

INSTITUTE OF SCIENCE, TECHNOLOGY & ADVANCED STUDIES (VISTAS)

(Deemed to be University Estd. u/s 3 of the UGC Act, 1956)

PALLAVARAM, THALAMBUR, PERIYAPALAYAM-CHENNAI

ACCREDITED BY NAAC WITH 'A++' GRADE

# POSTGRADUATE DEGREE PROGRAMME

## M.Sc. DATA SCIENCE & BUSINESS ANALYTICS

**(1 Year for PG Diploma / 2 Years for PG Degree)**

## CURRICULUM & SYLLABUS

**REGULATION 2024**

**Choice Based Credit System (CBCS)**

**&**

**Learning Outcomes Based Curriculum Framework (LOCF)**

**Effective from the Academic Year**

**2024 -2025**

## Department of Information Technology

## School of Computing Sciences




## DEPARTMENT OF INFORMATION TECHNOLOGY

| VISION OF THE DEPARTMENT |
|---|
| To produce globally competent graduates, skilled intellectual professionals and a successful entrepreneur in the field of Computer Science and Information Technology to meet the current challenges in the fast growing IT industry. |

| MISSION OF THE DEPARTMENT | |
|---|---|
| **M1** | To provide a strong theoretical and practical knowledge to the students in computer science discipline. |
| **M2** | To equip our students to attain a skill set imperative to harness the opportunities present in this dynamic environment. |
| **M3** | To provide technical & skill-based quality training to the students in the field of Information Technology. |
| **M4** | To empower the Youth in rural communities with computer knowledge. |
| **M5** | To inculcate knowledge about socio-economic duties & responsibilities, human & ethical values to become an optimistic global citizen and a life-long learner. |
| **M6** | To provide environmental education to empower knowledge on ecological issues and to protect and conserve invaluable natural resources for healthy and sustainable livelihoods. |

| PROGRAMME EDUCATIONAL OUTCOMES (PEO) | |
|---|---|
| **PEO1** | Graduates are prepared to be employed in IT industries by providing expected Domain Knowledge. |
| **PEO2** | Graduates are provided with practical training, hands-on and project experience to meet the industrial needs. |
| **PEO3** | Graduates are motivated in career and entrepreneurial skill development to become global leaders. |
| **PEO4** | Graduates are trained to demonstrate creativity, develop innovative ideas and to work in teams to accomplish a common goal. |
| **PEO5** | Graduates are addressed with social issues and guided to operate problems with Solution. |

| PROGRAMME OUTCOMES (PO) | |
|---|---|
| **PO1** | **Critical Thinking:** Apply knowledge of Data Science to identify, analyses problems and to provide effective solution in the area of Computing. |
| **PO2** | **Computing Skills and Ethics:** Analyze a problem, and identify and define the Computing requirements appropriate to its solution. |
| **PO3** | **Analytical skill:** Ability to design, develop algorithms and provide software solutions to cater the industrial needs. |
| **PO4** | **Modern Tool Usage:** Use current techniques, skills, and tools necessary for computing practices. |
| **PO5** | **Employability Skills:** Inculcate skills to excel in the fields of Information Technology and its Enabled services, Government and Private sectors, Teaching and Research. |
| **PO6** | **Ethics:** Insists ethical responsibilities, human and professional values and make their contribution to the society. |
| **PO7** | **Self-Directed and Life-long Learning:** Engaged in lifelong learning to equip them to the changing environment and be prepared to take-up mastering programmes. |

| | |
|---|---|
| **PO8** | **Individual and Team Work:** Function effectively as an individual, and as a member or a leader in diverse team and multidisciplinary settings. |
| **PO9** | **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development. |
| **PO10** | **Project Management:** Demonstrate knowledge and understanding of the problem and management principles and apply these to one's own work. |
| **PO11** | **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions. |
| **PO12** | **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations |

| **PROGRAMME SPECIFIC OUTCOMES (PSO)** | |
|---|---|
| **PSO1** | Professionally trained in the areas of programming, multimedia, animation, web designing, networking and to acquire knowledge in various domain-based electives. |
| **PSO2** | Abet students to communicate effectively and to improve their competency skills to solve real time problems. |
| **PSO3** | The ability to employ modern computer languages and applications for their successful career, to create platforms to become an entrepreneur and a relish for higher studies. |

# BOARD OF STUDIES

## List of Members

## Department of Information Technology

| S. No | Name & Designation | Address | Role |
|---|---|---|---|
| 1. | Dr. P. Magesh Kumar<br>Director | School of Computing Sciences, VISTAS | Chairperson |
| 2. | Dr.S.Gopinathan,<br>Associate Professor and Head | Department of Computer Science, University of Madras, Chennai | External Member (Academic Expert) |
| 3. | Mr. John Britto Chris,<br>Senior Manager-Projects | Cognizant Technologies (P) Ltd., Chennai | External Member (Academic Expert) |
| 4. | Dr. P.Sujatha,<br>Professor and Head | Department of Information Technology, VISTAS | Internal Member |
| 5. | Dr. T. Kamalakannan,<br>Professor | Department of Information Technology, VISTAS | Internal Member |
| 6. | Dr. Y. Kalpana, Professor | Department of Information Technology, VISTAS | Internal Member |
| 7. | Dr. K.Rohini,<br>Professor | Department of Information Technology, VISTAS | Internal Member |
| 8. | Dr. B.Booba,<br>Professor | Department of Information Technology, VISTAS | Internal Member |
| 9. | Dr. R.Devi,<br>Associate Professor | Department of Information Technology, VISTAS | Internal Member |
| 10. | Dr. K. Dharmarajan<br>Associate Professor | Department of Information Technology, VISTAS | Internal Member |
| 11. | Dr.S Sathya<br>Assistant Professor | Department of Information Technology, VISTAS | Internal Member |
| 12. | Dr.A Vidhya<br>Assistant Professor | Department of Information Technology, VISTAS | Internal Member |
| 13. | Mr. PV Vishnu,<br>Chief Innovation Officer | Xerago, Digital Impact Consulting & Services, Chennai | Alumni, VISTAS |

