ST ALOYSIUS COLLEGE, MANGALURU

(AUTONOMOUS)

SCHOOL OF INFORMATION TECHNOLOGY

ALOYSIUS INSTITUTE OF MGMT & INFORMATION TECHNOLOGY

BEERI CAMPUS, MANGALURU

RE-ACCREDITED BY NAAC "A" GRADE

COURSE STRUCTURE AND SYLLABUS

OF

M.Sc. BIG DATA ANALYTICS

CHOICE BASED CREDIT SYSTEM

SYLLABUS – 2019 ONWARDS

Preamble: M.Sc. Big Data Analytics course has been started after the Mangalore University approval in 2018, with a curriculum given below. Tata Consultancy Services, India's Leading Services and Research Company in IT Services under the Academia Interface Programme has partnered with M.Sc. Big Data Analytics programme. (This is the Fifth Institute in India to have such a programme and the First one with the New Syllabus 2019). With the MoU with TCS, we are able to deliver a TCS's Academia Interface Programme designed curriculum jointly with Industry Experts and Indian Statistical Institute, Kolkatta& Bangalore. This Curriculum gives stress on Statistics, Applied Mathematics, Information Technology and Big Data Technology

Proposed New Syllabus for M.Sc Big Data Analytics from Tata Consultancy Services under the Academia Interface Programme 2019.

SCHEME OF TEACHING: L-T-P-S Scheme

- The Board of Studies of M.Sc. Big Data Analytics had series of Interactions internally and with Tata Consultancy Services over the Syllabus
- The Board of Studies of M.Sc. Big Data Analytics is able to set this new Curriculum from TCS with the Mangalore University's new Choice Based Credit System; and suggestions made in the new system for Hard core , Soft core and Open Electives.
- Falling in line with these new guidelines the BOS of M.Sc. Big Data Analytics has outlined the new curriculum. The syllabus has Hard Core, Soft Core and Open Electives proposed in the curriculum. The proportionate ratio of all these as per the guidelines of Mangalore University.
- In the First semester the students have to study Three hard core Papers and a Hard Core Practical; and Two soft Core paper along with the Practical Lab.
- In the second semester the student has to study Three hard core papers; with a Hard Core Lab; Three Soft Core papers with a Soft Core Lab. There is a Open Elective paper offered.
- In the Third semester the student has to study Two hard core papers, with a Hard core Lab and Three Soft core Papers with Electives and Two Soft Core Labs.

- The final semester students have to do a Industry Internship / Project work / Dissertation at the Research & Development Labs / Institute of National Importance / Govt& Ministry funded Labs / Institutions

Key Features of the M.Sc. Big Data Analytics Syllabus from Tata Consultancy Services:

- Focus on Strong Foundations in Statistics and Mathematics and Logic Building
- Information Technology & Big Data Technology Applications for the Data Science
- TCS Case studies in understanding the Big Data Analytics in a better way.
- Refresher course / Pre-Requisite course to be conducted to strengthen the foundations
- Stress on Solid Inter-Disciplinary applications development, applying the Big Data Analytics Techniques in several domains such as Financial & Marketing Analytics, Health, Insurance, Travel Analytics etc
- Continuous Faculty Training through the Train the Trainer at TCS Labs
- More stress given applications in Data Science and Big Data Analytics.
- Study Materials, Case studies being provided by TCS. (This content to be used beyond the Text and Reference books suggested in the syllabus)
- Student Internship Support, Placement Support from the Academia Partner

Note : The Syllabus from TCS and other study materials are protected under the Copy Rights Law. The content is being kept at the Dept and the Registrars office

Programme Educational objectives

PEO1	To practice big data analytics and machine learning approaches, which include the study of modern computing using big data technologies and machine learning techniques focusing on industry applications.
PEO2	To develop Numerical and Statistical skills that will play an important role in their Job role as data Scientist / data analytics in analyzing the problem at hand and give the appropriate and efficient solution.
PEO3	Apply the concepts of Analytics to the real world problems by converting datasets to models in order to make better business decisions.
PEO4	Apply the skills gained in the course to improve the research which would have a great impact on the societal development by emphasizing on how data can be collected and used in ethical and socially sensitive ways.

Program Outcomes

PO1	**Statistical computing:** Ability to understand the basic concepts of how to explore the datasets using statistical analysis techniques in Python and R.
PO2	**Mathematical Skills:** Ability to understand and implement various algorithms which require strong hold on the mathematical skills
PO3	**Database management:** Ability to Execute queries, implement views and joins, use MongoDB for various operations on unstructured data. Ability to Optimize business decisions and create competitive advantage with Big Data analytics and understand architectural concepts of Hadoop and map reduce paradigm
PO4	**Implementation using various software:** This enables the students to develop strong programming skills required to handle complex data and build algorithms that will provide efficient solutions to the problem at hand.
PO5	**Machine learning:** Understand a wide variety of learning algorithm, how to evaluate models generated from data and apply the algorithms to a real problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models.
PO6	**Enabling technologies:** Learn about the relationship between data science and natural language and audio-visual content processing
PO7	**Natural language processing:** Understand approaches to syntax, semantics in *NLP*, to discourse, generation, dialogue and summarization within *NLP and* Understand current methods for statistical approaches to machine translation.
PO8	**Value thinking:** Recognize important ethical issues that arise in various business contexts and professional practice; To Demonstrate an understanding of the ethical, social and economic environments in which those occur.
PO9	**Advanced Statistical Analysis:** Mastering of a suite of methods and workflow styles that will enable the student to produce several new statistical analysis correctly and efficiently present the results from those analyses.
PO10	**Societal development:** Identify the information security models and their characteristics, by analysing the different types of cryptographic and forensic methods. Identify and solve different cyber security threats that hamper the society.
PO11	**Application of Skills:** Provide the knowledge and necessary skills to accomplish various analytics with respect to areas like health, HR,Travel, … so that they are able to provide efficient analysis and interpretation.

Syllabus - Scheme of Teaching

I Semester

Sl No	Subject Code	Subject Title	Type	Credits	Total No of Hours	Lec ture Hrs / week	Tutorial Hrs / week	Practical Hrs / week	Self Study Hrs / week
HARDCORE THEORY & PRACTICALS									
1	PH 801.1	Statistical Methods	HC	3	60	3	--	2	2
2	PH 802.1	Probability and Stochastic Process	HC	3	60	3	--	2	2
3	PH 803.1	Linear Algebra and Linear Programming	HC	3	60	3	--	2	2
4	PH 804.1 P	Computing for Data Sciences Lab	HCL	2	60	--	(1 x 2) 2	(3 x 2) 6	--
SOFTCORE THEORY & PRACTICALS									
5	PS 805.1	Database Management System	SC	3	45	3	--	2	2
6	PS 806.1	Python Programming	SC	3	45	3	--	2	2
7	PS 807.1 P	DBMS & Python Programming Lab	SCL	3	60	--	(1 x 2) 2	(4 x 2) 8	--
8	---	Project Work & Seminar	---	--	30	--	2	--	4
		Total		**20**		**15**	**06**	**24**	**14**

* HC – Hard Core SC – Soft Core HCL – Hard Core Lab SCL – Soft Core Lab OE – Open Elective

II Semester

Sl No	Subject Code	Subject Title	Type	Credits	Total No of Hours	Lecture Hrs / week	Tutorial Hrs / week	Practical Hrs / week	Self Study Hrs / week
HARDCORE THEORY & PRACTICALS									
1	PH 801.2	Machine Learning – I	HC	3	45	3	--	--	2
2	PH 802.2	Enabling Technologies for Data Science – I	HC	3	45	3	--	--	2
3		HC Electives: (Choose ONE)	HC	3	60	3	2	2	2
	PH 803.2 E1	Operations Research							
	PH 803.2 E2	Cloud Computing							
	PH 803.2 E3	Natural Language Processing							
	PH 803.2 E4	Unix Programming							
	PH 803.2 E5	Operating Systems							
	PH 803.2 E6	Multivariate Statistics							
4	PH 804.2 P	Machine Learning and Data Science Lab - I	HCL	3	90	--	(1 x 2) 2	(4 x 2) 8	--
SOFTCORE THEORY & PRACTICALS									
5	PS 805.2	Foundations of Data Science	SC	3	45	3	--	--	2
6	PS 806.2	Advanced Statistical Methods	SC	3	45	3	--	--	2
7	PS 807.2	Value Thinking	SC	1	30	2	--	2	4
8	PS 808.2 P	Programming for Big Data and Advanced Statistical Methods Lab	SCL	3	90	--	(1 x 2) 2	(4 x 2) 8	2
OPEN ELECTIVE – Offered to other Depts									
9	OE 809.2	Statistical Data Analysis using R	SC	3 *	45	3	--	--	2
10	---	Project Work & Seminar	---	--	30	--	2	--	4

		Total		**25**		**20**	**08**	**24**	**20**

* HC – Hard Core SC – Soft Core HCL – Hard Core Lab SCL – Soft Core Lab OE – Open Elective

III Semester

Sl No	Subject Code	Subject Title	Type	Credits	Total No of Hours	Lecture Hrs / week	Tutorial Hrs / week	Practical Hrs / week	Self Study Hrs / week
HARDCORE THEORY & PRACTICALS									
1	PH 801.3	Machine Learning – II	HC	3	45	3	--	--	2
2	PH 802.3	Enabling Technologies for Data Science – II	HC	3	45	3	--	--	2
3	PH 803.3 P	Machine Learning and Data Science Lab - II	HCL	3	90	--	(1 x 2) 2	(4 x 2) 8	--
SOFTCORE THEORY & PRACTICALS									
4	PS 804.3	Data Visualization with Tableau & Modeling in Operations Management	SC	3	45	3	--	--	2
5		SC Electives: (Choose ONE)	SC	3	45	3	--	--	2
	PS 805.3 E1	Introduction to Econometrics and Finance							
	PS 805.3 E2	Time Series Analysis and Forecasting							
	PS 805.3 E3	Bioinformatics							
	PS 805.3 E4	Big Data Technologies and Architecture							
6		SC Electives: (Choose ONE)	SC	3	45	3	--	--	2
	PS 806.4 E1	Intellectual Property Rights							
	PS 806.4 E2	Cyber Security							
	PS 806.4 E3	Text Mining							
	PS 806.4 E4	Advanced Analytics							
7	PS 807.3 P	Data Visualization with Tableau & Operation Management Lab	SCL	3	60	--	(1 x 2) 2	(3 x 2) 6	2
8	PS 808.3 P	Lab on Elective-1 & 2	SCL	3	60	--	(1 x 2) 2	(3 x 2) 6	2

OPEN ELECTIVE – Offered to other Depts									
9	OE 809.3	Big Data & Design Thinking	SC	3 *	45	3	--	--	2
10	---	Project Work & Seminar	---	--	30	--	2	--	4
		Total		**27**		**18**	**08**	**20**	**20**

* HC – Hard Core SC – Soft Core HCL – Hard Core Lab SCL – Soft Core Lab OE – Open Elective

IV Semester

Sl No	Subject Code	Subject Title	Credits	Max. Marks for Dissertation / Valuation		External Evaluation		Total
				Dissertation I	Dissertation II	Dissertation III	Max. Marks for Viva-voce (Final)	
1	PH 801.4	Industry Internship / Project Work/Dissertation	16	SRS & SDD (100)	Mid Term Evaluation (100)	Record Reading (100)	200	500

Sl No	Subject Code	Subject Title	Credits	Internal Evaluation	External Evaluation	Total
2	PS 802.4	Domain Knowledge Project	04	Assignments / Paper Presentation(50)	Documentation | Viva Voce (100)	150
	Total		**20**			**650**

SYLLABUS - SCHEME OF EXAMINATION

I Semester

Sl No	Subject Code	Subject Title	Type	Credits	Theory Exam duration (hours)	Practical Exam duration (hours)	Max Marks for Internal Assessment	Max. Marks for Term End Exam	Total Marks
HARDCORE THEORY & PRACTICALS									
1	PH 801.1	Statistical Methods	HC	3	3	--	30	70	100
2	PH 802.1	Probability and Stochastic Process	HC	3	3	--	30	70	100
3	PH 803.1	Linear Algebra and Linear Programming	HC	3	3	--	30	70	100
4	PH 804.1 P	Computing for Data Sciences Lab	HCL	2	--	3	25	50	75
SOFTCORE THEORY & PRACTICALS									
5	PS 805.1	Database Management System	SC	3	3	--	30	70	100
6	PS 806.1	Python Programming	SC	3	3	--	30	70	100
7	PS 807.1 P	DBMS & Python Programming Lab	SCL	3	--	3	25	50	75
8	---	Project Work & Seminar	---	--	--	---	---	---	---
		Total		**20**	**---**	**----**	**200**	**450**	**650**

* HC – Hard Core SC – Soft Core HCL – Hard Core Lab SCL – Soft Core Lab OE – Open Elective

II Semester

Sl No	Subject Code	Subject Title	Type	Credits	Theory Exam duration (hours)	Practical Exam duration (hours)	Max Marks for Internal Assessment	Max. Marks for Term End Exam	Total Marks
HARDCORE THEORY & PRACTICALS									
1	PH 801.2	Machine Learning – I	HC	3	3	--	30	70	100
2	PH 802.2	Enabling Technologies for Data Science – I	HC	3	3	--	30	70	100
3		HC Electives: (Choose ONE)	HC	3	3	--	30	70	100
	PH 803.2 E1	Operations Research							
	PH 803.2 E2	Cloud Computing							
	PH 803.2 E3	Natural Language Processing							
	PH 803.2 E4	Unix Programming							
	PH 803.2 E5	Operating Systems							
	PH 803.2 E6	Multivariate Statistics							
4	PH 804.2 P	Machine Learning and Data Science Lab – I	HCL	3	--	3	25	50	75
SOFTCORE THEORY & PRACTICALS									
5	PS 805.2	Foundations of Data Science	SC	3	3	--	30	70	100
6	PS 806.2	Advanced Statistical Methods	SC	3	3	--	30	70	100
7	PS 807.2	Value Thinking	SC	1	---	---	50	----	50
8	PS 808.2 P	Programming for Big Data and Advanced Statistical Methods Lab	SCL	3	--	3	25	50	75
OPEN ELECTIVE – Offered to other Depts									
9	OE 809.2	Statistical Data Analysis using R	SC	3 *	3	--	30	70	100
10	---	Project Work & Seminar	---	--	--	---	---	---	---
		Total		**25**	**---**	**----**	**280**	**520**	**800**

* HC – Hard Core SC – Soft Core HCL – Hard Core Lab SCL – Soft Core Lab OE – Open Elective

III Semester

Sl No	Subject Code	Subject Title	Type	Credits	Theory Exam duration (hours)	Practical Exam duration (hours)	Max Marks for Internal Assessment	Max. Marks for Term End Exam	Total Marks
HARDCORE THEORY & PRACTICALS									
1	PH 801.3	Machine Learning – II	HC	3	3	--	30	70	100
2	PH 802.3	Enabling Technologies for Data Science – II	HC	3	3	--	30	70	100
3	PH 803.3 P	Machine Learning and Data Science Lab – II	HCL	3	--	3	25	50	75
SOFTCORE THEORY & PRACTICALS									
4	PS 804.3	Data Visualization with Tableau & Modelling in Operations Management	SC	3	3	--	30	70	100
5		SC Electives: (Choose ONE)	SC	3	3	--	30	70	100
	PS 805.3E1	Introduction to Econometrics and Finance							
	PS 805.3E2	Time Series Analysis and Forecasting							
	PS 805.3E3	Bioinformatics							
	PS 805.3E4	Big Data Technologies and Architecture							
6		SC Electives: (Choose ONE)	SC	3	3	--	30	70	100
	PS 806.3E1	Intellectual Property Rights (IPR)							
	PS 806.3E2	Cyber Security							
	PS 806.3E3	Text Mining							
	PS 806.3E4	Advanced Analytics							
7	PS 807.3 P	Data Visualization with Tableau & Operation Management Lab	SCL	3	--	3	25	50	75
8	PS 808.3 P	Lab on Electives-1 & 2	SCL	3	--	3	25	50	75
OPEN ELECTIVE – Offered to other Depts.									
9	OE 809.3	Big Data & Design Thinking	SC	3	3	--	30	70	100
10	---	Project Work & Seminar	---	--	---	--	---	--	---
		Total		**27**	**---**	**----**	**255**	**570**	**825**

