M.Sc. Data Science & Business Analytics

Curriculum and Syllabus

Regulation 2022

Based on Choice Based Credit System (CBCS)

and

Learning Outcomes based Curriculum Framework (LOCF))

Effective from the Academic year

2022-2023

M.Sc. DATA SCIENCE & BUSINESS ANALYTICS CURRICULUM

Total number of Credits: 90

Code No.	Course	Lecture	Tutorial	Practical	Credits	CA	SEE	Total
		Hours/Week				Maximum Marks		
Core	Fundamentals of Data Science	4	0	0	4	40	60	100
Core	Java & Data Structures	4	0	0	4	40	60	100
Core	Advanced DBMS	3	0	0	3			
Core – Practical	Data Structures using Java Lab	0	0	4	2	40	60	100
Core –Practical	RDBMS Lab	0	0	4	2	40	60	100
DSE	DSE 1	3	0	0	3	40	60	100
DSE	DSE 2	3	0	0	3	40	60	100
SEC	Soft Skill 1	2	0	0	2	40	60	100
		19	**0**	8	**23**			
SEMESTER II								
Core	Regression Analysis	4	0	0	4	40	60	100
Core	Big Data Mining & Analytics	4	0	0	4	40	60	100
Core	Programming Logic using Python	3	0	0	3	40	60	100
Core - Practical	Python Programming Lab	0	0	4	2	40	60	100
Core - Practical	Big Data Analytics Lab	0	0	4	2			
DSE	DSE 3	3	0	0	3	40	60	100
DSE	DSE 4	3	0	0	3	40	60	100
SI	Internship	0	0	4	2	40	60	100
SEC	Soft Skill 2	2	0	0	2	40	60	100
		19	**0**	**12**	**25**			
SEMESTER III								
Core	Applied Statistics	4	0	0	4	40	60	100
Core	Big Data Tools & Techniques	4	0	0	4	40	60	100
Core	Natural Language Processing	4	0	0	4	40	60	100
Core Practical	Big Data Tools & Techniques Lab	0	0	4	2	40	60	100
Core Practical	NLP Lab	0	0	4	2			
DSE	DSE 5	4	0	0	4	40	60	100
DSE	DSE 6	4	0	0	4	40	60	100
SEC	Soft Skill 3/ Sector Skill Course	2	0	0	2	40	60	100
		22	**0**	8	**26**			

SEMESTER IV								
Core	Cloud Computing	4	0	0	4	40	60	100
GE	GE	2	0	0	2	40	60	100
Core	Project Work	0	0	20	10	40	60	100
		6	**0**	**20**	**16**			

DISCIPLINE SPECIFIC ELECTIVE

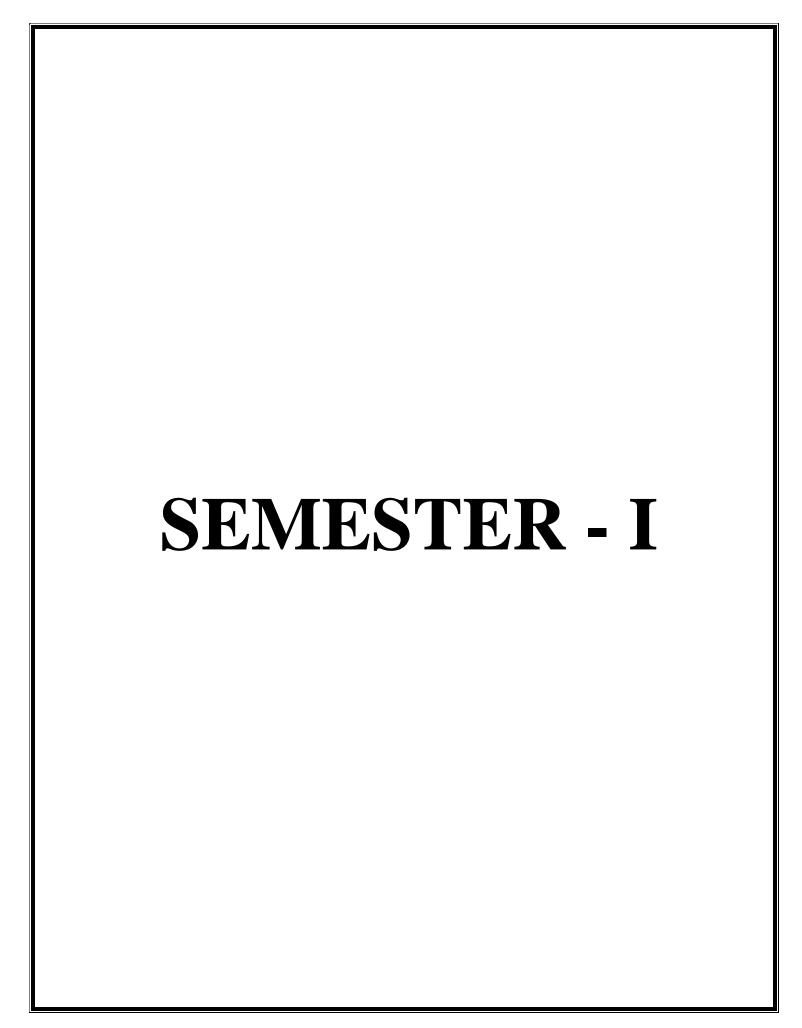|	
DSE	Machine Learning
DSE	Data Mining
DSE	Multi-core Architecture
DSE	Managerial Economics
DSE	Management Information System
DSE	Business Environment
DSE	Marketing Management
DSE	Statistical Analysis & Operation Research
DSE	Data Visualization
DSE	Predictive Analysis
DSE	Deep Learning

GENERIC ELECTIVE

22MDSBAGE104	ERP Systems
22MDSBAGE105	Ethical Hacking

SEMESTER - I

| 22CMDSBA11 | **FUNDAMENTALS OF DATA SCIENCE** | 4 | 0 | 0 | 4 |

COURSE OBJECTIVE

- Gain experience in using the tools and techniques of data science to structure and complete projects focused on obtaining actionable insights from complex data
- Dive deeply into a chosen area of practice to fully prepare to use knowledge gained in the program to add significant value in a professional setting
- Be able to utilize knowledge and skills to continue learning and adapting to new data science technologies

UNIT – I INTRODUCTION   **12**

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

UNIT – II DATA COLLECTION AND DATA PRE PROCESSING   **12**

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

UNIT – III EXPLORATORY DATA ANALYTICS   **12**

Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots –Pivot Table – Heat Map – Correlation Statistics – ANOVA.

UNIT – IV MODEL DEVELOPMENT   **12**

Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot –Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

UNIT – V MODEL EVALUATION   **12**

Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Over fitting –Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

Total Hours: 60

COURSE OUTCOME

At the End of this course, the Student will be able to:

CO1: Evaluate Fundamentals of Data Science.

CO2: Analyze different Collection Strategies and Data Pre-Processing.

CO3: Apply Descriptive Statistics and Multiple Regression.

CO4: Understand the Model Development using Visualization.

CO5: Understand the Model Evaluation and Testing Multiple Parameters

REFERENCES:

1. Jojo Moolayil, "Smarter Decisions: The Intersection of IoT and Data Science", PACKT, 2016.
2. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics", EMC 2013
4. Raj, Pethuru, "Handbook of Research on Cloud Infrastructures for Big Data Analytics", IGI Global.

WEB SITES

1. https://www.coursera.org/browse/data-science
2. https://www.codecademy.com/catalog/subject/data-science
3. https://builtin.com/data-science

WEB SOURCES

1. https://www.tutorialsduniya.com/notes/data-science-notes/
2. https://nathancarter.github.io/MA346-course-notes/_build/html/chapter-1-intro-to-data-science.html
3. https://www.guru99.com/data-science-tutorial.html

22CMDSBA12	JAVA & DATA STRUCTURES	4	0	0	4

COURSE OBJECTIVE

- This course is to develop programming skills in Java.
- It will focus on more sophisticated features such as design of classes, interfaces, packages and APIs.
- To design and implementation of both graphical applets and standalone applications.
- To learn about organizing data, linear & non - linear data structures.
- Operations on Data Structures

UNIT I   INTRODUCTION   12

Introduction to Java – Features of Java – Object Oriented Concepts – Lexical issues – Data Types – Variables – Arrays – Operators – Control Statements. Classes – Objects – Constructors – Overloading methods – Access control – Static and fixed methods – inner class – string Class – Inheritance – Overriding methods – using super – Abstract class- Dynamic Method Dispatch- using Final.

UNIT II   PACKAGES   12

Packages – Access Protection – Importing packages – Interfaces – extending an Interface- Exception Handling – Try catch, Nested Try, Multicatch Statements, Throw, Throws and finally – Exception Classes – user defined exception- Thread – states of a Thread- Thread methods-creation using Thread class- creation using Runnable Interface – Synchronization – Thread Priorities- Multithreading. - Inter thread communication- Deadlock.

UNIT III   STREAMS   12

I/O Streams – File Streams – Applets – String Class– Methods in String class-String Buffer class– Methods in String Buffer class – Java Util Package – Java Lang Package- Wrapper classes- Collection classes-

UNIT IV   STACKS & QUEUES   12

Definition of a Data Structure – Primitive and Composite data types,Asymptotic notations – Arrays – operations on arrays – ordered list.Stacks –Application of stack – Infix to postfix conversion, Recursion, queues – operation on Queue , Queue application ,circular queues-Singly linked list – Operation, Application –representation of a polynomial, polynomial addition, doubly linked list-Operation

UNIT V TREES & GRAPHS   **12**

Trees and Graphs: Binary Trees –Operation, tree traversals – Graph Implementation –Definition, Types of graph, Traversal– Shortest Path Problems, Dijikstra's algorithm

                                              Total Hours: 60

COURSE OUTCOME

On Completion of this course, the students can able to

CO1: Design Applet programs using AWT classes and utilize Controls and Layout Managers

CO2: Evaluate java program to solve specified problems and use Java SDK

Environment to create, debug and run simple java programs

CO3: Develop the Java Classes make use of Constructors and Inheritance

CO4: Apply linear & Non-linear data structures to problem solving.

CO5: List the Relationship between Linear & Non-Linear data structures like tress and graphs

TEXT BOOKS

1. Cay S.Horstmann, Gary Cornell-Core Java 2 Volume 1 – Fundamentals,5th PHI,2000.
2. Java The complete reference, 8th editon, Herbert Schildt, TMH
3. E. Horowitz, S. Sahni and Mehta – "Fundamentals of Data Structures in C++" - 2ndEdition, Universities Press – 2008.
4. Horowitz, S.Shani, and S.Rajasekaran, "computer algorithms", golgotia pub. Ltd., 2000.

REFERENCE BOOKS

1. P.Naughton and H.Schildt –Java2 (The Complete References)-Seventh Edition,TMH 2004.
2. K.Arnold and J.Gosling- The Java Programming Language – Second Edition,Addison Wesley,2002.

WEB SITES

1. **www.javapoint.com**
2. **www.geeksforgeeks.org**
3. **www.w3schools.com**
4. **www.tutorialspoint.com**

WEB SOURCES

1. **https://www.iitk.ac.in/esc101/share/downloads/javanotes5.pdf**
2. **https://www.tutorialspoint.com/java/java_tutorial.pdf**
3. **https://www.slideshare.net/intelligotech/java-tutorial-ppt-7189933**

22CMDSBA13	ADVANCED DBMS	3	0	0	3

COURSE OBJECTIVE

- This course aims to give students in depth information about system implementation techniques, data storage, representing data elements, database system architecture, the system catalog, query optimization, centralized DB concepts, Normalization, distributed databases and client server architecture, advanced database concepts.

UNIT-1 COMPARISON BETWEEN DIFFERENT DATABASES   9

Significance of Databases, Database System Applications, Advantages and Disadvantages of different Database Management systems, Comparison between DBMS, RDBMS, Distributed and Centralized DB.

UNIT-II RDBMS   9

Relational Query Languages, the SQL Query Language, Querying Multiple Relations, Creating Relations in SQL, Destroying and Altering Relations, Adding and Deleting Tuples, Integrity Constraints (ICs), Primary and Candidate Keys in SQL, Foreign Keys, Referential Integrity in SQL, Enforcing Referential Integrity.