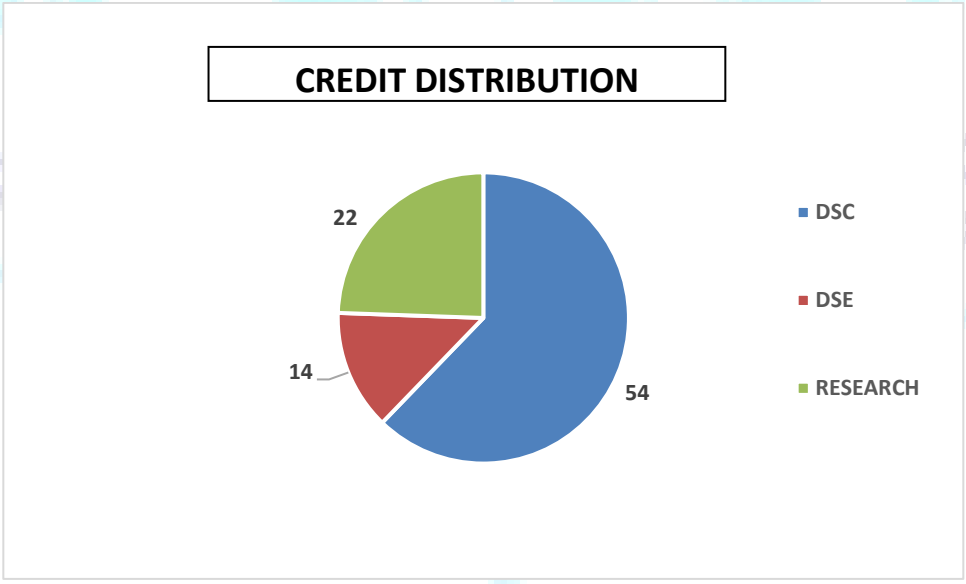
## CREDIT DISTRIBUTION

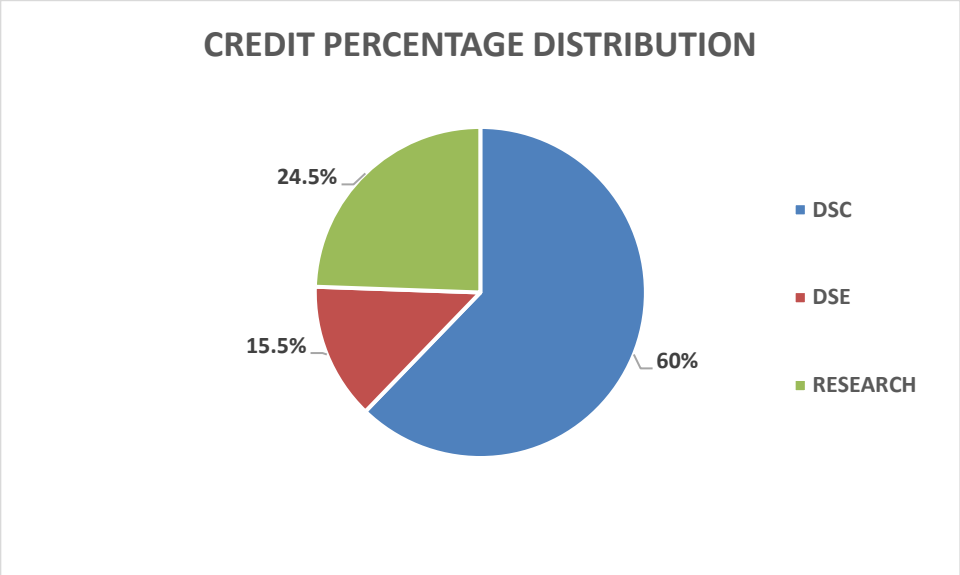
**M.Sc. Data Science & Business Analytics (PG Diploma) minimum credits to be earned : 46**

**/**

**M.Sc. Data Science & Business Analytics (PG Degree) maximum credits to be earned: 90**

| Component | I Sem | II Sem | III Sem | IV Sem | Total Credits |
|---|---|---|---|---|---|
| DSC | 16 | 16 | 16 | 6 | 54 |
| DSE | 3 | 3 | 4 | 4 | 14 |
| RESEARCH | 4 | 2 | 4 | 12 | 22 |
| Total Credits | 23 | 21 | 24 | 22 | 90 |






## ABBREVIATIONS

| | |
|---|---|
| **DSC** | Disciplinary Specific Core |
| **DSE** | Disciplinary Specific Elective |
| **RM** | Research Methodology, Internship, Mini project & Research Project |

## CURRICULUM STRUCTURE

| | M.Sc Data Science & Business Analytics (PG Diploma) minimum credits to be earned : 46<br>M.Sc Data Science & Business Analytics (PG Degree) maximum credits to be earned : 90 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SEMESTER 1 | | | | | | | | | |
| | | | Hours/Week | | | | | Maximum Marks | | |
| Category | Code | Course | L | T | P | O | C | CIA | SEE | Total |
| DSC 1 | 24CMDS11 | Programming Logic using Python | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| DSC 2 | 24CMDS12 | Advanced DBMS | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| DSC 3 | 24CMDS13 | Fundamentals of Data Science | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| DSC 1 (Lab) | 24PMDS11 | Python Programming Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| DSC 2 (Lab) | 24PMDS12 | Advanced DBMS Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| DSE 1 | 24DMDS12 | DSE 1 | 3 | 0 | 0 | 2 | 3 | 40 | 60 | 100 |
| RM 1 | 24RMDS11 | Research Methodology | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| | | | 19 | - | 8 | - | 23 | - | - | - |

CIA - Continuous Internal Assessment SEE - Semester End Examination

*L – Lecture, *T- Tutorial, *P- Practical, *O - Outside the class effort / self-study

| SEMESTER 2 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Hours/Week | | | | | Maximum Marks | | |
| **Category** | **Code** | **Course** | **L** | **T** | **P** | **O** | **C** | **CIA** | **SEE** | **Total** |
| **DSC 4** | 24CMDS21 | Machine Learning | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 5** | 24CMDS22 | Data Visualization | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 6** | 24CMDS23 | Probability and random variable | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 4 (Lab)** | 24PMDS21 | Machine Learning Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| **DSC 5 (Lab)** | 24PMDS22 | Data Visualization Lab using PowerBI | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| **DSE 2** | 24DMDS23 | DSE 2 | 3 | 0 | 0 | 2 | 3 | 40 | 60 | 100 |
| **RM 2** | 24INPG21 | Internship | 2 | 0 | 0 | 2 | 2 | | | |
| | | | **17** | **-** | **8** | **-** | **21** | **-** | **-** | **-** |

KNOWLEDGE IS POWER

| SEMESTER 3 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Hours/Week | | | | | Maximum Marks | | |
| **Category** | **Code** | **Course** | **L** | **T** | **P** | **O** | **C** | **CIA** | **SEE** | **Total** |
| **DSC 7** | 24CMDS31 | Big Data Analytics | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 8** | 24CMDS32 | Natural Language Processing | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 9** | 24CMDS33 | Applied Statistics | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 7 (Lab)** | 24PMDS31 | Big Data Analytics Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| **DSC 8 (Lab)** | 24PMDS32 | Natural Language Processing Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| **DSE 3** | 24DMDS32 | DSE 3 | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **RM 3** | 24RMDS31 | Mini Project | 0 | 0 | 4 | 2 | 4 | 40 | 60 | 100 |
| | | **TOTAL** | **16** | **-** | **12** | **-** | **24** | **-** | **-** | **-** |

| SEMESTER 4 | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Hours/Week | | | | | Maximum Marks | | |
| **Category** | **Code** | **Course** | **L** | **T** | **P** | **O** | **C** | **CIA** | **SEE** | **Total** |
| **DSC 10** | 24CMDS41 | Cloud Computing | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **DSC 10 (Lab)** | 24PMDS41 | Cloud Computing Lab | 0 | 0 | 4 | 1 | 2 | 40 | 60 | 100 |
| **DSE 4** | 24DMDS41 | DSE 4 | 4 | 0 | 0 | 2 | 4 | 40 | 60 | 100 |
| **RM 3** | 24RMDS41 | Research Project | 0 | 0 | 12 | 2 | 12 | 40 | 60 | 100 |
| | | | **8** | **-** | **16** | **-** | **22** | **-** | **-** | **-** |