IV Semester

Sl No	Subject Code	Subject Title	Credits	Max. Marks for Dissertation / Valuation		External Evaluation		Total
				Dissertation I	Dissertation II	Dissertation III	Max. Marks for Viva-voce (Final)	
1	PH 801.4	Industry Internship / Project Work/ Dissertation	16	SRS & SDD (100)	Mid Term Evaluation (100)	Record Reading (100)	200	500

Sl No	Subject Code	Subject Title	Credits	Internal Evaluation	External Evaluation	Total
2	PS 802.4	Domain Knowledge Project	04	Assignments / Paper Presentation(50)	Documentation | Viva Voce (100)	150
	Total		**20**			**650**

Pre-course Work

1. Microsoft Excel for Data Analysis
    a. *Excel Tables, Filters, Sorting*
    b. *Pivot Tables and Charts*
    c. *Formats, Formulas, Dates*
    d. *Functions – Mathematical, Statistical, Text, Date*

Reference:

On-line courses/Tutorials:

i. Microsoft Virtual Academy:
    a. *Analyzing and Visualizing Data with Excel https://mva.microsoft.com/en-US/training-courses/analyzing-and-visualizing-data-with-excel-11157*
    b. *Data Analysis with Excel https://mva.microsoft.com/en-US/training-courses/data-analysis-with-excel-16654*

ii. Edx.Org:
    e. *Introduction to Data Analysis using Excel https://www.edx.org/course/introduction-to-data-analysis-using-excel-0*

iii. Coursera.org:
    f. *Introduction to Data Analysis Using Excel https://www.coursera.org/learn/excel-data-analysis*

2. Basic Unix Programming
    a. *Basic Unix Commands*
    b. *Handling files and folders*
    c. *Concatenation, find and replace, modify file & texts*
    d. *Basic summary commands*

Reference:

On-line courses/Tutorials:

Data Camp:
    e. *Introduction to Shell for Data Science https://www.datacamp.com/courses/introduction-to-shell-for-data-science*

Linux.Org:
    f. *Linux Beginner Tutorials https://www.linux.org/forums/linux-beginner-tutorials.123/*
    g. *Github - Organizing with Unix: https://rafaLabgithub.io/dsbook/organizing-with-unix.html*

Book:
    h. *Data Science at the Command Line, JeroenJanssens, https://www.datascienceatthecommandline.com/*

SYLLABUS

SEMESTER – I

PH 801.1: STATISTICAL METHODS

Total No. of Lectures: 60 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 0 - 2 - 2]**

Course Objectives:

- To describe the critical elements of data collection by using var5ious techniques.
- To gain knowledge of the exploratory data analysis using visualization.
- To be able to understand how to present data effectively.
- To know the appropriate distribution of the data at hand.
- To gain insight on analysis of qualitative data using contingency table.

Course Outcomes: On completion of the course the student will be capable

- To design appropriate instruments to collect data effectively.
- To provide effective data visualization that will provide new insights from the data.
- To Organize, manage and present data effectively.
- To analyze statistical data graphically using frequency distributions.
- To Construct and interpret Contingency Tables

a) **Data Collection & Visualization** (25hrs. – Theory 17hrs. + Lab 8 hrs.)
Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date collection systems with examples from business, cleaning and treatment of missing data.

b) Principles of data visualization, and different methods of presenting data in business analytics.

c) **Basic Statistics:**(25 hrs. – Theory 17hrs. + Lab 8hrs.)
Frequency table, histogram, measures of location, measures of spread, skewness, kurtosis, percentiles, box plot, correlation and simple linear regression, partial correlation,

d) Probability distribution as a statistics model, fitting probability distributions, empirical distributions, checking goodness of fit through plots and tests.

e) **Contingency Tables:** *(*10 hrs. – Theory 8 hrs. + Lab 2 hrs.)
Two way contingency tables, measures of association, testing for dependence.

SUGGESTED BOOKS:

1. *Statistics: David Freedman, Pobert Pisani & Roger Purves, WW.Norten & Co. 4th Edition 2007.*

2. *The visual display of Quantitative Information: Edward Tufte, Graphics Press, 2001.*

3. *Best Practices in Data Cleaning: Jason W. Osborne, Sage Publications 2012.*

Evaluation:Theory: 70% + Practical/Lab : 30%

PH 802.1: PROBABILITY & STOCHASTIC PROCESS

Total No. of Lectures: 60 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 0 - 2 - 2]**

Course Objectives:

- To Understand the basic concepts of probability and their importance
- To Understand the concept of random variables their associated probability distributions,
- To gain insights on the need for continuous distribution and its applications.
- To know the importance of stochastic process and its application.
- To know equip students with various forecasting techniques and knowledge on modern statistical methods for analyzing time series data.

Course Outcomes: On successful completion of the course students will be able:

- To calculate the probabilities and identify the various types.
- To express the features of discrete random variables and formulate the distribution functions.
- To express the features of continuous random variables and formulate the distribution functions
- To Classify a stochastic process according to whether it operates in continuous or discrete time and whether it has a continuous or a discrete state space. To Understand the concept of Markov chains and study the transition diagram.
- To apply the concept of stationarity to the analysis of time series data in various contexts.

a) **Basic Probability:** (20 hrs. – Theory 18 hrs. + Lab 2 hrs.)
Concepts of experiments, Outcomes, Sample space, Events, Combinatorial probability, Birthday paradox, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem.

b) **Probability Distribution:** (20 hrs. – Theory 16hrs. + Lab 4hrs.)
Random Variables: discrete and continuous probability models, some probability distributions: Binomial, Poisson, Geometric, Hypergeometric.

c) Normal, exponential, Chi-square, expectation, variance and other properties of the distribution.

d) **Stochastic Process:**(10 hrs. – Theory 4 hrs. + Lab 6 hrs.)
Markov Chains, Classification of states, Stationery distribution, limit theorems, Poisson process, illustrations and applications.

e) **Introduction to Time Series*:***(10 hrs. – Theory 4 hrs. + Lab 6 hrs.)
Components of time series, Smoothing auto correlation, stationarity, concepts of AR, MA, ARMA & ARIMA models with illustrations.

SUGGESTED BOOKS:

1. *A First Course in Probability: Shelden M. Ross, 2014.*

2. *Introduction to Stochastic Process : Paul G. Hoel, Sydney C. Port & Charles J. Stone, Waveland Press, 1987.*

3. *Time Series Analysis and Its Applications: Robert H. Shumway and David S. Stoffer, Springer 2010.*

Evaluation:Theory: 70% +Practical/Lab: 30%

PH 803.1: LINEAR ALGEBRA & LINEAR PROGRAMMING

Total No. of Lectures: 60	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 0 - 2 - 2]**

Course Objectives:

- To Learn Matrix Operations and its properties
- To Learn Random Number generations and its properties
- To learn solving set of linear equations
- To learn mathematically formulate an applied word problem involving revenue, costs, and
- constraints as a linear program and convert a linear programing problem into standard form
- To apply the simplex algorithm to solve a linear programming problem

Course Outcomes:

On successful completion of the course students will be able:

- Understand the basic concepts of linear Algebra
- Understand the concept of Random Numbers and its properties.
- Understand the principles of solving a set of linear equations,
- Familiarize with the methods involved in solving a set of linear equations.
- To model a problem as a linear programming problem
- Use the simplex method to solve small linear programming models by hand, given a basic feasible point.

a) **Linear Algebra:**(40 hrs. – Theory 28hrs. + Lab 12 hrs. using R)
Linear equations and matrices, matrix operations.

b) solving system of linear equations, Gauss-Jordan method, Concept & Computation of determinant and inverse of matrix, Eigen values and Eigen vectors.

c) Illustrations of the methods, Positive semi definite and position definite matrices, illustrations

Lab – using R programming

d) **Linear** Programming (20 hrs. – Theory 14hrs. + Lab 6hrs.)
Definition of the problem, convex sets, corner points.

e) feasibility, basic feasible solutions, Simplex method

SUGGESTED BOOKS:

1. *Linear Algebra and Its Application: Gilbert Strang, 4th Edition, Academic Press.*
2. *Hands-On Matrix Algebra Using R (Active and Motivated Learning with Applications), Hrishikesh D Vinod, World Scientific*
3. *Linear Programming: G. Hadley, Addison-Wesley.*

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 804.1P: COMPUTING FOR DATA SCIENCES LAB

Total No. of Practical's : 90 **Total Marks : 75** **[L – T – P – S]**

No. of Labs / Week : 3 **Credits : 2** **[0 – 2 - 6–0]**

Course Objectives:

- To apply the theoretical concepts and gain practical knowledge using the R and Python. Here the students will learn the various data analysis that could be performed using the packages.
- To be able to understand the concept of mathematical convergence using the numerical techniques.
- To understand the concept of random numbers and their use.

Course outcomes: on completion of this course the students will be able

- To perform data analysis using the appropriate techniques.
- To know how convergence, takes place and use the appropriate methods**.**
- To generate random numbers and understand how a system can be simulated using them.

a) **Computer Packages**– R and Python: *(20 hrs. – Theory 10hrs. + Lab 10 hrs.)*
Usage of R and Python – data handling, data analysis, statistical modeling with illustration in python and R (10 hrs. R + 10 hrs. Python).

b) Data Structure &**Concepts of Computation** using Java(20 hrs. – Theory 6 hrs. + Lab 14hrs.)
Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods e.g. Newton-Raphson, Steepest ascent using Java

c) **Computing Methodologies:** (20 hrs. – Theory 8 hrs. + Lab 12 hrs.)
Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data.

SUGGESTED BOOKS:

1. *Introduction to Data Science (Data Analysis and Prediction Algorithms with R), Rafael A. Irizarry, https://rafaLabgithub.io/dsbook/*
2. *Hands-On Programming with R - Write Your Own Functions and Simulations, Grolemund Garrett, O'Reilly*
3. *Data Structures and Algorithm using Java, 6th Ed. Michael T. Goodrich and Roberto Tamassia, John Wiley & Sons, Inc*
4. *Python Data Science Handbook - Essential Tools for Working with Data, Jake VanderPlas, O'Reilly*
5. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, WES MCKINNEY, O'Reilly*

Evaluation:Theory: 40% + Practical/Lab : 60%

PS 805.1: DATABASE MANAGEMENT SYSTEM

Total No. of Lectures: 45	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 0 - 2 - 2]**

Course objectives: The objectives of this course are:

- Understanding the basic concepts of DBMS and representing the same with an ER diagram
- Learning the structure of Relational Database and its various operations , Normalizing the tables, purpose of NoSQL
- Overall understanding of Graph database, Parallel and Distributed database
- Practical implementation of ORACLE SQL, MongoDb
- Understanding the basic of Hadoop Ecosystem and implementing database security

Course Outcomes: At the end of the course, students should be able to:

- Draw an ER Diagram for a given system by analysing the requirements
- Normalize the tables atleast to 3N form and perform various operations on tables that are thus created
- Appreciate and apply Graph database
- Execute queries, implement views and joins, use MongoDB for various operations on unstructured data
- Work with Hadoop Ecosystem and also implement database security in SQL, NoSQL and Hadoop

a) **Basic Concepts:** (10 hrs. – Theory 10hrs.)
Different data models, ER and EER diagram, schema, table, Big Data Concepts and Hadoop Ecosystem

b) **Relational and Non-Relational Databases:** (25 hrs. – Theory 10 hrs. + Lab 15 hrs.)
Structure, various operations, normalization, SQL, No-SQL,

c) Graph Database, Parallel and distributed data base, Map-Reduce.

Lab using SQL/Oracle/MySql for Relational databases; Hadoop(any), MangoDB, GraphDB for Big Data

d) **Implementation:** (25 hrs. – Theory 10hrs. + Lab 15 hrs.)
ORACLE SQL/MS SQL/MySQL,

e) Hadoop Ecosystem, Concept of database security.

SUGGESTED BOOKS

1. *Database system concepts: Abraham Silberschartz, Henry F. Korth and S. Surarshan, McGraw Hill, 2011.*
2. Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem, Douglas Eadline, Addison-Wesley, Pearson Education India; First edition (1 March 2016)
3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, 2015

Evaluation: *Theory: 60% + Practical/Lab: 40%* **(***Oracle SQL, MS SQL, Hadoop Ecosystem, MangoDB)*

PS 806.1:PYTHON PROGRAMMING

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week : 3 **Credits : 3** **[3 – 0 - 2 - 2]**

Course objectives:

- To Introduce Python as a Programming language and understand the concept of the collections module in python
- To understand how control structures work in python and also to write and call functions appropriately
- To perform File based operations and manipulating Packages or Libraries available in python; Understand how math, statistics, regular expression and time modules work
- To handle exceptions and errors in a program and also to raise user-defined exceptions
- To introduce the concept of Classes and Objects in an Object Oriented Programming environment and also to integrate Python with a Database

Course outcomes: At the end of the course, students must be able to

- Choose the right data type or Collection module for any given set of data.
- Use conditional statements and loops to manipulate; Create, use & reuse functions created from python
- Open, Read and Write a File from Python and also to import and use various logical modules in python
- Handle any type of exceptions that might be raised from a typical program
- Create classes and objects to perform operations and also to perform CRUD Operations on a SQLite Database

a) Introduction to Python interpreter
Python program; Python's execution model; Everything is an object –Numbers, Immutable sequences, Mutable sequences, Set types, Mapping types – dictionaries, The collections module

b) Control statements, functions
Functions: Scopes and name resolution, Input parameters, Return values, Recursive functions, Anonymous functions,
Function attributes, Built-in functions, Documenting code, importing objects.

c) I/O, File handling, Packages/Libraries,
File handling functions: opening a file, read functions, write functions,
Modules and Packages: Modules, Packages, math module, statistics module, regular expression module, time module

d) Exception Handling,
Exception handling in python: try block, except block, finally block, else block. Excpetion Handling in functions, Types of Errors
Raising Exceptions, User Defined Exceptions, Handing user defined exceptions

e) OO Programming.
Abstraction and Entity, Encapsulation and Data Hiding, Classes and Object, constructor and destructor in python, classmethods and staticmethods, Inheritance & Polymorphism, Relationships
Database in Python: Python Database Integration using sqlite

PS 807.1 P: DBMS & PYTHON PROGRAMMING LAB

Total No. of Lectures: 60	**Total Marks: 75**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[0 – 2 - 8 - 0]**

Course objectives:

- To understand the working of python interpreter, various data types and collections, controls structures and looping.
- To create and call functions based on operations, file handling, manipulating Packages or Libraries available in python
- To use Classes and Objects, perform exception handling and also to integrate Python with a Database
- Create an understanding of various DDL, DML commands in depth using Oracle SQL along with the purpose of Cursors and Triggers
- Create an understanding of using NoSQL and Hadoop Ecosystem

Course outcomes: At the end of the course, students must be able to

- Solve real world problems using python as a programming language
- Create applications that handle files and include various packages to solve complex issues
- Create a completely data driven application that includes exception handling and perform all database related operations.
- Create a table, Execute complex and nested queries , create views and joins and also execute cursors and triggers using Oracle SQL
- Use MongoDb to create Database, Collection, Document etc. and also understand Hadoop Ecosystem

DBMS:

1. Creating a table and performing various operation on the table.
2. Simple queries using WHERE clause, ORDER BY and Aggregate functions
3. Subqueries using single table
4. Subqueries using multiple tables
5. Creating and working with Views
6. Demonstrating the use of Joins
7. Write a PL/SQL block to find the maximum and minimum number among two numbers. Also find if maximum number is divisible by minimum number.
8. Write a PL/SQL block to find the sum of individual digits in a number.
9. Write a PL/SQL block to create an explicit cursor to hold records from customers table and display them.
10. Write a PL/SQL block to implement implicit cursor to hold the sum of salary of employees from Electronics dept.
11. Write a PL/SQL block to create a Trigger to calculate amount in item table upon insertion or updation (Row level).

