with Pandas, Time-series analysis with Pandas, Using Pandas, the Python data analysis library, Series and Data Frames, Grouping, aggregating and applying, Merging and joining.

Text Books:

1. McKinney Wes, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly Media, 2012.
2. Hauck Trent, "Instant Data Intensive Apps with Pandas How-To", Packt Publishing Ltd, 2013.
3. Beazley David M.,"Advanced Python Programming", Pearson Education,2009.

4. Chun Wesley , Core Python Programming, 3rd Edition,Prentice Hall Professional, 2012.
5. Telles Matt "Python Power!: The Comprehensive Guide", Cengage Learning, 2008.
6. McKinney Wes &PyData Development Team, "pandas: powerful Python data analysis toolkit", Release 0.13.1, Feb 2014.
7. https://docs.python.org/3.4/tutorial/
8. http://www.tutorialspoint.com/python/python_quick_guide.htm

Course Outcomes:

- The student will learn core data types of python.
- The student will learn conditional and looping operations in python.
- The student will be able to work with Object-oriented concepts and Database connectivity in python.
- The student will be able to analyze data using Pandas and Numpy.
- The student will be able to visualize the data using seaborn and matplotlib.

DS5B-522: Web Mining Credits: 3 (2-0-2)

Objective: The main objective of this course s to student will learn how to apply web mining to derive data driven results.

COURSE DESCRIPTION:

UNIT I: Introduction to Web Mining: Web content, web usage mining and web structure mining. Web crawling: Crawling Basics; Indexing, Text analysis and classification.

UNIT II: Similarity and Clustering/community algorithms: Partitioning Approaches, Geometric Embedding Approaches, and Probabilistic Approaches; Topical locality,

UNIT III: Supervised Text Learining: Evaluating Text Classifiers, Nearest Neighbor Learners, Greedy Inclusion Algorithms, Truncation Algorithms, Exploiting Hierarchy among Topics, Discriminative Classification, Regression, Support Vector Machines, Hypertext Classification. SEMISUPERVISED LEARNING: Expectation Maximization, Labeling Hypertext Graphs.

UNIT IV: Link analysis: PageRank and HITS ranking methods, Ranking algorithms; Web search and retrieval, Web growth models, Web traffic models;Social tagging, Social networks and social media, Information diffusion.Applications of Web mining in Recommendtion system.

Text Books:

1. Christopher D. Manning, PrabhakarRaghavan and HinrichSchütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
2. Michael Berry and Gordon Linoff, “Data Mining Techniques for Marketing, Sales, and Customer Relationship Management”, Third Edition, John Wiley, 2011.
3. SoumenChakrabarti,"Mining the Web: Discovering Knowledge from Hypertext Data",Morgan Kaufmann, 2003.

4. Bing Liu “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data”,2011 edition, Springer.
5. Anthony Scime,"Web Mining: Applications and Techniques", Idea Group Inc (IGI), 2005.

Course Outcomes:

At the end of the course the student will be able to:

- understand basics of web mining.
- scrap the web content.
- analyzing web content.
- create NLP models to analyzing the web content.
- understand link analysis and page rank.

DS5B-524: Scientific Computing Credits: 3 (2-1-0)

Objective:

The course provides an appreciation of the need for numerical methods for solving different types of problems, and discusses basic approaches.It developsthe understanding of numerical mathematics or scientific computing - whether in mathematics, the sciences, engineering, or economics.

COURSE DESCRIPTION:

Unit I: Foundation of Scientific Computing, Quantum computing, Wentzel-Kramer-Brillouin Method, Runge-Kutta method, Trapezoidal method.

Unit II: Quasi-linear, Laplace equation, wave packets, Pressure fluctuation, wave phenomena, linearized shallow water wave equation, 1D convection equation, Upwinding, Numerical amplification factor, Stiff differential equation.

UNIT III: Numerical amplification factor, Heat equation, Parabolic partial differential equation, Tridiagonal matrices, Error propagation, Elliptic partial differential equations, Ordinary differential equation, Convergence properties, General elliptic equation, Multigrid method.

Unit IV: Spectral analysis of explicit and implicit, Highlight the scientific and high performance, Taylor series analysis, Buffer domain technique, Aliasing error, Accuracy compact schemes, CCD schemeStabilizing effects of filters, Properties of filters, Scientific elements of a FEM, Lagrange and hermite interpolations, Elliptic equation with linear basis function.