## DISCIPLINE SPECIFIC CORE COURSES

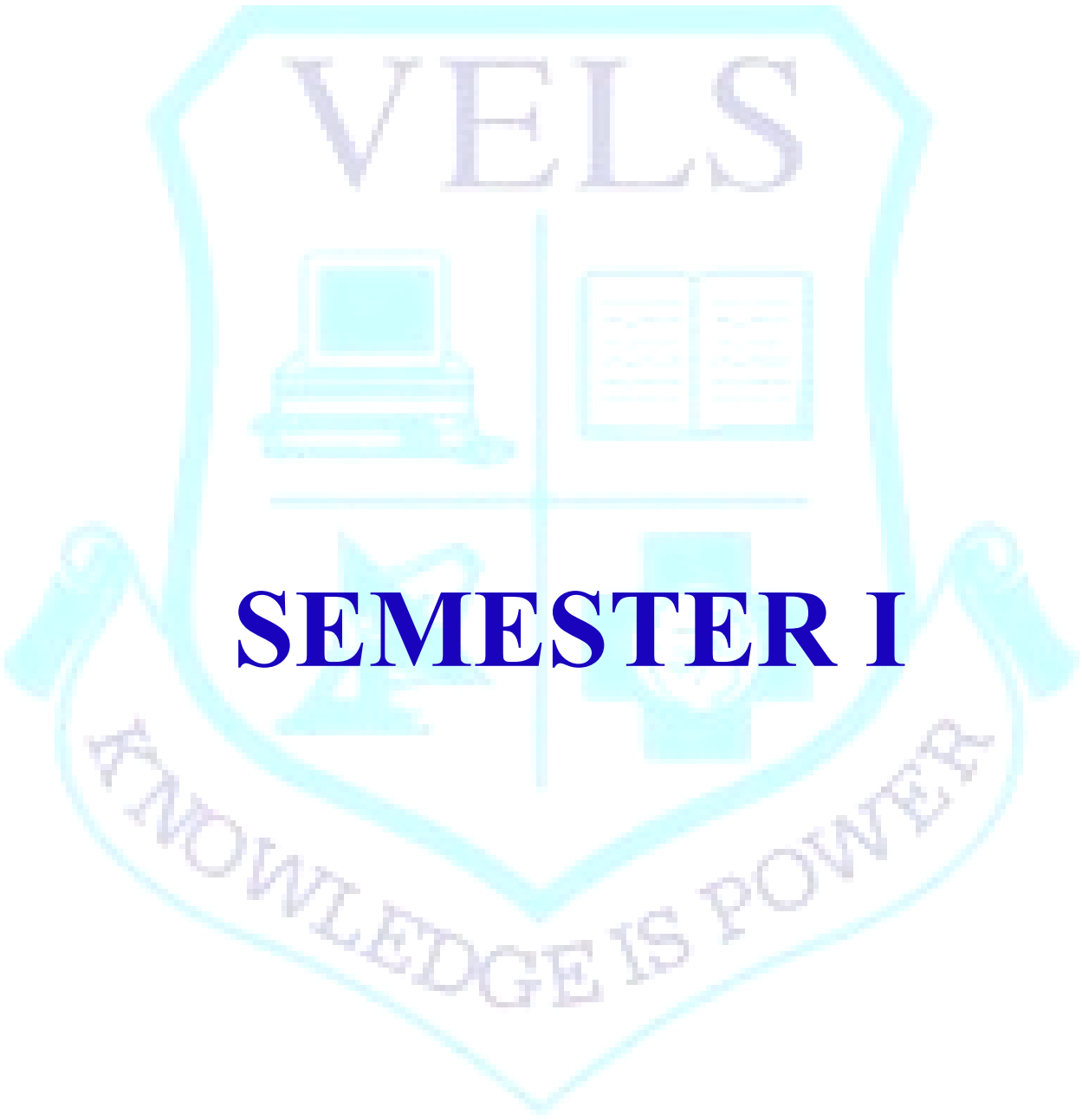
| Category | Code | Course | L | T | P | O | C |
|---|---|---|---|---|---|---|---|
| DSC 1 | 24CMDS11 | Programming Logic using Python | 4 | 0 | 0 | 2 | 4 |
| DSC 2 | 24CMDS12 | Advanced DBMS | 4 | 0 | 0 | 2 | 4 |
| DSC 3 | 24CMDS13 | Fundamentals of Data Science | 4 | 0 | 0 | 2 | 4 |
| DSC 1 (Lab) | 24PMDS11 | Python Programming Lab | 0 | 0 | 4 | 1 | 2 |
| DSC 2 (Lab) | 24PMDS12 | Advanced DBMS Lab | 0 | 0 | 4 | 1 | 2 |
| DSC 4 | 24CMDS21 | Machine Learning | 4 | 0 | 0 | 2 | 4 |
| DSC 5 | 24CMDS22 | Data Visualization | 4 | 0 | 0 | 2 | 4 |
| DSC 6 | 24CMDS23 | Probability and Random variable | 4 | 0 | 0 | 2 | 4 |
| DSC 4 (Lab) | 24PMDS21 | Machine Learning Lab | 0 | 0 | 4 | 1 | 2 |
| DSC 5 (Lab) | 24PMDS22 | Data Visualization Lab using PowerBI | 0 | 0 | 4 | 1 | 2 |
| DSC 7 | 24CMDS31 | Big Data Analytics | 4 | 0 | 0 | 2 | 4 |
| DSC 8 | 24CMDS32 | Natural Language Processing | 4 | 0 | 0 | 2 | 4 |
| DSC 9 | 24CMDS33 | Applied Statistics | 4 | 0 | 0 | 2 | 4 |
| DSC 7 (Lab) | 24PMDS31 | Big Data Analytics Lab | 0 | 0 | 4 | 1 | 2 |
| DSC 8 (Lab) | 24PMDS32 | Natural Language Processing Lab | 0 | 0 | 4 | 1 | 2 |
| DSC 10 | 24CMDS41 | Cloud Computing | 4 | 0 | 0 | 2 | 4 |
| DSC 10 (Lab) | 24PMDS41 | Cloud Computing Lab | 0 | 0 | 4 | 1 | 2 |

## DISCIPLINE SPECIFIC ELECTIVE COURSES

| Category | Code | Course code &Title | L | T | P | O | C |
|---|---|---|---|---|---|---|---|
| DSE 1 | 24DMDS11 | Reinforcement Learning for AI | 3 | 0 | 0 | 2 | 3 |
| | 24DMDS12 | Data Mining | | | | | |
| | 24DMDS13 | Introduction to Business Analytics | | | | | |
| | 24DMDS14 | Soft Computing | | | | | |
| DSE 2 | 24DMDS21 | Managerial Economics | 3 | 0 | 0 | 2 | 3 |
| | 24DMDS22 | Web Analytics | | | | | |
| | 24DMDS23 | Business Environment | | | | | |
| | 24DMDS24 | Social media and Text Analytics | | | | | |
| DSE 3 | 24DMDS31 | Marketing Management | 4 | 0 | 0 | 2 | 4 |
| | 24DMDS32 | Deep Learning | | | | | |
| | 24DMDS33 | Mathematics for Data Science | | | | | |
| | 24DMDS34 | Research and Publication Ethics | | | | | |
| DSE 4 | 24DMDS41 | Cyber Security | 4 | 0 | 0 | 2 | 4 |
| | 24DMDS42 | Predictive Analysis | | | | | |
| | 24DMDS43 | Digital Marketing Analytics | | | | | |
| | 24DMDS44 | Statistical Analysis & Operation Research | | | | | |

## RESEARCH METHODOLOGY (RM)

| Category | Code | Course | L | T | P | O | C |
|---|---|---|---|---|---|---|---|
| RM 1 | 24RMDS11 | Research Methodology | 4 | 0 | 0 | 2 | 4 |
| RM 2 | 24INPG21 | Internship | 2 | 0 | 0 | 2 | 2 |
| RM 3 | 24RMDS31 | Mini Project | 0 | 0 | 4 | 2 | 4 |
| RM 4 | 24RMDS41 | Research Project | 0 | 0 | 12 | 2 | 12 |




# SEMESTER I

## 24CMDS11 PROGRAMMING LOGIC USING PYTHON

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

### COURSE OBJECTIVES:

- To understand the NumPy, Pandas for Data Analysis
- To understand the Matplotlib for Python Plotting
- To learn the data aggregation and group operations
- To apply the Seaborn for statistical plots

### UNIT I: PARTS OF PYTHON PROGRAMMING LANGUAGE 12

Python Program Execution Procedure – Variables and Simple data types- Statements Expressions – Flow of Controls –Functions – Sequences – Strings – Tuples – Lists Dictionaries-Class – Constructors – Object Creation – Inheritance – The Polymorphism.

### UNIT II: PYTHON FOR DATA ANALYSIS – NUMPY, PANDAS 12

Introduction to Numpy - Numpy Array Indexing and Slicing - Numpy Operations-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 12

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

### UNIT IV: DATA AGGREGATION AND GROUP OPERATIONS 12

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

### UNIT V: PYTHON FOR DATA VISUALIZATION 12

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

**Total Hours: 60**

## COURSE OUTCOMES:

CO1: Able to apply the python principles in problem solving.

CO2: Analyses the numerical data using NumPy

CO3: Interpret the data using Pandas.

CO4: Implementing database using SQLite.

CO5: Access database using python programming

CO6: Translate the dataset into data Frame

CO7: Summarize the data transformations

CO8: Develop data visualization using python

## TEXT BOOKS:


1. Gowrishanker and Veena, "Introduction to Python Programming", Special Indian Edition, CRC Press, 2019.
2. Eric Matthes, "Python Crash Course", 3rd Edition, Lay flat binding, December 2022.