MongoDB:

1. Creating a database
2. Creating a collection
3. Performing simple queries
4. Performing a multiple queries.
5. Using HDFS and writing commands

Python Programs:

1. Program to accept a password and check for the following conditions: a. Length greater than 8. b. First letter capital. c. Alphanumeric
2. Program to count the no. of votes for 3 candidates standing for the election and display the winner at the end of voting.
3. Program to perform multiplication of two matrices
4. Write a function which accepts a sequence of comma separated 4 digit binary numbers as its input and then check whether they are divisible by 5 or not. Return the no's that are divisible by 5 in a comma separated sequence
5. Program that accepts a sequence of whitespace separated words as input and prints the words after removing all duplicate words and sorting them alphanumerically
6. Function where Given a string, find the length of the longest substring without repeating characters.
7. Program to create dictionaries with even, odd or prime, enter 8 numbers and identify the number as even, odd or prime and add it to the dictionary
8. Program to take a dictionary having stationary items as keys and their rates as values. Ask the user which items he wants to buy, after buying the items display the exact bill
9. Program to make a list whose elements are intersection of the above given lists
10. Write a function subprogram to input and display the contents of an array . Add a function to find a particular no in an array . Display the relevant message
11. Write a function to process n lists and get the median for all the list combined . Use variable length argument as parameter
12. Program to create functions faculty, students and outsiders. Ask the user whether he/she is a faculty, student or outsider and display the respective details according to the users answer
13. Create a class as Statistics which has a list as a variable. Create some statistical functions to manipulate the list and at last have a function called summary
14. Create a class called Student with variables as register number and marks in 3 separate subjects. Use a shared variable to calculate the total no. of students. Create objects of Student class and display student details
15. Program which can map() and filter() to make a list whose elements are square of even number in [1,2,3,4,5,6,7,8,9,10]
16. Use Enumerate - By using list comprehension, write a program to print the list after removing the 0th, 2nd, 4th,6th numbers in [12,24,35,70,88,120,155]

17. Write a script to read data from a text file (paragraph from hamlet). Get the word count of each word in the text

18. Write a script to enter the name, age, phone number of 5 people and write it to a file. Read the file and display only the phone numbers

19. Write a menu driven program to explain the various types of exceptions

20. Write a script to read a file containing data about "Hamlet" (3 paragraphs from the story). Replace all the occurrences of Hamlet in the original file with ******* using a regex

SEMESTER – II

PH 801.2: MACHINE LEARNING - I

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week : 4 **Credits : 3** **[3 – 1 - 0 - 2]**

Course Objectives:

- To understand the fundamental concepts of Linear regression and Multiple linear regression, its application in machine learning
- To understand the fundamental concepts of Logistic regression and problem of overfitting, its application in machine learning
- To understand the concept of Perceptron algorithm and backpropagation.
- To learn how to evaluate machine learning model, handling skewed data and large datasets; to understand SVM with and without kernels
- To understand the concept of unsupervised learning, dimensionality reduction, Anomaly detection and its usage in machine learning

Course Outcomes:

- To implement machine learning models with linear regression
- To design applications using Logistic regression by using the methodology to avoid overfitting
- To design systems using Perceptron algorithm
- To implement machine learning systems using SVM
- To implement machine learning models using k-means clustering by applying dimensionality reduction and anomaly detection

a) **Linear Regression** (10 hrs. – Theory 6hrs. + Lab 4 hrs.)
Linear Regression with Multiple variables, applications.

b) **Logistic Regression:**(10 hrs. – Theory 4hrs. + Lab 6hrs.)
Model, Classification, Problem of over-fitting, Applications.

c) **Neural Networks:** (9 hrs. – Theory 3hrs. + Lab 6 hrs.)
Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application.

d) **Machine Learning System Design:** (8 hrs. – Theory 2hrs. + Lab 6 hrs.)
Evaluating a learning algorithms, handling skewed data, using large data sets. (5 hrs.)

Support Vector Machines:(5 hrs. – Theory 3hrs. + Lab 2 hrs.)
Model, Large Margin Classification, Kernels, SVMs in practice. (5 hrs.)

e) **Unsupervised Learning**. (5 hrs. – Theory 3hrs. + Lab 2 hrs.)
Dimensionality Reduction.(8 hrs. – Theory 6hrs. + Lab 2 hrs.)
Anomaly Detection. (5 hrs. – Theory 3hrs. + Lab 2 hrs.)

SUGGESTED BOOKS:

1. *Machine Learning: Tom Mit chell*

Evaluation:Theory: 50% + Practical/Lab: 50% (*Lab: using R and/or Python*)

PH 802.2: ENABLING TECHNOLOGIES FOR DATA SCIENCE – I

Total No. of Lectures : 45	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week : 4	**Credits : 3**	**[3 – 1 - 0 - 2]**

Course Objectives: Students will learn

- To identify the scope and essentiality of datamining. Identify the application areas of programs and apply them.
- Online Transaction Processing and OnLine Analytical processing and Multidimensional schemas suitable for data warehousing. Learn to study the methodology of engineering legacy database for data warehousing to describe business rules for decision support system.
- Learn size reduction of the input space, smoother relationships, data normalization, noise reduction, and feature extraction, data pre-processing and data cleaning.
- Learn classification on real world data and also predict the possible areas where classification can be applied. Learn to characterize the kinds of patterns that can be discovered by association rule mining, classification.
- Learn different clustering methods and learn to assign observations to groups (clusters), group variables into homogeneous and distinct groups. Understand spatial data, discover the relation between pace and the non-space data, set up the spatial knowledge base, excel the query, reorganize spatial database and obtain concise total characteristic.

Course Outcome: Students will be able to

- To understand data mining principles and will identify appropriate datamining algorithms to solve real-world problems. To understand the strength and weakness of algorithms.
- To design a data mart or data warehouse for any organization. To design data warehouse with dimensional modelling and apply OLAP operations.
- To learn methods in integrating and interpreting the data sets and improving effectiveness, efficiency and quality for data analysis.
- To predict categorical class labels (discrete or nominal) and classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data and also predicts unknown or missing values.
- To identify clusters in multivariate data, apply normalization techniques, and correctly interpret the output of different clustering procedures. And to describe complex data types with respect to spatial and temporal data mining.

(Lab work is must) : NoSQL database, Oozie)

DATA MINING: (60 Hrs.)

a) **Introduction:** (5 hrs. – Theory 5hrs.)
Knowledge discovery from databases, scalability issues.

b) **Data Warehousing:** (8 hrs. – Theory 2 hrs. + Lab 6 hrs.)
General principles, modeling, design, implementation and optimization, Cloud Computing, OLAP.

c) **Data Preparation**: (5 hrs. – Theory 1hrs. + Lab 4 hrs.)
Pre-processing, sub-sampling, feature selection.

d) **Classification and Prediction:** (18 hrs. – Theory 10 hrs. + Lab 8 hrs.)
Bayes learning, decision trees, CART, neural learning, support vector machines, associations, dependence analysis, rule generation.

e) **Cluster Analysis and Deviation Detection:** (14 hrs. – Theory 6hrs. + Lab 8 hrs.)
Partitioning algorithms, Density bases algorithm, Grid based algorithm, Graph theoretic clustering.

Temporal and spatial data mining. (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

SUGGESTED BOOKS

1. *Data Mining Techniques: A. K. Pujari, Sangam Books Ltd., 2001*
2. *Mastering Data Mining: M. Berry and G. Linoff, John Wiley & Sons., 2000*
3. *Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management, Olivia Parr Rud, Wiley*

Evaluation: Theory: 50% + Practical/Lab: 50% (R programming 15 hrs. + Python 15hrs.*)*

Electives (Choose 1)

PH 803.2 (E1):OPERATIONS RESEARCH

Total No. of Lectures: 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objective:

- To formulate a given simplified description of a suitable real-world problem as a linear programming model in general, standard and canonical forms and to sketch a graphical representation of a two-dimensional linear programming model.
- To introduce the fundamentals of nonlinear optimization theory and methods.
- To understand the different types of assignment problems and the use of assignment models in industry and business.
- To comprehend the concept of a Transportation Model and develop the initial solution for the same;
- To find out the optimum service rate and the number of servers so that the average cost of being in queuing system and the cost of service are minimised.

Course Outcome: On completion of the course the student will be able

- To Proficiently deal with the tools for optimization.
- To Develop an understanding of the foundation of classic continuous optimization problems and to identify the convexity, smoothness, feasible region and dual reformulation.
- To proficiently allocate scarce resources to optimize and maximize profit or minimize loss and facilitates the optimal method of allocating jobs to persons.
- To facilitate with mathematical and computational modeling of real decision-making problems.
- To construct and analyse priority queuing systems.

a) Review of Linear Programming. (5 hrs. – Theory 5 hrs.)

b) Non-Linear Programming. (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

c) Assignment Models. (5 hrs. – Theory 1 hrs. + Lab 4 hrs.)

d) Transportation Models. (15 hrs. – Theory 11 hrs. + Lab 4 hrs.)

e) Queuing Models: Characteristics of Queuing Process, Poisson Process, Birth-Death Process, Single-Server Queues, Multi-Server Queues, Queues with Truncation, Finite-Source Queues, Numerical Techniques & Simulation.

(25 hrs. – Theory 19 hrs. + Lab 6 hrs.)

SUGGESTED BOOKS:

1. *Operations Research: PREM Kumar Gupta & D. S. Hira*

2. *Fundamentals of Queuing Theory: Donald Gross, John F. Shortle, James M. Thompson & Carl M. Harris, Fourth Edition, Wiley*

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 803.2 (E2) : CLOUD COMPUTING

Total No. of Lectures : 45	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To understand the fundamental ideas behind Cloud Computing, the evolution of the paradigm, its applicability;
- To understand the cloud benefits, as well as current and future challenges;
- To understand the basic ideas and principles in data center design and management;
- To understand the cloud storage technologies and relevant distributed file systems;
- To understand the variety of programming models and develop working experience in one of them.

Course Outcome:

After successfully completing the course delegates will have an understanding of:

- Apply the fundamental concepts in data centers to understand the trade-offs in power, efficiency and cost.
- Discuss system virtualization and outline its role in enabling the cloud computing system model.
- Illustrate the fundamental concepts of cloud storage and demonstrate their use in storage systems
- Illustrate the fundamental concepts of web services.
- Analyze various cloud programming models and apply them to solve problems on the cloud.

a) Introduction to Cloud computing, Cloud service methods, IaaS, PaaS, SaaS, fundamentals of cloud Architecture (load distribution, resource pooling, scalability, load balancing, redundancy, etc), Introduce DevOps, CICD. Hands-on practice either on AWS/Azure/Google Cloud Platform

b) Cloud computing: General Benefits and Architecture, Business Drivers, Main players in the Field, Overview of Security Issues, XaaS Cloud Based Service Offerings **Key Amazon offerings**: EC2, SimpleDB, S3, Simple Queue, Simple Relational Database, Elastic MapReduce, Virtual Amazon Cloud. S3 Command Line tool

c) **Bundling Amazon instances**: We will learn how to create and manipulate Amazon instances with command line tools, transfer application software to instances and bundle them into new AMI-s that could be offered to the public.

d) **Amazon's Elastic Block Storage (EBS)** provides persistence storage in the cloud. We will learn how to move application code and data from non-EBS instance into EBS volumes, and create our own EBS based AMI-s

Amazon's AWS Identity Management **and** Security in the Cloud, Amazon's Virtual Private Cloud(VPC) and Directory Service, Java AWS SDK, S3 API, Relational Database Service, SimlpeDB Service , NoSQL Databases, **Amazon's Messaging in the Cloud**

d) **Amazon's RESTFul WebServices** AWS APIs are sufficiently rich to allow you easy interaction with AWS service. However, in order to establish connectivity between your own modules in the Cloud you should use RESTFul Web Services

SUGGESTED BOOKS:

[1] Rajkumar Buyya, James Broberg, Andrez M Goscinski, "Cloud Computing: Principles and Paradigms", Wiley International.

[2] Barrie Sosinsky, "Cloud Computing Bible", Wiley India

[3] Thomas Erl, Ricardo Puttini, Zigham Mummod, "Cloud Computing : Concepts, Technology & Architecture", Pearson Education Asia

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 803.2 (E3):NATURAL LANGUAGE PROCESSING

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To understand about NLP. To know about syntax, semantics, and pragmatics. To understand some of the NLP applications like MT, IE, question-answering system. To learn about the role of language models and N-gram model.
- To learn parts-of-speech tagging, Hidden Markov models and LSTM neural networks.
- To learn about Context-Free Grammar (CFG). Understand about syntactic parsing and PCFG.
- To Understand about semantic parsing, word-sense disambiguation, IE using sequence labelling.
- To learn about issues in MT, word alignment, phrase-based translation

Course Outcomes:

- Analyse syntax, semantics, and pragmatics of NLP. Ability to develop simple N-gram models
- Perform POS tagging on simple English sentences using Hidden Markov model
- Develop grammars for some simple English sentences, ability to draw parse trees. Apply different parsing techniques
- Analyse syntactic, semantic and pragmatic ambiguities, learn to apply supervised and unsupervised word-sense disambiguation.
- Analyse different Machine translation approaches.

a) **Introduction** - 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.**N-gram Language Models** - The role of language models. Simple N-gram models. Estimating parameters and smoothing. Evaluating language models.

b) **Part Of Speech Tagging and Sequence Labeling** - Lexical syntax. Hidden Markov Models (Forward and Viterbi algorithms and EM training).LSTM Recurrent Neural Networks, optionally the original paper Long Short Term Memory.

c) **Syntactic parsing** - Grammar formalisms and treebanks. Efficient parsing for context-free grammars (CFGs). Statistical parsing and probabilistic CFGs (PCFGs). Lexicalized PCFGs. Neural shift-reduce dependency parsing

d) **Semantic Analysis**- Lexical semantics and word-sense disambiguation. Compositional semantics. Semantic Role Labeling and Semantic Parsing.; **Information Extraction (IE)** -amed entity recognition and relation extraction. IE using sequence labeling.

e) **Machine Translation (MT)** -Basic issues in MT. Statistical translation, word alignment, phrase-based translation, and synchronous grammars.

SUGGESTED BOOKS:

1. *Daniel Jurafsky and James H Martin. Speech and Language Processing, Pearson Education, 2016*
2. *James A.. Natural language Understanding 2e, Pearson Education, 2010*
3. *2. Bharati A., Sangal R., Chaitanya V.. Natural language processing: a Paninian perspective, PHI, 2012*
4. *Siddiqui T., Tiwary U. S.. Natural language processing and Information retrieval, OUP, 2018*

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 803.2 (E4): UNIX PROGRAMMING

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To learn the basics of Unix operating system with introduction of file system and shell commands using Vi editor.
- To learn basics and advanced topics of file attributes with some filters and grep command applications or usages.
- To learn interfaces of Unix with some system calls and its usages.
- To learn some advance system calls which handle signaling, multiplexing and Inter process communications.
- To learn advanced inter-process communication concepts using pipes and streams.

Course Outcomes:

- Students are able to know an overview of Unix operating system and uses of shell commands**.**
- Students will able to understand the concept of I-node and its use with applications of grep commands.
- Students get know about user and program interface with some system calls requirement and its applications.
- Students are able to know use of signaling and importance of Inter process communications.

Students will understand the importance and application of inter-process communications

a) Overview of The UNIX Operating System General Purpose Utilities. File system & Handling ordinary Files. Shell commands & simple programming. (Bourne Shell) Vi editor advanced Vi Editor.

b) Basic & More File attributes Concept of I-Node. Simple filters. grep command. Overview of process. Overview of sed & awk.

c) Organisation of Unix. User interface, Programmer interface. The environment of Unix process System calls. Process control, File related system calls.

d) Process related system calls. Signals programming using system calls. Advanced I/O multiplexing. Memory mapped I/O. Inter-process communication: Pipes, shared memory, semaphores, messages.

e) Advanced inter-process communications. Streams, Pipes, Open server. Basics of Visual programming tools like X-windows.

SUGGESTED BOOKS:

1. *UNIX: Concepts & Applications, Sumitava Das, TMH*
2. *2.Your UNIX –The Ultimate Guide, Sumitava Das, TMH*
3. *3.Design of UNIX Operating System,Maurice Bach, PHI*
4. *UNIX Systems Administration, Maxwell, TMH*
5. *4.UNIX Power Tools,Powers,SPD/O'REILLY*

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 803.2(E5): OPERATING SYSTEMS

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To learn the fundamentals operating systems – need, characteristics and functionalities with an overview of working of operating system.
- To learn the fundamentals UNIX operating systems with some system calls and working of UNIX with interrupts and scheduling.
- To learn about concurrent process and need of mutual exclusion in case of a shared resource access by different process.
- To learn some concepts of Monitors, Inter process communications and cause of deadlocks – prevention and avoidance.
- To learn need and importance of virtual memory and file organization or structure of UNIX.