UNIT- III   CATEGORIES OF SQL COMMANDS   9

Data Definition, Data Manipulation Statements: SELECT - The Basic Form Subqueries, Functions, GROUP BY Feature, Updating the Database, Data Definition Facilities, Views, Embedded SQL *, Declaring Variables and Exceptions, Embedding SQL Statements, Transaction Processing, Consistency and Isolation, Atomicity and Durability, Dynamic SQL.

UNIT-IV NORMALIZATION   9

Functional Dependency, Anomalies in a Database, The normalization process: Conversion to first normal form, Conversion to second normal form, Conversion to third normal form, The boyce-code normal form(BCNF), Fourth Normal form and fifth normal form, normalization and database design, Denormalization.

UNIT-V   QUERY OPTIMIZATION   9

Algorithm for Executing Query Operations: External sorting, Select operation, Join operation, PROJECT and set operation, Aggregate operations, Outer join, Heuristics in Query Optimization, Semantic Query Optimization, Converting Query Tree to Query Evaluation Plan, multi query optimization and application, Efficient and extensible algorithms for multi-query optimization, execution strategies for SQL sub queries, Query Processing for SQL Updates.

COURSE OUTCOME:

At the End of this course, the Student will be able to:

CO1: Evaluate hierarchy of DBMS.

CO2: Analyze different types of SQL statement.

CO3: Apply normalization in the database & understand the internal data structure.

CO4: Understand the transaction system & could extract data efficiently.

CO5: Understand the notion of transaction and its ACID properties

TEXT BOOKS:

1. Date C. J, "An Introduction to Database Systems", Addison Wesley Longman, 8th Edition, 2003.
2. Catell, R.G.G., Barry, D.K., Berler, M., et al, "The Object Data Standard: ODMG 3.0", Morgan Kaufmann, 2000.
3. Silberschatz A., Korth H., and Sudarshan S, "Database System Concepts", McGraw- Hill, 6th Edition, 2010.

REFERENCE BOOKS:

1. Charles F. Goldfarb, Paul Prescod, "The XML Handbook, Prentice Hall", 5th Edition, 2004.
2. Thomas M. Connolly, Carolyn Begg, "Database Systems: Practical approach to Design, Implementation and Management", Pearson Education Limited, 6th edition, 2012.

WEB SITES:

1. www.oracle-tutorial.com
2. www.studytonight.com

WEB SOURCES:

1. **https://lecturenotes.in/subject/1287/advanced-database-system**
2. **https://www.exploredatabase.com/p/blog-page.html**
3. **https://studentnet.cs.manchester.ac.uk/pgt/2011/COMP60732/ADBMS-introduction.pdf**
4. **https://edutechlearners.com/advance-database-management-system-notes/**

22PMDSBA11   **DATA STRCUTURES USING JAVA LAB**   **0 0 4 2**

OBJECTIVE

1. To familiarize various data structure implementations.
2. To implement heap and various tree structures like AVL, Red-black, B- Tree and segment trees.
3. To understand efficient implementation of line segment intersection.
4. To understand various search structures.
5. To get understanding of problem to program mapping

LIST OF EXPERIMENTS

1. Stack
2. Queue
3. Bubble sort
4. Linear Search
5. Binary Search
6. Quick sort
7. Heap Sort
8. AVL Tree

Total Hours: 30

COURSE OUTCOME

On Completion of this course, the students can able to

CO1: Achieve programming skill to convert a problem to a programming logic.

CO2: Apply suitable data structure for the problem in hand.

CO3: Implement heap and various tree structures like AVL, Red-black, B- Tree and segment trees.

CO4: Understand the usage of data structures for geometric problems.

CO5: Understand the importance of height balancing in search structures.

WEB SOURCES

1. **https://www.iitk.ac.in/esc101/share/downloads/javanotes5.pdf**
2. **https://www.tutorialspoint.com/java/java_tutorial.pdf**
3. **https://www.slideshare.net/intelligotech/java-tutorial-ppt-7189933**

| 22PMDSBA12 | RDBMS LAB | 0 | 0 | 4 | 2 |

OBJECTIVE

- This course gives training in design and implementation of data bases for the selected problems.
- To provide oriented data-processing and frameworks.
- To give a good formal foundation on the relational model of data.

LIST OF EXPERIMENTS

1. Learning basic DDL, DML, DCL and TCL commands
2. Working with dual table.
3. Use of Joins and Sub queries.
4. Views, sequences and indexes.
5. Managing users, privileges and roles.
6. PL/SQL-Data types, control structures.
7. Creating procedures with PL/SQL.
8. Error handling in PL/SQL.
9. Cursor Management in PL/SQL.
10. Writing Programs on Packages & triggers.
11. Embedding PL/SQL in high level language.
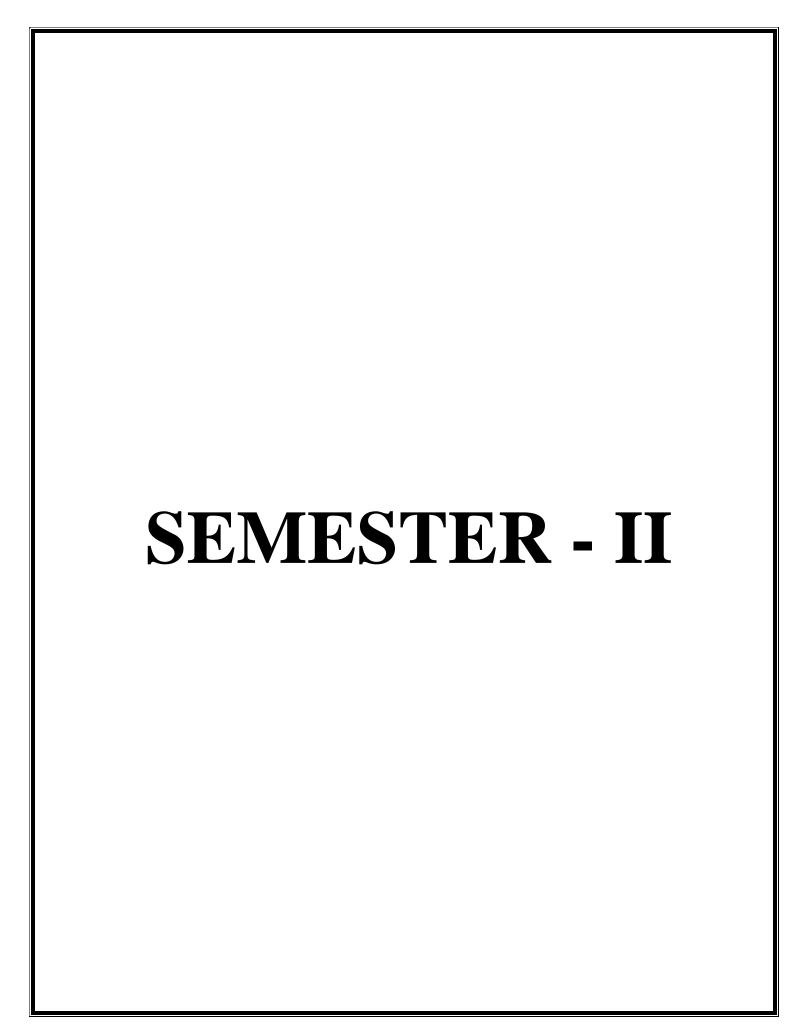12. Implementation of Triggers & Assertions for Bank Database.

COURSE OUTCOMES:

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

CO-1: Create an application to check user defined exception using PL/SQL.

CO-2: Evaluate the functionalities of trigger and cursor.

CO-3: Analyze different types of built-in function in PL/SQL.

CO-4: Apply DDL, DML and DCL statement using SQL.

CO-5: Apply various types of joins in tables.

WEB SOURCES:

1. **www.guru99.com/dbms-tutorial.html**
2. **www.oracletutorial.com**

SEMESTER - II

22MDSBA21   REGRESSION ANALYSIS   4   0   0   4

COURSE OBJECTIVES:

The focus of this course is on a careful understanding and of regression models and associated methods of statistical inference, data analysis, and interpretation of results, statistical computation and model building.

UNIT I 12

Linear Regression with One Predictor Variable.Variable Types. Regression models. SLR. Error term

UNIT II 12

Linear Regression with One Predictor Variable- Estimation coefficients. Maximum Likelihood Estimator, Least Squares Estimator. LSE with R

UNIT III 12

Inferences in Regression and Correlation Analysis. Inference about the regression coefficients β_0 and β_1. Confidence intervals with R

UNIT IV 12

Inferences in Regression and Correlation Analysis. (cont.) Estimation of $E(Y_h)$ for a given X_h. Prediction of Y_h(new) for a given X_h. Estimation and Prediction with R.

UNIT V 12

Inferences in Regression and Correlation Analysis. (cont.) Analysis of Variance approach to regression analysis. Descriptive Measures of Linear Association between X and Y. ANOVA table and R2 with .Matrices and Special Types of Matrices. Matrix Addition and Subtraction. Inverse of a Matrix. Matrix multiplication. Matrix Representation of Linear Regression Model. Regression Analysis with Matrices

Total No of Hours : 60

COURSE OUTCOMES:

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

CO-1: Can able to understand-linear Regression with One Predictor Variable

CO-2: Can able to understand Estimation coefficients

CO-3: Analyze Inferences in Regression and Correlation Analysis

CO-4: Apply Inferences in Regression and Correlation Analysis

CO-5: Analyze the Association between X and Y. ANOVA table and R2.

TEXT BOOKS:

1. Regression Analysis by ьy J. Holton Wilson Released August 2012,Publisher(s): Business Expert Press,ISBN: 9781606494356
2. Introduction to Linear Regression Analysis, Douglus,Elizebeth A Peck

WEB SITES

5. **www.javapoint.com**
6. **www.geeksforgeeks.org**
7. **www.w3schools.com**
8. **www.tutorialspoint.com**

WEB SOURCE

1. Interactive Lecture Notes 12-Regression Analysis (umich.edu)

| 22MDSBA22 | BIG DATA MINING & ANALYTICS | 4 | 0 | 0 | 4 |

COURSE OBJECTIVES:

- To understand the computational approaches to Modeling, Feature Extraction.
- To understand the need and application of Map Reduce.
- To understand the various search algorithms applicable to Big Data.
- To analyze and interpret streaming data.
- To learn how to handle large data sets in main memory.
- To learn the various clustering techniques applicable to Big Data.

UNIT I   DATA MINING AND LARGE SCALE FILES   12

Introduction to Statistical modeling – Machine Learning – Computational approaches to modeling – Summarization – Feature Extraction – Statistical Limits on Data Mining – Distributed File Systems– Map-reduce – Algorithms using Map Reduce – Efficiency of Cluster Computing Techniques.

UNIT II   SIMILAR ITEMS   12

Nearest Neighbor Search – Shingling of Documents – Similarity preserving summaries   Locality sensitive hashing for documents – Distance Measures – Theory of Locality   Sensitive Functions – LSH Families – Methods for High Degree of Similarities.

UNIT III   MINING DATA STREAMS   12

Stream Data Model – Sampling Data in the Stream – Filtering Streams – Counting   Distance Elements in a Stream – Estimating Moments – Counting Ones in Window –   Decaying Windows

UNIT IV   LINK ANALYSIS AND FREQUENT ITEMSETS   12

Page Rank –Efficient Computation – Topic Sensitive Page Rank – Link Spam – Market   Basket Model – Apriori algorithm – Handling Larger Datasets in Main Memory –   Limited Pass Algorithm – Counting Frequent Item sets.

UNIT V   CLUSTERING   12

Introduction to Clustering Techniques – Hierarchical Clustering –Algorithms – K-Means   –   CURE   – Clustering in Non – Euclidean Spaces – Streams and Parallelism – Case Study: Advertising on the Web – Recommendation Systems

Total Hrs: 60

COURSE OUTCOMES:

Upon Completion of the Course, the Student will be able to

- CO1: Design algorithms by employing Map Reduce technique for solving Big Data problems.
- CO2: Identify similarities using appropriate measures.
- CO3: Point out problems associated with streaming data and handle them.
- CO4: Discuss algorithms for link analysis and frequent item set mining.
- CO5: Design solutions for problems in Big Data by suggesting appropriate clustering techniques.

TEXT BOOKS:

1. White, "Hadoop: The Definitive Guide", Third Edition - 2012 – O'Reilly – ISBN: 9789350237564.
2. Cay S. Horstmann, Gary Cornell, "Core Java™ 2: Volume II–Advanced Features", Prentice Hall, 9th edition, ISBN: 978-0137081608.
3. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, Second Edition, 2014. 2. Jiawei Han, MichelineKamber,
4. Ian H.Witten, Eibe Frank "Data Mining – Practical Machine Learning Tools and Techniques", Morgan Kaufman Publications, Third Edition, 2011.