## REFERENCE BOOKS:


1. Leo Chin and Tanmay Dutta, "NumPy Essentials", By Kindle Edition, April 2016
2. Joel Grus, "Data Science from scratch", 2 Edition, O'Reilly, 2015.
3. Wes Mc Kinney, "Python for Data Analysis", 2 Edition, O'Reilly Media, 2012.
4. Kenneth A. Lambert, (2011), "The Fundamentals of Python: First Programs", Cengage
5. Jake Vanderplas. Python Data Science Handbook: Essential Tools for Working with Data 1st Edition.


## WEB SOURCES:

1. https://docs.python.org/3/tutorial/index.html
2. https://www.learnpython.org/

**24CMDS12** **ADVANCED DBMS**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

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 12

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

## UNIT-II RELATIONAL ALGEBRA AND CALCULUS 12

Relational algebra: introduction, Selection and projection, set operations, renaming, Joins, Division, syntax, semantics. Operators, grouping and ungrouping, relational comparison. Calculus: Tuple relational calculus, Domain relational Calculus.

## UNIT- III QUERY BASICS 12

Query Basics - Computation Using Queries - aggregate function- Subtotals and GROUP BY Command - Queries with Multiple Tables – Sub queries - Joins - DDL & DML - Testing Queries.

## UNIT-IV NORMALIZATION AND TRANSACTION MANAGEMENT 12

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. Transaction Processing, Consistency and Isolation, Atomicity and Durability, Dynamic SQL.

## UNIT-V QUERY OPTIMIZATION 12

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.

**Total Hours: 60**

## COURSE OUTCOMES:

CO1: Evaluate hierarchy of DBMS.

CO2: Analyze different types of SQL statement.

CO3: Apply relational algebra and calculus compare with relational database

CO4: Understand the transaction system could extract data efficiently.

CO5: Understand the notion of transaction and its ACID properties

CO6: Apply normalization in the database understand the internal data structure.

CO7: Analyze sub query and its implementation

CO8: Implement query optimization using SQL statement

## 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.


## WEBSITES:

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

## WEBSOURCES:

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/

**24CMDS13** **FUNDAMENTALS OF DATA SCIENCE**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

- 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 TO DATA SCIENCE 12

Introduction To Data Science: Definition, Big Data and Data Science Hype, Datafication , Data Science Profile, Meta-Definition, Data Scientist, Statistical Inference, Populations and Samples, Populations and Samples of Big Data, Big Data Can Mean Big Assumptions, Modeling, Philosophy of Exploratory Data Analysis, The Data Science Process , A Data Scientist's Role in this Process Case Study: RealDirect.

## UNIT –II MATHEMATICAL PRELIMINARIES 10

Mathematical Preliminaries: Probability, Descriptive Statistics, Correlation Analysis - Data Munging: Properties of Data, Languages for Data Science, Collecting Data, Cleaning Data, Crowdsourcing.

## UNIT – III SCORES AND RANKINGS 13

Scores and Rankings: Developing Scoring Systems, Z-scores and Normalization, Advanced Ranking Techniques Statistical Analysis: Sampling from Distributions, Statistical Distributions, Statistical Significance, Permutation Tests and P-values

## UNIT- IV VISUALIZING DATA 13

Visualizing Data: Exploratory Data Analysis, Developing a Visualization Aesthetic, Chart Types, Great Visualizations - Mathematical Models: Philosophies of Modeling, A Taxonomy of Models, Baseline Models, Evaluating Models, Evaluation Environment.

## UNIT-V SUPERVISED LEARNING 12

Supervised Learning: Linear Regression, Better Regression Models, Regression as Parameter Fitting, Simplifying Models through Regularization Classification and Logistic Regression, Issues in Logistic Classification, Naive Bayes, Decision Trees Classifiers.

**Total Hours: 60**

## COURSE OUTCOMES:

CO1: Describe the significance of data science and understand the Data Science process.

CO2: Explain how data is collected, managed and stored for data science.

CO3: Build, and prepare data for use with a variety of statistical methods and models.

CO4: Analyze Data using various Visualization techniques.

CO5: Choose contemporary models, such as machine learning, AI, techniques to solve practical problems

CO6: Evaluate Fundamentals of Data Science

CO7: Analyze different Collection Strategies and Data Pre-Processing

CO8: Apply Descriptive Statistics and Multiple Regression

## TEXT BOOKS:


1. Steven S. Skiena, "The Data Science Design Manual", Springer 2017.
2. Rachel Schutt & O'neil, "Doing Data Science", Straight Talk from The Frontline O'REILLY, ISBN:978-1-449-35865-5, 1st edition, October 2013.


## REFERENCE BOOKS


1. Joel Grus," Data Science from Scratch" First Edition, April 2015
2. Gareth James, Daniela Witten, Trevor Hatie, RoberstTibhirani , "An Introduction to Statistical Learning-with Applications in R", 2013
3. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2 edition (30 September 2014)
4. R Programming for Data Science, Roger D. Peng, LeanPub, 2015.


## WEB REFERENCES:

1. "Data science for engineers" https://nptel.ac.in/noc/courses/noc20/SEM1/noc20-

**24PMDS11 PYTHON PROGRAMMING LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

## OBJECTIVE

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

## LIST OF EXPERIMENTS

1. Programs using Flow Controls.
2. Programs using 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. Program for Object Creation and Usage
8. Program involving Inheritance
9. Program involving Polymorphism
10. Reading and Writing with Text Files and Binary Files
11. Combining and Merging Data Sets
12. Program for Selection and Filtering in Series and Data Frames
13. Data Aggregation and GroupWise Operations
14. Program for Plotting Graph

**Total Hours: 30**

## COURSE OUTCOMES

**On Completion of this course, the students can able to**

CO1: Create applications using PYTHON.

CO2: Create applications using python collections.

CO3: Analyze the utilities of object oriented programming in Python.

CO4: Analyze the numerical data using NumPy.

CO5: Perform data analysis using Pandas

CO6: Test data processing using Text files and Binary files

CO7: Test the different graph plotting

CO8: Analyze different type of data file.

**WEB SOURCES:**

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

**24PMDS12** **ADVANCED DBMS LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

## COURSE OBJECTIVES

- 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. Renaming a Table, Copying another Table
3. Working with Aggregate functions
4. Use of Joins and Sub queries.
5. Views, sequences and indexes.
6. PL/SQL-Data types, control structures
7. Creating procedures with PL/SQL.
8. PL/SQL Procedure for application using exception handling
9. PL/SQL Procedure for application using exception handling
10. Writing Programs on Packages & triggers.
11. Embedding PL/SQL in high level language.
12. Implementation of Triggers & Assertions for Bank Database.

**Total Hours: 30**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

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

CO-2: Evaluate the functionalities of trigger

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.

CO-6: Implement procedure concept using PL/SQL

CO-7: Analyze different types Aggregate functions

CO-8: Apply Sub queries concept in SQL

## WEB SOURCES:

1. www.guru99.com/dbms-tutorial.html

**DSE –I**

| L | T | P | O | C |
|---|---|---|---|---|
| 3 | 0 | 0 | 2 | 3 |

**24DMDS12** **DATA MINING**

**COURSE OBJECTIVES:**

- Learn about data mining functionalities, applications and issues.
- To learn Data mining techniques and algorithms.
- Comprehend the data mining environments

**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** **9**

Clustering Techniques: Clustering Paradigms, Partitioning Algorithms, k-Medoid Algorithms, CLARA, CLARANS, Hierarchical Clustering, DBSCAN, BIRCH, CURE, Categorical Clustering Algorithms, STIRR, ROCK, CACTUS.

**UNIT – IV: DECISION TREES** **9**

What is a Decision Tree?, Tree Construction Principle, Best Split, Splitting Indices, Splitting Criteria, Decision Tree Construction Algorithms, CART, ID3, C4.5, Decision Tree Construction with Presorting, RainForest, Approximate Methods, CLOUDS, BOAT, Pruning Techniques, Integration of Pruning and Constructio Ideal Algorithm

**UNIT – V: OTHER TECHNIQUES** **9**

What is a Neural Network?, Learning in NN, Unsupervised Leaning, Data Mining Using NN: A Case Study, Genetic Algorithms, Rough Sets, Support Vector Machines Web Mining: Web Mining, Web Content Mining, Web Structure Mining, Web Usage Mining, Text Mining, Unstructured Text, Episode Rule Discovery for Texts, Hierarchy of Categories, Text Clustering

**Total Hours: 45**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

CO1: Can able to understand steps involved in data mining.

