2019

MSc DATA SCIENCE

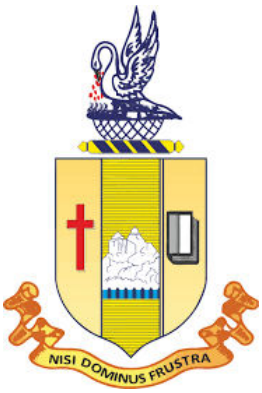
Applicable to Candidates admitted from the
Academic Year 2019-2020 onwards

DEPARTMENT OF DATA SCIENCE

BISHOP HEBER COLLEGE (AUTO), TIRUCHIRAPPALLI

MSc Data Science

Applicable to Candidates admitted from the
Academic Year 2019-2020 onwards



Department of Data Science
Bishop Heber College (Autonomous)
Tiruchirappalli 620017

BISHOP HEBER COLLEGE (AUTONOMOUS), TIRUCHIRAPPALLI-620 017
M. Sc., Data Science
(Applicable to Candidates admitted from the Academic Year 2019-2020 onwards)

Sem	Course	Course Title	Course Code	Hours / Week	Credits	Marks		
						CIA	ESE	Total
I	Core I	Mathematical Foundation for Data Science	P19DS101	5	4	25	75	100
	Core II	Problem Solving using Python and R	P19DS102	5	4	25	75	100
	Core III	NoSQL Database Management	P19DS103	5	4	25	75	100
	Elective I	Probability and Statistical Methods Design and Analysis of Algorithms Software Engineering	P19DS1:1 P19DS1:A P19DS1:B	5	4	25	75	100
	Core Practical I	Problem Solving using Python and R Lab	P19DS1P1	5	3	40	60	100
	Core Practical II	NoSQL Database Management Lab	P19DS1P2	5	3	40	60	100
II	Core IV	Regression Analysis	P19DS204	4	4	25	75	100
	Core V	Data and Visual Analytics	P19DS205	4	4	25	75	100
	Core VI	Practical Machine Learning	P19DS206	4	4	25	75	100
	Elective II	Natural Language Processing Multivariate Analysis Basics of Bioinformatics	P19DS2:2 P19DS2:A P19DS2:B	4	4	25	75	100
	Elective III	Health Care Data Analytics Customer Relationship Management	P19DS2:3 P19DS2:C	3	3	25	75	100
	Core Practical III	Data and Visual Analytics Lab	P19DS2P3	3	3	40	60	100
	Core Practical IV	Practical Machine Learning Lab	P19DS2P4	3	3	40	60	100
	Core Practical V	Natural Language Processing Lab	P19DS2P5	3	3	40	60	100
	VLO	RI/MI	P19VL2:1 P19VL2:2	2	2	25	75	100
III	Core VII	Time Series Analysis and Forecasting	P19DS307	5	4	25	75	100
	Core VIII	Big Data Management and Analytics	P19DS308	5	4	25	75	100
	Core IX	Social Media Analytics	P19DS309	4	4	25	75	100
	Elective IV	Image and Video Analytics Computational Genomics	P19DS4:1 P19DS4:A	4	3	25	75	100
	Core Practical VI	Big Data Management and Analytics Lab	P19DS3P6	5	3	40	60	100
	Core Practical VII	Social Media Analytics Lab	P19DS3P7	5	3	40	60	100
	Core Project-I	Project Preparation	P19DS3PJ	2	---	---	---	30
IV	Core X	Principles of Deep Learning	P19DS410	5	4	25	75	100
	Core XI	Web Development using Python	P19DS411	5	4	25	75	100
	Elective V	Supply Chain Management Internet of Things	P19DS5:1 P19DS5:A	5	4	25	75	100
	Core Project-II	Project Implementation	P19DS4PJ	15	5	---	---	70
			Total Credits		90			

Core-I: Mathematical Foundation for Data Science

Course Objectives
Linear Algebra, calculus and graph theory play a fundamental role in the theory of Data Science. This course aims at introducing the basic notions of vector spaces, Linear Algebra, basic notions of multivariable calculus and graph theory in the applications to Data Science.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Understand the properties of Vector spaces
LO2: Use the properties of Linear Maps in solving problems on Linear Algebra
LO3: Demonstrate proficiency on the topics Eigenvalues, Eigenvectors and Inner ProductSpaces
LO4: Understand the properties of multivariable calculus and graphs
LO5: Apply mathematics for some applications in Data Science

Unit-1. Introduction to Vector Spaces
Vector Spaces: R^n and C^n, lists, $F^n$and digression on Fields, Definition of Vector spaces, Subspaces, sums of Subspaces, Direct Sums, Span and Linear Independence, bases, dimension.

Unit-2. Linear Maps
Definition of Linear Maps - Algebraic Operations on - Null spaces and Injectivity - Range and Subjectivity - Fundamental Theorems of Linear Maps - Representing a Linear Map by a Matrix - Invertible Linear Maps - Isomorphic Vector spaces - Linear Map as Matrix Multiplication - Operators - Products of Vector Spaces - Product of Direct Sum - Quotients of Vector spaces.

Unit-3. Eigen Values, Eigen Vectors and Inner Product Spaces
Eigenvalues and Eigenvectors - Eigenvectors and Upper Triangular matrices – Eigenspaces and Diagonal Matrices - Inequalities on Linear Spaces - Norms on Linear Spaces - Inner products - Orthogonality - Unitary and Orthogonal Matrices - Norms for matrices

Unit-4. Calculus of several variablesand basic Graph Theory
Functions of Several Variables - Limits and continuity in Higher Dimensions – Partial Derivatives - The Chain Rule - Directional Derivative and Gradient vectors - Tangent Planes and Differentials - Extreme Values and Saddle Points - Lagrange Multipliers. Graphs - subgraphs - factors - Paths - cycles - connectedness - trees - Euler tours -Hamiltonian cycles - Planar Graphs - Digraphs.

Unit-5. Mathematics applied to Data Science
Singular value decomposition - Handwritten digits and simple algorithm - Classification of handwritten digits using SVD bases - Tangent distance - Text Mining.

Text Books
1. S. Axler, Linear algebra done right, Springer, 2017.
2. Eldén Lars, Matrix methods in data mining and pattern recognition, Society for Industrial and Applied Mathematics, 2007.
3. M. D. Weir, J. Hass, and G. B. Thomas, Thomas' calculus. Pearson, 2016.
4. D. Jungnickel, Graphs, networks and algorithms. Springer, 2014.

References
1. E. Davis, Linear algebra and probability for computer science applications, CRC Press,2012.
2. J. V. Kepner and J. R. Gilbert, Graph algorithms in the language of linear algebra, Society for Industrial and Applied Mathematics, 2011.
3. D. A. Simovici, Linear algebra tools for data mining, World Scientific Publishing, 2012.
4. P. N. Klein, Coding the matrix: linear algebra through applications to computer science, Newtonian Press, 2015.
5. J. Patterson and A. Gibson, Deep learning: a practitioner's approach. O'Reilly Media, 2017.
6. S. Sra, S. Nowozin, and S. J. Wright, Optimization for machine learning. MIT Press, 2012.

Core-II: Problem Solving Using Python and R

Course Objectives
This course introduces students the language features of both Python and R programming languages. Specifically, data structures, regular expressions and data visualization features are introduced.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Write simple Python programs using Python data structures
LO2: Develop object oriented programs in Python
LO3: Manipulate files using Python
LO4: Access internet and database data
LO5: Write R programs for data visualization

Unit-1. Python Basics, Functions, Loops and Strings
Variables – Getting Inputs – Conditions – Catching exceptions – Function calls – Built-in functions – Type conversion functions and math functions – Parameters and arguments –While statement – Infinite loops – Continue statement – For loops – Strings – Slice –– The in operator – String comparison – String methods – parsing strings – Format operator.

Unit-2. Files and Lists
Opening files – Text files – Reading files – Searching through files – Writing files – Traversing list – List operations – List slice – List methods – Deleting elements – Built-in list functions – Objects, value and aliasing – List arguments.

Unit-3. Dictionaries, Tuples and OOP
Dictionaries – Files and dictionaries – Looping and dictionaries – Tuples – Comparing tuples – Tuple assignments – Dictionaries and tuples – Tuples as keys in dictionaries – Creating objects – Encapsulation – Classes as types – Object lifecycle – Instances – Inheritance.