REFERENCES:

1. Jean Dollimore, Tim Kindberg, George Coulouris, "Distributed Systems Concepts and Design", 4th Edition, Jun 2005, Hardback, 944 pages, ISBN: 9780321263544.

2. Y. Daniel Liang, Introduction to Java Programming, Tenth Edition, Pearson, 2015.

3. Jian Pei, "Data Mining Concepts and Techniques", Morgan Kaufman Publications, Third Edition, 2011

4. David Hand, HeikkiMannila and Padhraic Smyth, "Principles of Data Mining", MIT Press,2001

WEB SITES:

1. **https://www.simplilearn.com/what-is-big-data-analytics-article**
2. **https://www.coursera.org/courses?query=big%20data%20analytics**
3. **https://www.mygreatlearning.com/academy/learn-for-free/courses/mastering-big-data-analytics**

WEB SOURCES:

1. **https://mrcet.com/downloads/digital_notes/CSE/IV%20Year/(R17A0528%20)%20Big%20Data%20Analytics%20Digital%20notes.pdf**
2. **https://www.iare.ac.in/sites/default/files/NEW%20LECHURE%20NOTES.pdf**
3. **https://www.studocu.com/in/document/galgotias-university/big-data-technology/big-data-analytics-lecture-notes-1/17645396**

22MDSBA23   PROGRAMMING LOGIC USING PYTHON   3   0   0   3

COURSE OBJECTIVES:

- To learn the fundamentals of natural language processing
- To understand the use of CFG and PCFG in NLP
- To understand the role of semantics of sentences and pragmatics
- To apply the NLP techniques to IR applications

UNIT I   INTRODUCTION   9

Python Program Execution Procedure – Statements – Expressions – Flow of Controls –Functions – Numeric Data Types – Sequences – Strings – Tuples – Lists – Dictionaries. Class – Constructors – Object Creation – Inheritance – Overloading.

UNIT II   NUMPY AND PANDAS PACKAGES   9

NumPy ndarray - Vectorization Operation - Array Indexing and Slicing - Transposing Array and Swapping Axes - Saving and Loading Array - Universal Functions - Mathematical and Statistical Functions in Numpy . Series and DataFrame data structures in pandas - Creation of Data Frames – Accessing the columns in a Data Frame - Accessing the rows in a Data Frame - Panda's Index Objects -Reindexing Series and Data Frames - Dropping entries from Series and Data Frames -Indexing, Selection and Filtering in Series and Data Frames - Arithmetic Operations between Data Frames and Series - Function Application and Mapping.

UNIT III   DATA WRANGLING   9

Combining and Merging Data Sets – Reshaping and Pivoting – Data Transformation – String manipulations – Regular Expressions.

UNIT IV   DATA AGGREGATION AND GROUP OPERATIONS   9

GroupBy Mechanics – Data Aggregation – GroupWise Operations – Transformations – Pivot Tables – Cross Tabulations – Date and Time data types.

UNIT V   VISUALIZATION IN PYTHON   9

Matplotlib and Seaborn Packages – Plotting Graph - Controlling Graphs – Adding Text –More Graph Types – Getting and Setting Values – Patches.

Total Hours: 45

COURSE OUTCOME

On Completion of this course, the students can able to

    CO1: Able to apply the principles python programming.

    CO2: Write clear and effective python code

    CO3: Create applications using python programming.

    CO4: Implementing database using SQLite.

    CO5: Access database using python programming

    CO6: Develop web applications using python programming

    TEXT BOOKS

      1.   1. Gowrishanker and Veena, "Introduction to Python Programming", CRC Press, 2019.

      2.   Python Crash Course, 2nd Edition, By Eric Matthes, May 2019

REFERENCE BOOKS:

1. NumPy Essentials, By Leo Chin and Tanmay Dutta, April 2016

2. Joel Grus, "Data Science from scratch", O'Reilly, 2015.

3. Wes Mc Kinney, "Python for Data Analysis", O'Reilly Media, 2012.

4. Kenneth A. Lambert, (2011), "The Fundamentals of Python: First Programs", Cengage

4.Jake Vanderplas. Python Data Science Handbook: Essential Tools for Working with   Data

1st Edition.

WEB SITES:

    1. **https://www.coursera.org/lecture/python-programming/logical-expressions-f5wH1**

    2. **https://www.udemy.com/course/python-37-for-beginners-logic-building-practice-questions/**

    3. **https://www.edx.org/professional-certificate/introduction-to-python-programming**

WEB SOURCES

    1. **https://docs.python.org/3/tutorial/index.html**

    2. **https://www.learnpython.org/**

22MDSBAP21 **BIG DATA ANALYTICS LAB** **0 0 4 2**

COURSE OBJECTIVES:

- To set up single and multi-node Hadoop Clusters.

- To solve Big Data problems using Map Reduce Technique.

- To design algorithms that uses Map Reduce Technique to apply on Unstructured and structured data.

- To learn NoSQL query.

- To learn Scalable machine learning using Mahout.

LIST OF EXPERIMENTS:

1. Set up a pseudo-distributed, single-node Hadoop cluster backed by the Hadoop Distributed File System, running on Ubuntu Linux. After successful installation on one node, configuration of a multi-node Hadoop cluster (one master and multiple slaves).

2. MapReduce application for word counting on Hadoop cluster

3. Unstructured data into NoSQL data and do all operations such as NoSQL query with API.

4. K-means clustering using map reduce

5. Page Rank Computation

6. Mahout machine learning library to facilitate the knowledge build up in big data analysis.

7. Application of Recommendation Systems using Hadoop/mahout libraries

Total Hrs: 30

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

CO1: Set up single and multi-node Hadoop Clusters.

CO2: Apply Map Reduce technique for various algorithms.

CO3: Design new algorithms that uses Map Reduce to apply on Unstructured and structured data.

CO4: Develop Scalable machine learning algorithms for various Big data applications using Mahout.

CO5: Represent NoSQL data

22MDSBAP22　　　　　**PYTHON PROGRAMMING LAB**　　　**0**　　**0**　　**4**　　**2**

OBJECTIVE

To provide practical experience in software development using open source tools like PYTHON & PHP.

LIST OF EXPERIMENTS

1. Editing and executing Programs involving Flow Controls.

2. Editing and executing Programs involving Functions.

3. Program in String Manipulations

4. Creating and manipulating a Tuple

5. Creating and manipulating a List

6. Creating and manipulating a Dictionary

7. Object Creation and Usage

8. Program involving Inheritance

9. Program involving Overloading

10. Reading and Writing with Text Files and Binary Files

11. Combining and Merging Data Sets

12. Program involving Regular Expressions

13. Data Aggregation and GroupWise Operations

Total Hours: 30

COURSE OUTCOMES

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

CO1: Create applications using PYTHON.

CO2: Analyze and apply the role of languages like HTML, DHTML.

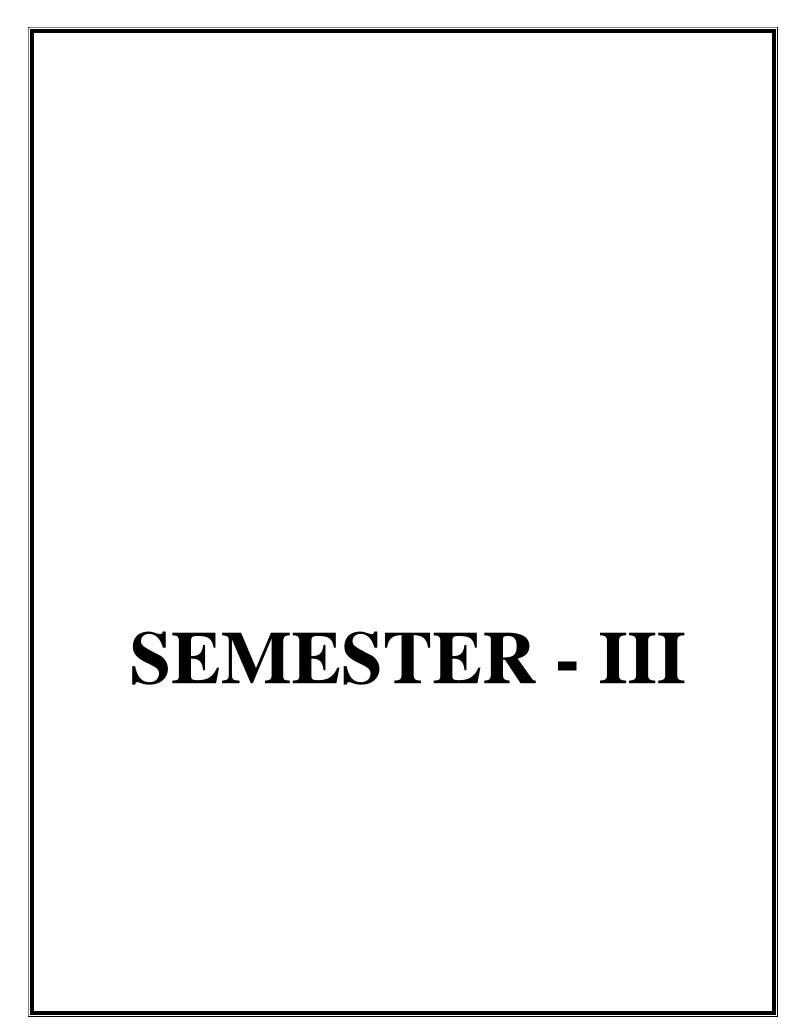CO3: Analyze the utilities of Python with other object oriented programming Languages.

CO4: Analyze and Object Creation and Usage

Using PYTHON.

CO5: Explain Reading and Writing with Text Files and Binary Files

WEB SOURCES:

1. **https://www.javatpoint.com/python-programs.**
2. **https://www.datacamp.com/community/tutorials/mysql-python**

SEMESTER - III

22MDSBA31   **APPLIED STATISTICS**   **4   0   0   4**

COURSE OBJECTIVES:

Statistics plays an important role in data analytics. The main aim of this course is to help the students to read, classify and then interpret the data given to them and draw conclusions.

UNIT I   **12**

Data and Representation Introduction to Statistics, Collection of data, classification and tabulation of data, Types of data: Primary data, Secondary data, Presentation of data Diagrammatic and Graphical Representation: Histogram, frequency curve, frequency polygon, Ogive curves, stem and leaf chart.

UNIT II   **12**

Measures of Central Tendency and Dispersion Arithmetic Mean (A.M.) Definition, Mode, Median, Partition Values: Quartiles, Deciles and Percentiles, Box Plot, Percentile ranks. Means of transformed data, Geometric Mean (G.M.) Definition, Harmonic Mean (H.M.), Weighted Mean : Weighted A.M., G.M. and H.M. Range, Mean deviation Mean square deviation, Variance and Standard Deviation, Combined variance (derivation for 2 groups), Combined standard deviation.

UNIT III   **12**

Correlation and Regression Bivariate normal distribution, types, importance, methods of measuring correlation-scatter diagram, Karl Pearson's Coefficient of Correlation and Spearman's rank Correlation. Regression lines, Difference between regression and correlation, uses of Regression.

UNIT IV   **12**

Sampling theory and tests of significance Methods of sampling (Description only): Simple random sampling with and without replacement (SRSWR and SRWOR)

UNIT V   **12**

Stratified random sampling, systematic sampling. Tests of significance – z, t, chi-square and F.

                                                              Total No of Hours: 60

TEXT BOOKS:

Mood, A. M., Graybill, F. A. And Boes, D.C. : Introduction to the Theory of Statistics, McGraw Hill. Biswas and Srivastava – A textbook, mathematical Statistics, Ist Edition, Narosa Publishing House, New Delhi.

Gupta, S.C. and V. K. Kapoor – Mathemathical Statistics, Sultan Chand and sons. Hogg,R.V. and Craig, A.T: Introduction to Mathematical Statistics, McMillan. S. C. Gupta – Fundamentals of Statistics, Himalaya Publishing House.