Course Outcomes:

- Students are able to understand the basics of operating systems with need and working**.**
- Students will able understand the fundamentals of UNIX operating system with signals and system class.
- Students will able to understand fundamentals of concurrent process and concept of mutual exclusion and implementation of semaphores.
- Students are able to understand importance of Inter process communications resulting deadlocks which can be prevented or avoided with some algorithms.
- Students will understand the importance and benefits of virtual memory. The file structure of UNIX operating system.

a) Overview of operating systems, functionalities and charateristics of OS; Hardware concepts related to OS, CPU states, I/O channels, memory hierarchy, microprogramming; The concept of a process, operations on processes, process states, concurrent processes, process control block, process context.

b) UNIX process control and management, PCB, signals, forks and pipes;Interrupt processing, operating system organisation, OS kernel FLIH, dispatcher; Job and processor scheduling, scheduling algorithms, process hierarchies.

c) Problems of concurrent processes, critical sections, mutual exclusion, synchronisation, deadlock.; Mutual exclusion, process co-operation, producer and consumer processes; Semaphores: definition, init, wait, signal operations.; Use of semaphores to implement mutex, process synchronisation etc., implementation of semaphores.

d) Critical regions, Conditional Critical Regions, Monitors, Interprocess Communication (IPC), Message Passing, Direct and Indirect; Deadlock: prevention, detection, avoidance, banker's algorithm.Memory organisation and management, storage allocation.

e) Virtual memory concepts, paging and segmentation, address mapping.; Virtual storage management, page replacemant strategies.; File organisation: blocking and buffering, file descriptor, directory structure; File and Directory structures, blocks and fragments, directory tree, inodes, file descriptors, UNIX file structure.

SUGGESTED BOOKS:

1. *Avi Silberschatz, Peter Baer Galvin, Greg Gagne, Operating System Concepts, 9th Edition, John wiley & Sons,*
2. *William Stallings, Operating Systems: Internals and Design Principles, 8th edition Pearson Education Limited, 2014*
3. *D.M Dhamdhere: Operating systems - A concept based Approach, 3rd Edition, Tata McGraw- Hill, 2012. 2010.*
4. *Harvey M Deital: Operating systems, 3rd Edition, Pearson Education, 2011.*

Evaluation:Theory: 70% + Practical/Lab: 30%

PH 803.2 (E6): MULTIVARIATE STATISTICS:

Total No. of Lectures: 45	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To Demonstrate knowledge and understanding of the basic ideas behind several common statistical techniques for analysing multivariate data.
- To demonstrate the use of PCA in dimension reduction and analysis of multivariate data.
- To consolidate the volume of data in such a way that similarities and differences can be quickly understood.
- To understand the terminology of factor analysis, including the interpretation of factor loadings, specific variances, and communalities.
- To understand the importance of clustering of data which will help to identify patterns in the data.

Course Outcome: On successful completion of the course the student will be able

- To identify the most appropriate statistical techniques for a multivariate dataset and carry out and apply commonly used multivariate data analysis techniques, and interpret results
- To carry out a principal components analysis Assess how many principal components are needed and Interpret principal component scores.
- To classify data using appropriate algorithms.
- To describe the difference between Factor Analysis (FA) and Principal Component Analysis (PCA) and will be able to extract factors that describe the data.
- To Create a document retrieval system using k-nearest neighbors. -Identify various similarity metrics for text data.

a) Representation of multivariate data, bivariate and multivariate distributions, multinomial distribution, multivariate normal distribution, sample mean & sample dispersion matrix, concepts of location & depth in multivariate data.
(20 hrs. – Theory 12 hrs. + Lab 8 hrs.)

b) Principal Component Analysis (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

c) Classification (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

d) Factor Analysis (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

e) Clustering (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

SUGGESTED BOOKS:

1. *Applied Multivariate Statistical Analysis: Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002*

Evaluation:Theory: 60% + Practical/Lab: 40%

PH 804.2P: MACHINE LEARNING AND DATA SCIENCE LAB - I

Total No. of Lectures: 45	**Total Marks : 75**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Learning Objectives: This subject enables student to

- master the basics in business intelligence (BI), data mining (DM), and knowledge discovery in databases;
- learn the role that software tools/applications play in BI and DM, with emphasis on industrial case studies and practical applications;
- Have an overall understanding of the major issues and applications in business intelligence and data mining, including a basic grasp of the algorithm classes and best practices for building successful BI projects

Learning Outcome: Upon completion of the subject, students will be able to

- examine the concepts of data warehousing and OLAP;
- apply the concepts of BI and DM techniques for clustering, association, and classification;
- understand the operation procedures of BI projects in an organization;
- select appropriate DM tools and methods to manipulate and achieve data;
- apply DM concepts for formulating business strategies and programs to enhance business intelligence.

Machine learning:

1. Linear regression Models.
    a. Simple and multiple models with checking the assumptions.
2. Logistic regression Models and problems of over and under fitting.
3. Neural Networks Models.
    a. Single layer
    b. Multi-layer
4. Support vector machine algorithm.
5. K-means and Hierarchical clustering.
6. Unsupervised learning algorithms.
7. Dimension reduction methods
    a. PCA
    b. Factor Analysis
8. Checking for anomaly using anomaly detection method.
9. Comparing performance of classifier of a classifier before and after dimensionality reduction by applying LDA.

Enabling technologies:

1. Extracting and staging of data from sources
2. Cleaning and aligning of data and Summarizing data
3. Data transformation
4. Classification algorithms
    a. Bayesian approach
    b. Decision tree
    c. CART
    d. Neural learning
    e. SVM
    f. Association and rule generation.
5. Clustering
    a. Partitioning Algorithms
    b. Density based algorithms

c. Grid based algorithms.

6. Temporal and spatial data mining techniques

PS 805.2: FOUNDATIONS OF DATA SCIENCE

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objective: The objective of this course is to make the students

- Understand and apply the fundamental concepts of graph theory
- To understand mathematical concepts relevant to high-dimensional data
- To learn the complex understanding of large structures, like the web and social networks, and also building models to capture essential properties of these structures.
- Learning the concept of singular value decomposition (SVD) for dimension reduction of high-dimensional data sets, and multi-dimensional scaling and its connection to principle component analysis.
- Learning the concept of frequency moments of data streams and matrix algorithms in streaming model

Course outcomes: Upon successful completion of this course, the student will able to

- Solve problems using basic graph theory
- Applying various concepts relevant with high-dimensional data.
- Understanding large structures, like the web and social networks, in building models.
- Applying the use of singular value decomposition (SVD) for dimension reduction of high-dimensional data sets, and multi-dimensional scaling and its connection to principle component analysis.
- Applying the concept of frequency moments of data streams and matrix algorithms in streaming model

a) **Graph Theory:** (10 hrs. – Theory 4hrs. + Lab 6 hrs.)
Basic Concepts, Algorithms for connectedness, shortest path, Minimum Sampling Tree, Lab: Graph Databases, Java/Python Programming

b) **High Dimensional Space:**(12 hrs. – Theory 6hrs. + Lab 6 hrs.)
Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications.
Lab: Graph Databases, Java/Python Programming

c) **Random Graphs:**(12 hrs. – Theory 6hrs. + Lab 6 hrs.)
Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.
Lab: Graph Databases, Java/Python Programming

d) **Singular Value Decomposition (SVD):** (5 hrs. – Theory 1hrs. + Lab 4 hrs.)
Best rank k approximation, Power method for computing the SVD, Applications.
Lab: R and Python Programming (Optional: Matlab/Octave)

Random Walks: (5 hrs. – Theory 1hrs. + Lab 4 hrs.)
Reflection Principle, Long leads, Changes of Sign, Illustrations.
Lab: R and Python Programming (Optional: Matlab/Octave)

e) **Algorithm for Massive Data Problems:**(16 hrs. – Theory 6hrs. + Lab 10 hrs.)
Frequency Moments of data streams, matrix algorithms.
Lab: R and Python Programming (Optional: Spark, Matlab/Octave)

SUGGESTED BOOK:

1. *Foundations of Data Science: John Hopcroft & Ravindran Kannan.*

Evaluation:Theory: 40% + Practical/Lab:60% (*GraphDB, R, Python, Java)*

PS 806.2: ADVANCED STATISTICAL METHODS

Total No. of Lectures:45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To Understand problem of statistical inference, problem of point estimation. Properties of point estimator such Consistency, Unbiasedness, Sufficiency.
- To understand hypothesis testing procedures based on type of outcome variable and number of sample.
- To understand methods based on linear models to data analysis, with proper attention to underlying assumptions and a major emphasis on the practical interpretation and communication of results.
- To gain a conceptual and practical introduction to the basic concepts and techniques of regression analysis.
- To understand the basic ideas behind modeling categorical data with binary logistic regression and to fit the model and interpret the results.

Course Outcomes: On successful completion of the course students will be able:

- To estimate population parameters using point and interval estimates.
- To recognize the logic behind a hypothesis test and how it relates to the P-value.
- To know the theoretical foundation of applied linear modeling, starting with the univariate models and then with multivariate data
- To apply multiple linear regression analysis, differentiate between simple linear regression analysis and multiple linear regression analysis and predict the model and interpret it.
- To apply the functional form of the logistic model and how to interpret model coefficients.

a) **Estimation:**(15 hrs. – Theory 13hrs. + Lab 2 hrs.)
Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. (15 hrs.)

b) **Test of Hypotheses**: (15 hrs. – Theory 13hrs. + Lab 2 hrs.)
Two types of errors, test statistic, parametric tests for equality of means & variances.

c) **Linear Model:**(15 hrs. – Theory 9hrs. + Lab 6 hrs.)
Gauss Markov Model, least square estimators, Analysis of variance.

d) **Regression**: (15 hrs. – Theory 9hrs. + Lab 6 hrs.)
Multiple linear regression, forward, backward & stepwise regression,

e) Logistic Regression.

SUGGESTED BOOKS:

1. *Statistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.*

2. *Introduction to Linear Regression Analysis: Douglas C. Montgomery*

Evaluation:Theory: 70% + Practical/Lab: 30% (*Lab work: Using R and Python***)**

PS 807.2: VALUE THINKING

Total No. of Lectures: 30	**Total Marks: 50**	**[L – T – P – S]**
No. of Lectures / Week: 2	**Credits : 1**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To develop skills in recognizing and analyzing ethical issues and present this analysis in written form
- To understand cross cultural variations and similarities in organizational practices in corporate social responsibility and business ethics
- To diagnose sources of organizational ethical culture and deviant behavior
- To design ethical programs designed to accomplish specific objectives in anizations
- To develop ethical leadership skills and practices

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

- Recognize important ethical issues that arise in various business contexts and professional practice;
- Demonstrate an understanding of the ethical, social and economic environments in which those occur;
- Demonstrate critical thinking skills required for the successful practice of management and the professions within the framework of societal values;
- Demonstrate confidence in introducing ethical considerations into professional and managerial decision making and explaining their importance to others; and
- Use their ethical imaginations in resolving dilemmas and enhancing business decision-making.

(It does not need full class contact hours. 30 Hours of contact classes. Students have to learn themselves through Movies and Books).

This course involves watching few movies (list provided below) and reading few books (list provided below) that deals mostly with argumentative logic, evidence, drawing inference from evidences. After watching the movies and reading the books, there will be general discussion amongst *the students. Couple of case studies that involve mostly logical thinking will also be presented. Each student will prepare a term paper. Evaluation will be on the basis of this term paper and participation in group discussion.*

Movies:

Twelve Angry Men
Roshoman by Kurosawa
Trial of Nuremberg
Mahabharata by Peter Brook

Books:

The Hound of the Baskervilles by Arthur Conan Doyle
Five Little Pigs by Agatha Christie
The Purloined Letter by Edger Allan Poe
The Case of the Substitute Face

Case Studies:

Case studies from TCS Academia Interface division.
Case studies from IDEO.ORG / Stanford University
Case studies from Social Sciences

PS 808.2P: PROGRAMMING FOR BIG DATA AND ADVANCED STATISTICAL METHODS LAB

Total No. of Lectures: 45	**Total Marks: 75**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To understand, and practice big data analytics and machine learning approaches, which include the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications.
- To conceptualize and summarize big data and machine learning, trivial data versus big data, big data computing technologies, machine learning techniques, and scaling up machine learning approaches.
- To demonstrate the use of various graph theory algorithms and study the various assumptions.
- To understand the various Hypothesis testing scenarios and apply them practically.
- To understand the importance of ANOVA and apply them based on the appropriate type of problems.
- To build tools for applying the linear regression model and its generalizations. To explore the workings of multiple regression and problems that arise in applying it, as well as going deeper into the theory of inference underlying regression and most other statistical methods.
- To build classes of models for binary and count data, emphasizing the need to fit appropriate models to the underlying processes generating the data being explained.

Course Outcome: Upon completion of the subject, students will be able to

- To perform machine learning techniques such as clustering and classification effectively.
- To apply the concepts of BI and DM techniques for clustering, association, and classification;
- To apply the graph theory algorithms to real data and analyze appropriately.
- To use appropriate statistical testing criteria based on the problem.
- To evaluate and apply ANOVA to the problem at hand.
- To identify and apply appropriate regression models considering all the assumptions.
- To perform binary output models using logistic regression.

Big data:

1. Algorithms of Graph theory
2. High dimensional space
3. Random graphs
4. Singular Value Decomposition (SVD)
5. Random walks
6. Algorithm for Massive Data Problems

Advanced Statistical methods:

1. Estimation
2. Testing of hypothesis
3. Linear models
4. Regression
5. Logistic regression

[OPEN ELECTIVE – OFFERED TO OTHER DEPTS]

OE 809.2: STATISTICAL DATA ANALYSIS USING R

Total No. of Lectures : 45	**Total Marks : 100**	**[L – T – P – S]**
No. of Lectures / Week : 4	**Credits : 3**	**[3 – 1 - 0 - 0]**

Course Objectives:

- To demonstrate the basic concepts of R, Installing the R , installing the packages and use of R for effective data analysis.
- To understand the concepts such as objects, matrix, vectors, data frames, tables along with practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code
- To demonstrate the summary commands with single value results and with multiple value results, cumulative statistics used in data analysis,
- To learn the concepts of data distribution and develop the skill in statistical tools in choosing appropriate distribution that fits the data.
- To understand hypothesis testing procedures based on type of outcome variable and number of sample.

Course Outcomes: After completing the course the student will have the

- Ability install R programming language on windows, Linux and Mac operating systems and able to program simple R programs.
- Ability to use inbuilt R functions to work on objects, matrix, vectors, data frames and tables.
- Ability to program summary and cumulative commands to apply it on tables and objects.
- Ability to use stem and leaf plot on the dataset , histograms to represent the data and ability to use sharpiro-wilk test, Kolmogorov-smirnov test etc.
- Ability to use students t-test, U-test , chi squared test monte carlo simulation and able apply these on different data sets.

a) **Introducing R:** The R Website, Downloading and Installing R from CRAN, Running the R Program, The Help Command in R, Command Packages, Standard Command Packages, Loading Packages; Storing the Results of Calculations; Reading and Getting Data into R 30; combineCommand; Reading Bigger Data Files - read.csv(), Viewing Named Objects, Converting Between Number and Text Data, The Structure of Data Items; History Commands

b) **Starting Out: Working With Objects:** Manipulating Objects, Vectors, Sorting and Rearranging a Vector, Returning Logical Values from a Vector, Manipulating Matrix and Data Frames, Selecting and Displaying Parts of a Matrix or Data Frame, Sorting and Rearranging a Matrix or Data Frame, Manipulating Lists, Viewing Objects within Objects, Looking Inside Complicated Data Objects; Opening Complicated Data Objects, Viewing and Setting Names, Rotating Data Tables, Constructing Data Objects, Making – Lists, Data Frames, Matrix Objects; Re-ordering Data Frames and Matrix Objects; Forms of Data Objects: Testing and Converting; Converting from One Object Form to Another.

c) **Data: Descriptive Statistics and Tabulation:** Summary Commands, Summarizing Samples, Summary Statistics for Vectors, Summary Commands With Single Value Results & With Multiple Results; Cumulative Statistics – Simple, Complex; Generic Summary Commands for Data Frames; Special Row and Column Summary Commands, Summary Tables, Contingency Tables, Cross Tabulation

d) **Data: Distribution:** Stem and Leaf Plot, Histograms, Density Function, Data Distribution -Normal Distribution, other Distributions, Random Number Generation and Control, Sampling; Shapiro-Wilk Test for Normality, Kolmogorov-Smirnov Test; Quantile-Quantile Plots; Adding a Straight Line to a QQ Plot

e) **Simple Hypothesis Testing:** Using the Student's t-test, t-Test with Unequal Variance, t-Test with Equal Variance, One-Sample t-Testing, Directional Hypotheses, Formula Syntax and Subsetting Samples in the t-Test, Wilcoxon U-Test, Two-Sample & One Sample U-Test; Directional Hypotheses; Paired t- and U-Tests; Correlation

and Covariance; Testing in Correlation Tests; Tests for Association; Multiple Categories: Chi-Squared Tests; Monte Carlo Simulation; Yates' Correction for 2 n 2 Tables

Text Book :

[1]. Mark Gardner, "Beginning R: The Statistical Programming Language", 1st Ed, 2017, Wiley

[2]. Garret G, "Hands on Programming with R", 2nd Edition, 2016, O'Reilly Publications

PROJECT WORK & SEMINAR

Total No. of Demo hrs : 30 **Total Marks : --- [L – T – P – S]**

No. of Hrs / Week: 2 **Credits : --- [0 – 2 - 0–4]**

Objective: This course will help you develop technical writing skills that you will use in a variety of professional contexts. Using a reader-centered approach, you will learn to plan, draft, write and revise documents common to the workplace. Please remember that this is an intensive writing course, so be prepared to do a lot of drafting, revising and critiquing in group writing workshops.