Unit-4. Internet Programming
Regular expressions – Character matching – Extracting data – Escape character – Designing simple web browser using sockets – Retrieving images using HTTP – Retrieving web pages using urllib – Reading binary files using urllib – Accessing data from databases

Unit-5. Programming with R
Variables - Vector, matrix, arrays – List – Data Frames – Functions – Strings – Factors – Loops – Packages – Date and Time – Files - Making packages

Text Books
1. Allen B. Downey, —Think Python: How to Think like a Computer Scientist, 2ndedition, Updated for Python 3, O Reilly Publishers, 2016
2. Charles R. Severance, Python for Everybody: "*Exploring data using Python 3*",Schroff Publishers, 1ed, 2017, ISBN 978-9352136278.
3. Richard Cotton, "Learning R", O'Reilly, 2013

References
1. Zed Shaw's , Learn Python the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful World of Computers and Code, Addison-Wesley Professional; 3 edition, 2013
2. Robert Sedgewick, Kevin Wayne, Robert Dondero, Introduction to Programming in Python: An Inter - disciplinary Approach, Pearson India Education Services Pvt. Ltd., 2016.
3. Wesley J Chun, Core Python Programming , 2nd edition, Prentice Hall ,2009
4. Colin Gillespie, Robin Lovelace, and Efficient R Programming: A Practical Guide to Smarter Programming, "O'Reilly Media, Inc.", 2016
5. Paul Teetor, R Cookbook-Proven Recipes for Data Analysis, Statistics, and Graphics, O'Reilly Media, 2011

Core-III: NoSQL Database Management

Course Objectives

The widespread emergence of big data storage needs has driven the development and adoption of a new class of non-relational databases commonly referred to as NoSQL databases. This course will explore the origins of NoSQL databases and the characteristics that distinguish them from traditional relational database management systems. Core concepts of NoSQL databases will be presented, followed by an exploration of how different database technologies implement these core concepts.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Model data using ER diagrams
LO2: Demonstrate competency in designing NoSQL database management systems.
LO3: Demonstrate competency in describing how NoSQL databases differ from relational databases from a theoretical perspective.
LO4: Demonstrate competency in selecting a particular NoSQL database for specific use cases.
LO5: Implement databases using MongoDB

Unit-1. Data Modeling

DBMS: terminologies, components, roles, advantages and disadvantages – Database architectures: teleprocessing, file server, 2-tier, 3-tier, N-tier, middleware and Transaction processing monitor – Software components of DBMS and Database Manager – Data modeling using ER diagram: Entity, relationship, attributes, keys, strong and weak entities, attributes on relationships, relationship types, cardinality and participation

Unit-2. Structured Query Language

SQL statements: SELECT, WHERE, ORDERBY, GROUPBY and HAVING clauses - Sub Queries – ANY and ALL – JOIN – inner and outer joins – EXISTS and NON EXISTS – UNION, INTERSECT and EXCEPT – Updating databases: INSERT, UPDATE and DELETE – SQL data types – Creating, altering and removing tables – Indexes and views: CREATE and REMOVE

Unit-3. NoSQL Database Theory

Why NoSQL – Value of Relational Database – Emergence of NoSQL – Aggregate data models – More details on data models: Relationships, Graphs DB, Schemaless DB, Materialized views – Distribution models: Single server, shrading, replication – Consistency: Update, read, relax consistency

Unit-4. NoSQL Databases

Key value databases: What is Key Value store, Features of Key value DB, Suitable use cases, When not to use it – Document databases: Definition, features, Suitable use cases, when not to use – Column family stores: Definition, features, suitable use cases, when not to use – Graph databases: Definition, features, use case, when not to use – Schema migration – Polyglot persistence - Beyond NoSQL – Choosing your database

Unit-5. MongoDB

Document – Collection – Database - Datatypes – Creating, deleting, updating documents – Querying – Indexing – Aggregation: Pipeline, Aggregation commands – Application design

Text Books

1. Thomas M. Connolly and Carolyn E. Begg. Database Systems: *"A Practical Approach to Design, Implementation, and Management",* 6th Edition, Pearson, 2015.
2. Pramod J. Sadalage; Martin Fowler. *NoSQL Distilled: A Brief Guide to the Emerging World of PolyglotPersistence*. Addison-Wesley. 2012 ISBN: 0321826620
3. Kristina Chodorow, MongoDB: The Definitive Guide, 2ed, Oreilly Publishers

References

1. Eric Redmond; Jim R. Wilson. *Seven Databases in Seven Weeks: A Guide to Modern Databases and theNoSQL Movement*. Pragmatic Bookshelf. 2012. ISBN: 1934356921

Elective-1A: Probability and Statistical Methods

Course Objectives
This course introduces the concepts of probability and statistics that are required to design and implementthe Data Science applications and systems.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Explain basic probabilistic and statistical models and illustrate their related applications
LO2: Estimate the likelihood of events from population
LO3: Propose, test and evaluate hypothesis

Unit-1. Probability
Probability Spaces-Combinatorial methods (or) Counting techniques-Elementary Theorem –Conditional Probability –Bayes‟ theorem-Probability Distributions and Probability Densities.

Unit-2. Expectation and Regression
Mathematical Expectation: Expected value-Moments-Chebyshev's theorem-Moment Generating functions-Product Moment-Conditional Expectation-Special Probability Distributions and Probability Densities-Functions of Random Variable.Multiple regression-Linear models-Logistic regression-Rates and Poisson regression-Nonlinear curve fitting –correlation.

Unit-3. Distribution
Descriptive Statistics & Sampling Distributions: Population-Sampling-Measures of Central tendency, variations and position –Sampling distributions: Standard Normal Distribution-Chi-Square Distribution-t-Distribution –F-Distribution -The Central Limit Theorem.

Unit-4. Estimation
Estimation: Point Estimation: the method of moments and the method of maximum likelihood estimation-Interval estimation: estimation of mean, estimation of difference of means, estimation of variance and estimation of ratio of variances.

Unit-5. Hypothesis Testing
Test of Hypothesis-Testing for Attributes –Mean of Normal Population –One-tailed and two-tailed tests, F-test and Chi-Square test –Analysis of Variance-Nonparametric test.

Text Books
1. John.E.Freund, Irwin Miller, Marylees Miller "Mathematical Statistics with Applications ", 8th, Prentice Hall of India, 2014
2. Yannisviniotis, "Probability and Random Processes for electrical engineers", McGraw-Hill International Edition, 1997.
3. Ross, Sheldon. M, "Introduction to Probability and Statistics for Engineers and Scientists", Academic Press, 2009

References
1. D. C. Montgomery and G. C. Runger, Applied Statistics and Probability for Engineers, 6thEdition, Wiley, 2014
2. G.K. Bhattacharya and R.A. Johnson, Statistical Concepts & Methods, 6thEdition, Wiley, 2010

Elective-1B: Design and Analysis of Algorithms

Course Objectives
This course introduces students the concepts of designing and analysing algorithms, sorting data, greedy algorithms and graph algorithms.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Analyse the asymptotic performance of algorithms.
LO2: Solve problems using key techniques of algorithm design.
LO3: Develop optimal solution by applying various methods and differentiate polynomial and non-polynomial problems.

Unit-1. Basics of Algorithm Analysis
Basics of Algorithms and Mathematics: Introduction to an algorithm, Mathematics for Algorithmic Sets, Linear Inequalities and Linear Equations. The efficient algorithm, Average, Best and worst-case analysis of Time Complexity and Space Complexity. **Analysis of Algorithm:** Amortized analysis, Asymptotic Notations, analysing control statement, Loop invariant and the correctness of the algorithm, Master's Theorem. **Sorting Algorithms and analysis:** Bubble sort, Selection sort, Insertion sort, Shell sort, Heap sort, sorting in linear time: Bucket sort, Radix sort and Counting sort

Unit-2.Divide and Conquer Algorithm
Introduction, Recurrence and different methods to solve recurrence, multiplying large Integers Problem, Problem Solving using divide and conquer algorithm.Binary Search, Max-Min problem, Sorting (Merge Sort, Quick Sort), Matrix Multiplication.

Unit-3. Dynamic Programming and Greedy Algorithm
Dynamic Programming: Introduction, The Principle of Optimality, Problem Solving using Dynamic Programming–Calculating the Binomial Coefficient, Assembly Line-Scheduling. Matrix chain multiplication, Longest Common Subsequence, All Points Shortest path.**Greedy Algorithm:** General Characteristics, Problem solving using Greedy Algorithm- Activity selection problem. Minimum Spanning trees (Kruskal's algorithm, Prim's algorithm).

Unit-4. Graphs
Graphs: Shortest paths, The Knapsack Problem, Job Scheduling Problem, Huffman code.Exploring Graphs Algorithms: - Applications of DFS- bi-connectivity, topological sort, Articulation point, Connected components. **Backtracking and Branch and Bound:**Introduction, The Eight queen's problem, Minimax principle.

Unit-5. Selected Problems in Algorithms
String Matching: Introduction, The naive string matching algorithm, The Rabin Karp algorithm, String Matching with finite automata, The Knuth Morris-Pratt algorithm. **Approximation algorithms:** Travelling Salesman problem, Hamiltonian problem, Vertex Cover Problem. **Introduction to NP-Completeness:** The class P and NP, Polynomial reduction, NP Completeness Problem, NP-Hard Problems.

Text Books
1. Cormen, Leiserson, Rivest, and Stein, Introduction to Algorithms (3rd Edition), MIT Press, 2009. ISBN-978-0262033848.
2. Jon Kleinberg and Eva Tardos, Algorithm Design, 1ed, Pearson, 2005. ISBN- 978-0321295354.

References
1. V. Aho, J. E. Hopcroft and J. D. Ullman, Data Structures and Algorithms, (Addison Wesley), Pearson, 1983. ISBN- 978-0201000238.

Elective-1C: Software Engineering

Course Objectives

This course gives students an exposure to software engineering models and methods, version controling using GitHub, designing UML diagrams, creating Software Requirements Specification (SRS) in both plan based as well as in agile development environment.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Demonstrate understanding of Software Engineering as an iterative and systematic process.
LO2: Apply development life cycle through version control system, IDE and UML
LO3: Design software development process to complement technical understanding of SW products.