CO2: Analyze different classification & clustering techniques.

CO3: Understand the basic functionalities & issues of clustering techniques

CO4: Understand the basic components & benefits of decision tree.

CO5: Understand the concept of neural networks

CO6: Able to get employability in data analysis.

CO7: Summarize Web mining techniques

CO8: Recall the Text mining techniques

## TEXT BOOKS:

1. Data Mining Techniques, Arun K Pujari, University Press
2. Data Mining: Concepts and Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei

## REFERENCE BOOKS:

1. .Jiawei Han and Micheline Kamber, "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

**24RMDS11** **RESEARCH METHODOLOGY**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

**OBJECTIVES:**

To familiarize the students to the principles of scientific methodology in business enquiry, to develop analytical skills of business research, and to develop the skills for scientific communications.

**UNIT I: INTRODUCTION: 12**

Meaning, objectives and motivations in research, Characteristics and limitations of research – Components of research work - Criteria of good research, Research process – Types of Research, Fundamental, Pure or Theoretical Research –Applied Research –Descriptive Research – Evaluation Research –Experimental Research – Survey Research – Qualitative Research – Quantitative Research – Historical Research.

**UNIT II: RESEARCH DESIGN & LITERATURE REVIEW: 12**

Research Design – definition – essentials and types of research design – errors and types of errors in research design. Research problem: Selecting and analyzing the research problem – problem statement formulation – formulation of hypothesis. Literature review: purpose, sources, and importance - literature review procedure. Objectives: Learning Objectives; Definitions; Formulation of the research objectives.

**UNIT III: MEASUREMENT. SCALING AND SAMPLING: 12**

Variables in Research – Measurement and scaling – Different scales – Construction of instrument – Validity and Reliability of instrument. Data Collection methods – primary and secondary data – Construction of questionnaire and instrument – validation of instruments. Sample size determination - Sample design and sampling techniques.

**UNIT IV: DATA ANALYSIS AND TOOLS: 12**

Processing of Data: Editing of Data – Coding of Data – Classification of Data – Statistical Series. Qualitative vs Quantitative data analyses – Univariate, Bivariate and Multivariate statistical techniques Measures of Central Tendency, Dispersion, correlation and Regression, Chi-square test: Applications, Steps, characteristics, limitations, Analysis of Variance and Covariance, Factor analysis – Discriminant analysis – cluster analysis – multiple regression and correlation – multidimensional scaling – Conjoint Analysis - Application of statistical software for data analysis.

**UNIT V: RESEARCH REPORT WRITING: 12**

Research report – Different types – Contents of report –executive summary – chapterization – contents of chapter – report writing – the role of audience – readability – comprehension – tone – final proof – report format – title of the report – Ethical issues in research: Code of Ethics in Research – Ethics and Research

-Process – Importance of Ethics in Research.

**Total Hours: 60**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

1. Demonstrate knowledge of research processes (reading, evaluating, and developing);
2. Perform literature reviews using print and online databases;
3. Employ American Psychological Association (APA) formats for citations of print and electronic materials;
4. Identify, explain, compare, and prepare the key elements of a research proposal/report;
5. Define and develop a possible HIED research interest area using specific research designs;
6. Compare and contrast quantitative and qualitative research paradigms, and explain the use of each in HIED research;
7. Describe sampling methods, measurement scales and instruments, and appropriate uses of each;
8. Analyze various research writing Techniques

## TEXT BOOK:

1. Kothari, C.R., 2004. Research methodology: Methods and techniques. New Age International.

2. Cooper, D.R., Schindler, P.S. and Sun, J., 2006. Business research methods (Vol. 9). New York: McGraw-Hill Irwin.

3. Creswell, J.W. and Creswell, J.D., 2017. Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

## REFERENCE BOOKS:

1) Business Research Methods- Donald Cooper & Pamela Schindler, TMGH, 9th editions.

2) Business Research Methods- Alan Bryman & Emma Bell, Oxford University Press.

## WEBSITE:

1.https://www.geeksforgeeks.org/introduction-to-research-methodology/

# SEMESTER 2

**24CMDS21** **MACHINE LEARNING**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

- To introduce the fundamental concepts of machine learning and its applications.
- To learn the classification, clustering and regression based machine learning algorithms.
- To understand the deep learning architectures.
- To understand the methods of solving real life problems using the machine learning techniques.
- To understand the multiple learners, boosting and stacked generalization.
- To provide theoretical knowledge on setting hypothesis for pattern recognition.

## UNIT I – INTRODUCTION 12

Machine Learning Foundations – Overview – Design of a Learning System – Types of Machine Learning – Supervised Learning and Unsupervised Learning – Applications of Machine Learning – Tools Overview for ML.

## UNIT II – CLASSIFICATION ALGORITHMS 12

Perceptron and back-propagation neural network - k-nearest neighbor rule. Support vector machine: multi category generalizations – Regression Decision trees: classification and regression tree – random forest.

## UNIT III – COMPONENT ANALYSIS AND CLUSTERING ALGORITHMS 12

Principal component analysis - Linear discriminant analysis - Independent component analysis. K-means clustering - fuzzy k-means clustering Expectation-maximization algorithm-Gaussian mixture models –auto associative neural network.

## UNIT IV–SUPERVISED AND UNSUPERVISED LEARNING 12

Convolution neural network (CNN) -Layers in CNN - CNN architectures. Recurrent Neural Network -Applications: Speech-to-text conversion-image classification time series prediction.

## UNIT V – COMBINING MULTIPLE LEARNERS 12

Generating diverse learners - model combination schemes - voting - error-correcting output codes -bagging - boosting - mixture of experts revisited - stacked generalization - fine-tuning an ensemble –cascading .

**Total Hours: 60**

## COURSE OUTCOMES

**On Completion of this course, the students can able to**

1. Understand the basic concepts of Bayesian theory and normal densities.
2. Implement different classification algorithms used in machine learning.
3. Implement clustering and component analysis techniques.
4. Design and implement deep learning architectures for solving real life problems.
5. Combine the evidence from two or more models/methods for designing a system.
6. Recognize the characteristics of machine learning strategies.
7. Apply various supervised learning methods to appropriate problems.
8. Identify and integrate more than one technique to enhance the performance of learning.

## TEXT BOOKS


1. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
2. R. O. Duda, E. Hart, and D.G. Stork, "Pattern Classification", Second Edition, John Wiley & Sons, Singapore, 2012.
3. Francois Chollet, "Deep Learning with Python", Manning Publications, Shelter Island, New York, 2 018.


## REFERENCE BOOKS


1. Ethem Alpaydin, "Introduction to Machine Learning", 3rd Edition, MIT Press, 2014.
2. C. M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
3. Navin Kumar Manaswi, "Deep Learning with Applications using Python", A press, New York, 2018.


## WEBSITES

1) www.javatpoint.com
2) www.geeksforgeeks.org
3) www.aitude.com
4) www.educative.io

## WEB SOURCES

1) https://www.javatpoint.com/basic-concepts-in-machine-learning
2) https://www.geeksforgeeks.org/machine-learning-algorithms
3) https://www.simplilearn.com/10-algorithms-machine-learning
4) https://www.aitude.com/supervised-vs-unsupervised-vs-reinforcement

**24CMDS22** **DATA VISUALISATION**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES:

To understand the concepts of different types of visualization and how humans perceive information. Learn the processes, mechanics in implementing principles of design and color to make effective visualizations, learn how to visualize graphs, be exposed to visualization dashboards, and to apply techniques from user-interface design to create an effective visualization system.

## UNIT I: THE COMPUTER AND THE HUMAN 12

Defining data visualization - Visualization workflow-describing data visualization- Work flow process in practice

## UNIT II: VISUALIZATION OF NUMERICAL DATA 12

Data Representation: chart types: categorical, hierarchical, relational, temporal & spatial; 2-D: bar charts, Clustered bar charts, dot plots, connected dot plots, pictograms, proportional shape charts, bubble charts, radar charts, polar charts, Range chart.