22MDSBA32   **BIG DATA TOOLS & TECHNIQUES**   **4   0   0   4**

COURSE OBJECTIVES:

1. To familiarize with R Programming.

2. To understand data analysis using R and HADOOP Integrated Programming Environment.

3. To Understand Analytics for Big data 'at Rest' and Real-Time Analytical Processing for Big data 'in Motion'.

4. To understand the Pig Data model and Pig scripts.

5. To learn way of Querying Big Data using Hive.

UNIT I INTRODUCTION TO R PROGRAMMING   12

Introduction to R – Vectors – Filtering – Matrices – Creating Matrices – Applying Functions to Matrix Rows and Columns – Lists – Creating List – General List Operations – Data Frames –Creating Data Frames – Matrix like Operations in Frames – Applying Functions to Data Frames –Factors and Tables – Math and Simulations in R – Input/Output – Reading and Writing Files –Graphics – Creating Three-Dimensional Plots – Linear Models – Non-linear models – Clustering

UNIT II DATA ANALYSIS USING R AND HADOOP   12

Features of R Language - HADOOP Features – HDFS and Map Reduce architecture – R and Hadoop Integrated Programming Environment- RHIPE Introduction – Architecture of RHIPE –RHIPE function reference. RHADOOP Introduction – Architecture of RHADOOP – RHADOOP function reference, SQL on HADOOP.

UNIT III ANALYTICS FOR BIG DATA STREAMS   12

IBM Pure Data Systems – Netezza‟s Design Principles – The Netezza Appliance – Extending the Netezza Analytics – Real-Time Analytical Processing – Info Sphere Streams Basics – InfoSphere Streams Working – enterprise class – industry use cases – Indexing Data from Multiple Sources –Creating Information Dashboards

UNIT IV PROGRAMMING WITH PIG   12

Introduction – installation and execution – PIG Data Model – PIG Latin – Input, Output- Relational Operators – User Defined Functions – Join Implementations – Integrating Pig with Legacy Code and Map Reduce – Developing and Testing Pig Latin Scripts – Embedding Pig Latin in Python –Evaluation Function in Java- Load Functions – Store Functions.

UNIT V PROGRAMMING WITH HIVE   12

Introduction – Data Types and File Formats – Databases in Hive – HiveQL: Data Definition – Data Manipulation – Queries – Views – Indexes – Schema Design.

Total Hours: 60

COURSE OUTCOMES:

Upon completion of the course, the students will be able to

CO1: Design applications using R, HADOOP.

CO2: Design applications using RHADOOP& RHIPE.

CO3: Develop analytic applications for data Streams.

CO4: Develop Pig scripts for Big data applications.

CO5: Design Big data applications schema and use HIVE QL.

REFERENCES:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.

2. Tom White " Hadoop: The Definitive Guide" Third Edition, O"reilly Media, 2011.

3. Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D.V Harness the

Power of Big Data The IBM Big Data Platform. McGraw Hill Professional, 2012.

4. Prajapati, V, "Big Data Analytics with R and Hadoop", Packt Publishing Ltd, 2013.

5. Gates, A. Programming Pig." O'Reilly Media, Inc.", 2011.

6. Capriolo, E., Wampler, D., &Rutherglen, J., "Programming Hive", O'Reilly Media, Inc.",2012

7. Norman Matloff , "The Art of R Programming: A Tour of Statistical Software Design",

NoStarch Press, 2011.

8. Jared P. Lander, "R for Everyone: Advanced Analytics and Graphics", Addison-Wesley

Data & Analytics Series, 2013

22MDSBA33   **NATURAL LANGUAGE PROCESSING**   **4**   **0**   **0**   **4**

COURSE OBJECTIVES:

• To learn the fundamentals of natural language processing

• To understand the use of CFG and PCFG in NLP

• To understand the role of semantics of sentences and pragmatics

• To apply the NLP techniques to IR applications

UNIT I   **INTRODUCTION**   **12**

Origins and challenges of NLP – Language Modeling: Grammar-based LM, Statistical LM - Regular Expressions, Finite-State Automata – English Morphology, Transducers for lexicon and rules, Tokenization, Detecting and Correcting Spelling Errors, Minimum Edit Distance

UNIT II   **WORD LEVEL ANALYSIS**   **12**

Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part-of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models.

UNIT III   **SYNTACTIC ANALYSIS**   **12**

Context-Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing, Ambiguity, Dynamic Programming parsing – Shallow parsing – Probabilistic CFG, Probabilistic CYK, Probabilistic Lexicalized CFGs - Feature structures, Unification of feature structures.

UNIT IV   **SEMANTICS AND PRAGMATICS**   **12**

Requirements for representation, First-Order Logic, Description Logics – Syntax-Driven Semantic analysis, Semantic attachments – Word Senses, Relations between Senses, Thematic Roles, selectional restrictions – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods.

UNIT V   **DISCOURSE ANALYSIS AND LEXICAL RESOURCES**   **12**

Discourse segmentation, Coherence – Reference Phenomena, Anaphora Resolution using Hobbs and Centering Algorithm – Coreference Resolution – Resources: Porter Stemmer, Lemmatizer, Penn Treebank, Brill's Tagger, WordNet, PropBank, FrameNet, Brown Corpus, British National Corpus (BNC).

                                                  Total Hours: 60

COURSE OUTCOME

On Completion of this course, the students can able to

CO1: To tag a given text with basic Language features

CO2: To design an innovative application using NLP components

CO3: To implement a rule based system to tackle morphology/syntax of a language

CO4: To design a tag set to be used for statistical processing.

CO5: To compare and contrast the use of different statistical approaches

or different types of NLP applications.

TEXT BOOKS

1. Daniel Jurafsky, James H. Martin―Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.

2. Steven Bird, Ewan Klein and Edward Loper, ―Natural Language Processing with Python‖, First Edition, O‗Reilly Media, 2009.

REFERENCE BOOKS

1.Breck Baldwin, ―Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015.

2.Richard M Reese, ―Natural Language Processing with Java‖, O‗Reilly Media, 2015.

3.Nitin Indurkhya and Fred J. Damerau, ―Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.

4.Tanveer Siddiqui, U.S. Tiwary, ―Natural Language Processing and Information Retrieval‖, Oxford University Press, 2008.

WEB SITES

1. https://www.cl.cam.ac.uk/teaching/2002/NatLangProc/nlp1-4.pdf
2. https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP
3. https://www.cs.jhu.edu/~phi/ai/slides/lecture-natural-language-processing.pdf

22PMDSBA31 **BIG DATA TOOLS & TECHNIQUES LAB** 0 0 4 2

COURSE OBJECTIVES:

1. To understand the basic programming constructs of R and understand the use of R in Big Data.

2. To solve Big data problems using Map Reduce Technique in R, HADOOP.

3. To develop Pig scripts for analyzing large un-structured and semi-structured data.

4. To develop program for Query processing using Hive.

5. To perform analytics on Big data streams using Hadoop Streaming API.

6. To learn to work on Sqoop.

LIST OF EXPERIMENTS:

1. Perform descriptive and predictive analytics using "R programming".

2. Map Reduce application for word counting on R HADOOP after successful installation of three R packages(rhdfs, rmr, and rhbase).

3. Understand data pipeline using Pig Interactive Shell Commands after successful "Pig" Installation.

4. Develop Pig Scripts and call UDF's to accomplish functionalities to meet the problem Objectives .

5. Embedding PIG Latin in Python.

6. Log analysis using "Pig" on semi structured data.

7. Perform query processing on data warehousing after successful installation of "Hive".

8. Perform adhoc query on HDFS data using Hive Query Language (HQL).

9. Accomplish MapReduce Job by using Hadoop Streaming API.

10. Perform various HDFS commands.

11. Loading data into HDFS using Sqoop.

Total Hours: 60

COURSE OUTCOMES:

Upon completion of the course, the students will be able to

CO1: Set up R packages and develop a program using R Programming constructs.

CO2: Solve Big Data problems using RHADOOP.

CO3: Understand setting up of Pig and solve Big Data problems.

CO4: Understand setting up of Hive and perform query processing.

CO5: Apply Hadoop Streaming API for Big Data problems.

CO6: Apply Sqoop for data loading into HDFS.

| 22PMDSBA32 | NLP LAB | 0 | 0 | 4 | 2 |

COURSE OBJECTIVE:

- Make use of Data sets in implementing the machine learning algorithms

- Implement the machine learning concepts and algorithms in any suitable language of choice.

LIST OF EXPERIMETS

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.

2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

4. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.

5. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.

8. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment 15CSL76 ML LAB Dept. of CSE,MSEC Page 3 on the quality of clustering. You can add Java/Python ML library classes/API in the program.

9. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

30

COURSE OUTCOME

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

CO1: Understand the implementation procedures for the machine learning algorithms.

CO2: Design Java/Python programs for various Learning algorithms.

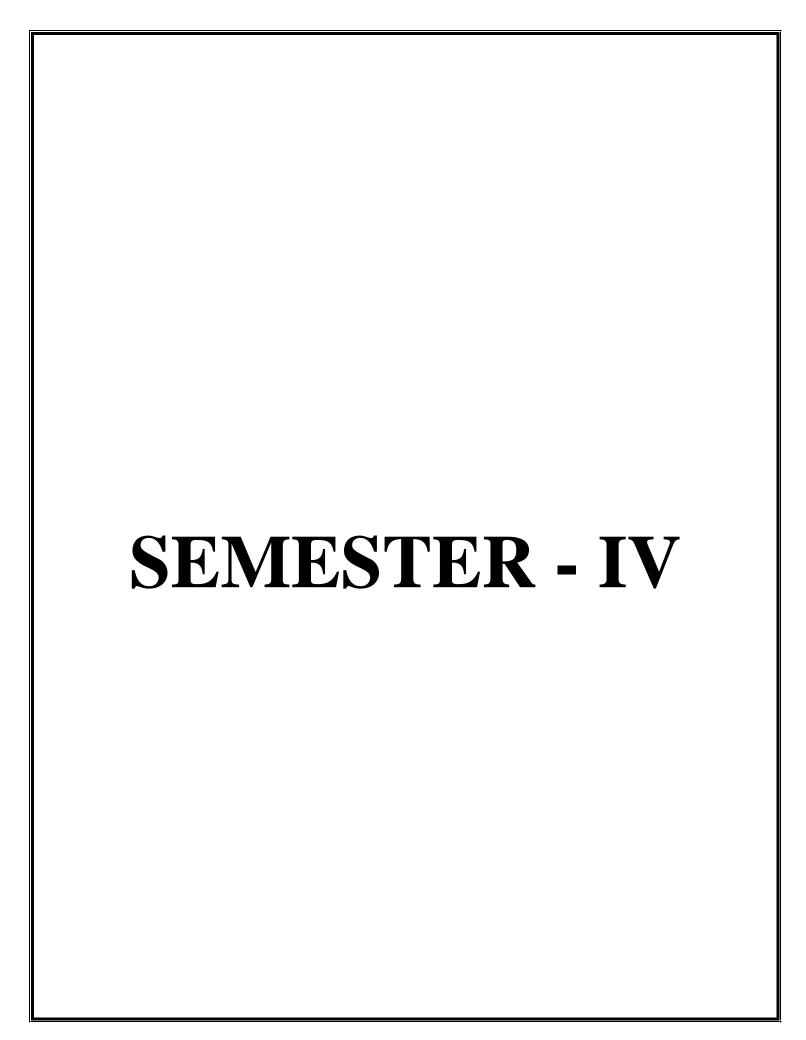CO3: Apply appropriate data sets to the Machine Learning algorithms.

CO4: Identify and apply Machine Learning algorithms to solve real world problems

TEXT BOOKS:

1. Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.

2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, h The Elements of Statistical Learning, 2nd edition, springer series in statistics.

3. Ethem Alpaydın, Introduction to machine learning, second edition, MIT press

WEB SOURCES:

1. **https://nlp-lab.org/**
2. **https://nlp-iiith.vlabs.ac.in/**
3. **https://www.srmist.edu.in/natural-language-processing-machine-learning-lab**

SEMESTER - IV

22MDSBA41 **CLOUD COMPUTING** 4 0 0 4

COURSE OBJECTIVE

1. To understand the concept of cloud and utility computing
2. To understand the various issues in cloud computing
3. To familiarize themselves with the lead players in cloud
4. To appreciate the emergence of cloud as the next generation computing paradigm
5. To be able to set up a private cloud

UNIT I   INTRODUCTION   12

Evolution of Cloud Computing –System Models for Distributed and Cloud Computing–NIST Cloud Computing Reference Architecture -IaaS–On-demand Provisioning -Elasticity inCloud –E.g.of IaaS Providers -PaaS –E.g.of PaaS Providers –SaaS –E.g. of SaaS Providers–Public , Private and Hybrid Clouds.