Unit-1. Software Process Models

Importance of Software Engineering, Discipline of Software Engineering; Eclipse Introduction, Overview, and Demo; Lifecycle models: Requirements Engineering, Design and Implementation, Maintenance, Software Process Model Introduction, Waterfall Process, Spiral Process, Evolutionary Prototyping Process, Agile Process, Choosing a Model, Lifecycle Documents; Version Control System: Introduction to Git, Git Demo: Git + Eclipse, Git.

Unit-2. Requirements Analysis and UML

Requirements Engineering: General RE Definition, Functional and Non-functional Requirements, User and System Requirements, Modelling Requirements, Analysing Requirements, Requirements Prioritization, Requirements Engineering Process and steps; Creating SRS and performing requirements inspections. **OO Software and UML:** Object Orientation Introduction, UML Structural Diagrams: Class Diagrams, Component Diagram, UML Structural Diagram: Deployment Diagram. UML creation tips; UML Behavioural Diagram: Use Case, Use Case Diagram: Creation Tips, UML Behavioural Diagrams: Sequence, UML Behavioural Diagrams: State Transition Diagram. UML creation tips;

Unit-3. Software Architecture

Software Architecture: What is Software Architecture? Advantages and use of architectural models. Architectural patterns. Designing architectural patterns. Design Patterns: Patterns Catalogue, Pattern Format, Factory Method Pattern, Strategy Pattern, Choosing a Pattern, Negative Design Patterns.

Unit-4. Software Testing

Software Testing: Black Box Testing Failure, Fault and Error, Verification Approaches, Pros and Cons of Approaches, Testing Introduction, Testing Granularity Levels, Alpha and Beta Testing, Black-Box Testing, Systematic Functional Testing Approach; Test Data Selection, Equivalence Partitioning and Boundary Value Analysis, Create and Evaluate Test Case Specifications, Generate Test Cases from Test Case Specifications. White-Box Testing: Coverage Criteria Intro, Statement Coverage, Control Flow Graphs, Test Criteria, MC/DC Coverage.

Unit-5. Agile Development

Agile Development Methods: Cost of Change, Agile Software Development, Extreme Programming (XP), XP's Values and Principles, Test First Development, Refactoring, Pair Programming, Continuous Integration, Testing Strategy, High Level Scrum Process. Unified Software Process: Use-Case Driven, Inception Phase, Elaboration Phase, Construction Phase, Transition Phase, Phases and Iterations. **Software Evolution:** Evolution processes, Legacy Systems, Software Maintenance. Situations during software evolution and maintenance. Software **Reengineering and Refactoring:** Reasons to Reengineer and Refactor, Advantages, Refactoring Demo, Refactoring Risks, Cost of Refactoring, When Not to Refactor.

Text Books

1. R. Pressman, Software Engineering, A Practitioner's Approach, 7ed, McGraw Hill, 2014. ISBN-9789339212087.
2. Sommerville, Software Engineering, 10ed, Pearson, 2015. ISBN- 978-0133943030.

References

1. *Software Engineering Essentials, https://www.edx.org/course/software-engineering-essentials-tumx-seecx*
2. Tutorials *Point for Design Pattern, https://www.tutorialspoint.com/design_pattern/*

Core Practical I: Problem Solving Using Python and R Lab

Course Objectives

This lab course enables students to master the language features of both Python and R programming languages. Specifically, data structures, regular expressions, files and data visualization features are introduced.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Write simple Python programs using Python data structures
LO2: Develop object oriented programs in Python
LO3: Manipulate files using Python
LO4: Access internet and database data
LO5: Write R programs for data visualization

Lab Exercises

Develop applications that will demonstrate the following Python and R programming features

- Functions
- Strings
- Lists
- Dictionaries
- Tuples
- Files
- Regular Expressions
- OOP
- Internet Programming
- Data visualization in R

Core Practical II: NoSQL Database Management Lab

Course Objectives
This lab course will explore the features of SQL and NoSQL languages for solving business applications.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Model data using ER diagrams
LO2: Create queries of SQL and NoSQL
LO3: Implement databases using any RDBMS and MongoDB

Lab Exercises

Develop applications that will demonstrate the following concepts using SQL and NoSQL

- Drawing E-R Diagrams for the given business problem
- Designing tables from E-R Diagrams
- Creating tables using SQL and NoSQL features
- Developing queries and subqueries
- Performing Join operations
- Creating Indexes
- Solving real world business use cases

Core-IV: Regression Analysis

Course Objectives
This course introduces students the concepts of regression and how to buildsimple and multiple regression models.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Develop a deeper understanding of the linear regression model.
LO2: Learn about R-square criteria for model selection
LO3: Understand the forward, backward and stepwise methods for selecting the variables
LO4: Understand the importance of multicollinearity in regression modeling
LO5: Ability to use and understand generalizations of the linear model to binary and count data

Unit-1. Simple Linear Regression
Introduction to regression analysis: Modelling a response, overview and applications of regression analysis, major steps in regression analysis. Simple linear regression (Two variables): assumptions, estimation and properties of regression coefficients, significance and confidence intervals of regression coefficients, measuring the quality of the fit.

Unit-2. Multiple Linear Regression
Multiple linear regression model: assumptions, ordinary least square estimation of regression coefficients, interpretation and properties of regression coefficient, significance and confidence intervals of regression coefficients.

Unit-3. Criteria for Model Selection
Mean Square error criteria, R2 and $\hat{R}^2$ criteria for model selection; Need of the transformation of variables; Box-Cox transformation; Forward, Backward and Stepwise procedures.

Unit-4. Residual Analysis
Residual analysis, Departures from underlying assumptions, Effect of outliers, Collinearity,Non-constant variance and serial correlation, Departures from normality, Diagnostics and remedies.

Unit-5. Non Linear Regression
Introduction to nonlinear regression, Least squares in the nonlinear case and estimation of parameters, Models for binary response variables, estimation and diagnosis methods for logistic and Poisson regressions. Prediction and residual analysis.

Text Books
1. D.C Montgomery, E.A Peck and G.G Vining, Introduction to Linear Regression Analysis, John Wiley and Sons,Inc.NY, 2003.
2. S. Chatterjee and AHadi, Regression Analysis by Example, 4th Ed., John Wiley and Sons,Inc, 2006
3. Seber, A.F. and Lee, A.J. (2003) Linear Regression Analysis, John Wiley, Relevantsections from chapters 3, 4, 5, 6, 7, 9, 10.

References
1. Iain Pardoe, Applied Regression Modeling, John Wiley and Sons, Inc, 2012.
2. P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman & Hall, 1989.

Core-V: Data and Visual Analytics

Course Objectives

This course introduces students the concepts of data and visual analytics such as array processing, data wrangling and time series plotting using Python language and its libraries.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Use Jupyter Notebooks for interactive computation
LO2: Apply NumPy functions for array processing
LO3: Create and use Series and Data Frames for data wrangling
LO4: Create various plots using Matplotlib
LO5: Perform data aggregation and group operations
LO6: Use Date and Time classes to create time series plots

Unit-1. NumPy and Pandas Basics

Why Python for Data analysis – Essential Python libraries – ndarray – Universal functions – Data processing using arrays – File I/O with arrays – Random number generation – Series, Data Frames – Indexing, re-indexing, sorting, ranking –Summarizing descriptive statistics – Handling missing data – Hierarchical indexing

Unit-2. Data Loading and Wrangling

Data Loading: reading and storing data in text format, binary format – Data Wrangling: Combining and merging data sets – Reshaping – Pivoting – Data transformation – String manipulation

Unit-3. Plotting and Visualization using Matplotlib

Figures – Subplots – Colors – Ticks – Label – Legends – Annotation – Saving plots to file – Plots: Line, Bar, Histogram, Density Plots – Scatter Plots

Unit-4. Data Aggregation and Group Operations

Iterating over groups – Selecting columns – Grouping with Series and functions –Data aggregation: Column wise aggregation, returning aggregated data – General-Split-Apply-Combine – Quantile and bucket analysis – Pivot table and cross tabulation

Unit-5. Time Series

Date and Time – Time Series – Date Range, Frequencies and Shifting – Periods and period arithmetic – Resampling and frequency coversion – Time Series Plotting

Text Books

1. Wes. Mc Kinney, "*Python for Data Analysis"*, 2nd Edition, Schroff Publishers, 2013. ISBN 9789352136414

References

1. CyrilleRossant. "Learning I*Python for interactive Computing and data visualization"*, First edition [Packt]
2. Jake VanderPlas ,Python Data Science Handbook - Essential Tools for Working with Data, O'Reily, 2016
3. Zhang.Y ,An Introduction to Python and Computer Programming, Springer Publications,2016

Core-VI: Practical Machine Learning

Course Objectives
This course introduces students the various ML models for supervised and unsupervised ML problems. Also, it explains all methods for pre-processing, dimensionality reduction and feature selection. Various methods for evaluating the developed ML model are discussed.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Learn and practice Scikit-Learn Python Library
LO2: Apply pre-processing methods and reduce dimensions of data
LO3: Design a ML model for the given business problem
LO4: Perform training, testing and evaluation of the designed ML model

Unit-1. ML Basics and Perceptron
Three types Machine Learning – Three steps of ML process – Perceptron neural network – Adaline neural network – Stochastic gradient descent neural network

Unit-2. Supervised Learning classifiers
Logistic regression – Support vector machines – Kernel SVM – Decision Trees – K-Nearest Neighbour classifier – Random Forest – Linear Regression–Sentiment Analysis of Movie Reviews using Logistic Regression - Developing a web application with Flask

Unit-3. Pre-processing and Dimensionality Reduction
Pre-processing: Missing data, categorical data, feature scaling, feature selection. **Dimensionality reduction:** Principal Component Analysis, Linear Discriminant Analysis, Kernel PCA

Unit-4. Model evaluation
Pipelines - K-fold cross validation - Grid search - Confusion matrix, Precision, Recall, ROC curves, Scoring metrics –Majority vote classifier – Bagging, Bootstrapping, Adaptive Boosting

Unit-5. Unsupervised Learning classifiers and Multilayer NN
K-Means, K-Means++, Finding optimal no. of classifiers - Agglomerative Hierarchical clustering, Density based clustering -Multilayer Neural Network: Feed forward, Back Propagation Training, Multilayer Perceptron

Text Books
1. Sebastian Raschka, ***"Python Machine Learning"***, First Edition, [PACKT], 2015.