The Standard IEEE Format

ALL MANUSCRIPTS MUST BE IN ENGLISH.

1) PAGE AND COLUMN LAYOUT. Start the first page in a one-column format. Center your title about 3 lines down from the normal top of the print area. Follow the title with two blank lines. The author name(s) and affiliation(s) are next, centered beneath the title and followed by two blank lines.

2) Your Abstract and the remainder of the paper are to be in a two-column format (except for figures or tables that may span both columns, if necessary). If the last page is not filled, please divide the data into two equal columns.

3) Columns in the two-column format are to be 3-1/4 inches wide (approx. 8.5 cm), with a 3/8-inch (approx. 1.0 cm) space between columns, for a total print area width of 6-7/8 inches (approx. 17.5 cm). The length of the print area of any page must not exceed 8-7/8 inches (approx. 22.5 cm).

4) TYPE STYLE AND SIZE OF TEXT. Normal text is to be single-spaced in 10-point Times or Times Roman (or similar font), with 12-point interline spacing, in the two-column format. The first line of each paragraph is to be indented approximately 1/4 inch (approx. 0.7 cm), and the entire text is to be justified -- that is, flush left and flush right. Please do not place additional line spacing between paragraphs. Figure and table captions should be Helvetica 10-point boldface; callouts should be Helvetica 9-point non-boldface.

5) ABSTRACTS. The Abstract should be approximately 150 words or fewer, italicized, in 10-point Times (or Times Roman.) Please leave two spaces between the Abstract and the heading of your first section.

6) TITLE AND HEADINGS. The main title should be in Times (or Times Roman) 14-point boldface centered over both columns. In the main title, please initially capitalize nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, and prepositions (unless the title begins with such a word).

- Initially capitalize only the first word in first-, second-, and third-order headings. Leave two blank lines before author names(s)/affiliation(s).
- AUTHOR NAME(S)/AFFILIATION(S) are to be centered in Times (or Times Roman) 12-point non-boldface. Leave two blank lines before your Abstract.
- ABSTRACT HEADING. The abstract heading is to be 11-point boldface, initially capitalized and centered within the column.
- FIRST-ORDER HEADINGS. First-order headings (for example, 1: Introduction) are to be Times 12-point boldface, flush left, with one blank line before, and one blank line after.
- SECOND-ORDER HEADINGS. Second-order headings (for example, 1.1: Database elements) are to be Times 11-point boldface, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), then it is to be in Times 10-point boldface, preceded by one blank line, and followed by a period and text on same line.

7) ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS.
Illustrations, graphs, and photographs may fit across both columns, if necessary. Your artwork must be in place in the article.

8) FOOTNOTES. Use footnotes sparingly and place them at the bottom of the column in which they are referenced (not full width across two columns). Use Times 8-point type with 10-point interline spacing for footnotes. To help readers, avoid footnotes altogether and include necessary peripheral observations in your text (within parentheses, if you prefer, as in this sentence).

9) REFERENCES. List and number all bibliographical references at the end of your paper in 9-point Times, with 10-point interline spacing. When referenced within the text, enclose the citation number in square brackets, for example [1].

SEMESTER – III

PH 801.3: MACHINE LEARNING - II

Total No. of Lectures: 45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To understand the fundamental concepts of decision tree and regression tree classifiers, know about probabilistic and generative classifiers
- To understand the fundamental concepts of hyperplane classifiers like SVM, Perceptron; learn about loss functions and stochastic gradient descent
- To understand the concept of clustering, EM algorithm and Collaborative filtering
- To learn about Ensemble models and graphical models
- To understand the concept of large scale machine learning, gradient descent with large datasets and genetic algorithm

Course Outcomes:

- To implement classification models with decision tree and probabilistic classifiers; regression models with regression tree classifiers
- To implement predictive models using SVM and Perceptron with usage of loss functions and gradient descent
- To implement machine learning models with k-means clustering; models with collaborative filtering and implement EM algorithm
- To implement machine learning systems using Ensemble models and graphical models
- To implement models with genetic algorithm and working out gradient descent for large datasets

a) Decision Tree Classification: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Entropy, Gini index, Algorithms, Regression Trees.

Probabilistic Classifiers: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Generative and Conditional classifiers.

b) Hyper plane classifiers: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Loss functions, stochastic gradient algorithms, Perceptron algorithms.
Application of Pattern Recognition Problems. (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

c) Clustering: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Performance criteria, K-means clustering, EM algorithm
Collaborative filtering (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

d) Combining models (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Probabilistic graphical models (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

e) Large Scale Machine Learning: 6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Gradient descent with large data sets
Genetic Algorithm. (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

SUGGESTED BOOKS

1. *Machine Learning: Tom Mitchell*

Evaluation:Theory: 50% + Practical/Lab : 50%

PH 802.3: ENABLING TECHNOLOGIES FOR DATA SCIENCE - II

Total No. of Lectures: 45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To learn how to invoke Spark Shell and use it for various common operations.
- To Examine external data sets and Query existing data sets using Spark SQL.
- To develop scalable and fault-tolerant streaming applications by analyzing structured and unstructured data using Spark SQL and Spark structured streaming
- To Learn the foundations of the language for developers and data scientists interested in using Scala for data analysis and Tackle data analysis problems involving Big Data & Scala.
- To learn how to apply the functional programming style in the design of larger applications

Course Outcomes: Upon successful completion of this course, the student will be able to:

- Read data from persistent storage and load it into Apache Spark, - manipulate data with Spark
- Understand working of spark sessions, functions to manipulate and analyze data using Spark data frames
- Warehouse your data efficiently using Hive, Spark SQL and Spark Data Frames
- Manipulate data using Scala and write programs that effectively use parallel collections to achieve performance
- Recognize and apply design principles of functional programs

(Lab is Must) : Spark, Scala, Mahout.

a) **Spark – REPL & RDDS:** Dig deeper into Apache Spark, Apache Spark installation, Introduction to RDDs, Using the Spark shell, Actions and Transformations, Caching, Loading and saving data. **Special RDD Operations:** Types of RDDs, Aggregations, Partitioning and shuffling, Broadcast variables, Accumulators.

b) **Spark SQL:** Introducing Spark Session, Understanding Spark SQL concepts, Using Spark SQL in streaming applications; Spark SQL and DataFrames, DataFrame API and SQL API, Aggregations, Joins.

c) **Spark SQL for Processing Structured and Unstructured Data:** data sources in Spark applications, Spark with relational databases, Spark with MongoDB (NoSQL database), Spark with JSON data, Spark with Avro files, Spark with Parquet files, Defining and using custom data sources in Spark

d) **Scala:** Purposes of Scala, Platforms and editors, Installing and setting up Scala, Scala: the scalable language, Scala for Java programmers, Scala for the beginners

Object Oriented Scala: Variables in Scala, Methods, classes, and objects in Scala, Packages and package objects, Java interoperability, Pattern matching, Implicit in Scala, Generic in Scala, SBT and other build systems.

e) **Functional Programming Concepts:** Introduction to functional programming, Functional Scala for the data scientists, FP and Scala for learning Spark; Pure functions and higher-order functions; Using higher-order functions; Error handling in functional Scala; Functional programming and data mutability.

Evaluation:Theory: 50% + Practical/Lab : 50%

PH 803.3 P: MACHINE LEARNING AND DATA SCIENCE LAB - II

Total No. of Lectures: 45 **Total Marks: 75** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives: These subject enables students to

- To learn advanced cutting edge and state-of-the-art knowledge and implementation in big data.
- To read and understand research publications in the technical area of big data..
- To conduct independent project and to equip for scholarly research in big data.
- To explore the next generation of big data tools and applications, and other advanced topics if time permits.

Course Outcomes: Upon completion of the subject, students will be able to

- Demonstrate the knowledge of big data, data science, data analytics, distributed file systems, parallel Map Reduce paradigm, NoSQL, machine learning, etc.
- Program and implement examples of big data and NoSQL applications using open source Hadoop, HDFS, Spark, Scala, etc.
- Read current research papers and implement example research group project in big data.

Programs related to machine learning

1. Decision tree classification
2. Probabilistic classifiers
3. Hyper plane classifiers
4. Clustering
5. Combining models
6. Probabilistic graphical models
7. Large scale machine learning Algorithms
8. Genetic algorithm

Data Science:

1. Spark – REPL & RDDS: Spark SQL
2. Special RDD Operations
3. Spark SQL
4. Spark SQL for Processing Structured and Unstructured Data
5. Scala:
6. Functional Programming Concepts**:**

PS 804.3: DATA VISUALIZATION WITH TABLEAU & MODELLING IN OPERATIONS MANAGEMENT

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objective: The objective of this course is to make the students

- Develop skills to design and critique visualizations
- Understand the effective visualization using Tableau
- Use the power of visual analytics to tackle the complex challenges in banking industry
- Identify and demonstrate the processes involved in revealing patterns and insights to reduce costs and improve patient care and experience
- Understand the visualization technique and analysis of retail data

Course outcomes: Upon successful completion of this course, the student will able to

- Understand and apply the fundamental concepts and techniques in data visualization
- Design, develop, and evaluate effective visualizations and dashboards using various development tools
- Solve specific real-world problems related to the Visualization and interpretation of data analysis results
- Making use of patterns and insights in healthcare analytics
- Visualize the analyzed data pertaining to retail industry

3A. (30 hours). Data Visualization with Tableau:

a) Learn about design principles, human perception and effective story telling with data
b) dashboards, modern visualization tools and techniques (cover Tableau).

Hands-on practice on Tableau is must.

3B. MODELLING IN OPERATIONS MANAGEMENT: (30 hrs.)

c) Banking analytics (10 hrs.– Lab 10 hrs.)
d) Healthcare analytics (10 hrs. – Lab 10 hrs.)
e) Retail analytics (10 hrs. – Lab 10 hrs.)

SUGGESTED BOOKS: NONE

1. *David Baldwin, "Mastering Tableau - Master the intricacies of Tableau to create effective data visualizations", 1st Ed, 2017, PACKT*
2. *Acharya, Seema, Chellappan, Subhashini, "Pro Tableau A Step-by-Step Guide", 2017, Apress*
3. *Joshua N. Milligan, "Learning Tableau 10", 2nd Edition, 2016, PACKT Publishers*

Evaluation: Practical / Lab / Report: 100%

Electives (Choose 1)

PS 805.3 (E1): INTRODUCTION TO ECONOMETRICS & FINANCE

Total No. of Lectures: 45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To introduces various econometric models and techniques that can be applied to panel data.
- To describe generalized method of moments (GMM) estimation for linear and non-linear models with applications in economics and finance.
- To acquire knowledge of various advanced econometric models - ordinary least squares model estimation methods and to conduct basic panel regression analysis.
- To have a deeper understanding of the concept of cointegration and establish whether there exists a correlation between several time series in a long term.
- To gain insight on ARCH/GARCH/SV models of econometrics.

Course Outcome: On successful completion of the course the students will be able

- To apply the above theories to empirical data or be able to develop new econometric theory
- To apply the generalized method of moments (GMM) estimation and interpret the results.
- To Use various economic models and methods to interpret and analyze real data in economics and finance.
- To test cointegration among times series data using appropriate tests.
- To perform Autoregressive conditional heteroscedasticity model and interpret the coefficients.

a) Analysis of Panel Data. (19 hrs. – Theory 16 hrs. + Lab 3 hrs.)

b) Generalized Method of Moments (GMM). (18 hrs. – Theory 16 hrs. + Lab 3 hrs.)

c) Simultaneous Equations System: (7 hrs. – Theory 4 hrs. + Lab 3 hrs.)
Least Squares, Bias Problem, Estimation Method.

d) Cointegration: (8hrs. – Theory 2 hrs. + Lab 6 hrs.)
Concept, two variable model, Engle-Granger Method, Vector auto regressions (VAR), Vector error correlation model (VECM).

e) ARCH/GARCH/SV models, some important generalizations like EGARCH & GJR models, ARCH –M models. (8 hrs. – Theory 2 hrs. + Lab 6 hrs.)

SUGGESTED BOOKS:

1. *The Econometrics of Financial Markets: J. Campbell, A.Lo and C. Mackinlay*

2. *Econometric Analysis: William H. Greene*

Evaluation:Theory:70% + Practical/Lab:30%

PS 805.3 (E2): TIME SERIES ANALYSIS & FORECASTING

Total No. of Lectures: 45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives:

- To learn basic concepts in time series data and describe the data in qualitative and quantitative terms, in preparation for forecasting.
- To understand the concept of auto-regressive and model averaging models used in time series data and estimate the parameters.
- To know the various tests used for checking non stationarity that exists in the data.
- To gain knowledge on the various models used to models time series data.
- To identify missing data patterns in time series data and deal appropriately.

Course Outcome: On successful completion of the course the students will be able to

- Know the basic time series structure and identify patterns.
- Define the concept of stationarity and describe its importance in time series analysis
- Test for non-stationarity that exists in the time series data by applying suitable tests.
- Model times series data and and use them efficiently to forecast.
- Identify and deal with the missing data values in time series data.

a) Exploratory Analysis of Time Series. (10 hrs. – Theory 4 hrs. + Lab 6 hrs.)

Stationary and Non-Stationary Time Series. (5 hrs. – Theory 5 hrs.)

b) AR, MA, ARMA, ARIMA models, their properties, estimation of parameters. (20 hrs. – Theory 16 hrs. + Lab 4 hrs.)

c) Tests of Non-Stationarity – Unit Root tests. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

Forecasting, Smoothing, Minimum MSE Forecast, Forecast Error. (10 hrs. – Theory 8 hrs. + Lab 2 hrs.)

d) Modelling Seasonal Time Series. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

e) Missing Data Problem in Time Series. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

SUGGESTED BOOKS

1. *Introduction to Statistical Time Series: W. A. Fuller*

2. *Introduction to Time Series Analysis: P. J. Brockwell and R. A. Davis*

Evaluation:Theory:70% + Practical/Lab:30%

PS 805.3 (E3): BIOINFORMATICS

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To understand the fundamental concepts of Sequence alignment and Multiple Sequence alignment; Global and local alignment
- To understand the fundamental concepts of Gibbs sampling and genetic mapping
- To understand the concept of gene recognition and transcriptomics
- To learn about protein structure, motifs and HMM
- To understand the concept of lattice models

Course Outcomes:

- Gain knowledge in using tools for implementing sequence alignment (BLAST, FASTA), MSA (ClustalW, T-Coffee etc), variants of BLAST
- To implement Gibbs sampling and genetic mapping using tools available
- Gain knowledge in using tools for implementing gene recognition and Transcriptomics
- Gain knowledge in using tools for implementing HMM, finding motifs
- Gain knowledge in using tools for implementing lattice models

a) **Sequence Alignments.**(4 hrs. – Theory 2 hrs. + Lab 2 hrs.*)*

Advance Alignment Methods.(4 hrs. – Theory 2 hrs. + Lab 2*hrs.)*

b) **Gibbs Sampling.***(*8 hrs. – Theory 2 hrs. + Lab 6 hrs.*)*

Population Genomics*. (*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Genetic Mapping.*(*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

c) **Disease Mapping***. (*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Gene Recognition.(4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Transcriptome & Evolution*. (*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

d) **Protein Structure***.* (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Protein Motifs.(4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Hidden Markov Model.*(*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

e) **Lattice Model.***(*4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

Algorithms*. (*8 hrs. – Theory 6 hrs. + Lab 2 hrs.)