UNIT II   VIRTUALIZATION   12

Basics of Virtualization -Types of Virtualization -Implementation Levels of Virtualization - Virtualization Structures -Tools and Mechanisms -Virtualization ofCPU, Memory, I/O Devices -Desktop Virtualization –Server Virtualization.

UNIT III   CLOUD INFRASTRUCTURE   12

Architectural Design of Compute and Storage Clouds –Layered Cloud Architecture Development –Design Challenges -Inter Cloud Resource Management–Resource Provisioning and Platform Deployment –Global Exchange of Cloud Resources.

UNIT IV   PROGRAMMING MODEL   12

Parallel and Distributed Programming Paradigms –Map Reduce, Twister and Iterative MapReduce –Hadoop Library from Apache –Mapping Applications Programming Support -Google App Engine, Amazon AWS -Cloud Software Environments -Eucalyptus, Open Nebula, OpenStack.

UNIT V   SECURITY IN THE CLOUD   12

Security Overview –Cloud Security Challenges –Software-as-a-Service Security –Security Governance –Risk Management –Security Monitoring –Security Architecture Design –Data Security –Application Security –Virtual Machine Security.

Total No of hours 60

COURSE OUTCOME:

At the end of this course the students will be able to,

CO1: Analyze various cloud programming models and apply them to solve problems on the cloud.

CO2: Build cloud architecture.

CO3: Explain the core concepts of the cloud computing paradigm.

CO4: Get clear knowledge of various cloud models and their
         services, characteristics, advantages and Challenges.

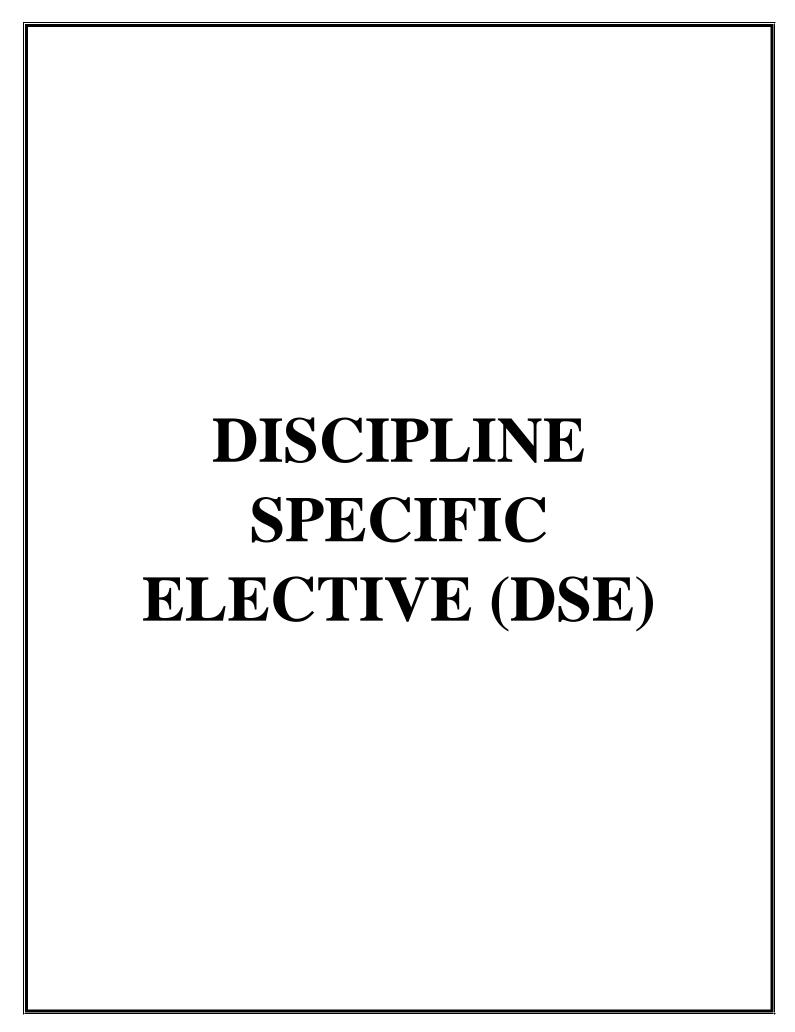CO5: Interpret the security issues in cloud.

TEXT BOOK:

1. Kai Hwang, Geoffrey C Fox, Jack G Dongarra, "Distributedand Cloud Computing, From Parallel

Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.

REFERENCE BOOKS:

1. John W.Rittinghouse and James F.Ransome, "Cloud Computing: Implementation, Management, and Security", CRC Press,2010.

2. Toby Velte, Anthony Velte, Robert Elsenpeter, "Cloud Computing, A Practical Approach",TMH, 2009.

3. George Reese, "Cloud Application Architectures: Building Applications and Infrastructure in the Cloud"O'Reilly, 2009.

4. James E. Smith, Ravi Nair, "Virtual Machines: Versatile Platforms for Systems and Processes", Elsevier/Morgan Kaufmann, 2005.

WEBSITES

1. **www.w3schools.com**
2. **www.tutorialspoint.com**
3. **www.geeksforgeeks.org**

WEB SOURCES

1. **https://www.smartzworld.com/notes/cloud-computing-pdf-notes-cc/**
2. **https://lecturenotes.in/subject/366/cloud-computing-cc/note**

DISCIPLINE SPECIFIC ELECTIVE (DSE)

21DMDSBA11 MACHINE LEARNING 3 0 0 3

COURSE OBJECTIVES:

- To understand the concepts of Machine Learning.
- To appreciate supervised learning and their applications.
- To appreciate the concepts and algorithms of unsupervised learning.
- To understand the theoretical and practical aspects of Probabilistic Graphical Models.
- To appreciate the concepts and algorithms of advanced learning.

UNIT I INTRODUCTION 8

Machine Learning–Types of Machine Learning –Machine Learning process- preliminaries, testing Machine Learning algorithms, turning data into Probabilities, and Statistics for Machine Learning Probability theory – Probability Distributions – Decision Theory.

UNIT II SUPERVISED LEARNING 10

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

UNIT III UNSUPERVISED LEARNING 9

Clustering- K-means – EM Algorithm- Mixtures of Gaussians –Dimensionality Reduction, Linear Discriminant Analysis, Factor Analysis, Principal Components Analysis, Independent Components Analysis.

UNIT IV PROBABILISTIC GRAPHICAL MODELS 9

Graphical Models – Undirected Graphical Models – Markov Random Fields – Directed Graphical Models –Bayesian Networks – Conditional Independence properties – Markov Random FieldsHidden Markov Models – Conditional Random Fields(CRFs).

UNIT V ADVANCED LEARNING **9**

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

Total Hours: 45

COURSE OUTCOMES:
- Upon completion of the course, the students will be able to
- Design a learning model appropriate to the application.
- Design a Neural Network for an application of your choice.
- Implement Probabilistic Discriminative and Generative algorithms for an application of your choice and analyze the results.
- Use a tool to implement typical Clustering algorithms for different types of applications.
- Design and implement an HMM for a Sequence Model type of application.
- Identify applications suitable for different types of Machine Learning with suitable justification.

TEXT BOOKS

1. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2007.

2. Stephen Marsland, "Machine Learning – An Algorithmic Perspective", Chapman and Hall, CRC Press, Second Edition, 2014.

3. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.

4. EthemAlpaydin, "Introduction to Machine Learning", MIT Press, Third Edition, 2014.

5. Tom Mitchell, "Machine Learning", McGraw-Hill, 1997.

21DMDSBA12   DATA MINING   3   0   0   3

COURSE OBJECTIVE:

- Learn about data mining functionalities, applications and issues.
- Demonstrate the data mining classification and clustering analysis.
- To understand the components and benefits of data ware housing for employability as data analyst or scientist

UNIT I INTRODUCTION   9

Data mining – Functionalities – Knowledge Discovery Process- Applications of Data mining– Issues in Data mining -Classification of Data mining- Tasks Primitives- Data Preprocessing-Cleaning, Reduction, Transformation.

UNIT 2 CLASSIFICATION AND PREDICTION   9

Classification: Introduction to Classification-Issues Regarding Classification and Prediction-Classification by Decision Tree Induction- Rule based Classification-Support Vector Machine-Other Classification Methods-Prediction: Introduction- Regression Analysis.

UNIT 3   CLUSTERING ANALYSIS, ASSOCIATION MINING, TEXT MINING   9

Introduction- Applications of Cluster Analysis- Requirements of Clustering in Data Mining-Categorization of Major Clustering Methods- Partitioning Methods-Hierarchical Methods-Association Rule Mining-Text Mining-Web Mining.

UNIT 4   DATA WAREHOUSING   9

Data warehousing Components- Benefits of data Warehousing-Operational and informational Data-Data Warehouse Characteristics- Data Warehouse Architecture and its components-Benefits of data warehousing- Mapping the data warehouse architecture to Multiprocessor architecture.

UNIT 5 ON-LINE ANALYTICAL PROCESSING (OLAP) and Data Mining Tool   9

OLAP-Introduction-Need for OLAP- Categories of OLAP Tools- OLAP Tools and the Internet-List of tools for Data Mining.

Total No of Hours 45

COURSE OUTCOME:

At the end of this course the students will be able to,

CO1: Analyze different classification & clustering techniques

CO2: Can able to understand steps involved in data mining.

CO3: Understand the basic functionalities & issues of data mining

CO4: Understand the basic components & benefits of Data Warehousing.

CO5: Understand the concept of OLAP tools

CO6: Able to get employability in Data warehouse engineer.

TEXT BOOKS

1. Jiawei Han and MichelineKamber, "Data Mining Concepts and Techniques", Third Edition, Elsevier, 2012.

2. Alex Berson and Stephen J.Smith, ―Data Warehousing, Data Mining & OLAP‖, Tata McGraw – Hill Edition, 35th Reprint 2016

WEBSITES

1. **www.w3schools.com**
2. **www.tutorialspoint.com**
3. **www.geeksforgeeks.org**

WEB SOURCES

1. **https://lecturenotes.in/subject/32/data-mining-and-data-warehousing-dmdw/note**
https://www.smartzworld.com/notes/data-warehousing-and-data-mining-pdf-notes-dwdm/

21DMDSBA12   MULTICORE ARCHITECTURE   3 0 0 3

COURSE OBJECTIVES:

- To introduce the students to the recent trends in the field of Computer Architecture and
- Identify performance related parameters.
- To understand the different multiprocessor issues.
- To expose the different types of multicore architectures.
- To understand the design of the memory hierarchy.
- To understand how the various forms of parallelism are exploited by the architecture.

UNIT I   FUNDAMENTALS OF COMPUTER DESIGN AND ILP   9

Fundamentals of Computer Design – Measuring and Reporting Performance – Instruction Level Parallelism and its Exploitation – Concepts and Challenges – Limitations of ILP – Multithreading –SMT and CMP Architectures – The Multicore era.

UNIT II   MEMORY HIERARCHY DESIGN   9

Introduction – Optimizations of Cache Performance – Memory Technology and Optimizations Protection: Virtual Memory and Virtual Machines – Design of Memory Hierarchies – Case Studies.

UNIT III MULTIPROCESSOR ISSUES   9

Symmetric and Distributed Shared Memory Architectures – Cache Coherence Issues –

Performance Issues – Synchronization Issues – Models of Memory Consistency – Interconnection Networks – Buses, Crossbar and Multi-stage Interconnection Networks.

UNIT IV MULTICORE ARCHITECTURES   9

Homogeneous and Heterogeneous Multi-core Architectures – Intel Multicore Architectures – SUN CMP architecture – IBM Cell Architecture. Introduction to Warehouse- Scale computers, Cloud Computing – Architectures and Issues – Case Studies.

UNIT V VECTOR, SIMD AND GPU ARCHITECTURES   9

Vector Architecture – SIMD Extensions for Multimedia – Graphics Processing Units – Case Studies – GPGPU Computing – Detecting and Enhancing Loop Level Parallelism – Introduction to Domain Specific Architectures.

Total Hrs : 45

COURSE OUTCOMES

Upon Completion of the Course, the Student will be able to

CO1: Identify the limitations of ILP and the need for multicore architectures.

CO2: Discuss the issues related to multiprocessing and suggest solutions.

CO3: Point out the salient features of different multicore architectures and how they exploit parallelism.