References
1. Andreas C Muller and Sarah Guido, ***Introduction to Machine Learning with Python***, Shroff Publishers, ISBN 978935213451
2. Joel Grus, ***"Data Science from Scratch"***, First Edition, O'Reilly,2015
3. Gavin Hackeling, "***Mastering machine learning with scikit-learn"***, First Edition, [PACKT] , 2014

Elective-2: Natural Language Processing

Course Objectives
This course provides an introduction to the field of natural language processing, synthesizing research from linguistics and computer science. The course will cover formal models for representing and analyzing syntax and semantics of words, sentences, and documents.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Understand NLP pipeline
LO2: Categorize and tag words
LO3: Classify text using supervised classifiers
LO4: Extract entities and named entities
LO5: Analyze sentences and their meanings

Unit-1.NLP Basics
What is NLP - Frequency distributions, Collocations – Unigram, Bigrams – word sense disambiguation – Pronoun resolution – Machine Translation – Textual entilement – Limitations of NLP –Conditional frequency distributions – Plotting distributions – Lexicons – WordNet – Semantic similarity

Unit-2. Text Processing
Accessing text from web and disk – NLP Pipeline – String processing – Text processing with UNICODE – Regular expressions: Metacharacters, Ranges and Closure – Useful applications of Regular applications – Stemming and Lemmatization – Text tokenization using regular expressions – Segmentation –Introduction to Dynamic programming – NetworkX package. **Categorizing and Tagging Words:** POS tagging – Part of speech Tagset – Reading corpora – Exploring corpora – Regular expression tagging – Look up tagging – Ngram Tagging – Transformation based tagging –Determining category of a word.

Unit-3. Learning to Classify Text and Information Extraction
Document classification – Sequence classification: Greedy approach, Hidden Markov Models and Conditional Random Fields –Recognizing text entailment –Text classifiers: Decision Trees, Naïve Bayes and Maximum Entropy classifiers – Generative vs conditional classifiers. **Information Extraction:** Architecture – Entity Recognition: Chunking, Chinking – Named Entity Recognition – Relation Extraction

Unit-4. Analysing Sentence Structure
Ambiquity – Context Free Grammar: Simple grammar, writing your own grammar – Parsing with CFGs – Dependence grammar – Valency and Lexicon – Probabilistic CFG – Feature Based Grammars

Unit-5. Analysing Meaning of Sentences
Propositional logic – First order logic – First order theorem proving – Model checking – Quantification – Discourse Processing

Text Books
1. Steven Bird, Ewan Klein, and Edward Loper , Natural Language Processing with Python– Analyzing Text with the Natural Language Toolkit.

References
1. Jurafsky and Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 2ed, Pearson India. ISBN-10: 9789332518414, ISBN-13: 978-9332518414
2. Indurkhya, Nitin and Fred Damerau, Handbook of Natural Language Processing, 2ed, 2010, Chapman & Hall/CRC.
3. Emily M Bender, Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax, Synthesis Lectures on Human Language Technologies, June 2013, Vol. 6, No. 3, Pages 1-184.
4. Christopher Manning and HinrichSchutze, Foundations of Statistical Natural Language Processing, MIT Press.

Elective-2A: Multivariate Analysis

Course Objectives
This course introduces the concepts of Multivariate data analysis to understand the research data, its presentation and analysis. Topics such as multivariate data structure, multinomial and multivariate normal distribution, estimation and testing of parameters and data reduction methods, are covered in this course.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Understand multivariate data structure, multinomial and multivariate normal distribution
LO2: Apply Multivariate analysis of variance (MANOVA) of one and two- way classified data.

Unit-1. Introduction
Basic concepts on multivariate variable - Multivariate normal distribution, Marginal and conditional distribution - Concept of random vector: Its expectation and Variance-Covariance matrix - Marginal and joint distributions - Conditional distributions and Independence of random vectors - Multinomial distribution

Unit-2. Distribution
Sample mean vector and its distribution - Likelihood ratio tests: Tests of hypotheses about the mean vectors and covariance matrices for multivariate normal populations - Independence of sub vectors and sphericity test.

Unit-3. Multivariate Analysis
Multivariate analysis of variance (MANOVA) of one and two- way classified data - Multivariate analysis of covariance. Wishart distribution - Hotelling's T2 and Mahalanobis'D2 statistics - Null distribution of Hotelling's T2 - Rao's U statistics and its distribution

Unit-4. Classification and Discriminant Procedures
Bayes, minimax, and Fisher's criteria for discrimination between two multivariate normal populations - Sample discriminant function - Tests associated with discriminant functions - Probabilities of misclassification and their estimation - Discrimination for several multivariate normal populations

Unit-5. Principal Component and Factor Analysis
Principal components, sample principal components asymptotic properties – Canonical variables and canonical correlations: definition, estimation, computations - Test for significance of canonical correlations - Factor analysis: Orthogonal factor model, factor loadings, estimation of factor loadings, factor scores - Applications

Text Books
1. Anderson, T.W. An Introduction to Multivariate Statistical Analysis, 3ed,John Wiley. 2009
2. Everitt B, Hothorn T. An Introduction to Applied Multivariate Analysis with R,Springer. 2011
3. Barry J. Babin, Hair, Rolph E Anderson, and William C. Blac, Multivariate DataAnalysis, Pearson New International Edition, 2013

References
1. Giri, N.C. Multivariate Statistical Inference. Academic Press. 1977.
2. Chatfield, C. and Collins, A.J. Introduction to Multivariate analysis. Prentice Hall.1982.
3. Srivastava, M.S. and Khatri, C.G.An Introduction to Multivariate Statistics. NorthHolland. 1979.

Elective-2B: Basics of Bioinformatics

Course Objectives
This course will offer basic understanding of Bioinformatics and its applications to molecular biology, clinical medicine and other disciplines. It enables students to create profound advances in their understanding of life and improvements in the health of humans and other living organisms.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Design their own databases according to their data of research.
LO2: Acquire knowledge on collecting latest and updated biological information from the browsers.

Unit-1. Introduction to Bioinformatics
Overview- Definition and History. Milestones in Bioinformatics. Methods in Bioinformatics. Role of Bioinformatics in various fields. Useful Bioinformatics web sites. Dogmas: Central and Peripheral. Introduction to single letter code of amino acids, Symbols used in nucleotides.

Unit-2. Biological Data and Databases
Introduction to Biological Databases- Nucleotide sequence database, Protein sequence & Structure Databases, Organism specific databases, Metabolic pathway databases, Bibliographic databases, Biodiversity databases and Specialized databases.

Unit-3. Sequence formats and Information Retrieval
Sequence Formats in Biological databases- FASTA, Phylip, Clustal, Genbank, EMBL, SWISS PROT. Data retrieval- Entrez, SRS, Protein identification resources (PIR), Expasy, Ensembl.

Unit-4. Database Searches
Similarity, homology, assessing significance of sequence similarity-Z score, P value, E value, Similarity search programs- fast searching methods-BLAST, FASTA, Dynamic programming searching methods, profile based methods-PSI-BLAST, Sensitivity and Specificity.

Unit-5. Applied Bioinformatics
Commercial bioinformatics, Survey of bioinformatics companies in India and abroad –Economics prospects, pharama informatics, combinatorial chemistry, HT screening – in silico screening - from lead to commercialization.

Text Books
1. Arthur M. Lesk, Introduction to Bioinformatics, Oxford University Press, New Delhi, 2003.

References
1. Attwood, T.K and D.J. Parry-Smith, Introduction to Bioinformatics, Pearson Education Ltd., New Delhi, 2004.
2. Westhead D.R., J.H Parish and R.M. Twyman, Instant notes in Bioinformatics, Viva Books Pvt. Ltd., 2003.
3. Manju Bansal, Basic Bioinformatics, Atlantic Publishers & Distributors, 2009.

Elective-3: Health Care Data Analytics

Course Objectives
This course will introduce students the various forms of electronic health care information. It helps students to learn the techniques to analyse health care data and to create predictive models for clinical data.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Analyse health care data using analytical techniques.
LO2: Apply analytics for decision making in healthcare services.
LO3: Integrate health data from multiple sources and develop fficient clinical decision support systems.