SUGGESTED BOOKS

1. *Introduction to Computational Molecular Biology: C. Setubal & J. Meidanis, PWS Publishing, Boston, 1997*

Evaluation:Theory:50% + Practical/Lab: 50%

PS 805.3 (E4): BIG DATA TECHNOLOGIES AND ARCHITECTURE

Total No. of Lectures: 45	**Total Marks: 100**	**[L – T – P – S]**
No. of Lectures / Week: 3	**Credits : 3**	**[3 – 2 - 2 - 2]**

Course Objectives: The course aims to cover the following objectives

- Understanding the Big Data Fundamentals, including Hadoop life cycle management ,Hadoop cluster and the Hadoop distributed system architecture.
- Introducing to No-SQL databases and different properties and characteristics associated with it, and getting started with Hive and Hbase.
- Understanding Map Reduce programming and learning to write programs using mapper and reducer.
- Learning to apply Map Reduce programs for K-means clustering and using Sort comparator and Group comparator.
- Exploring Hadoop to be used as a Database, and getting introduced to Search engines, inverted index etc.

Course Outcomes: Upon Completion of the course, the students will be able to

- Identify the use of Hadoop for processing the data, configuring Hadoop cluster and exploring Hadoop distributed file system.
- Describe No SQL databases and understanding different concepts related to No SQL and its applications using Hive and Hbase.
- Writing map reduce programs using mapper and reducer.
- Writing map-reduce programs to perform K-Means clustering customizing partitioner and sort comparator.
- Learned the role of Inverted Index and usage of hadoop as a database.

a. Enterprise Data Architecture Principles, Hadoop Life cycle Management, Hadoop Design Consideration: Understanding data structure principles; Installing Hadoop cluster; Exploring HDFS architecture; Introducing YARN;

b. Data Movement Techniques: Data Modeling in Hadoop:Apache Hive, Hive architecture, Hive data model management, JSON documents using Hive, Apache HBase; NoSQL:NoSQL Databases. Application; NoSQL approach; NoSQL Storage types – Storage types; Comparing the models.

c. MapReduce: MapReduce - Visualized And Explained; The Reducer; MapReduce - Combiners, Shuffle, Sort & Streaming API: HDFS & YARN: HDFS - Yarn - Submitting a job to Yarn, Plug in scheduling policies, Configure the scheduler.

d. MapReduce Customization:ToolRunner and GenericOptionsParser; Configuring properties of the Job object; Customizing the Partitioner, Sort Comparator, and Group Comparator.K-Means Clustering:MapReduce job for K-Means Clustering;

e. Inverted Index, Custom Data Types: Search engines ; Input-Output Formats and Customized Partitioning:Hadoop as Database

SUGGESTED BOOKS

[1]. Naresh Kumar, Prashant Shindgikar, "Modern Big Data Processing with Hadoop", 1st Ed, 2018, PACKT Publishers International

[2]. Gaurav Vaish, "Getting Started with NoSQL", 1st Edition, 2017, PACKT Publishers

Evaluation:Theory: 50% + Practical/Lab: 50%

PS 806.3 (E1): INTELLECTUAL PROPERTY RIGHTS IPR

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- Introducing the concept of intellectual Property and making the learners aware about different types of Intellectual property rights and how they can be protected.
- Understanding the importance of Patents and Copyrights
- Familiarizing the importance of Industrial design and also understanding what are the remedies for piracy of designs
- Identifying the Importance of Digital signatures and digital contracts
- Exposing the learners to the area of cybercrimes and also to focus on the remedies under cyber law for such unauthorized activities in the cyber domain.

Course Outcomes: Upon successful completion of this course, a student will be able to:

- Understand and distinguish between different Intellectual properties and also identify the procedures to protect Intellectual property
- Protect his own invention under patent and copyright specifically related to software. And also understand how one can derive revenue from protection of patents/copyrights
- Identify the importance of industrial design and its protection
- Recognizes the importance of different types of digital contracts and also finds mechanisms to protect digital documents
- Identify different types of cybercrimes and also will understand what are the remedies available under cyber law in the case of such unlawful activities

.

a. Introduction to Intellectual Property Rights: Concepts of IPR, The economics behind development of IPR: Company perspective; International protection of IPR 2;
Trademark Law and Geographical Indication ; Concept of trademarks; Importance of brands and the generation of "goodwill"; Trademark: A marketing tool; Trademark registration procedure ; Infringement of trademarks and Remedies available; Assignment and Licensing of Trademarks; Trademarks and domain names; Concept of Geographical Indication

b. Patent Law : Introduction to Patents, Procedure for obtaining a Patent, Licensing and Assignment of Patents, Software Licensing, General public Licensing, Compulsory Licensing, Infringement of Patents.
Copyrights: Concept of Copyright Right; Assignment of Copyrights; Registration procedure of Copyrights; Infringement (piracy) of Copyrights and Remedies

c. Designs: Concept of Industrial Designs; Registration of Designs: Piracy of registered designs and remedies 6. IP Management; Concept of IP Management; Intellectual Property and Marketing; IP asset valuation.

d. Digital Signature: Technical Issues & Legal Issues; Digital signature; Digital signature Certificate; Certifying Authorities and liabilities in the event of Digital Signature Compromise; Contract in the InfoTech World: Status of Electronic Contracts; Click –Wrap And Shrink – Wrap Contract; Contract Formation In The Internet Vis- A –Vis Contract Law.

e. Cyber Crime and the IT Act, 2000 : Cyber Crimes: Technical Issues; Cyber Crimes : Legal Issues; Cyber Crimes : Legal Issues [Penalty under the IT Act]; Cyber Crimes : Legal Issues [Offences under the IT Act]; Cyber Crimes : Legal Issues [Offences under IPC]; Cyber Crimes & Investigation; Cyber Crimes & Adjudication.

SUGGESTED BOOKS

[1] Neeraj Pandey, Khushdeep Dharni, "Intellectual Property Rights", 1st Ed, 2014, PHI Publishers

[2] N.S. Sreenivasulu, "Intellectual Property Rights", 1st Edition, 2011, Regal Publications

Evaluation:Theory: 50% + Practical/Lab: 50%

PS 806.3 (E2): CYBER SECURITY

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To understand basics of Cryptography and Network security
- To learn about computer and network security threats, classify the threats and develop a security model to prevent, detect and recover from the attacks.
- To learn about Encrypt and decrypt messages using block ciphers, sign and verify messages using well known signature generation and verification algorithms
- To learn about how to maintain the Confidentiality, Integrity and Availability of a data.
- To understand various protocols for network security to protect against the threats in the networks

Course Outcomes: Upon the successful completion of the course.,

- Understand the basics of security attacks and threat model
- Appreciate the vulnerabilities and threats posed by criminals, terrorist and nation states to national infrastructure
- Have a strong understanding of different cryptographic protocols and techniques and be able to use them.
- Apply methods for authentication, access control, intrusion detection and prevention.
- Identify and mitigate software security vulnerabilities in existing systems

a. Introduction: Security attacks to information systems. Threat model. Security services. Mechanisms for providing confidentiality, authentication, integrity, non-repudiation, and access control. Cryptography in data and communication security.

b. Conventional Cryptography: Cryptosystems and cryptanalysis. Block ciphers and stream ciphers. Confidentiality using encryption. Key distribution. Random number generation.

c. Public-Key Cryptography: One-way functions. Trapdoor one-way functions. Public-key cryptosystems. RSA, Diffie-Hellman, ElGamal, and elliptic curve cryptosystems.

d. Message Authentication and Hash Functions: Cryptographic checksums. Message authentication codes. Hash functions. Security issues. Attacks and countermeasures.
Digital Signatures and Strong Authentication: Digital signatures. One-way and mutual authentication protocols.

e. Network Security Practice: Authentication applications. Kerberos. Directory services. PGP and S/MIME. IP Security and Web security.
Wireless Network Security: Wireless networking protocols and security issues. 801.11 protocols, attacks, and countermeasures.

SUGGESTED BOOKS

1. *Pfleeger, C.P., Security in Computing 5th Edition, Prentice Hall, 2012*
2. *Schneier, Bruce. Applied Cryptography, Second Edition, John Wiley & Sons, 2016*

Evaluation:Theory: 50% + Practical/Lab: 50%

PS 806.3 (E3): TEXT MINING

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- To understand the background and motivation for text mining. Know the differences between structured, unstructured and semi-structured data. Understand the evolving information needs and knowledge management.
- To understand about IR, IE and data mining. Understand the fundamentals of NLP, and various approaches to text mining.
- To understand about text types, document formats and conversion, low-level processes like sentence splitting and parts-of-speech tagging
- To understand some of the main tasks of text extraction – keyword extraction, named entity recognition and feature extraction.
- To understand the resources for text mining, issues in processing of large scale text.

Course Outcomes:

- The ability to analyse structured, unstructured and semi-structured data. Understand about user experience of information seeking behaviour.
- Ability to analyse linguistic foundations, and various approaches to text mining.
- To analyse various text types, document formats and conversion, character encodings. Perform parts-of-speech tagging for simple English sentences.
- To distinguish few tasks of text extraction – keyword extraction, named entity recognition. Perform simple extraction from small text.
- To understand computational grammars, design and construction.

a. Introduction: background, motivation, dealing with information overload and information overlook, unstructured vs. (semi-)structured data, evolving information needs and knowledge management issues, enhancing user experience of information provision and seeking, the business case for text mining.

b. The text mining pipeline: information retrieval, information extraction and data mining.; Fundamentals of natural language processing: linguistic foundations, levels of linguistic analysis. Approaches to text mining: rule-based vs. machine learning based vs. hybrid; generic vs. domain specific; domain adaptation.;

c. Dealing with real text: text types, document formats and conversion, character encodings, markup, low-level processes (sentence splitting, tokenisation, part of speech tagging, chunking)

d. Information extraction: term extraction, named entity recognition, relation extraction, fact and event extraction; partial analysis vs. full analysis.; Data mining and visualisation of results from text mining; Evaluation of text mining systems: evaluation measures, role of evaluation challenges, usability evaluation

e. Resources for text mining: annotated corpora, computational lexica, ontologies, computational grammars; design, construction and use issues.; Issues in large scale processing of text: distributed text mining, scalable text mining systems.; A sampler of text mining applications and services; case studies.

SUGGESTED BOOKS

1. *Mining Text Data. Charu C. Aggarwal and ChengXiang Zhai, Springer, 2012.*
2. *Speech & Language Processing. Dan Jurafsky and James H Martin, Pearson Education India,*
3. *Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze, Cambridge University Press, 2017.*

Evaluation:Theory: 50% + Practical/Lab: 50%

PS 806.3 (E4): ADVANCED ANALYTICS

Total No. of Lectures: 45 **Total Marks: 100** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- Learn the basic concepts of internet of things, network and communication, security privacy trust
- Learn the concepts of IoT reference architecture, Identity Management Models, trust management in IoT
- Learn different IoT protocols in different layers of OSI Layers. Also learn IoT analytics for cloud
- Learn how IoT data is explored and how data is displayed using visualization techniques
- Learn how machine learning techniques are used to extract the useful data and how the data is organized for data analytics

Course Outcomes – upon successful completion of this course, the participant will be able to:

- Understand why IoT is used and how it is implemented and how networks and communication is used to implement IoT
- Understand how identity management models are used in IoT, also understand why trust management is important for IoT environment
- Understand the use of protocols which are used in different layers and how it is combined with other protocols down the layers to carry out the communication
- Understand how data is stored in cloud and how it is represented using different application to carry out or execute different data analytics tools
- Understand the concepts of data science for IoT analytics, how to organize data for analytics, and how to get benefits from IoT analytical tools.

a. IoT Web Technology, Internet of Things Privacy, Security and Governance -Projects, Security, Privacy and Trust in IoT-Data-Platforms

b. Introduction, Defining a Common Architectural Ground, M2M Service Layer Standardisation, OGC Sensor Web for IoT, IEEE, IETF and ITU-T standardization activities; Identity Management Models in IoT :Trust Management in IoT:Identity and trust, Third party approach, Public key infrastructure, Attribute certificates, Web of trust models, Web services security, SAML approach, Fuzzy approach

c. Defining IOT Analytics and challenges; IoT Devices and Networking Protocols**:** IoT Analytics for Cloud:The AWS Cloud Formation overview; The AWS Virtual Private Cloud (VPC)

d. Collecting the Data, Strategies and Techniques:Exploring IoT Data: Visualizing and Dashboarding :

e. Applying Geospatial Analytics to IoT Data: Data Science for IoT Analytics:Machine learning (ML); Anomaly detection using R; Forecasting using ARIMA; Deep learning; Strategies to Organize Data for Analytics:Linked Analytical Datasets; Managing data lakes; The data retention strategy.; Economics of IoT Analytics:

SUGGESTED BOOKS

[1]. Andrew Minteer, "Analytics for the Internet of Things (IoT)", 2nd Ed, 2017, PACKT

[2]. Arsheep Bhaga, Vijay Madasetti, "Internet of Things : A Hands on Approach", 1st Edition, 2016, University Press India Ltd

[3]. Donald Norris, "The Internet of Things: Do-It-Yourself at Home Projects for Arduino, Raspberry Pi and BeagleBone Black", 1st Edition, 2015, McGraw Hill Publishers

Evaluation:Theory: 50% + Practical/Lab: 50%

PS 807.3 P: DATA VISUALIZATION WITH TABLEAU & OPERATION MANAGEMENT LAB

Total No. of Lectures: 45 **Total Marks: 75** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objective: The objective of this course is to make the students

- Develop skills to design and critique visualizations
- Understand the effective visualization using Tableau
- Use the power of visual analytics to tackle the complex challenges in banking industry
- Identify and demonstrate the processes involved in revealing patterns and insights to reduce costs and improve patient care and experience
- Understand the visualization technique and analysis of retail data

Course outcomes: Upon successful completion of this course, the student will able to

- Understand and apply the fundamental concepts and techniques in data visualization
- Design, develop, and evaluate effective visualizations and dashboards using various development tools
- Solve specific real-world problems related to the Visualization and interpretation of data analysis results
- Making use of patterns and insights in healthcare analytics
- Visualize the analyzed data pertaining to retail industry

Data visualization with Tableau

1. Fundamentals of visualizations
2. Overriding defaults, building a dashboard, adding interactivity to a dashboard
3. Observing metadata differences, connecting to Tableau Server, creating a local data
4. Measure Names and Measure Values and shortcuts
5. Dimensions and measures
6. Row level, aggregate level, table level
7. Surveying data, extrapolating data and cleaning the data
8. Tableau Prep
9. Other topics based on theory coverage

Modelling in operations management

1 Banking analytics
2 Healthcare analytics
3 Retail analytics

PS 808.3: LAB ON ELECTIVES 1 & 2

Total No. of Lectures: 45 **Total Marks: 75** **[L – T – P – S]**

No. of Lectures / Week: 3 **Credits : 3** **[3 – 2 - 2 - 2]**

Course Objectives:

- *To understand the practical concepts related to econometrics and finance*
- *To understand the importance of analysing time series data and the model used to represent them*
- To understand the fundamental concepts of bioinformatics model that can be applied in analytics
- To understand how to architect big data solutions by assembling various big data technologies
- To Learn how machine learning techniques are used to extract the useful data and how the data is organized for data analytics
- To understand about text types, document formats and conversion, low-level processes like sentence splitting and parts-of-speech tagging

Course outcomes

- *Apply the econometric models to the real time data and interpret it.*
- Model times series data and and use them efficiently to forecast.
- *Use various models/ algorithms to gain information from the data and use it for better decision making*
- Architect multiple real life use cases
- Apply the concepts of data science for IoT analytics, how to organize data for analytics, and how to get benefits from IoT analytical tools.
- Analyze various text types, document formats and conversion, character encodings. Perform parts-of-speech tagging for simple English sentences

Elective 1

1. Introduction to Econometrics and Finance
    a. Analysis of Panel Data.
    b. Generalized Method of Moments (GMM).
    c. Simultaneous Equations System:
    d. Least Squares, Bias Problem, Estimation Method., Cointegration:
    e. Concept, two variable model, Engle-Granger Method, Vector auto regressions (VAR), Vector error, correlation model (VECM).
    f. ARCH/GARCH/SV models

2. Time Series Analysis and Forecasting
    a. Exploratory Analysis of Time Series.
    b. Stationary and Non-Stationary Time Series.
    c. AR, MA, ARMA, ARIMA models, their properties, estimation of parameters.
    d. Tests of Non-Stationarity – Unit Root tests.
    e. Forecasting, Smoothing, Minimum MSE Forecast, Forecast Error.
    f. Modelling Seasonal Time Series.
    g. Missing Data Problem in Time Series.