## UNIT III: VISUALIZATION OF NON-NUMERICAL DATA 12

Range chart, Box-and-whisker plots, univariate scatter plots, histograms word cloud, pie chart, waffle chart, stacked bar chart, back-to-back bar chart.

## UNIT IV: VISUALIZATION DASHBOARD 12

Visualization system for large datasets and dashboards, Tools for dashboard, graph data visualization; Annotation.

## UNIT V: 2-D & 3-D DATA VISUALISATION 12

Relevant 2-D charts. 3-D: surfaces, contours, hidden surfaces, 3d coloring, 3D mapping: multi dimensional data visualization.

**Total Hours: 60**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

CO1: Discuss the concepts of what data visualization is, how it's used, and how Computers display information.

CO2: Implement principles of design and color to make data visualizations more Effective

CO3: Learn how to visualize graphs that depict relationships between data items.

CO4: Use and designing own visualization system for large datasets and dashboards

CO5: To interpret visualization system.

CO6: Understand the representation of complex and voluminous data

CO7: Design and use various methodologies present in data visualization

CO8: Use interactive data visualization to make inferences.

**TEXT BOOKS:**

1. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016

2. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.

**REFERENCE BOOKS:**

1. Information Dashboard Design-Displaying data for at-a-glance monitoring-Stephen Few

2. Beautiful Visualisation, Looking at data through the eye experts by Julie

**WEBSITE:**

1.https://www.geeksforgeeks.org/introduction-to-data-visualisation

## 22CMDS23 PROBABILITY AND RANDOM VARIABLES

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

### COURSE OBJECTIVES:

- To provide necessary basic concepts of probability in data analysis.
- To provide the concepts of random variables in data analysis.
- To acquaint the knowledge about mathematical expectation.
- To acquaint the knowledge of discrete distributions which plays an important role in data analysis.
- To acquaint the knowledge of continuous distributions which plays an important role in data analysis.

### UNIT I: PROBABILITY 12

Introduction-Different Approaches to Probability-Finite Sample, Conditional Probability and Independent Events-Total Probability - Baye's Theorem.

### UNIT II: RANDOM VARIABLES 12

Types of Random Variables-Discrete Random Variables-Probability Mass Function- Cumulative Distribution function-properties-Continuous Random Variables-Probability Density Function Cumulative Distribution function.

### UNIT III: MATHEMATICAL EXPECTATION 12

Expectation of a discrete random variable- Expectation of a continuous random Variable Properties-Variance-Moments- Moment generating function.

### UNIT IV: DISCRETE DISTRIBUTIONS 12

Binomial Distribution-Poisson Distribution-Geometric Distribution-Properties-Moment Generating Functions-Mean and Variance-Problems

### UNIT V: CONTINUOUS DISTRIBUTIONS 12

Uniform Distribution-Exponential Distribution-Normal Distribution-Properties- Moment Generating Functions-Mean and Variance-Problems

**Total Hours: 60**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

CO1: Understand the fundamental knowledge of the concepts of probability K3

CO2: Understand the basic concepts of random variables K5

CO3: Apply the concept of moment generating function K4

CO4: Analyze the knowledge of discrete distributions K5

CO5: Utilize the knowledge of continuous distributions K4

## TEXT BOOKS:


1. Kolmogorov. A. N Foundations of the Theory of Probability, Dover Publication English Edition, 2018.
2. Sheldon Ross; A First Course in Probability, Pearson, 9th Edition, 2014.
3. Sivaramakrishna Das. P and Vijayakumari. C, Probability and Statistics, Pearson


## REFERENCE BOOKS:


1. F.M. Dekking C. Kraaikamp H.P. Lopuhaa¨ L.E. Meester, A Modern Introduction to Probability and Statistics, Springer London Ltd, 2nd printing, 2007.
2. Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, 2nd Edition, 2015.


## WEBLINKS:

1.https://www.cse.iitk.ac.in/users/piyush/courses/pml_fall17/material/probabilty_tutorial.pdf

2. https://web.stanford.edu/~oas/SI/QM/notes/ProbabilityText.pdf

3. https://onlinecourses.nptel.ac.in/noc22_ma81/preview

**24PMDS 21** **MACHINE LEARNING LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

## OBJECTIVES:

- Understand the mathematical and statistical perspectives of machine learning algorithms through python programming.
- Understand the basic concepts of deep neural network models and design the same.

## LIST OF EXERCISES:

1) Write a program to perform Data Pre-processing:
2) Data cleaning b. Data Transformation c. Data Reduction d. Feature extraction
3) Write a program to implement the Adaboost Methods
4) Write a program to implement Logistic Regression algorithm
5) Write a program to demonstrate the workflow of Decision Tree Classification
6) Write a program to implement the Random Forest Classification
7) Write a program to implement the SVM Classification
8) Write a program to perform the K Means Clustering
9) Write a program to perform the Density based Clustering
10) Write a program to implement the Apriori algorithm for market basket analysis
11) Write a program to compare the Supervised Machine Learning algorithms.

**Total Hours: 30**

## COURSE OUTCOMES:

- Design and evaluate the unsupervised models through python in built functions.
- Evaluate the machine learning model algorithms by python programming.
- Design and apply various reinforcement algorithms to solve real time complex problem
- Design and develop the code for the recommender system using Natural Language
- Implement different classification algorithms used in machine learning.
- Implement clustering and component analysis techniques.
- Combine the evidence from two or more models/methods for designing a system.
- Apply various supervised learning methods to appropriate problems.

## REFERENCE BOOK:

1. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012.

## 24PMDS22 DATA VISUALISATION LAB USING POWER BI

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

### COURSE OBJECTIVES

- To familiarize data visualization using Power BI implementations.
- To implement and designing own visualization system for large datasets and dashboards.
- To understand and interpret shapes, Textboxes and Images in Power BI..
- To understand various shapes in Power BI.
- To get understanding of problem to program mapping

### LIST OF EXERCISES:

1) Introduction to Visuals In Power BI
2) Visualization Charts in Power BI
3) Matrixes and Tables
4) Slicers and Map Visualizations
5) Gauges and Single Number Cards
6) Modifying Colors in Charts And Visuals
7) Shapes, Text Boxes, and Images
8) Custom Visuals
9) Page Layout and Formatting
10) Bookmarks and Selection Pane
11) KPI Visuals
12) Z-order

**Total Hours: 30**

### COURSE OUTCOMES:

**On Completion of this course, the students can able to**

CO1: Implement Visualization Charts program in Power BI

CO2: Write matrix and table program in Power BI

CO3: Learn how to visualize graphs, charts in Power BI

CO4: Use and designing own visualization system for large datasets and dashboards

CO5: To interpret shapes, Textboxes and Images in Power BI.

CO6: Understand the representation of complex and voluminous data

CO7: Design and use various methodologies present in data visualization

CO8: Use interactive data visualization to make inferences.

**TEXT BOOKS:**

1. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016

2. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.

**REFERENCE BOOKS:**

1. Information Dashboard Design-Displaying data for at-a-glance monitoring-Stephen Few

2. Beautiful Visualisation, Looking at data through the eye experts by Julie

**WEBSITE:**

1.https://www.geeksforgeeks.org/introduction-to-data-visualisation

DSE -II

24DMDS23 BUSINESS ENVIRONMENT

| L | T | P | O | C |
|---|---|---|---|---|
| 3 | 0 | 0 | 2 | 3 |

## COURSE OBJECTIVES:

The basic objective of this course is to familiarize the participants with various aspects of economic, social, political and cultural environment of India. This will help them in gaining a deeper understanding of the environmental factors influencing Indian business organizations.

## UNIT I: AN OVERVIEW OF BUSINESS ENVIRONMENT 9

The concept of Business Environment – Its nature and significance - Type of Environment - Internal, External, Micro and Macro Environment, Competitive structure of Industries, Environmental analysis and strategic management - Process and limitations of environmental analysis.

## UNIT II: ECONOMIC ENVIRONMENT 9

Nature of Economic Environment - Economic Factors - Basic Economic System, Economic Planning, Economic Policies- New Industrial Policy - FEMA – NITI Ayog – National Development Council – Five Year Plans - Patents and trade mark.