Unit-1. Introduction
Introduction to Healthcare Data Analytics- Electronic Health Records–Components of EHR- Coding Systems- Benefits of EHR- Barrier to Adopting HER Challenges-Phenotyping Algorithms.

Unit-2. Analysis
Biomedical Image Analysis- Mining of Sensor Data in Healthcare- Biomedical Signal Analysis- Genomic Data Analysis for Personalized Medicine.

Unit-3. Analytics
Natural Language Processing and Data Mining for Clinical Text- Mining the Biomedical- Social Media Analytics for Healthcare.

Unit-4. Advanced Data Analytics
Advanced Data Analytics for Healthcare– Review of Clinical Prediction Models- Temporal Data Mining for Healthcare Data- Visual Analytics for Healthcare- Predictive Models for Integrating Clinical and Genomic Data- Information Retrieval for Healthcare- Privacy-Preserving Data Publishing Methods in Healthcare.

Unit-5. Applications
Applications and Practical Systems for Healthcare– Data Analytics for Pervasive Health- Fraud Detection in Healthcare- Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems- Computer-Assisted Medical Image Analysis Systems- Mobile Imaging and Analytics for Biomedical Data.

References
1. Chandan K. Reddy and Charu C Aggarwal, "Healthcare data analytics", Taylor &Francis, 2015
2. Hui Yang and Eva K. Lee, "Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, Wiley, 2016.

Elective-3C: Customer Relationship Management

Course Objectives
This course will introduce students the concepts of customer relationship management with industry case studies and strategies for implementing them in any organization. It helps students to better understand customer needs and to maintain long-term customer relationships.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Understand concept of Customer Relationship Management (CRM) and implementation of Customer Relationship Management
LO2: Provide insight into CRM marketing initiatives, customer service and designing CRM strategy
LO3: Understand new trends in CRM, challenges and opportunities for organizations

Unit-1. Introduction
Introduction to Customer Relationship Management: Concept- Evolution of Customer Relationships: Customers as strangers- acquaintances- friends and partners. Objectives- Benefits of CRM to Customers and Organizations- Customer Profitability Segments- Components of CRM: Information- Process- Technology and People-Barriers to CRM. Relationship Marketing and CRM: Relationship Development Strategies: Organizational Pervasive Approach- Managing Customer Emotions- Brand Building through Relationship Marketing- Service Level Agreements- Relationship Challenges.

Unit-2. Marketing and Data Management
CRM Marketing Initiatives- Customer Service and Data Management: CRM Marketing Initiatives: Cross-Selling and Up-Selling- Customer Retention- Behaviour Prediction-Customer Profitability and Value Modeling- Channel Optimization- Personalization and Event-Based Marketing. CRM and Customer Service: Call Center and Customer Care: Call Routing- Contact Center Sales-Support- Web Based Self Service- Customer Satisfaction Measurement- Call-Scripting- Cyber Agents and Workforce Management. CRM and Data Management: Types of Data: Reference Data- Transactional Data- Warehouse Data and Business View Data- Identifying Data Quality Issues- Planning and Getting Information Quality- Using Tools to Manage Data- Types of Data Analysis: Online Analytical Processing (OLAP) - Clickstream Analysis- Personalization and Collaborative Filtering- Data Reporting.

Unit-3. Strategy
CRM Strategy- Planning: Understanding Customers: Customer Value- Customer Care-Company Profit Chain: Satisfaction- Loyalty- Retention and Profits. Objectives of CRM Strategy- The CRM Strategy Cycle: Acquisition- Retention and Win Back- Complexities of CRM Strategy.

Unit-4. Implementation and Evaluation
CRM Implementation and Evaluation: Planning and Implementation of CRM: Business to Business CRM- Sales and CRM- Sales Force Automation- Sales Process/ Activity Management- Sales Territory Management- Contact Management- Lead Management-Configuration Support- Knowledge Management CRM Implementation: Steps- Business Planning- Architecture and Design- Technology Selection- Development- Delivery and Measurement.

Unit-5. Evaluation
CRM Evaluation: Basic Measures: Service Quality- Customer Satisfaction and Loyalty-Company 3E Measures: Efficiency- Effectiveness and Employee Change. CRM New Horizons: e-CRM: Concept- Different Levels of E- CRM- Privacy in E-CRM -Software App for Customer Service:# Activity Management- Agent Management- Case Assignment- Contract Management- Customer Self Service- Email Response Management- Escalation- Inbound Communication Management- Invoicing- Outbound Communication Management- Queuing and Routing- Scheduling - Social Networking and CRM - Mobile-CRM - CRM Trends- Challenges and Opportunities - Ethical Issues in CRM.

References

1. Anderrson Kristin and Carol Kerr. Customer Relationship Management. Tata McGraw-Hill, 2002.
2. Ed Peelen. Customer Relationship Management. Prentice Hall, 2005.
3. BhasinJaspreet Kaur. Customer Relationship Management. Dreamtech Press, 2012
4. Valarie A Zeithmal, Mary Jo Bitner, Dwayne D Gremler and Ajay Pandit. Services Marketing Integrating Customer Focus Across the Firm. Tata McGraw Hill, 2010.
5. UrvashiMakkar and Harinder Kumar Makkar. CRM Customer Relationship Management. McGraw Hill Education, 2013.
6. Baran Roger J. & Robert J. Galka. Customer Relationship Management: The Foundation of Contemporary Marketing Strategy, Routledge Taylor & Francis Group. 2014

Core Practical III: Data and Visual Analytics Lab

Course Objectives
This lab course enables students to implement the concepts of data and visual analytics such as array processing, data wrangling and time series plotting using Python language and its libraries.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Apply NumPy functions for array processing
LO2: Create Series and Data Frames for data wrangling
LO3: Create various plots using Matplotlib
LO4: Perform data aggregation and group operations
LO5: Use Date and Time classes to create time series plots

Lab Exercises

Develop applications that will demonstrate data analysis features of the following technologies.

- NumPy arrays
- Pandas data wrangling features such as Series and DataFrame
- MatplotLib's Bar, Scatter plot, Line, Histogram, Box plots
- Time series plots

Core Practical IV: Practical Machine Learning Lab

Course Objectives
This lab course enables students to practice various ML models for supervised and unsupervised Machine Learning problems.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Practice Scikit-Learn Python Library
LO2: Apply pre-processing methods and reduce dimensions of data
LO3: Design a ML model for the given business problem
LO4: Perform training, testing and evaluation of the designed ML model

Lab Exercises

Develop applications that will demonstrate the following Machine Learning models for the given business use case.

- Pre-processing methods
- Dimensionality reduction and feature selection using PCA, Kernel PCA and SVD as well as Entropy, Information Gain and Gini Index.
- Supervised Machine Learning Models
  - Logistic Regression
  - Support Vector Machine
  - Decision Trees
  - K-Nearest Neighbour
  - Neural Network
- Unsupervised Machine Learning Models
  - KMeans and KMeans++
  - Agglomerative Clustering
  - DBSCAN
- Evaluation of ML Models
  - K-Fold Cross Validation
  - Grid Search
- Performance Evaluation
  - Accuracy
  - MSE and SSE
  - Confusion Matrix
  - Precision, Recall and F-Score

Core Practical V: Natural Language Processing Lab

Course Objectives
This lab course enables students to represent and analyze syntax and semantics of words, sentences, and documents.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Design NLP pipeline
LO2: Categorize and tag words
LO3: Classify text using supervised classifiers
LO4: Extract entities and named entities
LO5: Analyze sentences and their meanings

Lab Exercises

Develop NLP applications that will demonstrate the following NLP concepts and feature.

- Frequency Distribution Histogram Plotting using WordNet
- NLP preprocessing such as tokenization, stemming, lemmatization
- POS Tagging
- Bigrams and nGram Tagging
- Text Classification using Naïve Bayes Classifier
- Named Entity Recognition
- Context Free Grammars

Core-VII: Time Series Analysis and Forecasting

Course Objectives
This course will provide students a basic introduction to modern time series analysis. It covers time series regression and exploratory data analysis, ARMA and ARIMA models, model identification, estimation and linear operators, Fourier analysis, spectral estimation, and state space models.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Learn basic analysis of time series data
LO2: Learn basic concepts in time series regression
LO3: Learn auto-regressive and model averaging models
LO4: Learn basic concepts of spectral analysis and space-time models

Unit-1. Basis Time Series Models
Examples of Nature of Time series data – Time series statistical models – Measures of dependence - Stationary. Time series regression – Detrending and differencing – Smoothing a time series

Unit-2. AR models, Forecasting and Estimation
Auto Regressive models – Moving Average models - ARMA models – Auto Correlation Function - Partial Auto Correlation Function – Forecasting algorithms – Estimation: Yule-Walker, Method of moments, MLE and LSE

Unit-3. ARMA and GARMA Models
Basics of ARIMA models: random models with drift, Steps to fitting ARMA model – Multiplicative Seasonal ARIMA models: Mixed, SARMA – Generalized Auto Regressive Conditionally Heteroscedastic (GARCH) models

Unit-4. Spectral Analysis
Cyclical Behaviour and Periodicity: concepts, Periodic Series, Star Magnitude - The Spectral Density: Periodic stationary process–Periodogram: Spectral analysis as ANOVA, Principal Component Analysis

Unit-5. State Space Models
Dynamic Linear Models – Examples of DLMs – Filtering DLM – Smoothing DLM: Kalman, Lag One covariance – Forecasting DLM – Maximum Likelihood Estimator for DLMs

Text Books
1. Shumway and Stoffer. Time Series Analysis and its applications, with examples in R. 4ed, Springer. 2016.