3. Bioinformatics
    a. Sequence Alignments
    b. Advance Alignment Methods.
    c. Gibbs Sampling.
    d. Population Genomics.

e. Genetic Mapping.
f. Disease Mapping.
g. Gene Recognition.
h. Transcriptome & Evolution.
i. Protein Structure. Protein Motifs.
j. Hidden Markov Model.
k. Lattice Model.
l. Algorithms.

4. Big Data Technologies and Architecture
    a. Practical aspects with respect to the theory.

Elective 2

1. Intellectual Property Rights
    a. Practical aspects with respect to the theory.

2. Cyber Security
    a. Practical aspects with respect to the theory.

3. Text Mining
    a. Practical aspects with respect to the theory.

4. Advanced Analytics
    a. Practical aspects with respect to the theory.

[OPEN ELECTIVE – OFFERED TO OTHER DEPTS]

OE 809.3: BIG DATA & DESIGN THINKING

Total No. of Lectures : 45 **Total Marks : 100** **[L – T – P – S]**

No. of Lectures / Week: 4 **Credits : 3** **[3 – 1 - 0 - 0]**

Course Objectives:

- Understand the terminology and conceptual models used in design disciplines
- Recognize the ethical and social dilemmas and obligations of the practice of design
- Diagnose common adoption barriers in individuals, groups and organizations.
- Develop a design theory from independent and qualitative research and observations and lead innovation in creative and collaborative settings
- Undertake complex and unstructured problem-solving challenges in unfamiliar domains

Course Objectives: Upon the successful completion of the course, the student will

- Develop viable solutions to user challenges using the design thinking and hypothesis-driven innovation processes.
- Gain user empathy through observation and interviewing, and develop user insights to identify unmet needs.
- Use multiple brainstorming techniques to find innovative solutions.
- Prototype a solution to a user challenge.
- Develop and test a business model or business case to support the viability of the solution.

a. Introduction to Design Thinking - Design Thinking Tools; Personas : Powerful Tool for Designers : Defining Personas, The Importance of Personas, Creating Personas, Illustrative Application of Personas; Customer Experience Mapping : Inputs to the Experience Map, The Experience Mapping Process, The Experience Map as a Springboard to Innovative Solutions.; Design Thinking to bridge Research and Concept Design: The Visualize, Empathize, and Ideate Method

b. Boosting Creativity in Idea Generation using Design Heuristics:Design Heuristics, The Evidence Base, Key Roles of Stories and Prototypes:Integrating Design Into Fuzzy Front End of the Innovation Process: Role of Design in Early stage Ventures: Design Thinking for Non-Designers:Non-Designers Need to Learn, Challenges Teams Face with Design Thinking, Three Team Strategies for Success,

c. Developing Design Thinking; Leading for a corporate Culture in Design Thinking:Knowledge Management for Innovation; Embedding Design Thinking in a Firm:

d. Design thinking for services:Capturing Context through Service Design :Optimal Design for New Products :Business Model Design : Lean Method for Innovation :

e. Consumer Responses and Values:Drivers of Diversity in Consumers:Future-Friendly Design: User interface and Industrial Design: Intellectual Property Protection for Designs:

SUGGESTED BOOKS

[1]. Michael G. Luchs, Scott Swan, Abbie Griffin, "Design Thinking: New Product Development Essentials", 1st edition, 2016, Wiley International

Evaluation:Theory: 70% + Practical/Lab: 30%

TECHNICAL WRITING & PRESENTATION

Total No. of Demo hrs : 30	**Total Marks : ---**	**[L – T – P – S]**
No. of Hrs / Week: 2	**Credits : ---**	**[0 – 2 - 0–4]**

Course objective:

- To learn the technical writing skills which will form the basic requirements of writing a research paper.
- To understand and analyze a given situation of dataset that will help in the analysis
- To be able to identify the appropriate algorithms to be applied by understanding the importance of literature review.

Course Outcomes: This course will help the students to

- Develop technical writing skills that you will use in a variety of professional contexts.
- Use a reader-centered approach, learn to plan, draft, write and revise documents common to the workplace.
- Know the importance of literature review and build on the research area

The Standard IEEE Format

ALL MANUSCRIPTS MUST BE IN ENGLISH.

10) PAGE AND COLUMN LAYOUT. Start the first page in a one-column format. Center your title about 3 lines down from the normal top of the print area. Follow the title with two blank lines. The author name(s) and affiliation(s) are next, centered beneath the title and followed by two blank lines.

11) Your Abstract and the remainder of the paper are to be in a two-column format (except for figures or tables that may span both columns, if necessary). If the last page is not filled, please divide the data into two equal columns.

12) Columns in the two-column format are to be 3-1/4 inches wide (approx. 8.5 cm), with a 3/8-inch (approx. 1.0 cm) space between columns, for a total print area width of 6-7/8 inches (approx. 17.5 cm). The length of the print area of any page must not exceed 8-7/8 inches (approx. 22.5 cm).

13) TYPE STYLE AND SIZE OF TEXT. Normal text is to be single-spaced in 10-point Times or Times Roman (or similar font), with 12-point interline spacing, in the two-column format. The first line of each paragraph is to be indented approximately 1/4 inch (approx. 0.7 cm), and the entire text is to be justified -- that is, flush left and flush right. Please do not place additional line spacing between paragraphs. Figure and table captions should be Helvetica 10-point boldface; callouts should be Helvetica 9-point non-boldface.

14) ABSTRACTS. The Abstract should be approximately 150 words or fewer, italicized, in 10-point Times (or Times Roman.) Please leave two spaces between the Abstract and the heading of your first section.

15) TITLE AND HEADINGS. The main title should be in Times (or Times Roman) 14-point boldface centered over both columns. In the main title, please initially capitalize nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, and prepositions (unless the title begins with such a word).

- Initially capitalize only the first word in first-, second-, and third-order headings. Leave two blank lines before author names(s)/affiliation(s).
- AUTHOR NAME(S)/AFFILIATION(S) are to be centered in Times (or Times Roman) 12-point non-boldface. Leave two blank lines before your Abstract.
- ABSTRACT HEADING. The abstract heading is to be 11-point boldface, initially capitalized and centered within the column.
- FIRST-ORDER HEADINGS. First-order headings (for example, 1: Introduction) are to be Times 12-point boldface, flush left, with one blank line before, and one blank line after.
- SECOND-ORDER HEADINGS. Second-order headings (for example, 1.1: Database elements) are to be Times 11-point boldface, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), then it is to be in Times 10-point boldface, preceded by one blank line, and followed by a period and text on same line.

16) ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS.
Illustrations, graphs, and photographs may fit across both columns, if necessary. Your artwork must be in place in the article.

17) FOOTNOTES. Use footnotes sparingly and place them at the bottom of the column in which they are referenced (not full width across two columns). Use Times 8-point type with 10-point interline spacing for footnotes. To help readers, avoid footnotes altogether and include necessary peripheral observations in your text (within parentheses, if you prefer, as in this sentence).

18) REFERENCES. List and number all bibliographical references at the end of your paper in 9-point Times, with 10-point interline spacing. When referenced within the text, enclose the citation number in square brackets, for example [1].

SEMESTER – IV:

Internship based project

A real life project has to be undertaken at an industry for 20 weeks. Each student will have two supervisors: one from academic institution and one from the industry. The project shall involve handling data extensively and use of methodologies learnt during the course work to derive meaningful inferences. A final project report has to be submitted and an "open" presentation has to be made.

PH 801.4 : INDUSTRY INTERNSHIP / PROJECT WORK / DISSERTATION

Total Marks	**: 500**	**External**	**: 300**	**[L – T – P – S]**
Internal Assessment	**: 200**	**Credits**	**: 16**	**[0 – 2 - 0 – 10]**

Course Objectives:

- To provide students with an opportunity to gain work experience that will enhance and complement their academic learning.
- To help students learn, understand, and practice big data analytics and machine learning approaches, which include the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications.
- Conceptualization and summarization of big data and machine learning, trivial data versus big data, big data computing technologies, machine learning techniques, and scaling up machine learning approaches.
- To understand the use and application of all the classroom teaching with respect to all the algorithms.

Course outcomes

- The course requirements are designed to provide a structure that will enable students to make connections between what they learn in the classroom and on the job, to further develop analytical and interpersonal skills, and to practice business writing skills.
- Ability to select and implement machine learning techniques and computing environment that are suitable for the applications under consideration.
- Ability to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques.
- Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies like Hadoop and map reduce.

Course Requirements

1. The Internship / Project work / Dissertation for credit requires students to spend the majority of their time in technical, analytical, or administrative work that will contribute to their learning as outlined in the course objectives.
2. Work of a clerical nature must be limited to a maximum of 15 percent of the time spent on the job.
3. Prior to beginning an internship for credit, students must receive an internship orientation at the Training and Placement Cell of AIMIT..
4. A meeting with the faculty advisor / Guide to cover the ground rules and requirements.
5. Submission of the Final Report within seven days of the completion of the internship

The body of the Final Report will cover the following:

- **Job description:** Describe in detail your internship position duties and responsibilities. Discuss what duties were performed on a daily basis, periodic (e.g.,weekly) basis, and one-time special projects.
- **Knowledge gained:** Describe the knowledge gained or enhanced as a result of your internship experience. Relate this knowledge to what you learned in specific courses at AIMIT. Did your courses prepare you to handle the responsibilities of your position?
- **Skills learned:** Describe the skills that you learned or sharpened on the job. Discuss any skills that you learned as part of a course at AIMIT that were useful on the job. Consider a skill as the ability to do something like problem solving, analyze a problem, work in a groups, etc.
- **Attitudes/values:** Describe the attitudes or values that you found to be important for success in your job. Think of attitudes as a way of thinking or behavior, e.g., stubborn, patient, confrontational, etc. Consider values as the things you regard as important in life, e.g., dependability, integrity, hard work, etc.
- **Learning outcomes:** Identify the outcomes or results from the knowledge, skills and attitudes or values that you have described above. For example, what can you do for an organization today that you could not have done, or could not have done as well, before your internship

Guidelines to prepare the Project Document / Dissertation

The following format guidelines are intended to help you prepare your master's thesis or dissertation and should be used in conjunction with the specific style adopted by your IT program. It is your responsibility to conform to the following format requirements and ensure that your manuscript's presentation is of the highest quality. Because requirements may change over time, students should not use existing library or departmental copies of manuscripts as examples of proper format.

Format

All pages of your manuscript must be in 'Letter Size', 8 ½ X 11 inch, format.

Reproduction Quality

The copies of your manuscript must be clean, unshaded, and free of spots and smudges. Faint, streaked, or uneven copies are unacceptable.

Margins

The left margin of each page must be 1 1/2 inches, and the top, right, and bottom margins 1 inch.

Placement of Page Numbers

There are only two ways to paginate your manuscript; upper right hand corner and bottom center. If your page numbers are at the bottom, leave two blank line spaces between the last line of text and the line on which the page number is placed. Whether they are at the top or the bottom, page numbers should appear just outside the 1-inch margins (.5 to .8 inches from the top or bottom edge of the page. Whichever method of pagination is selected, it must be followed consistently. Use lower-case Roman numerals for the front matter (which is all pages before the body of research), create section break and continue (beginning again with page "1") with Arabic numerals for the remainder of the manuscript, including the text, illustrations, appendices and references.

Type and Font

Your manuscript should be double-spaced (one and a half spaces can be used if approved by the chair of your committee) and single-sided. References may be single-spaced with a double space between each reference. Indented quotations may also be single-spaced. Use a professional quality font (e.g., Arial, Times Roman, Courier or Helvetica). Font size for your text should be 12 point; headings may be up to 14 point. Fonts for tables, figures, and appendices may range from 8 to 12 point.

Style

The style of your thesis or dissertation may follow any one of many standard style guides, as preferred by your graduate unit, or the style considered standard in your particular discipline. You should consult your advisor for preferences or additional requirements your department may have. In any case, your manuscript must be internally consistent.

Format Guidelines

Thesis or dissertation manuscripts are generally divided into three sections – the Front Matter, the Body of Research, and the Reference Matter.

For a visual reference to format the Front Matter, utilize the Example of Completed Front Matter link at the Approval Template page. A Thesis Example and a Dissertation Example will help you format the Front Matter for pagination and the placement of the proper order of the front matter pages which can be generated at the "MS Word Templates" page.

Approval (Signature) Page

The Approval Page is the first page of the Front Matter and the manuscript. Count this as page "i". Type the names of the committee members on the lines provided at the center of the page.

Title Page

The Title page contains the title of your manuscript, your name, your previous degrees (including your majors, institutions and years centered on the page). Count the title page as page "ii". The degree you will be receiving at SAC, and the month and year of your graduation should be listed toward the bottom of the page (check with Graduate Studies for the graduation date). When listing the degree to be awarded, please refer to the Master's Degree List included with these guidelines. Do not use abbreviations.

Dedication Page

This is an optional page. If you use one, number it in appropriate sequence with a lower-case Roman numeral. The title must be in capital letters, centered just below the top margin of the page. The dedication itself may be single or double-spaced.

Acknowledgement Page

This is also an optional page. If you use one, number it in appropriate sequence with a lower-case Roman numeral. The title must be in capital letters, centered just below the top margin of the page. The acknowledgment itself may be single or double-spaced.

Abstract Page

On the Abstract page first list the title of your manuscript, your name, all degrees you have already earned (in chronological order), and the degree to be awarded. These lines should be centered on the page, beginning just below the top margin. Then double-space, and on the following line, center the word ABSTRACT. Double-space again before beginning the text of your abstract. Use paragraph indentation as appropriate. The text itself should be either one and a half spaces or double-spaced.

Number the Abstract Page in appropriate sequence with a lower-case Roman numeral.

Table of Contents

The Table of Contents page is counted and numbered with a lower-case Roman numeral. If you have used a List of Figures (Optional) and/or a List of Tables (Optional) , they must be included in your Table of Contents. Tab leaders should be used between the heading levels and the page numbers. (DO NOT simply type dots across the page--the spacing will not work out. Instead, set a dot leader tab. If you have multiple appendices, they must each be listed (see section on Appendices).

List of Figures

Placed on separate page after the Table of Contents. This page is counted and numbered with a lower-case Roman numeral. It may be single or double-spaced. Tab leaders should be used between the title of the figures and page the numbers. This page should be listed at the beginning of the Table of Contents.

List of Tables

Placed on separate after the Table of Contents. This page is counted and numbered with a lower-case Roman numeral. It may be single or double-spaced. Tab leaders should be used between the title of the tables and the page numbers. This page should be listed after the List of Figures at the beginning of the Table of Contents.

Preface Page

This page is counted and numbered with a lower-case Roman numeral.

The Text

Beginning with the first page of the text (begin again with page "1"), pages are numbered with sequential Arabic numerals through the end of your manuscript. Each chapter/major division of the text must start on a new page. Each of these lead pages should be counted and numbered.

Figures

Figures may include diagrams, charts, drawings, schematics, photographs, etc. Each of the two required copies of your manuscript must contain its own original figures, with the exception of photographs, in which case both may be high-quality copies either in color or black and white. Figures should be inserted as near as reasonably possible to the text to which they relate. They should be numbered consecutively with Arabic numerals as part of the continuing text.