Text Books:

1. Scientific Computing by Michael T Heath, Mc Graw Hill, 2001
2. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 2007
3. Guide to Scientific Computing by Peter R Turner, CRC Press, 2001

Course Outcomes:

After successful completion of the course the students will be able to-

- transform scientific problems into generic computational models;

- have an overview of advanced algorithms for solving a wide range of problems;
- solve mathematical problems by using elementary algorithms, and compute solutions using a structured computer program.
- display and analyse data appropriately, including the results of numerical calculations.
- plan and develop efficient numerical programs.

DS5B-542: Technical Communication Credits: 3 (2-1-0)

Objective:

This course is designed to help the students to develop skills that will enable them to produce clear and effective scientific and technical documents.

Unit I:

Technical writing: Definition, Similarities to other writings, Unique features, importance, technical writing as profession, qualifications for technical writing. Identifying audience. Problems involving content, words and phrases, punctuation, unity, coherence, logic, etc. Being concise.

Unit II:

Techniques of Technical Communication. Analysing-Division, Classification, Partition. Defining-Formal, informal, expanded. Describing-subjective versus objective, spatial description, description of mechanism, process, selected details. Illustrating-Tables, graphs, charts, pictorials. Researching-Basic types of research, original research, searching the literature. Abstracting of your own reports, the works of others, precautions. Oral communication-one to one reporting, participation in conferences, speaking to large audiences, organising the speech.

Unit III:

Basic forms of Technical Writings. The memorandum, The business letter, Formal report.

Unit IV:

Technical reports. Justification reports and proposals, the progress reports, periodic reports, status reports, trip reports. Laboratory reports, Feasibility reports, State-of-the-Art reports. Instructions and manuals.

Text Books:

1. James Sherlock: A Guide to Technical Communication

Course Outcome:

After completion of this course student will be able to-

- Document the knowledge about products, services, technology, or concepts into well-crafted and organised information collateral.
- Write technical reports, memorandum, business letters, manuals, proposals, progress reports etc.
- Develop document involving spatial description, description of mechanism, process, illustrations, etc.

Third Semester:

DS5B-601: Cloud Computing Credits: 3 (2-1-0)

Objective: The main objective of this subject is to help student to understand Cloud and It's Services, Architecture, Deployment, Core Issues, Strengths and limitations of cloud computing.

UNIT I Introduction to Cloud Computing: Overview of Cloud Computing, History of Cloud Computing, Importance of Cloud Computing, advantages and disadvantages of Cloud Computing, Applications, Cloud computing vs. Cluster computing vs. Grid computing, Future of Cloud Computing; Cloud Computing Architecture: Cloud computing stack, Comparison with traditional computing architecture (client/server), Cloud Service Models(XaaS), Deployment Models (Public cloud, Private cloud, Hybrid cloud).

Unit II: Cloud Service Models and Virtualization: Infrastructure as a Service (IaaS): Introduction, Introduction to virtualization, Different approaches to virtualization, Hypervisors, Virtual Machine(VM), Resource Virtualization(Server, Storage, Network),VMware vSphere, Machine Image ,Porting Applications Case study on Amazon EC2;Platform as a Service(PaaS): Introduction to PaaS, advantages and disadvantages of PaaS, case study on Microsoft Azure; Software as a Service(SaaS): Introduction to SaaS, Web services, Web 2.0, Web OS; Development Services and Tools : Amazon Ec2, Google App Engine, IBM Clouds.

UNIT III: Capacity Planning: Defining Baseline and Metrics, Baseline measurements, System metrics, Load testing, Resource ceilings, Server and instance types, Network Capacity, Scaling. Understanding Service Oriented Architecture, Moving Applications to the Cloud, Working with Cloud-Based Storage, Working with Productivity Software.

UNIT IV: Managing the Cloud: Administrating the Clouds, Cloud Management Products; Cloud Security: Security Overview, Cloud Security Challenges and Risks, Security Governance, Risk Management: Security Monitoring, Security Architecture Design, Data Security, Application Security, Virtual Machine Security, Identity Management and Access Control, Authentication in cloud computing, Client access in cloud, Cloud contracting Model. Using the Mobile Cloud: Working with Mobile Devices and Working with Mobile Web Services;

Case Studies on Various Clouds: Google Web Services, Amazon Web Services, Microsoft Cloud Services, IBM Clouds, Eucalyptus.