References
1. Brockwell& Davis. Introduction to Time Series and Forecasting, 3rd edition, Springer. 2016
2. Cryer& Chan. Time Series Analysis with Applications in R, Springer. 2008
3. Prado & West. Time Series: Modeling, Computation, and Inference Chapman & Hall. 2010
4. Petris, Petrone, Campagnoli. Dynamic Linear Models with R, Springer. 2009
5. Ruppert& Matteson. Statistics and Data Analysis for Financial Engineering with R examples, 2ed, Springer. 2016

Core-VIII: Big Data Management and Analytics

Course Objectives

This course will cover concepts, algorithms and standard tools used to analyze Big Data such as MapReduceover modern distributed analysis platforms such as Hadoop.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Understand Big Data concepts and technologies
LO2: Develop applications using Hadoop
LO3: Store and manipulate data using HDFS
LO4: Explore very large datasets using Pig
LO5: Perform Data Warehousing operations using Hive

Unit-1. Introduction to Big Data

What is Big data – Industrial examples of Big Data: Digital Marketing, fraud, risk, trading, healthcare, medicine, advertising – Big Data Technology: Hadoop, cloud, BI, crowdsourcing analytics – Business Analytics:

Unit-2. MapReduce-I and HDFS

MapReduce model: Weather dataset, Analyzing data with Hadoop, Combiner functions, Hadoop streaming with Python. **Hadoop Distributed File System**: Block, Namenode, Datanode, Caching – File system operations in command line – Java Interface to Basic Hadoop - Reading data and writing data – Anatomy of File Write

Unit-3. MapReduce-II

Steps of developing MapReduce application - Working of MapReduce: Running Jobs, failure, Shuffle and sort, Task execution - MapReduce Types: Input formats - Output formats - MapReduce features: Counters, Sorting, Joins

Unit-IV. Exploring large datasets using Pig

Structure, Statements, Expressions, Types, Schemas, Functions, Macros - User-Defined Functions: Filter UDF, Eval UDF, Load UDF - Data Processing Operators: Loading and Storing Data, Filtering Data, Grouping and Joining Data, Sorting Data, Combining and Splitting Data

Unit-5. Data Warehousing using Hive

Comparison with Traditional Databases - HiveQL: Data Types, Operators and Functions - Tables: Managed Tables and External Tables, Partitions and Buckets, Storage Formats, Importing Data, Altering Tables, Dropping Tables - Querying Data: Sorting and Aggregating, MapReduce Scripts, Joins, Subqueries, Views - User-Defined Functions: Writing a UDF, Writing a UDAF

Text Books

1. Michael Minelli, Michele Chambers and AmbigaDhiraj. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, 1ed, Wiley CIO Series, 2013. ISBN 9781118147603
2. Tom White Hadoop: The Definitive Guide, Fourth Edition, O'reilly Media, 2015.

References

1. Nathan Marz and James Warren, Big Data Principles and Practice of Scalable Real Time Data Systems, Manning Publications. 2015
2. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007
3. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley& sons, 2012.
4. Glenn J. Myatt, Making Sense of Data, Volume I and II. John Wiley & Sons, 2007.
5. Mark Grover, Ted Malaska, Jonathan Seidman, Gwen Shapira. Hadoop Application Architecture, Shroff Publishers.2015
6. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGrawHill Publishing, 2012.

Core-IX: Social Media Analytics

Course Objectives

This course will introduce students the necessary theories and the state-of-the art techniques in Web mining, networks analysis, and predictive modeling to study emerging problems with social media. These problems include information diffusion, information retrieval, recommendations, behavior analysis, and event analytics in social media.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Understand the concepts of Social Media Mining
LO2: Understand and analyse social media graphs
LO3: Apply supervised and unsupervised learning algorithms on social media
LO4: Discover community and follow information diffusion in social networks
LO5: Develop recommender systems

Unit-1. Introduction to SMM and Graph Mining

What is social media mining – New challenges for mining. **Graph Essentials:** Graph basics – Graph representation – Types of graphs – Connectivity in graphs – Special graphs – Graph algorithms

Unit-2. Social Network Models

Network Measures: Centrality – Transitivity, reciprocity – Balance and status – Similarity. **Network Models:** Properties – Random graphs – Small world models – Preferential attachment model

Unit-3. Data Mining Basics and Community Discovery

Data Mining Essentials: Data Preprocessing – Supervised Learning Algorithms – Unsupervised Learning Algorithms. **Community Analysis:** Community detection – Community evolution – Community evaluation

Unit-4. Information Diffusion and Influence in Social Media

Information Diffusion: Herd behaviour – Information cascades – Diffusion of innovations – Epidemics. **Influence and Homophily:** Measuring Assortativity – Measuring and modelling influence – Measuring and modelling homophily – Distinguishing influence and homophily

Unit-5. Recommendation and Behaviour Analysis in Social Media

Recommendation in Social Media: Challenges – Classical recommendation algorithms – Recommendation using social context – Evaluating recommendations. **Behaviour Analysis:** Individual behaviour – Collective behaviour. **Events Analytics in Social Media**.

Text Books

1. Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. Social Media Mining: An Introduction, Cambridge University Press, 2014

References

1. Matthew A. Russell. Mining the Social Web. 2nd Edition. O'Reilly Media. 2013
2. Jennifer Golbeck. Analyzing the Social Web. Morgn Kaufmann. 2013. ISBN 978-0124055315
3. Ricardo Baeza-Yates and BerthierRibeiro-Neto. Modern Information Retrieval: The Concepts and Technology behind Search. 2ed. ACM Press Books, 2011. ISBN 978-0321416919
4. Charu C. Aggarwal. Social Network Data Analytics. Springer. 2011

Elective-4: Image and Video Analytics

Course Objectives
This course will provide the basic techniques for digital image and video processing. Topics covered in the course includeimage representation, filtering, colors, compression, object detection & tracking and applications of video processing.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Learn the fundamentals of digital image processing, image and video analysis.
LO2: Understand and appreciate the real time use of image and video analytics.
LO3: Demonstrate real time image and video analytics applications.

Unit-1. Image Representation and Processing
Digital image representation- Visual Perception- Sampling and Quantization- Basic Relations between Pixels- Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matric Operations- Image Transforms (DFT, DCT,DWT, Hadamard).

Unit-2. Image Filtering
Fundamentals of spatial filtering: spatial correlation and convolution-smoothing, blurring-sharpening- edge detection - Basics of filtering in the frequency domain: smoothing-blurring- sharpening--Histograms and basic statistical models of image.

Unit-3. Colors and Compression
Colour models and Transformations – Image and Video segmentation-Image and video demonising- Image and Video enhancement- Image and Video compression.

Unit-4. Object Detection and Tracking
Object detection and recognition in image and video-Texture models Image and Video classification models-Object tracking in Video.

Unit-5. Applications
Applications and Case studies- Industrial- Retail- Transportation & Travel- Remotesensing-Video Analytics in WSN: IoT Video Analytics Architectures.

References
1. R.C. Gonzalez and R.E. Woods. Digital Image Processing. 3rd Edition. Addison Wesley, 2007.
2. Pratt, W.K. Digital image processing: PIKS scientific inside. 4ed. New York: John Wiley, 2007.
3. W. Härdle, M. Müller, S. Sperlich, A. Werwatz. Nonparametric and Semi parametric Models. Springer, 2004.
4. Rick Szelisk. Computer Vision: Algorithms and Applications. Springer 2011.
5. Jean-Yves Dufour. Intelligent Video Surveillance Systems. Wiley, 2013.
6. Caifeng Shan, FatihPorikli, Tao Xiang, Shaogang Gong. Video Analytics for Business Intelligence. Springer, 2012.
7. AsierPerallos, Unai Hernandez-Jayo, Enrique Onieva, Ignacio Julio GarcíaZuazola. Intelligent Transport Systems: Technologies and Applications. Wiley, 2015.
8. BasudebBhatta. Analysis of Urban Growth and Sprawl from Remote Sensing Data. Springer, 2010

Elective-4: Computational Genomics

Course Objectives

Genomics is a new and very active application area of computer science. This course aims to present some of the most basic and useful algorithms for sequence analysis, together with the minimal biological background necessary for a computer science student to appreciate their application to current genomics research. Sequence alignments, hidden Markov models, multiple alignment algorithms and heuristics such as Gibbs sampling, and the probabilistic interpretation of alignments will be covered.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Acquire knowledge about the computational problems in the emerging areas of Bioinformatics, Computational Biology, and Genomics.
LO2: Apply the algorithm related to sequence search and alignments, bioinformatics tools.
LO3: Synthesize of Markov Model building and searching of large databases.