CO4: Point out the various optimizations that can be performed to improve the memory hierarchy design.

CO5: Point out the salient features of vector, GPU and domain specific architectures.

TEXTBOOKS:

1. John L. Hennessey and David A. Patterson, "Computer Architecture – A Quantitative Approach", Morgan Kaufmann, Elsevier, 5th edition, 2012.

2. Darryl Gove, "Multicore Application Programming: For Windows, Linux, and Oracle Solaris",Pearson, 2011.

3. David B. Kirk, Wen-mei W. Hwu, "Programming Massively Parallel Processors", Morgan Kauffman, 2010.

4. Wen– mei W. Hwu, "GPU Computing Gems", Morgan Kaufmann, Elsevier, 2011.

21DMDSBA31

MANAGERIAL ECONOMICS   3   0   0   3

COURSE OBJECTIVES

- To introduce the economic concepts.
- To familiarize with the students the importance of economic.
- approaches in managerial decision making.
- To understand the applications of economic theories in business decisions.

UNIT – I   **9**

General Foundations of Managerial Economics - Economic Approach - Circular Flow of Activity - Nature of the Firm - Objectives of Firms - Demand Analysis and Estimation - Individual, Market and Firm demand - Determinants of demand - Elasticity measures and Business Decision Making -- Demand Forecasting.

UNIT-II   **9**

Law of Variable Proportions - Theory of the Firm - Production Functions in the Short and Long Run - Cost Functions – Determinants of Costs – Cost Forecasting - Short Run and Long Run Costs –Type of Costs - Analysis of Risk and Uncertainty.

UNIT-III   **9**

Product Markets -Determination Under Different Markets - Market Structure – Perfect Competition – Monopoly – Monopolistic Competition – Duopoly - Oligopoly - Pricing and Employment of Inputs Under Different Market Structures – Price Discrimination - Degrees of Price Discrimination.

UNIT-IV   **9**

Introduction to National Income – National Income Concepts - Models of National Income Determination - Economic Indicators - Technology and Employment - Issues and Challenges – Business Cycles – Phases – Management of Cyclical Fluctuations - Fiscal and Monetary Policies.

UNIT – V   **9**

Macro-Economic Environment - Economic Transition in India - A quick Review - Liberalization, Privatization and Globalization - Business and Government - Public-Private Participation (PPP) - Industrial Finance - Foreign Direct Investment (FDIs).

REFERENCES

1. Yogesh Maheswari, Managerial Economics, Phi Learning, Newdelhi, 2005 Gupta G.S.,

2. Managerial Economics, Tata Mcgraw-Hill, New Delhi Moyer &Harris,

3. Anagerial Economics, Cengage Learning, Newdelhi, 2005 Geetika, Ghosh & Choudhury, ,

4. Managerial Economics, Tata Mcgrawhill, Newdelhi, 2012

21DMDSBA31   MANAGEMENT INFORMATION SYSTEM   3   0   0   3

COURSE OBJECTIVE:

- Information Systems (IS) enables new approaches to improve efficiency and efficacy of business models.
- This course will equip the students with understanding of role, advantages and components of an Information System.
- The objective of the course is to help students integrate their learning from functional areas, decision making process in an organization and ole of Information Systems to have a vintage point in this competitive world.

UNIT I: BASIC CONCEPTS OF INFORMATION SYSTEM   9

Role of data and information, Organization structures, Business Process, Systems Approach and introduction to Information Systems.

UNIT II: TYPES OF IS   9

Resources and components of Information System, integration and automation of business functions and developing business models. Role and advantages of Transaction Processing System, Management Information System, Expert Systems and Artificial Intelligence, Executive Support Systems and Strategic Information Systems.

UNIT III: ARCHITECTURE & DESIGN OF IS   9

Architecture, development and maintenance of Information Systems, Centralized and Decentralized Information Systems, Factors of success and failure, value and risk of IS.

UNIT IV: DECISION MAKING PROCESS   9

Programmed and Non- Programmed decisions, Decision Support Systems, Models and approaches to DSS

UNIT V: INTROTO ENTERPRISE MANAGEMENT TECHNOLOGIES   9

Business Process Reengineering, Total Quality Management and Enterprise Management System viz. ERP,SCM, CRM and Ecommerce.

                                                      Total Hrs: 45

TEXTBOOKS

1. Management Information Systems, Effy OZ, Thomson Leaning/Vikas Publications
2. Management Information Systems, James A. O'Brein, Tata McGraw-Hill References:

3. Management Information System, W.S Jawadekar, Tata Mc Graw Hill Publication.

4. Management Information System, David Kroenke, Tata Mc Graw Hill Publication.

5. MIS: Management Perspective, D.P. Goyal, Macmillan Business Books.

6. MIS and Corporate Communications, Raj K. Wadwha, Jimmy Dawar, P. Bhaskara Rao, Kanishka Publishers.

7. MIS: Managing the digital firm, Kenneth C. Landon, Jane P. Landon, Pearson

21DMDSBA32   MARKETING MANAGEMENT   3   0   0   3

COURSE OBJECTIVES

The objective of this course is to facilitate understanding of the conceptual framework of marketing and its applications in decision making under various environmental constraints.

UNIT-I INTRODUCTION: 09

Concept, nature, scope and importance of marketing; Marketing concept and its evolution; Marketing mix; Strategic marketing planning – an overview. Market Analysis and Selection: Marketing environment – macro and micro components and their impact on marketing decisions; Market segmentation and positioning; Buyer behavior; consumer versus organizational buyers; Consumer decision making process.

UNIT-II   09

Product Decisions: Concept of a product; Classification of products; Major product Decisions; Product line and product mix; Branding; Packaging and labeling; Product life cycle – strategic implications; New product development and consumer adoption process. Pricing Decisions: Factors affecting price determination; Pricing policies and strategies; Discounts and rebates.

UNIT-III   09

Distribution Channels and Physical Distribution Decisions: Nature, functions, and types of Distribution channels; Distribution channel intermediaries; Channel management decisions; Retailing and wholesaling. Promotion Decisions: Communication Process; Promotion mix Advertising, personal selling, sales promotion, publicity and public relations; Determining advertising budget; Copy designing and testing; Media selection; Advertising effectiveness; Sales promotion – tools and techniques.

UNIT-IV   09

Marketing Research: Meaning and scope of marketing research; Marketing research Process. Marketing Organization and Control: Organizing and controlling marketing Operations

UNIT-V   09

Issues and Developments in Marketing: Social, ethical and legal aspects of marketing;
Marketing of services; International marketing; Green marketing; Cyber marketing;
Relationship marketing and other developments of marketing.

Total No of Hours: 45

TEXTBOOKS

1. Kotlar, Philip, Marketing Management, Prentice Hall, New Delhi.

2. Stanton, Etzel, Walker, Fundamentals of Marketing, Tata-McGraw Hill, New Delhi.

3. Saxena, Rajan, Marketing Management, Tata-McGraw Hill, New Delhi.

4. McCarthy, E.J., Basic Marketing: A managerial approach, Irwin, New York

DSE IV

21DMDSBA32   BUSINESS ENVIRONMENT   3   0   0   3

COURSE OBJECTIVES

The course aims to educate the student with the different environmental factors which effect business. This course aims to develop ability to understand and scan business environment in order to analyse the opportunities and take decisions under the uncertainty.

UNIT-I   09

Concept, Significance and nature of Business Environment; Elements of environment - internal and external; Different roles of government in relation to business; Social Responsibilities of Business.

UNIT-II   09

Broad profile of Indian Economy, Industrial Policy - its historical perspectice(inbrief) , Industrial Policy Liberalisation .Economic Planning in India ; Rationale of economic planning , latest five year plan .

UNIT-III   09

Public sector -its objectives and working, major problems of public sector enterprises; Privatisation of public sector enterprises - the issue involved. Role of Private and Joint Sectors.

Securities Exchange Board of India - Organisation and Role, Regulation of Mergers and Acquisitions.

UNIT-IV   09

Competition Act; Industries Development and Regulation Act; Consumer Protection Act -its main provisions.

UNIT-V   09

Latest Export -Import Policy; Foreign Exchange Management Act; Globalisation and Business practices, WTO- objectives and Role in International trade.

                                                                                           Total No of Hrs: 45

TEXTBOOKS:

1. Francis Cherunilam : Business Environment

2. K. Ashwathapa : Business Environment

3. Rudra Dutt: Indian Ecnomy

4. Kuchhal S.C.: Industrial Economy of India

DSE - V

21DMDSBA41 STATISTICAL ANALYSIS & OPERATIONS RESEARCH 4   0   0   4

COURSE OBJECTIVES:

The basic aim of this course is to impart knowledge of basic statistical tools & techniques and operations research with emphasis on their application in Business decision process and Management.

UNIT-I   **12**

Statistics: Concept, Significance and Limitations. Collection of Primary and Secondary Data, Classification and Tabulation, Frequency Distributions and their graphical representation.

UNIT-II   **12**

Measures of Central Tendency: Mean, Median and Mode. Measures of Dispersion: Range, Mean Deviation, Standard Deviation and Quartile Deviation. Moments, Measures of Skewness and Kurtosis.

UNIT-III   **12**

Sampling : Methods of Sampling ; Sampling and Non-Sampling Errors ; Law of Large Numbers and Central Limit Theorem (without proof) .Estimation , Point & Interval Estimates , Confidence Intervals .Statistical Testing - Hypothesis and Errors ; Large and Small One Sample and Two sample Tests - Z test, t -test and F-test. Chi-Square as a test of Independence and as a test of Goodness of Fit, Analysis of Variance.

UNIT-IV   **12**

Linear Programming: Problem formulation, Graphic Method, Simplex Method (including Big M Method), Duality, Transportation and Assignment problem.

UNIT-V   **12**

Decision Theory & Games Theory: Decision making under uncertainty - Laplace principle, Maximin or Minimax principle, Maximax or minimin principle, Hurwicz principle and Savage principle, Decisions under risk; Decision Tree Approach and its applications. Two person zero - sum game, Pure Strategy and Mixed Strategy

Total No of Hrs: 60

TEXTBOOKS

1. Levin , R.I. & Rubin , D.S.: Statistics for Management(PHI)

2. Gupta , S.P. & Gupta , M.P. : Business Statistics

3. Levin , R.I. : Quantitative Techniques

4. Sharma , J.K : Operations Research : Theory and Applications

21DMDSBA41   DATA VISUALIZATION   4   0   0   4

COURSE OBJECTIVES:

- To understand how to accurately represent voluminous complex data set in web and from other data sources.
- To understand the methodologies used to visualize large data sets.
- To understand the various process involved in data visualization.
- To get used to with using interactive data visualization.
- To understand the different security aspects involved in data visualization.

UNIT I INTRODUCTION   12

Context of data visualization – Definition, Methodology, Visualization design objectives. Key Factors – Purpose, visualization function and tone, visualization design options – Data Representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools.

UNIT II VISUALIZING DATA METHODS   **12**

Mapping - Time series - Connections and correlations – Indicator-Area chart-Pivot table- Scatter charts, Scatter maps - Tree maps, Space filling and non-space filling methods-Hierarchies and Recursion - Networks and Graphs-Displaying Arbitrary Graphs-node link graph-Matrix representation for graphs- Info graphics

UNIT III VISUALIZING DATA PROCESS   **12**

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

UNIT IV INTERACTIVE DATA VISUALIZATION   **12**

Drawing with data – Scales – Axes – Updates, Transition and Motion – Interactivity - Layouts Geomapping – Exporting, Framework – T3, .js, tablo.

UNIT V SECURITY DATA VISUALIZATION 12

Port scan visualization - Vulnerability assessment and exploitation - Firewall log visualization
Intrusion detection log visualization -Attacking and defending visualization systems – Creating
security visualization system.

Total No of Hrs: 60

COURSE OUTCOMES:

CO1: Upon completion of the course, the students will be able to

CO2: Understand the representation of complex and voluminous data.

CO3: Design and use various methodologies present in data visualization.

CO4: Understand the various process and tools used for data visualization.

CO5: Use interactive data visualization to make inferences.

CO6: Discuss the process involved and security issues present in data visualization.