## UNIT-III: ECONOMIC POLICIES 9

Economic Reforms and New Economic Policy – New Industrial Policy – Competition Law – Fiscal Policy – Objectives and Limitations - Impact of foreign culture on business-Business and society - Business ethics and corporate governance - Social responsibility of business.

## UNIT-IV: SOCIAL AND ECONOMIC ENVIRONMENT 9

Economic systems and their impact of business - Concept of Social Responsibility of Business towards Stakeholders - Demonetisation, GST and their Impact - Political Stability - Legal Changes.

## UNIT-V: NATURAL AND TECHNOLOGICAL ENVIRONMENT 9

Innovation - Globalization – Trend in globalisation - Impact of technology on globalization - transfer of technology Status of technology in India - Management of technology - Features and Impact of technology.

**Total Hours: 45**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

CO1: Understand the importance of motivation in building a strong and competitive Business Organization.

CO2: Familiarize with the nature of Business Environment and its components.

CO3: Identify and evaluate the complexities of business environment and their impact on the business.

CO4: Examine how different factors and trends in the external environment are likely to impact upon a proposed business venture.

CO5: Conduct a business analysis of the local and national environment.

CO6: Employ business models and tools to evaluate changes in an organization's business environment.

CO7: Understand the work techniques of Organizations to ensure success and timely completion of tasks.

CO8: Apply an understanding of the different modes of engagement and explore the interconnectedness between these and the economic, legal, governmental, political, regulatory, social and other environments in which expanding companies operate.

## TEXT BOOKS:


1. Francis Cherunilam, Francis; Business Environment - Text and Cases, Himalaya Publishing House.
2. Namitha Gopal, Business Environment, McGraw Hill
3. A C Fernando: Business Environment, Pearson


## REFERENCE BOOKS:


1. Aswathappa, K.; Essentials of Business Environment, Himalaya Publishing House, New Delhi.
2. Dr S Sankaran : Business Environment, Margham Publications.
3. Rosy Joshi and Sangam Kapoor: Business Environment, Kalyanai Publications.
4. Dr V Murali Krishna, Business Environment, Spectrum Publications.


## WEB SITES

1) www.toppr.com
2) www.unacademy.com

**WEB SOURCES**

1. https://www.youtube.com/watch?v=REm0aqPvRb8
2. https://www.bajajfinserv.in/business-environment
3. https://unacademy.com/content/cbse-class-11/study-material/business-studies/concept-of-business-environment
4. https://www.toppr.com/guides/business-environment

DSE 2

**24DMDS22** **WEB ANALYTICS**

| L | T | P | O | C |
|---|---|---|---|---|
| 3 | 0 | 0 | 2 | 3 |

## COURSE OBJECTIVES

To Assess that how website visitors view and interact with a site's pages and features, and business intelligence, which would allow using data on customer purchasing patterns, demographics, and demanding trends to make effective strategic decisions.

## UNIT I WEB ANALYTICS BASICS 9

Definition, Process, Key terms: Site references, Keywords and Key phrases; building block terms: Visit characterization terms, Content characterization terms, Conversion metrics; Categories: Offsite web, On site web; Web analytics platform, Web analytics evolution, Need for web analytics, Advantages, Limitations.

Data Collection: Clickstream Data: Web logs, Web Beacons, JavaScript tags, Packet Sniffing; Outcomes Data: E-commerce, Lead generation, Brand/Advocacy and Support; Research data: Mindset, Organizational structure, Timing; Competitive Data: Panel-Based measurement, ISP-based measurement, Search Engine data.

## UNIT II QUALITATIVE ANALYTSIS 9

Qualitative Analysis: Heuristic evaluations: Conducting a heuristic evaluation, Benefits of heuristic evaluations; Site Visits: Conducting a site visit, Benefits of site visits; Surveys: Website surveys, Post-visit surveys, Creating and running a survey, Benefits of surveys. Web Analytic fundamentals:

Capturing data: Web logs or JavaScripts tags, Separate data serving and data capture, Type and size of data, Innovation, Integration, Selecting optimal web analytic tool, Understanding clickstream data quality, Identifying unique page definition, Using cookies, Link coding issues.

## UNIT III WEB METRICS 9

Common metrics: Hits, Page views, Visits, Unique visitors, Unique page views, Bounce, Bounce rate, Page/visit, Average time on site, New visits; Optimization(e-commerce, non e-commerce sites): Improving bounce rates, Optimizing adwords campaigns; Real time report, Audience report, Traffic source report, Custom campaigns, Content report, Google analytics, Introduction to KPI, characteristics, Need for KPI, Perspective of KPI, Uses of KPI.

## UNIT IV WEB ANALYTICS 2.0 9

Web analytics 1.0, Limitations of web analytics 1.0, Introduction to analytic 2.0, Competitive intelligence analysis : CI data sources, Toolbar data, Panel data ,ISP data, Search engine data, Hybrid data, Website traffic analysis: Comparing long term traffic trends, Analyzing competitive site overlap and opportunities.

## UNIT V GOOGLE ANALYTICS 9

Brief introduction and working, Adwords, Benchmarking, Categories of traffic: Organic traffic, Paid traffic; Google website optimizer, Implementation technology, Limitations, Performance concerns, Privacy issues.

**Total hours: 45**

## COURSE OUTCOMES:

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

**CO1** Understand the Web analytics
**CO2** Identify Different Data Collection and Web Analytics Strategies
**CO3** Apply Different Web Analytics Tools
**CO4** Summarize Various Google Analytics and testing
**CO5** Understand web analytics 2.0 and its tasks
**CO6** Implementing data using JavaScript
**CO7**Analyze the Web metrics to measure performance
**CO8** Measure the Qualitative Analytis

## TEXT BOOKS:

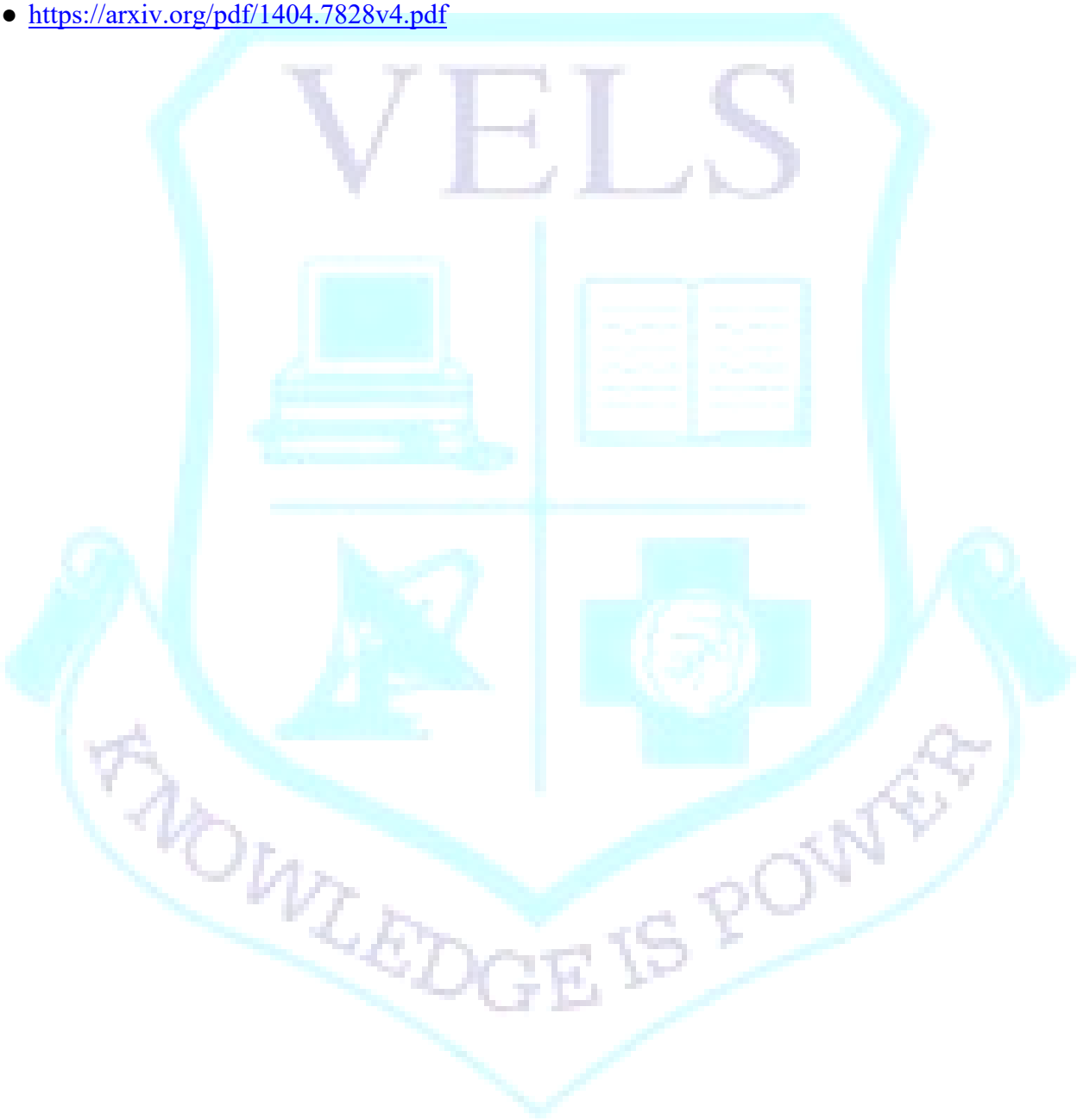
1. Clifton B., Advanced Web Metrics with Google Analytics, Wiley Publishing, Inc. (2010), 2nd ed.