Unit-1. Sequence Alignment

DNA, RNA and Proteins – Genetic Variations – High throughput sequencing - Sequence Alignments: Dynamic Programming, Homology, Global sequence alignment, Local Sequence alignment, Heuristic local alignment, BLAST algorithm, Biologically relevant scoring for alignment, RNA folding, Advanced alignment techniques: Linear space, Affine gaps, Banded linear time alignments, Time warping, Burrow Wheeler Index;

Unit-2. Hidden Markov Models

Hidden Markov Models: Markov Chains and Hidden Markov models, The Viterbi algorithm, Parameter estimation of HMMs, Connection between pair HMMs and alignments, Forward-Backward algorithm, CRFs, HMMs for motif-finding.

Unit-3. Applications of Alignment

Applications of alignments and HMMs: Analysis of a genome, The human genomes: Chromosomes, Repeats, Genes and SNPs, Gene Prediction, Suffix Trees, Comparative genomics, Efficient alignment algorithms, Cross species comparison based gene recognition, Microarrays and Gene Regulation;

Unit-4. Sequencing of Genome

Sequencing of a genome: Sequencing methods- Shotgun sequencing, BAC to BAC Sequencing and other modern methods, Computational assembly of a genome, Gene Expression Analysis, Clustering and trees, Parsimony problems-small and large, Sequencing by hybridization, SBH, Spectrum Graphs, Spectral Convolution, Spectral Alignment.

Unit-5. Building Phylogenetic Trees

Phylogenetic: Introduction to phylogeny, Limitations and workaround, Neighbor joining, Fitch's algorithm, Parsimonious trees, Fast alignment and tree building; Multiple Sequence Alignments: Multiple Sequence Alignments and their scoring methods, progressive alignment, CLUSTALW, Expectation maximization, Gibbs sampling.

Text Books

1. Durbin, Eddy, Krogh, Mitchison. Biological Sequence Analysis. Cambridge University Press.

References

1. Makinen, Belazzougui, Cunial, Tomescu. Genome Scale Algorithm Design. Cambridge University Press. 2015
2. Gusfield D., Algorithms on strings, trees, and sequences (1st Edition), Cambridge University Press, 1997, ISBN-13: 978-0521585194.

Core Practical VI: Big Data Management and Analytics Lab

Course Objectives
This lab course will enable students to analyzelarge amount of datausing modern distributed analysis platforms such as HDFS, MapReduce, Pig and Hive technologies

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Develop applications using Hadoop
LO2: Store and manipulate data using HDFS
LO3: Explore very large datasets using Pig
LO4: Perform Data Warehousing operations using Hive

Lab Exercises

1a). Perform setting up and Installing Hadoop in Pseudo distributed and Fully distributed operating modes.
1b). Use web based tools to monitor your Hadoop setup.

2a). Implement the following file management tasks in Hadoop:
- Adding files and directories
- Retrieving files
- Deleting files

2b). Benchmark and stress test an Apache Hadoop cluster

3). Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
- Find the number of occurrence of each word appearing in the input file(s)
- Performing a MapReduce Job for word search count (look for specific keywords in a file)

4). Stop word elimination problem:
Input:
- A large textual file containing one sentence per line
- A small file containing a set of stop words (One stop word per line)

Output: A textual file containing the same sentences of the large input file without the words appearing in the small file.

5). Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.

Data available at: https://github.com/tomwhite/hadoopbook/tree/master/input/ncdc/all.
- Find average, max and min temperature for each year in NCDC data set?
- Filter the readings of a set based on value of the measurement, Output the line of input files associated with a temperature value greater than 30.0 and store it in a separate file.

6). Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.

7). Write a Pig Latin scripts for finding TF-IDF value for book dataset (A corpus of eBooks available at: Project Gutenberg)

8). Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

Core Practical VII: Social Media Analytics Lab

Course Objectives
This lab course wil lenable students to practice techniques in Web mining, networks analysis, and predictive modelling for solving emerging problems from social media.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Understand the concepts of Social Media Mining
LO2: Understand and analyse social media graphs
LO3: Apply supervised and unsupervised learning algorithms on social media
LO4: Discover community and follow information diffusion in social networks
LO5: Develop recommender systems

Lab Exercises

1). Using TwitterAPI, crawl tweets at real time for the given hashtag and display
- Frequency distribution of words
- Plot histogram of frequency distribution of words

2). Using FacebookAPI and GraphAPI Explorer, query for
- information about you
- information about your friends, and
- the term '*social web*'

3). Using FacebookAPI and GraphAPI, examine the following friendship tasks.
- Print the top-10 likes among your friends
- Print the most popular categories for likes among your friends

4). Using Twitter SearchAPI, query twitter for the given word, collect the results and store them as JSON file. Retrieve the stored JSON file and display.

5). Using USA Airline flight dataset (social_media_analytics_lab_data.csv), perform the following tasks
- Install NetworkX package
- Display the head (top-5 rows) using DataFrame
- Display the nodes and edges
- Plot the graph
- Perform 5 other graph operations over this graph
- Find the shortest path based on the airtime between the airports AMA and PBI

6). Develop a Movie Recommender System that suggests movie IDs that are most similar to a particular movie ID. The dataset can be downloaded from the website (https://cdncontribute.geeksforgeeks.org/wp-content/uploads/file.tsv).
- Display the head (top-5 rows) of DataFrame
- Display mean rating of all movies
- Display count rating of all movies
- Plot the graph of ratings column
- Analyze the correlation of two movies
- Suggest similar movies for a given movie

Core Project-I: Project Preparation

Course Objectives

Students will do a masters project based on application scenario or research based problems. For all application based projects, they will follow software engineering methodologies. Students will follow a two phase project development. In the first project preparation phase, students will develop Software Requirements Specifications (SRS) document. The SRS will contain a problem statement, scope & justification, requirements specification, cost estimation, limitations, methodology identified and tools and programming languages to be used in the development of the project.

For research projects, students will define research questions, perform literature survey and submit the proposal document for their research problem.

Student Learning Outcomes

On successful completion of this course, students will be able to:

LO1: Identify and define the problem statement
LO2: Define and justify scope of the proposed problem
LO3: Gather and analyze application requirements
LO4: Propose an optimized solution among the existing solutions
LO5: Develop software design document for the chosen application

Text Books

1. Lynn E. Miner & Jeremy T. Miner, "Proposal Planning and Writing", Third Edition, Greenwood Publishing Group, 2003.

References

1. William Navidi, "Statistics for Engineers and Scientists", 2nd Edition, McGraw-Hill, 2007.

Core-X: Principles of Deep Learning

Course Objectives

This course will introduce the fundamental concepts of Deep Learning to solve problems related to classification, language modelling, machine translation and reinforcement learning.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Learn TensorFlow programming
LO2: Design and implement CNN for classification
LO3: Perform sequence analysis
LO4: Design machine translation networks using RNN
LO5: Implement a reinforcement learning system

Unit-1. TensorFlow Basics

TensorFlow: variables, operations, placeholder Tensors, sessions – Navigating variable scopes and shared variables – Managing models over CPU and GPU – LogisticRegression in TensorFlow–Training LogisticRegression model – Visualizing using TensorBoard – Building multilayer model in TensorFlow

Unit-2. Convolutional Neural Networks

Shortcomings of Feature Selection – Width, height and depth of layers – Filters and feature maps – Describing convolutional layer – Max pooling - Architectural Description of Convolution Networks – Recognizing handwritten digits using CNN for MNIST dataset –Image preprocessing pipelines - Training with Batch normalization

Unit-3. Autoencoders and Sequence Analysis

Embedding – Principal Component Analysis - Architecture of Autoencoders – Implementing autoencoders in TensorFlow–Denoising - Word2Vec framework for language modelling. Sequence Analysis: seq2seq problem – Dependency parsing – Beam search

Unit-4. Recurrent Neural Networks

Single neuron and fully connected recurrent layer – Challenges of vanishing gradients - LSTM architecture – TensorFlow primitives for RNN models – Implementing Sentiment analysis Model –Solving seq2seq tasks with RNN – Augmenting RNN with Attention – Designing Neural Translation Network

Unit-5. Deep Reinforcement Learning

Reinforcement Learning: Markov Decision Processes, Policy, Future return, Discounted future return, Balancing Explore-Exploit dilemma, Annealed e-Greedy – Policy learning and Value learning - Solving Pole Cart problem with Policy Gradients - QLearning -Deep QNetworks – Deep Q Recurrent Networks – UNREAL Learning

Text Books

1. Nikhil Buduma, Nicholas Locascio. Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms. O'Reilly Media. 2017.
2. Ian Goodfellow, YoshuaBengio, Aaron Courville. Deep Learning (Adaptive Computation and Machine Learning series). MIT Press, 2017.

References

1. Francois Chollet. Deep Learning with Python. 1ed, Manning Publications, 2017. ISBN 978-1617294433.

Core-XI: Web Development Using Python

Course Objectives
Flask is a popular Python framework known for its lightweight and modular design. This course will enable students to develop and deploy web applications using Python, Flask and NoSQL languages.

Student Learning Outcomes
On successful completion of this course, students will be able to:
LO1: Develop a Flask extension using best practices
LO2: Implement various authentication methods
LO3: Learn how to develop Jinja2 templates
LO4: Build tests for your applications and APIs
LO5: Develop RESTful APIs and secure REST API's
LO6: Deploy highly available applications that scale on Heroku and AWS using Docker or VMs

Unit-1. Models, Templates and Web Forms
Simple Application Structure. **Creating Models with SQLAlchemy:** CRUD operations, Relationships, Constraints and Indexes. **Creating Views with Templates**:Jinja, Creating views. **Web Forms**: Basics, Custom validation, Posting comments.