REFERENCES:

1. Scott Murray, "Interactive data visualizationfor the web", O"Reilly Media, Inc., 2013.

2. Ben Fry, "Visualizing Data", O"Reilly Media, Inc., 2007.

3. Greg Conti, "Security Data Visualization: Graphical Techniques for Network Analysis",
NoStarch Press Inc, 2007.

21DMDSBA42   **PREDICTVE ANALYSIS**   4   0   0   4

COURSE OBJECTIVES:

- To explain terminology, technology and applications of predictive analysis
- To apply data preparation techniques and generate appropriate association rules.
- To discuss various descriptive models, their merits, demerits and application
- To describe various predictive modelling methods.
- To introduce the text mining tools, technologies and case study which is used in day-to-day analytics cycle

UNIT I INTRODUCTION TO PREDICTIVE ANALYTICS   12

Overview of Predictive Analytics- Setting up the Problem - Data Understanding- Single Variable- Data Visualization in One Dimension- Data Visualization, Two or Higher Dimensions- The Value of Statistical Significance- Pulling It All Together into a Data Audit.

UNIT II DATA PREPARATION AND ASSOCIATION RULES   12

Data Preparation- Variable Cleaning- Feature Creation- Item sets and Association Rules- Terminology- Parameter Settings- How the Data Is Organized- Measures of Interesting Rules- Deploying Association Rules- Problems with Association Rules- Building Classification Rules from Association Rules.

UNIT III MODELLING   12

Descriptive Modeling- Data Preparation Issues with Descriptive Modeling- Principal Component Analysis- Clustering Algorithms- Interpreting Descriptive Models- Standard Cluster Model Interpretation

UNIT IV PREDICTIVE MODELLING   12

Decision Trees- Logistic Regression -Neural Network Model – K-Nearest Neighbours – Naive Bayes – Regression Models - Linear Regression - Other Regression Algorithms.

UNIT V TEXT MINING   12

Motivation for Text Mining- A Predictive Modeling Approach to Text Mining- Structured vs. Unstructured Data- Why Text Mining Is Hard- Data Preparation Steps- Text Mining Features- Modeling with Text Mining Features- Regular Expressions- Case Studies:- Survey Analysis.

Total No of Hrs: 60

COURSE OUTCOMES:

At the end of this course, the students will be able to:

CO1: Explain terminology, technology and applications of predictive analysis

CO2: Apply data preparation techniques to effectively interpret big data

CO3: Discuss various descriptive models, their merits, demerits and application.

CO4: Describe principles of predictive analytics and apply them to achieve real, pragmatic solutions.

CO5: Illustrate the features and applications of text mining.

TEXTBOOKS:

1. Dean Abbott, "Applied Predictive Analytics-Principles and Techniques for the Professional Data Analyst", Wiley, 2014

2. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Third Edition, Elsevier, 2012.

3. Conrad Carlberg, "Predictive Analytics: Microsoft Excel", 1st Edition, Que Publishing, 2012.

4. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning with Applications in R Springer 2013

5. Alberto Cordoba, "Understanding the Predictive Analytics Lifecycle", Wiley, 2014

6. Anasse Bari, Mohammad Chaouchi, Tommy Jung, Predictive Analytics for Dummies, 2nd Edition, 2017

 21DMDSBA42   **DEEP LEARNING**   4   0   0   4

COURSE OBJECTIVES

To understand the major technology trends driving Deep Learning and apply fully connected deep neural networks. This course is used to analyze the key parameters and hyper parameters in a neural network's architecture.

UNIT I PROBABILITY AND INFORMATION THEORY   **12**

Random Variables- Probability Distributions- Marginal Probability- Conditional Probability- Expectation- Variance and Covariance- Bayes' Rule-Information Theory - Numerical Computation- Overflow and Underflow- Gradient-Based Optimization- Constrained Optimization- Linear Least Squares.

UNIT II MACHINE LEARNING BASICS   **12**

Machine Learning Basics and under fitting, Hyper parameters and Validation Sets Estimators- Bayesian Statistics- Supervised and Unsupervised Learning-Stochastic Gradient Descent- Challenges Motivating Deep Learning. Deep Feed forward Networks: Learning XOR- Gradient- Based Learning- Hidden Units-Architecture Design- Back- Propagation and other Differentiation Algorithms.

UNIT III   REGULARIZATION FOR DEEP LEARNING   **12**

Regularization for Deep Learning: Parameter Norm Penalties- Norm Penalties as Constrained Optimization- Regularization and Under-Constrained Problems- Dataset Augmentation- Noise Robustness- Semi-Supervised Learning- Multi-Task Learning- Optimization for Training Deep Models: Pure Optimization- Challenges in Neural Network Optimization- Basic Algorithms- Algorithms with Adaptive Learning Rates- Optimization Strategies and Meta-Algorithms.

UNIT IV   CONVOLUTIONAL NETWORKS   **12**

Convolutional Networks: The Convolution Operation, Pooling- Convolution- Basic Convolution Functions -Structured Outputs, Data Types -Efficient Convolution Algorithms- Random or Unsupervised Features -Basis for Convolutional Networks.

 UNIT V SEQUENCE MODELLING   **12**

Sequence Modeling: Recurrent and Recursive Nets- Unfolding Computational Graphs-

Recurrent Neural Networks- Bidirectional RNNs-Deep Recurrent Networks - Recursive Neural Networks- Echo State Networks- LSTM –Gated RNNs- Optimization for Long- Term Dependencies.

Total No. of Hrs: 60 hours

COURSE OUTCOMES:

At the End of this course, the Student will be able to:

CO-1: Analyze mathematical foundation of neural network.

CO-2: Apply Efficient Convolution Algorithms.

CO-3: Understand the Concept of Convolutional Networks.

CO-4: Understand about Information theory.

CO-5: Understand Supervised and Unsupervised Learning.

TEXT BOOKS:
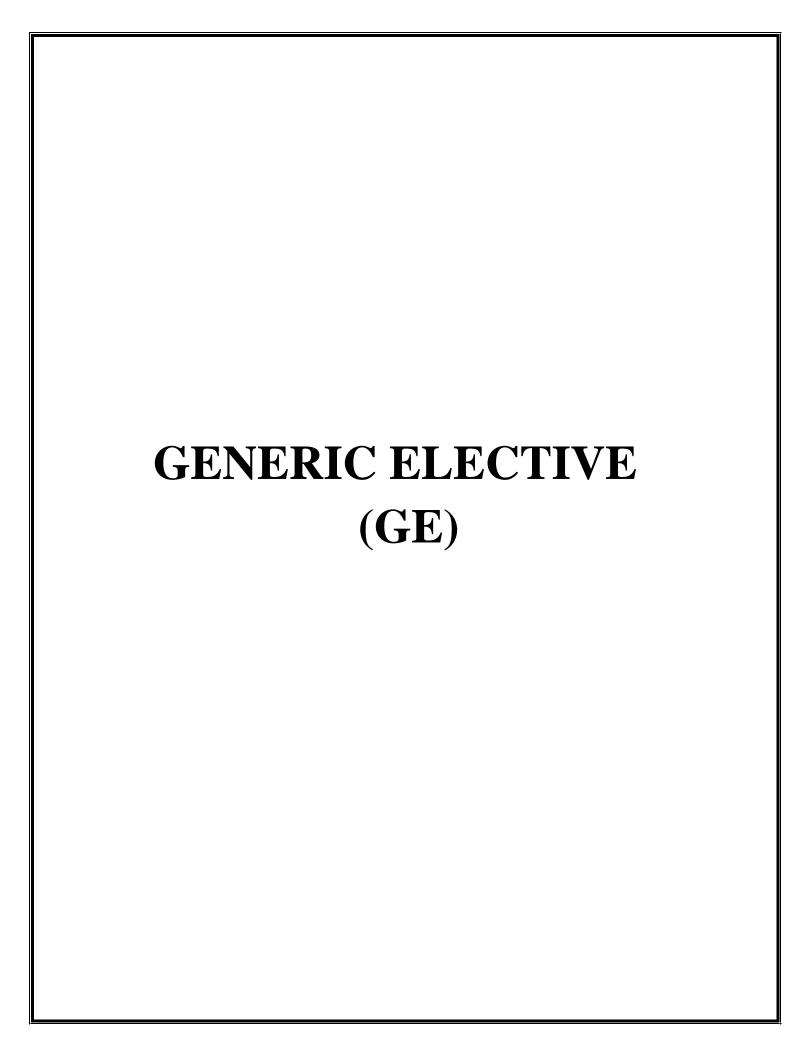
1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT

Press,2016.

2. Josh Patterson and Adam Gibson, "Deep learning: A practitioner's approach",

O'Reilly Media, First Edition, 2017.

REFERENCE BOOKS:

1. Nikhil Buduma, O'Reilly, Fundamentals of Deep Learning, Designing next-

generation machine intelligence algorithms, Shroff Publishers, 2019.

2. Douwe Osinga, O'Reilly,Deep learning Cook Book, Practical recipes to get started

Quickly, Shroff Publishers, 2019.

WEB SOURCES:

1. https://keras.io/datasets/

2. http://deeplearning.net/tutorial/deeplearning.pdf

3. https://arxiv.org/pdf/1404.7828v4.pdf

GENERIC ELECTIVE (GE)

| GE | ERP SYSTEMS | 2 | 0 | 0 | 2 |

COURSE OBJECTIVE

- To know the basics of ERP
- To understand the key implementation issues of ERP
- To know the business modules of ERP
- To be aware of some popular products in the area of ERP
- To appreciate the current and future trends in ERP

UNIT I   ERP Introduction   10

ERP Introduction, Benefits, Origin, Evolution and Structure: Conceptual Model of ERP, the Evolution of ERP, the Structure of ERP.

UNIT II   Business Process Reengineering   10

Business Process Reengineering, Data ware Housing, Data Mining, Online Analytic Processing(OLAP), Product Life Cycle Management(PLM),LAP, Supply chain Management.

UNIT III   ERP Marketplace and Marketplace Dynamics   10

ERP Marketplace and Marketplace Dynamics: Market Overview, Marketplace Dynamics, the Changing ERP Market. ERP- Functional Modules: Introduction, Functional Modules of ERP Software, Integration of ERP, Supply chain and Customer Relationship Applications.

Total No. of Hrs: 30

COURSE OUTCOME:

At the end of this course the students will be able to,

CO1: Describe basic concepts of ERP systems for manufacturing or service companies.

CO2: Analyze the strategic options for ERP identification and adoption.

CO3: Develop skills necessary for building and managing relationships with customers, and stakeholders.

CO4: Design the ERP implementation strategies.

CO5: Create reengineered business processes for successful ERP implementation.

TEXT BOOK:

1. Jagan Nathan Vaman, ERP in Practice, Tata McGraw-Hill, 2008

REFERENCE BOOKS:

1. Alexis Leon, Enterprise Resource Planning, second edition, Tata McGraw-Hill, 2008.

2. Mahadeo Jaiswal and Ganesh Vanapalli, ERP Macmillan India, 2006.

3. Vinod Kumar Grag and N.K. Venkitakrishnan, ERP- Concepts and Practice, Prentice Hall of India, 2006.

4. Summer, ERP, Pearson Education, 2008.

WEBSITES

1. www.w3schools.com

2. www.tutorialspoint.com

GE   **ETHICAL HACKING**   2   0   0   2

COURSE OBJECTIVE:

- To help students understand how ethical hacking is used as a method to prevent hacking.
- To make it possible for students to learn the process of identifying vulnerabilities and exploits of the technological ecosystem comprising of various hardware, software, network, OS and applications and identify suitable countermeasures.
- To facilitate students, appreciate the need for understanding non-technology aspects of ethical hacking such as legal frameworks, documentation and report writing.

UNIT   I   INTRODUCTION TO ETHICAL HACKING   6

Hacking Methodology, Process of Malicious Hacking, and Foot printing and scanning: Foot printing, scanning. Enumeration: Enumeration.

UNIT II   TYES OF HACKING   6

System Hacking and Trojans: System Hacking, Trojans and Black Box Vs. White Box Techniques.

UNIT   III   HACKING METHODOLOGY   6

Denial of Service, Sniffers, Session Hijacking and Hacking Web Servers: Session Hijacking, Hacking Web Servers.

UNIT IV   WEB APPLICATION   6

Web Application Vulnerabilities and Web Techniques Based Password Cracking: Web Application Vulnerabilities, Web Based Password Cracking Techniques.

UNIT   III   WEB AND NETWORK HACKING   6

SQL Injection, Hacking Wireless Networking, Viruses, Worms and Physical Security: Viruses and Worms, Physical Security. Linux Hacking: Linux Hacking. Evading IDS and Firewalls: Evading IDS and Firewalls.