Text Books:

1. Kris A Jamsa: Cloud computing : SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More 2013Jones & Bartlett Learning ISBN-13: 9781449647391.
2. Michael Miller, Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Que Publishing, August 2008.
3. Toby Velte, Anthony Velte, Robert Elsenpeter, "Cloud Computing, A Practical Approach", TMH, 2009.
4. Sosinsky B., "Cloud Computing Bible", Wiley India
5. Cloud Computing: Principles and Paradigms, Editors: Rajkumar Buyya,James Broberg, Andrzej M. Goscinski, Wiley, 2011.

6. Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Ronald L. Krutz, Russell Dean Vines, Wiley-India, 2010.

Course Outcomes:

- The student will be able to understand the basics of Cloud Computing.
- The student will be able to understand Cloud Computing Models.
- The student will be able to work with AWS Cloud Platform.
- The student will be able to manage cloud platform.
- The student will be able to understand various cloud services.

DS5B-603: Data Visualization Credits: 3 (2-0-2)

Objective: This course will help the student to understand data visualization and how data scientists/analysts use visualization technique to represent results.

COURSE DESCRIPTION:

UNIT I Introduction to Data Visualization, characteristics, goals, need of data visualization, comparison of data science and data visualization, Types of Data, Operations on datatypes, Data Dimensions, Designing Visuals, Visual attributes, Designing Visuals, Mackinlay Design Criteria, Retinal Variables: Size, texture, shape, orientation, color, color Saturation, color hue; Where to use which variable?Bertin's visual attributes, Seven Stages of Visualizing Data.

UNIT II Types of visualization chart: Scatter, line, pie, bar, histogram, Bubble, stacked area chart, pair plot, Heatmaps. Smart Charts: List, Process, Cycle, Hierarchy, Relationships, matrix, pyramid, pictures. Text Visualization.

UNIT III Dashboard, Characteristics, Types of dashboards, best practices and design issues, visual perception, limits of short-term memory, visual encoding data, Gestalt principles, principles of visual perception. Case studies: sales dashboard, CIO dashboard, Telesales dashboard, marketing analysis dashboard.

UNIT IV: Tableau Data Visualization Tools: Data loading, Connecting with Databases, Data Prep with Text and Excel Files, Drill Down and Hierarchies, Sorting, Grouping, Filters, Filtering for Top and Top N, Parameters Formatting; Tableau's Mapping, Custom Geocoding, Polygon Maps, WMS; Dashboards and Stories, Dashboard Layouts and Formatting

Text Books:

1. Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press, 2009.
2. Stephen Few, "Information dashboard design: The effective visual communication of data",O'Reilly, 2006.
3. Edward R. Tufte, "The visual display of quantitative information", Second Edition, GraphicsPress, 2001.
4. Nathan Yau, "Data Points: Visualization that means something", Wiley, 2013.

5. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment",O'Reilly, 2008.
6. Gert H. N. Laursen and JesperThorlund, "Business Analytics for Managers: Taking business intelligence beyond reporting", Wiley, 2010.

Course Outcomes:
- The student will learn basic principles of data visualization.
- The student will learn various types of charts to visualize data.
- The student will understand Dashboard concepts.
- The student will be able to use data visualization tools Tableau/Qlik view.
- The student will be able to tell stories using the dashboard and visualization.

DS5B-605: Deep Learning Credits: 3 (2-1-0)

OBJECTIVE:
This course aims to present the mathematical, statistical and computational challenges of building stable representations for high-dimensional data, such as images, text and data. We will delve into selected topics of Deep Learning, discussing recent models from both supervised and unsupervised learning. Special emphasis will be on convolutional architectures, invariance learning, unsupervised learning and non-convex optimization.

COURSE DESCRIPTION:

Unit-I:
Introduction: Feed forward Neural networks. Gradient descent and the backpropagation algorithm. Unit saturation, aka the vanishing gradient problem, and ways to mitigate it. RelU Heuristics for avoiding bad local minima. Heuristics for faster training. Nestors accelerated gradient descent. Regularization. Dropout. Convolutional Neural Networks: Architectures, convolution / pooling layers. Recurrent Neural Networks: LSTM, GRU, Encoder Decoder architectures.