Figures and their captions must appear on the same page, within the required margins. If captions are more than one line long, they may be single-spaced. They should be placed on paper of the same size and weight as the rest of the manuscript. Colored materials are acceptable, but since color does not reproduce on microfilm, an alternate key should be provided if it will be needed for interpretation of the figure in black and white.

Oversized figures may be presented in one of two ways: A horizontal figure that is too wide to fit on a regular manuscript page may be mounted on another piece of paper at the left hand margin, and folded like a fan. The folds must be within the right hand margin.

A figure that is both too long and too wide may be folded and inserted into a 6 ½ x 9 ‘’ envelope that is mounted on another sheet of paper.

Tables

Tables that are a half-page or shorter in size may be placed at the top or the bottom of the page, or in the center, with text above and below. Tables may also be placed alone on the page immediately following the page that refers to them. The placement of tables should be consistent throughout the manuscript. Tables continuing for more than one page should be labeled [e.g., Table 1 (cont.)], and oversized tables should be treated in the same way as oversized figures.

Reference matter

It may include notes to the text, appendices, a glossary, a list of references, and an index. The appendices and the list of references are the two most commonly used categories, and are discussed in detail below.

Appendices

The appendix (or a series of appendices) usually follows the main text, and contains material that is useful for a detailed review of the study, but is not essential to an understanding of the text. For example, an appendix is the appropriate section in which to place material such as raw data, the results of individual laboratory analyses, or sample forms. This may be useful supporting data even though it is not specifically utilized in the text. Note, however, that some reference should be made in the text to the fact that these materials appear in the appendix. The pages of the appendices are numbered in sequence with those of the text. Although it is preferable, you do not have to meet margin requirements for the material in your appendix, except for the left-hand margin.

After the last page of text, include a list of Appendices on a separate page. Individual appendices should start on a separate page and should be clearly marked.

The List of Appendices should also be included in the Table of Contents. To list appendices in your Table of Contents, use "Appendices" as your major heading. On the next line, indent five spaces, and begin listing each appendix with its title (for example, "Appendix A. Survey Data") and page number.

List of References

No manuscript is complete without a full listing of the necessary bibliographic information about the sources upon which the study is based. In most cases, only those sources actually cited in the study are listed. This section will be called "List of References", "References Cited", or simply "References", and all pages are

counted and numbered. The list is most often arranged alphabetically, although it may follow some other logical plan. It will follow the appendices.

Problems encountered :

Front Matter

Formatting on front material is very detailed. Make sure that formatting matches examples. Make sure that pages are numbered correctly.

Margins

Margins for the entire manuscript must be:

left – 1.5 inches; top, right, bottom – 1 inch

Exceptions: Top, right and bottom margins may vary in the Appendix pages, but the left margin must be 1.5 inches throughout the ENTIRE manuscript. Page numbers in the Appendices must also appear consistently as in the rest of the manuscript.

Landscape-oriented figures and tables

Margin requirements must be met, and page numbers must be placed consistently as in the rest of the manuscript.

Table of Contents/Headings & Subheadings within manuscripts

Your Table of Contents will be checked against the body of your manuscript.

Page numbers must be correct.

All headings/sub-headings that appear at the same level should have the same appearance in the body of the manuscript and be distinguished from other heading-levels by appearance or numbering.

All headings and subheadings (as many levels as you include in the Table of Contents) should appear in the Table of Contents. If you have some third-level headings in the Table of Contents, then ALL third-level subheadings should appear in the Table of Contents. The major section headings (usually Chapter headings) should each begin on a new page. Other levels should not be started on a new page.

Forms

Your manuscript must be accompanied by ALL the appropriate completed forms or it will not be accepted.

Please Note: The problems listed above are among the most common, but manuscripts can be returned for other problems as well.

PS 802.4 : DOMAIN KNOWLEDGE PROJECT

Total Marks : 150 **Final Viva : 100** **[L – T – P – S]**

Internal Assessment : 50 **Credits : 4** **[0 – 2 - 4–4]**

The purpose of a thesis is to enable the student to develop deeper knowledge, understanding, capabilities and attitudes in the context of the programme of study. The thesis offers the opportunity to delve more deeply into and synthesise knowledge acquired in previous studies. A thesis for a Domain Knowledge should place emphasis on the technical/scientific/artistic aspects of the subject matter.

The overall goal of the thesis is for the student to display the knowledge and capability required for independent work as a Master of Science in Software Technology.

Learning objectives for a thesis are based on the objectives for Master of Science in Software Technology has Specific learning outcomes: for the student to demonstrate:

- Considerably more in-depth knowledge of the major subject/field of study, including deeper insight into current research and development` work.
- Deeper knowledge of methods in the major subject/field of study.
- A capability to contribute to research and development work.
- The capability to use a holistic view to critically, independently and creatively identify, formulate and deal with complex issues.
- The capability to plan and use adequate methods to conduct qualified tasks in given frameworks and to evaluate this work.
- The capability to create, analyse and critically evaluate different technical/architectural solutions.
- The capability to critically and systematically integrate knowledge.
- The capability to clearly present and discuss the conclusions as well as the knowledge and arguments that form the basis for these findings in written and spoken English.
- The capability to identify the issues that must be addressed within the framework of the specific thesis in order to take into consideration all relevant dimensions of sustainable development.
- A consciousness of the ethical aspects of research and development work.

The same formulated learning objectives apply for theses encompassing 4 Credits (Hec). The course must be considerably more ambitious with respect to the scientific level or technical/architectural realisation. For example, the objective could be to author a scientific article as the result of the research or to further develop a technical solution for mass production and market introduction.

Guidelines for Min-Project development

I. Background:

- Every student is required to carry out Mini Project work under the supervision of a faculty member of the department.
- The internal supervisor shall monitor progress of the student continuously.
- A candidate is required to present the progress of the work time to time to the guide.
- There will a final presentation of theMini Project work at the end of the semester.
- The course Mini Project is one that involves requirement analysis, feasibility analysis, Database design coding, testing, implementation and maintenance.
- The Mini Project should be an Application of knowledge and techniques learnt in theoretical classes fordeveloping the s/w for real problems.

II. Motivation:

- Overview of the product design process and its relationship to the development of Web-based products and services has to be studied.
- Generate at least 10 market opportunities consistent with the course project guidelines
- Using the interviewing, pick one of your opportunities and interview at least one “customer” about their “customer needs” for your opportunity.
- Translate your interview notes into “needs” statements using the process and rules.
- Identify a well executed Web-based promotional site, product orservice that could be developed using Open source tools / web programming.
- The titles and descriptions of market opportunities that (1) you are passionate about (2) meet theproject guidelines (3) you believe can be refinedand prototyped over the course.
- Identify at least twodifferent organizational schemes.
- Sketch the site hierarchy for each of the organizational schemes thatyou identified
- Select an organizational structure and draw the corresponding sitehierarchy for your project idea using a computer online drawingpackage. (Smart Draw)
- Match technical specifications to each element of your information architecture.
- List the data that you will need for your project and identify the sources of that data.

- ***Submit a full specification document (no more than 15 pages including figures) for your mini-project.***
- ***Submit a project plan with dates, deliverables, and job assignments. Prepare your project plan and submit it as a Microsoft Excel file or as a PDF file.***
- ***Submit a one page sheet that includes a product name, an associated domain name that is open and available, and a graphical logo for your product.***

- ***Submit a one-page screen dump of your product home page incorporating the product name and graphical logo.***
- ***You need not actually purchase/reserve your domain name, but verify that the name is available in the .com space.***
- ***Submit in a single MS Word or PDF file, a semi-colon separated list of 20 email addresses that we can email about the roll-out of our online products. In the interest of reaching a wide audience, please do 10 consider reaching out to classmates at other academic institutions. Please try to avoid including names of individuals who are already familiar with your products and/or processes in order to provide a fresh look at the product pages.***
- ***Identify one aspect of your site design that you could evaluate using A/B testing.-- Draft the text and interfaces required to implement the two versions. --- Define hypotheses or expectations for evaluating results.***
- ***Based upon five days of user feedback and user log analysis***
  - ***Identify opportunities for architecture redesign***
  - ***Analyze navigation paths***
  - ***High/low traffic content***
  - ***Strategies for promoting low traffic content***

III. Mini-Project Guidelines

You are developing a web-based product or service. The definition of a "web-based product or service" is quite broad and includes retailing, financial services, healthcare, public service, etc. Additional constraints on your project:

- Free to users
- This is a desktop or laptop/notebook browser-based product or service
- You already have legal access to all content (images, catalog information, narratives, etc.) in electronic form.
- The product or service neither collects nor retains any personal or otherwise individually identifiable information about the user (e.g. no cookies, no user registration, etc.)
- All content conforms to one of a maximum of five page templates.
- No use of proprietary algorithms (whether implemented or not)
- Web-enabled product or service for the target consumers. This is not, for example, a corporate website for public relations.
- Design and deploy a desktop or laptop/netbook browser-based product or service. Plans to extend your service to the wireless medium are acceptable, but the demonstration must rely upon desktop or laptop/netbook browsers.
- You must design the Information Architecture backbone that drives the user experience (e.g. this is not a Facebook app, etc.)

- Use of open source or publicly exported Web services such as the Amazon API or WordPress blogging is permitted.
- Language specifications are optional (e.g. PHP, Perl, Python, Ruby, C++, etc.)

Key milestones and deliverables for the mini-project are detailed further in the syllabus. However, your mini-project will include:

- A name
- A concept description
- A hierarchy of user needs beginning with those that are explicitly addressed
- A set of design specifications including an information architecture and page templates.
- Electronic copies of all content (images, catalog information, narratives, etc. as well as boilerplate for About, Contact, and Privacy)
- Any programming language-specific specifications.

IV. Mini Project Proposal (Synopsis) Structure:

Mini Project proposal should be prepared in **consultation with the guide.** It should clearly state the objectives and environment of the proposed Mini Project to be undertaken. Ensure to include the following items while submitting your Mini Project synopsis. Mini Project synopsis may contain 10-20 pages and sequence of contents strictly should be in the following order:

Cover and Title page

Synopsis Approval Performa duly filled and signed by the student

1) Index.
2) Acknowledgement.
3) Introduction and objective of the Mini Project.
4) Analysis (Feasibility Study, DFD 0 Level, 1- Level and 2 Level/ER Diagram etc).
5) H/W and S/W Requirement.
6) Table and Structure, Number of Modules, Detail of Modules, Data Structure.
7) Types of Reports.
8) Scope of future application.

V. Communication of Synopsis Approval

The candidate has to submit the Synposis to the Project coordinator, which will be scrutinized by a panel of experts (Faculties). The Dean of the Faculty will release a list of approved synopsis. A list of approved synopsis will be put on the notice board of the Institute as per the dates mentioned in the activity schedule. In case of non approval, the suggestions for reformulating the Mini Project synopsis will be communicated to the student. Students can resubmit the modified synopsis to Project Coordinator Department as per the specified time given in activity schedule.

VI. Mini Project Report Structure.

The Mini Project should be prepared in consultation with the guide and may contain 100-150 pages (including coding). Mini Project Report should strictly follow the points given below:

1) Cover and Title page
2) Synopsis Approval Certificate/Company Certificate
3) Index
4) Acknowledgement
5) Certificate of Originality
6) Introduction/Aims and Objective
7) System Analysis
   - 7.1 Identification of Need
   - 7.2 Preliminary Investigation
8) Feasibility Study
   - 8.1 Technical Feasibility
   - 8.2 Economic Feasibility
   - 8.3 Operational Feasibility
9) Analysis (Feasibility Study, DFD 0 Level, 1- Level and 2 Level/ER Diagram, and Data structure, Table structure etc)
10) S/W Engineering. Paradigm applied
11) S/W & H/W Requirement Specification
12) System Design
13) Screen Shots
14) Coding
15) Validation Checks
16) Implementation and Maintenance
17) Testing (Testing techniques and Testing strategies)
18) System Security measures
19) Various types of Reports/Modules
20) Pert Chart/Gantt Chart
21) Future scope of the Mini Project
22) Bibliography/References/Glossary
23) Original Copy of the Approved Synopsis

VII. Formatting of Mini Project

- The entire Mini Project report should be nicely composed and presented.
- Students should use only A4 / Letter (8.27 x 11.69 inches) size paper.
- Students should leave at-least 3.00 mm margin from all edges of paper.
- The Mini Project report should be nicely typed in good word processor and should avoid spellings and grammatical mistakes.
- Students should use one side of paper for printing.

- The Mini Project should be Hardbound and double spacing should be preferred.
- Along with the Mini Project report contents and source code, students are required to follow the points given below.
- **Cover Page** – The cover page sample is provided Training & Placement Cell, the format should be strictly followed.
- **Letter of Authentication** - To be submitted by students declaring that the Mini Project Report is the original work of student and no reward had been attained for same Mini Project ever before. Students are advised not to **COPY** the Mini Project report from other students.
- **Authorization from Organization / Institute/ University/ where such Mini Project have been implemented** - With certificate showing the student name, Mini Project name with future recommendations of organization if any.
- **Certificate from Mini Project Guide** - Certificate from the Mini Project Guide certifying the Mini Project work done under his/her guidance along with course, student name & Mini Project details complete in all respects.
- If the student feels to add on any other topics as per the demand of the Mini Project or want to include the functionalities as per the SDLC or the Software Engineering model used, that can be done and included in the Mini Project Report.
- The Mini Project report must include all the components as per the SDLC (Software Development Life Cycle).
- It is highly recommended to follow the approaches of Software Engineering methodology.
- Design documents should be reviewed properly.
- Testing should have been taken at various levels followed by acceptance test based on SRS and user manual.
- Students are advised to implant and present the alternative methods and techniques reviewed during the Mini Project phases.
- Mini Project should not be close copy or resemblance of commercial / freely available software from Internet or market.
- Students should avoid concept introduction of various SE / SAD terminologies

VIII. Important Points while Submitting Mini Project Report

- The cover page must be hard bound in ***Black Color; with Gold Embossing***.
- The size of the report would depend on the Mini Project undertaken. However it must be 100 to 120 typed pages (Double space) on A4 size paper.
- All the students are required to use the uniform font and format (except in heading and subheadings) throughout the text of the report.
- The Mini Project report must accompany a certificate authenticating the originality of the work done in the prescribed format from the Guide.
- Each student will make Two copies of Mini Project report in the recommended format. One copy is to be submitted to the Project Coordinator of the Department of IT and students will keep one copy of the Mini Project for their further reference in future.

- Each student is required to make a copy of Mini Project in a USB /CD and submit along with his/her Mini Project report.
- There should not be any deviation from formats of the project supplied by the Training & Placement Dept of AIMIT.
- In case students require any letter from the Institute for doing a Mini Project in any organization (Only the R&D Labs Government / Private / Public Sector Undertakings / NGO etc/; not the IT Companies), they can get a "Mini Project Trainee" letter attached with this document attested by the Director/Executive Director/CRC of the Institute.
- Submission of Mini Project should reach the institute at-least one month prior to start of term end examinations.
- The Mini Project report should accompany various supporting documents as described in the guidelines. It is mandatory to submit the synopsis for the Mini Project chosen.
- Along with this, student should also keep the soft copy of the Mini Project and the Mini Project report, Power Point Slides ready for the viva-voce and presentation at the time of Viva-Voice by the Internal and External Examiners.

IX. Submission of a Mini Project Report

Only one Copy of the Mini Project report in bound form is to be submitted to the Project Coordinator of Department of IT, AIMIT. Another copy of the Mini Project Report must be retained by the student which should be produced before the examiner at the time of the Viva-voice.

X. Mini Project Evaluation

As per Mangalore University norms Mini Project Report shall be evaluated by Internal and External examiners at the end of the semester. However there will be continuous monitoring of the Mini-Project progress report during the semester and distribution of marks shall be as follows

Subject Code	Subject Title	Seminar	Final Evaluation of the Mini Project Work					Total
PS 801.4	Domain Knowledge & Mini Project	Presentation & Viva	Knowledge	Application	Presentation	Documentation	Viva	
		50	20	20	20	20	20	150

****Course Enhancement carried out with Inputs from Dr. Bimal Roy from ISI, ILP Head, AIP Head, A&I HR Head, A&I Business leaders and the 4 partner college professors.**