                                                      Total No of Hours: 30

COURSE OUTCOME
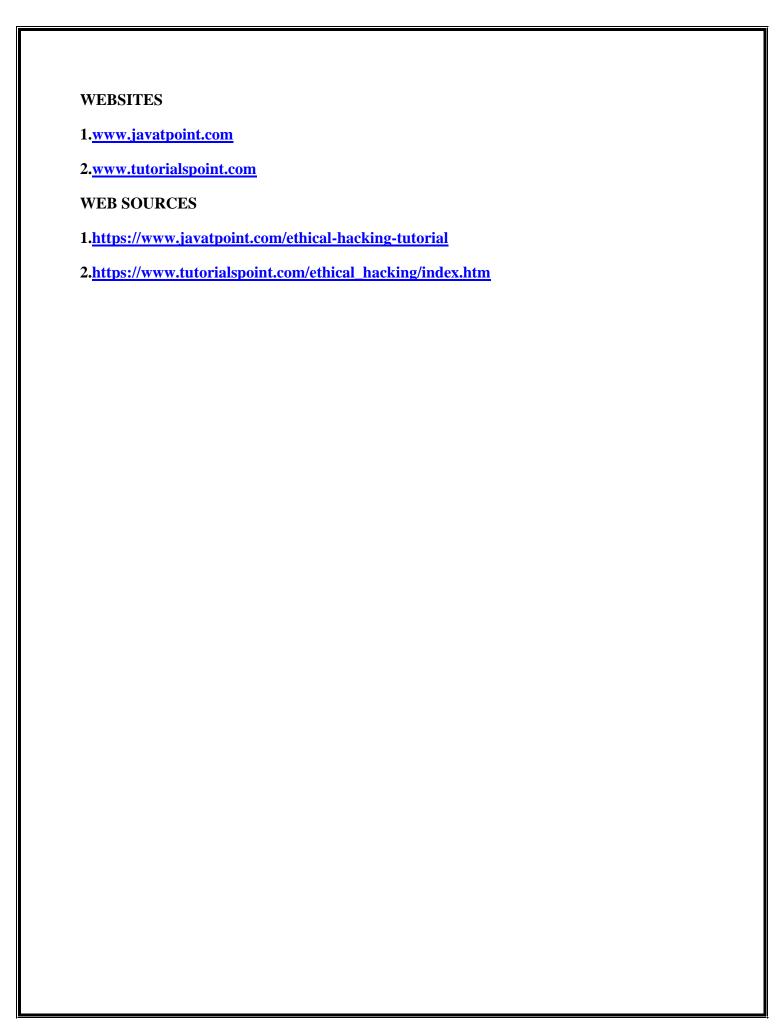
On Completion of this course, students can able to

CO1: Justify the need for meticulous documentation in writing reports for consumption of both technical and management audiences

CO2: Differentiate the processes of vulnerability assessment and ethical hacking from penetration testing.

CO3: Comprehend the importance of appropriate countermeasures for managing vulnerabilities.

CO4: Explain the importance of ethical hacking in achieving the goals of information security.

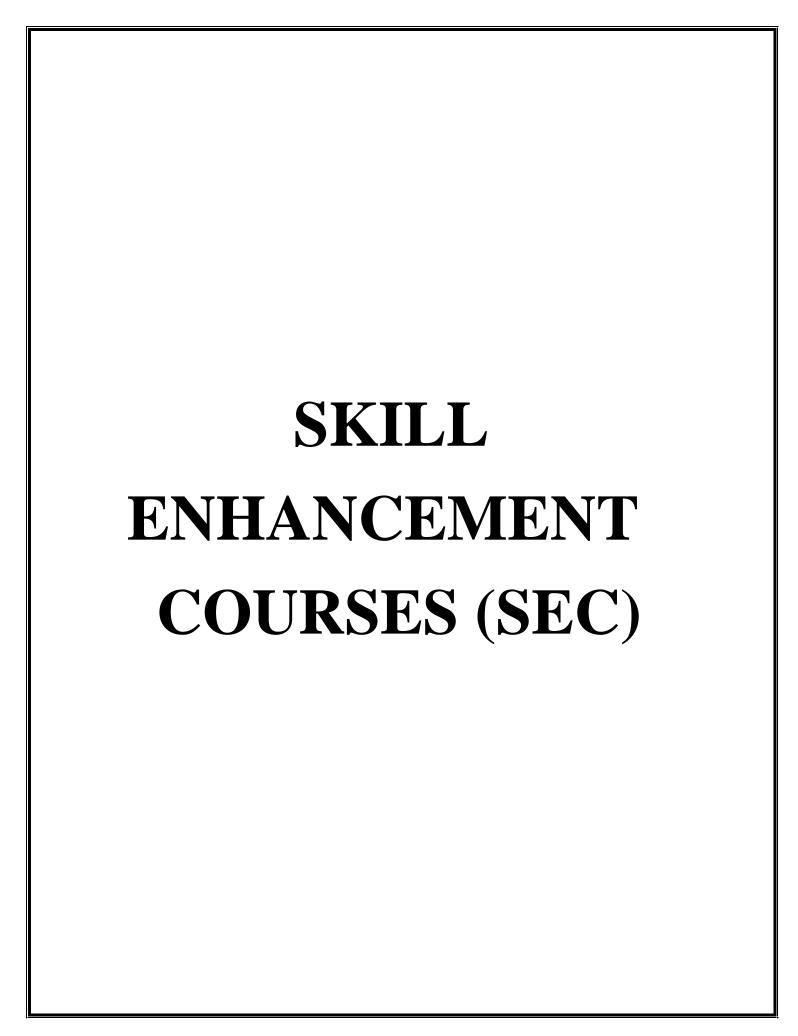CO5: Articulate the rationale for having an adequate legal framework for dealing with hacking and ethical hacking.

TEXT BOOKS

1. Gray Hat Hacking The Ethical Hackers Handbook, 3rd Edition Paperback – 1 Jul 2017 by Allen Harper, Shon Harris, Jonathan Ness, Chris Eagle, McGraw Hill Education; 3 ed (1 July 2017)

2. CEH v9: Certified Ethical Hacker Version 9 Study Guide by Sean-Philip Oriyano, Sybex; Stg edition (17 June 2016)

3. Hacking for Beginners: Ultimate 7 Hour Hacking Course for Beginners. Learn Wireless Hacking, Basic Security, Penetration Testing by Anthony Reynolds, CreateSpace Independent Publishing Platform (10 April 2017)

4. An Ethical Guide To WI-FI Hacking and Security by Swaroop Yermalkar, BecomeShakespeare.com; First edition (15 August 2014)

5. Hands-On Ethical Hacking and Network Defense by Michael T. Simpson | Kent Backman | James Corley, Cengage India 1st edition (2016)

REFERENCE BOOKS

1. The Basics of Hacking and Penetration Testing: Ethical Hacking and Penetration Testing Made Easy by Patrick Engebretson, Syngress; 2 edition (12 September 2013)

2. Hacking With Python: The Complete Guide to Ethical Hacking, Basic Security, Botnet Attack, Python hacking and Penetration Testing Kindle Edition by John C. Smalls

WEBSITES

1.www.javatpoint.com

2.www.tutorialspoint.com

WEB SOURCES

1.https://www.javatpoint.com/ethical-hacking-tutorial

2.https://www.tutorialspoint.com/ethical_hacking/index.htm

SKILL ENHANCEMENT COURSES (SEC)

<div align="center">**SOFTSKILLS – I**   2   0   0   2</div>

<u>**Course Objective:**</u>

- To enable participants Business Communication Skills
- To enhance participants E-mail writing skills
- To impart Leadership and Team Bonding skills

Credit Hours

1. READING COMPREHENSION AND VOCABULARY   **06**

Filling the blanks – Cloze Exercise – Vocabulary building – Reading and answering Questions.

2. LISTENING AND ANSWERING QUESTIONS.   **06**

Listening and writing – Listening and sequencing sentences – Filling in the blanks – Listening and answering questions.

3. GROUP DISCUSSIONS   **06**

Why GD part of a selection process – Structure of a GD – strategies in GD – Team Work – Body Language

4. CONVERSATION.   **06**

Face to face Conversation and Telephone conversation.

5. SELF- INTRODUCTION AND ROLE PLAY   **06**

Total No. of Hrs: 30

<u>**Course Outcome**</u>

At the end of this course the students will be able to,

CO1: Prioritize power of understanding and aids assimilation of vocables. Vocabulary to charge communication with educated words
CO2: Develop comprehensive knowledge through listening leading to answering questions
CO3: Build observation power and infuse self-confidence through group discussions
CO4: Identify methodology for befitting constructional ability
CO5: Experiments with inward looking and visualization of the 'otherness' of situations

<u>**Books Recommended**</u>

- Barun K. Mitra. Personality Development and Soft Skills. Oxford University Press. New Delhi.2011.
- S.P. Sharma. Personality Development. PustaqMahal. New Delhi. 2010. Meenakshi

Raman and Sangeetha Sharma. Technical Communication. Oxford University Press. New Delhi. 2009.

- Tiko, Champa& Jaya Sasikumar. Writing with a Purpose.OUP. New Delhi. 1979

Web Source:

- https://www.skillsyouneed.com/ips/communication-skills.html
- https://blog.smarp.com/top-5-communication-skills-and-how-to-improve-them
- https://blog.hubspot.com/service/phone-etiquette

<div align="center">**SOFT SKILLS – II**   2   0   0   2</div>

Course Objective:

- To enable students to develop their communication skills effectively
- To enhance students Reading, Writing, Listening and Speaking skills
- To develop their self-confidence through communication

Credit Hours

1. PRESENTATION SKILLS   06

Elements of an effective presentation – structure of presentation – voice modulation – Audience analysis – Body language

2. SOFT SKILLS   06

Time Management – Articulateness – Assertiveness – Stress management

3. RESUME / REPORT PREPARATION / LETTER WRITING   06

Structuring the resume / Report – Business letters – E-Mail Communication

4. INTERVIEW SKILLS   06

Kinds of Interviews – Required by Skills – Corporate Culture – Mock Interviews

5. 30 FREQUENTLY ASKED QUESTIONS   06

Total No. of Hrs: 30

Course Outcome

At the end of this course the students will be able to,

CO1: Illustrate the essential of presentation skills, thoughts, structure, voice modulation, audience analysis and body language

CO2: Utilize the psychological skills pertaining to time management, articulation, assertion and stress management

CO3: Construct methodology for preparation of resume, reports, business letters and email communication

CO4: Appraise learners with varied skills needed for expose to interviews

CO5: Categorize the nature of questions asked usually in interviews

Books Recommended

- BarunK.Mitra. Personality Development and soft skills. Oxford University Press. New Delhi. 2011.
- S P Sharma. Personality Development. PustaqMahal. New Delhi. 2010.

- Meenakshi Raman and Sangeetha Sharma. Technical Communication. Oxford University Press. New Delhi. 2009.

Web Sources:

- https://www.skillsyouneed.com/ips/communication-skills.html
- https://www.businessnewsdaily.com/5836-top-interviewing-skills.html
- https://gdpi.hitbullseye.com/Group-Discussion.php

SOFT SKILLS III   **2**   **0**   **0**   **2**

Course Objective:

- To enable students to develop their soft skills and Body Language
- To enhance students Reading, Writing, Listening and Speaking skills
- To develop their self-confidence to excel at Interviews

Credit Hours

UNIT–I   **06**

Powerful Presentation

UNIT–II   **06**

Reinforcement

UNIT–III   **06**

Using visual aids

UNIT–IV   **06**

Types and Methods of Presentations

UNIT–V   **06**

Obstacles to Presentation

Total 30 Hours

Course Outcome:

CO1: To develop participant's social and professional skills

CO2: To help participants manage time effectively

CO3: To build a strong resume to suit corporate requirements

CO4: To face interviews confidently

CO5: To enhance their aptitude abilities

Books Recommended:

- Roz Townsend: Presentation Skills for the Upwardly Mobile, Emerald, Chennai.
- Prasad, H. M. How to Prepare for Group Discussion and Interview. NewDelhi: Tata McGraw-Hill Publishing Company Limited, 2001.
- Pease, Allan. Body Language. Delhi: Sudha Publications, 1998.

Web Sources:

- https://www.skillsyouneed.com/ips/communication-skills.html
- https://venngage.com/blog/presentation-skills/
- https://gdpi.hitbullseye.com/Group-Discussion.php