Unit-II:
Deep Unsupervised Learning: Autoencoders (standard, sparse, denoising, contractive, etc), VariationalAutoencoders, Adversarial Generative Networks, Autoencoder and DBM. Attention and memory models. Dynamic memory networks.
Applications of Deep Learning to Computer Vision: Image segmentation, object detection, automatic image captioning, Image generation with Generative adversarial networks, video to text with LSTM models. Attention models for computer vision tasks.

Unit-III:
Applications of Deep Learning to NLP: Introduction to NLP and Vector Space Model of Semantics. Word Vector Representations: Continuous Skip-Gram Model, Continuous Bag-of-Words model (CBOW), Glove, Evaluations and Applications in word similarity, analogy reasoning. Named Entity Recognition, Opinion Mining using Recurrent Neural Networks.

Unit-IV:

Parsing and Sentiment Analysis using Recursive Neural Networks. Sentence Classification using Convolutional Neural Networks. Dialogue Generation with LSTMs. Applications of Dynamic Memory Networks in NLP. Recent Research in NLP using Deep Learning: Factoid Question Answering, similar question detection, Dialogue topic tracking, Neural Summarization, Smart Reply.

TEXTBOOKS

Bengio, Yoshua, Ian J. Goodfellow, and Aaron Courville. "Deep learning." MIT Press, (2015).
Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends in Machine Learning 2.1 (2009): 1127.
Hochreiter, Sepp, and JargenSchmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 17351780.

Course Outcomes:

After completing the study of the course the students are expected to:

- understand complexity of Deep Learning algorithms and their limitations;
- understand modern notions in data analysis oriented computing;
- be capable of confidently applying common Deep Learning algorithms in practice and implementing their own;
- be capable of performing distributed computations;
- be capable of performing experiments in Deep Learning using real-world data.

DS5B-607: Natural Language Processing Credits: 3 (2-0-2)

Objective: The main object of this course is to provide an introduction to the field of computational linguistics, and natural language processing (NLP).

COURSE DESCRIPTION:

UNIT I: Introduction to NLP, NLP tasks in syntax, semantics, and pragmatics. Applications such as information extraction, question answering, and machine translation. The problem of ambiguity. The role of machine learning. Brief history of the field.

UNIT II: The role of language models. Simple N-gram models. Estimating parameters and smoothing. Evaluating language models. Lexical syntax. Hidden Markov Models. Maximum Entropy Models. Conditional Random Fields.

UNIT III: Grammar formalisms and treebanks. Efficient parsing for context-free grammars (CFGs). Statistical parsing and probabilistic CFGs (PCFGs). Lexicalized PCFGs. Lexical semantics and word-sense disambiguation. Compositional semantics. Semantic Role Labeling and Semantic Parsing.

UNIT IV: Named entity recognition and relation extraction. IE using sequence labeling. Basic issues in MT. Statistical translation, word alignment, phrase-based translation, and synchronous grammars. Web 2.0 Applications : Sentiment Analysis; Text Entailment.

Text Books:

1. Allen, James, Natural Language Understanding, Second Edition, Benjamin/Cumming, 1995.
2. Charniack, Eugene, Statistical Language Learning, MIT Press, 1993.
3. Jurafsky, Dan and Martin, James, Speech and Language Processing, Second Edition, Prentice Hall, 2008.
4. Steven Bird, Ewan Klein, and Edward Loper,"Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit", O'Reilly Media, 2009.
5. Ian H. Witten and Eibe Frank. "Data Mining: Practical Machine Learning Tools and Techniques" , 3rd edition, Morgan Kaufmann, 2005.
6. Manning, Christopher and Heinrich, Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.

Course Outcomes:

After successful completion of this course, the student will be able to:

1. Broad understanding of the field of Natural Language Processing.
2. Understand mathematical concepts for NLP algorithms.
3. Understanding the capabilities and limitations of NLP technologies.
4. Apply the fundamental knowledge of various types of basic NLP techniques to analyze, design, formulate and implement solutions for any real time situation.
5. Understand the theoretical concepts of NLP in formal language theory.