Unit-2. Controllers and Databases
Creating Controllers and Advance Application Structure**. Using NoSQL with Flask:**NoSQL, RDBMS vsNoSQL, MongoDB: CRUD operations, Relationships. Email support **-** Large Application Structure

Unit-3. Authentication, Blog posts and Followers
User Authentication: Methods, Flask Login, OpenID, OAuth, Role Based Access Control. User Roles - User Profiles - Blog Posts – Followers - User Comments

Unit-4. REST and Extensions
Building RESTful API: REST, Authentication, Get, post, put and delete requests. **Creating Asynchronous Tasks:** Running, monitoring and remembering. **Flask Extensions:** Caching, Assets and Admin. **Building your own extensions:** Creating and Modifying

Unit-5. Testing, Deployment and Version Control
Testing and Performance: Unit Testing, Interface Testing and Test Coverage. **Deployment:** Deploying on Heruku, AWS and Docker, Version Control with Git

Text Books
1. Daniel Gaspar, Jack Stouffer. Mastering Flask Web Development: Build enterprise-grade, scalable Python web applications. 2ed. Packt Publishing Ltd. 2018. ISBN 978-1788995405.
2. Miguel Grinberg. Flask Web Development, 2ed. Shroff Publishers. 2018. ISBN 9789352136995

References
1. Italo Maia. Building Web Applications with Flask. Packt Publishing Ltd. 2015. ISBN 978-1784396152.
2. Shlabh Aggarwal. Flask Framework Cookbook. Packt Publishing Ltd. 2014.

Elective-5: Supply Chain Management

Course Objectives

This course teaches students supply chain strategy and concepts. It provides solid understanding on building framework to analyze supply chain, designing supply chain network, managing demand and supply and inventories, besides designing transportation networks.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Understand the importance of good supply chain design, planning and operation
LO2: Apply strategic drivers and metrics of supply chain performance
LO3: Provide recommendations for the given business problems

Unit-1. Building strategic framework

Supply chain: Definition, 3 streams of knowledge, objectives and importance - Decision phases and process views of a supply chain (SC) - Examples of supply chain - Competitive strategy and SC strategy - 3 steps of achieving strategic fit - Improving SC performance by expanding scope of strategic fit, challenges to achieving strategic fit - Financial measures and drivers of SC performance - Logistical drivers: Roles in SC and decision components - Cross functional drivers: Roles in SC and decision components - Role of infrastructure in SC performance.

Unit-2. Designing SC network

Key factors influencing distribution network design - Design options for a distribution network - Impact of online sales on customer service and cost - Network design decisions: Influencing factors, framework - Capacitated plant location model for network optimization - Gravity location model for network design - Model for demand allocation and locating plants - Global supply chain: Dimensions to evaluate total cost, SC risks, tailored risk mitigation strategies - Discounted cash flow analysis to evaluate network design decision - Decision tree analysis: Basics, Evaluating flexibility at Trip Logistics.

Unit-3. Planning and coordinating demand and supply

Demand forecasting: role, characteristics, components and methods - Static demand forecasting methods - Adaptive demand forecasting methods - Measures of demand forecasting error - Aggregate planning: role, identifying aggregate units, strategies - Aggregate planning using Linear programming - Managing supply and demand to improve synchronization in SC - Lack of SC coordination: Bullwhip effect, effect on performance - Obstacles to coordination in SC - Managerial levers to achieve coordination of demand and supply in SC.

Unit-4. Planning and managing inventories

Cycle inventory terminologies: Lot size, Average flow time, Inventory holding cost, Ordering cost - Computing optimal lot size for single product: Economic order quantity, for Production environment, with Capacity constraint - Lot size based discount schemes: All unit quantity discounts, Marginal unit quantity discount - Trade promotions: Goals, Forward buying, Impact on lot size and cycle inventory - Factors affecting the level of safety inventory - Evaluating required safety inventory: Given a replenishment policy, Desired cycle service level, Desired fill rate - Impact of desired product availability and uncertainty on safety inventory - Impact of supply uncertainty on safety inventory - Factors affecting optimal level of product availability - Managerial levers of inventory to improve SC profitability.

Unit-5. Transportation and cross functional drivers

Modes of transportation in SC - Design options for a transportation network - Transportation and inventory cost trade off - Transportation cost and customer responsiveness trade off - Tailored transportation - Sourcing decisions: In house or Outsource - Sharing risk and reward in SC - Pricing and revenue management for multiple customer segments - Pricing and revenue management for perishable assets - Pricing and revenue management for seasonal demand.

Text Books

1. Sunil Chopra, Peter Meindl and DV Karla. “Supply Chain Management: Strategy, planning and operation”, 6th edition, Pearson, 2016. ISBN 978-9332548237 (Excluding Excel Examples)

References

1. David Simchi-Levi and Philip Kaminsky. “Designing and managing the supply chain: Concepts, strategies and case studies”, 3rd edition, McGraw Hill, 2007.

Elective-5: Internet of Things

Course Objectives

Internet of Things is a new revolution of the Internet that is rapidly gathering momentum driven by the advancements in sensor networks, mobile devices, wireless communications, networking and cloud technologies.This course will cover IoT design, programming methodology and IoT applications over wireless sensor networks concepts.

Student Learning Outcomes

On successful completion of this course, students will be able to:
LO1: Learn the fundamentals of IoT and
LO2: Design IoT systems
LO3: Develop programs for IoT systems
LO4: Understand protocols and routing methods for wireless sensor networks

Unit-1. Introduction to IoT

Introduction to IoT - Definition and Characteristics, Physical Design Things- Protocols,Logical Design- Functional Blocks, Communication Models- Communication APIs-Introduction to measure the physical quantities, IoT Enabling Technologies - Wireless Sensor Networks, Cloud Computing Big Data Analytics, Communication Protocols- Embedded System- IoT Levels and Deployment Templates.

Unit-2. IoT Programming

Introduction to Smart Systems using IoT - IoT Design Methodology- IoT Boards (RasberryPi, Arduino) and IDE - Case Study: Weather Monitoring- Logical Design using Python, Datatypes & Data Structures- Control Flow, Functions- Modules- Packages, File Handling -Date/Time Operations, Classes- Python Packages of Interest for IoT.

Unit-3.IoT Applications

Home Automation – Smart Cities- Environment, Energy- Retail, Logistics- Agriculture, Industry- Health and Lifestyle- IoT and M2M.

Unit-4. Network of Wireless Sensor Nodes

Sensing and Sensors - Wireless Sensor Networks, Challenges and Constraints - Applications: Structural Health Monitoring, Traffic Control, Health Care - Node Architecture – Operating system.

Unit-5. Mac, Routing and Transport Control in WSN

Introduction – Fundamentals of MAC Protocols – MAC protocols for WSN – Sensor MACCase Study – Routing Challenges and Design Issues – Routing Strategies – Transport Control Protocols – Transport Protocol Design Issues – Performance of Transport Protocols

Text Books

1. ArshdeepBahga and Vijay Madisetti. Internet of Things: Hands-on Approach, Hyderabad University Press, 2015.
2. Michael Miller. The Internet of Things, Pearson Education, 2015.
3. KazemSohraby, Daniel Minoli and TaiebZnati. Wireless Sensor Networks: Technology.Protocols and Application, Wiley Publications, 2010.
4. WaltenegusDargie and Christian Poellabauer. Fundamentals of Wireless SensorNetworks: Theory and Practice, John Wiley and Sons Ltd., 2010.

References

1. Edgar Callaway. Wireless Sensor Networks: Architecture and Protocols, AuerbachPublications, 2003.
2. Holger Karl and Andreas Willig. Protocols and Architectures for Wireless SensorNetworks, John Wiley & Sons Inc., 2005.
3. ErdalÇayırcıandChunmingRong. Security in Wireless Ad Hoc and Sensor Networks, JohnWiley and Sons, 2009.
4. Carlos De MoraisCordeiro and Dharma Prakash Agrawal. Ad Hoc and Sensor Networks:Theory and Applications, World Scientific Publishing, 2011.
5. WaltenegusDargie and Christian Poellabauer. Fundamentals of Wireless Sensor Networks : Theory and Practice, John Wiley and Sons, 2010

Core Project-II: Project Implementation

Course Objectives

In the second implementation phase, students will implement the proposed software design outlined in SRS. Students will write code, debug source code, test them and deploy at the client site. As part of the deployment, they will develop various documentations such as user manuals.

For research projects, students will propose methodology based on the proposal report submitted during Phase-I in the previous semester and validate their research with various experimental results. This way, they are expected to advance the current state of the art in their chosen area of research.

Student Learning Outcomes

On successful completion of this course, students will be able to:

LO1: Apply coding, debugging and testing tools and construct a new software system
LO2: Prepare documentation and user manuals by following the standard guidelines
LO3: Learn technical report writing and oral presentation skills

Text Books

1. Lynn E. Miner & Jeremy T. Miner, "Proposal Planning and Writing", Third Edition, Greenwood Publishing Group, 2003.

References

1. William Navidi, "Statistics for Engineers and Scientists", 2nd Edition, McGraw-Hill, 2007.