DS5B-621: Cluster Analysis Credits: 4 (3-1-0)

Objective: The purpose of this course is to provide fundamental knowledge of Clustering Methods, and develop skill for reporting analysis studies.

UNIT I Introduction to Classification and Clustering, Requirements for Cluster Analysis, Overview of Basic Clustering Methods, Concept of Similarity, Similarity Measures Basic Clustering Methods: Iterative Partition Clustering: k-means, K-Modes, k-medoids algorithms, CLARA, CLARANS;

UNIT II Hierarchical Clustering: Agglomerative, Divisive and Multi-Phase:BIRCH, Chameleon;Density Based Clustering Method: DBSCAN, OPTICS, DENCLUE; Grid-Based Methods: STING, CLIQUE;

UNIT III Advanced Clustering Methods: Probabilistic Model-Based Clustering: Fuzzy Cluster, Probabilistic Model-Based Clusters, Expectation-Maximization Algorithm;Validation Techniques: Cophenetic Correlation, Significance Tests on Variables Used to Create Clusters, Replication,Significance Tests on External Variables, Monte Carlo Procedures.

UNIT IV

Determining the Number of Clusters: Elbow criterion; Measuring Cluster Quality and Robustness; Displaying cluster solutions graphically. Cluster Analysis Software, Statistical Packages Containing Clustering Software, Reporting Cluster Analysis Studies.

Text Books:

1. Charles Romesburg,"Cluster Analysis for Researchers", LULU Press, 2004.
2. Mark S. Aldenderfer, Roger K. Blashfield, "Cluster analysis", Sage Publications, 01-Nov-1984.
3. Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl, "Cluster Analysis", John Wiley & Sons, 14-Jan-2011.
4. Brian S. Everitt, Graham Dunn," Applied Multivariate Data Analysis", Wiley, 28-Jun-2010
5. Bryan F.J. Manly, "Multivariate Statistical Methods: A Primer, Third Edition", CRC Press, 06-Jul-2004.
6. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann.

Course Outcomes:

- The student will be able to understand basics of clustering.
- The student will be able to understand hierarchical clustering.
- The student will be able to understand advanced clustering approaches.
- The student will be able to assess the clustering performance.
- The student will be able to perform clustering operations in R/Python.

DS5B-623: Multivariate Analysis Credits: 4 (2-1-2)

Objective:

1. To understand the main features of multivariate data.
2. To be able to use exploratory and confirmatory multivariate statistical methods properly.
3. To be able to carry out multivariate statistical techniques and methods efficiently and effectively.

COURSE DESCRIPTION:

UNIT I: Analysis of categorical data. Loglinear models for two- and higher-dimensional contingency tables, Characterizing and Displaying Multivariate Data, Tests on one or two mean vectors.

UNIT II: Multivariate Analysis of Variance, Aspects of multivariate analysis, random vectors, sample geometry and random sampling, multivariate normal distribution, inferences about the mean vector, MANOVA.

UNIT III: Classification and grouping techniques: discrimination and classification, clustering. Logistic regression models.

UNIT IV: Analysis of covariance structures, Principal Component Analysis (PCA), Factor Analysis, Cluster Analysis. Use of statistical computer packages.

Text Books:

1. Applied Multivariate Statistical Analysis, by Richard A. Johnson and Dean W.Wichern (6^{th} edition), Prentice Hall.
2. Characterizing and Displaying Multivariate Data, by A.C.Rencher, John Wiley and Sons.
3. Multivariate Data Analysis, by Joseph F.Hair, William, Babin and Anderson.
4. Cluster Analysis, by Brian S.Everitt, Sabine Landau, Morven Leese. Wiley

Course Outcomes:

- To be able to understand the concept of analysing multivariate data.
- To be familiar with a basic minimum level of matrix competency and with general aspects of handling multivariate data.
- Perform exploratory analysis of multivariate data, such as plot multivariate data, calculating descriptive statistics, testing for multivariate normality;
- Conduct statistical inference about multivariate means including hypothesis testing, confidence ellipsoid calculation and different types of confidence intervals estimation;
- Undertake statistical analyses using appropriate multivariate techniques, which includes principal component, factor analysis, discriminate and clustering analysis
- Analyse multivariate data using the statistical software package.