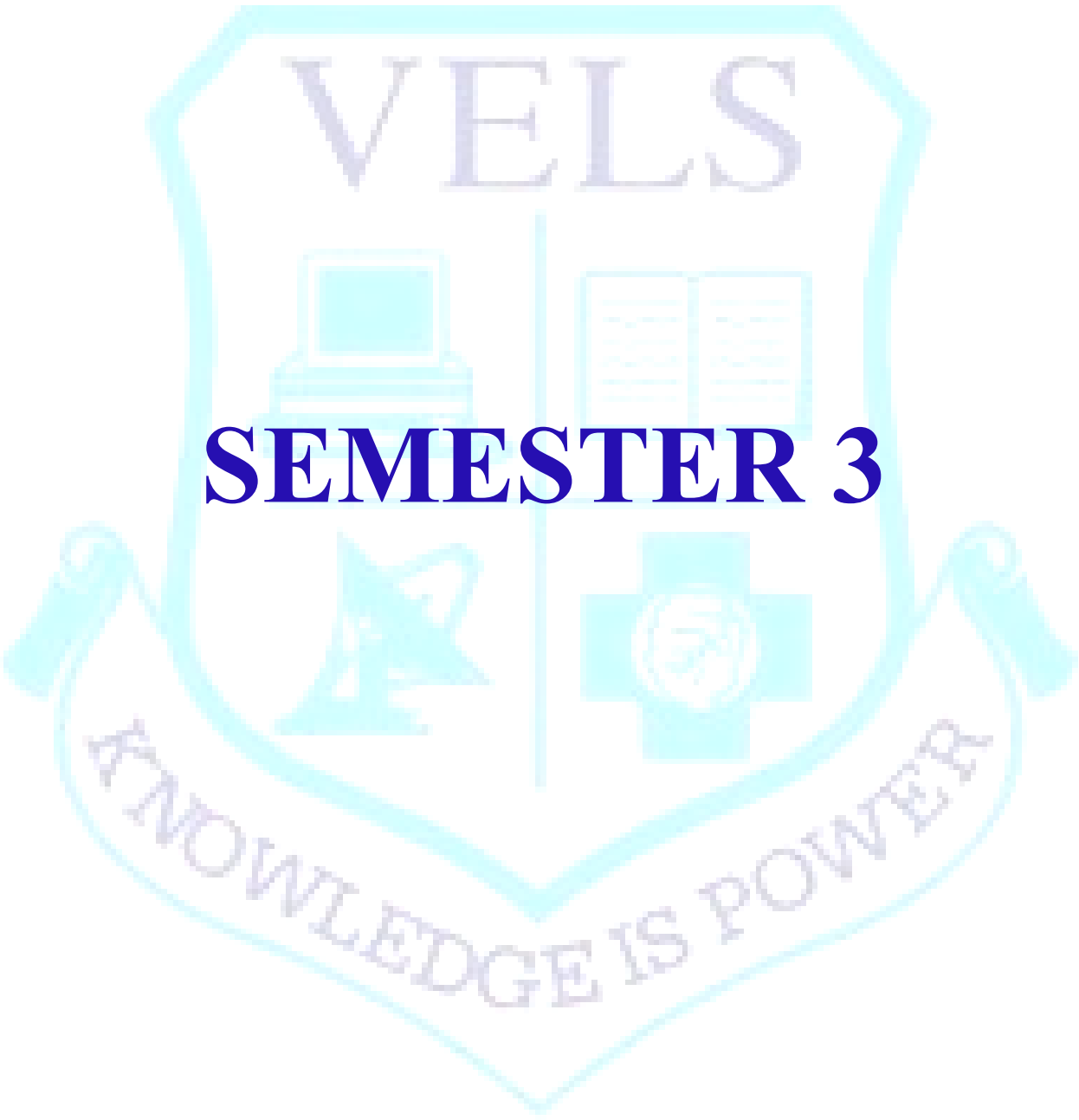
## REFERENCE BOOKS:

1. Kaushik A., Web Analytics 2.0 The Art of Online Accountability and Science of Customer Centricity, Wiley Publishing, Inc. (2010),1st ed.
2. Sterne J., Web Metrics:Proven methods for measuring web site success, John Wiley and Sons (2002),1stedJosh Patterson and Adam Gibson, “Deep learning: A practitioner & approach”, Reilly Media, First Edition, 2017.

**WEB SOURCES:**

- https://keras.io/datasets/
- http://deeplearning.net/tutorial/deeplearning.pdf
- https://arxiv.org/pdf/1404.7828v4.pdf






# SEMESTER 3

**24CMDS31** **BIGDATA ANALYTICS**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

- To optimize business decisions and create competitive advantage with Big Dataanalytics
- To explore the fundamental concepts of big data analytics.
- To learn to analyze the big data using intelligent techniques.
- To understand the various search methods and visualization techniques.
- To learn to use various techniques for mining data stream.
- To understand the applications using Map Reduce Concepts.
- To introduce programming tools PIG & HIVE in Hadoop echo system.

## UNIT I - INTRODUCTION TO BIG DATA 12

Introduction to big data : Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting.

## UNIT II – MINING DATA STREAMS 12

Mining data streams : Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications - CaseStudies - Real Time Sentiment Analysis- Stock Market Predictions.

## UNIT III – INTRODUCTION HADOOP 12

Unit III :Hadoop: History of Hadoop- the Hadoop Distributed File System – Components of Hadoop Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics.

## UNIT IV – MAP REDUCE 12

Map Reduce : Developing a Map Reduce Application-How MapReduce Works-Anatomy of a Map Reduce Job run-Failures-Job Scheduling-Shuffle andSort – Task execution - Map Reduce Types and Formats-Map Reduce Features- Hadoop environment.

## UNIT V – BIG DATA FRAMEWORK 12

Frameworks: Applications on Big Data Using Pig and Hive – Data processing operatorsin Pig – Hive

services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper - IBM InfoSphere Big Insights and Streams.

**Total hours: 60**

## COURSE OUTCOMES:

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

CO1 : Work with big data platform and explore the big data analytics techniques business applications.

CO2: Design efficient algorithms for mining the data from large volumes.

CO3: Understand the concepts of Map Reduce technologies associated with big Data analytics.

CO4: Explore on Big Data applications Using Pig and Hive.

CO5: Understand the fundamentals of various big data analytics techniques.

CO6: Analyze various Big data framework

CO7: Design Hadoop distributed file system

CO8: Understand mining data streams

## TEXTBOOKS:


1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media, 2012.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGrawHill Publishing, 2012.
4. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", CUP,2012.


## REFERENCE BOOKS:


1. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge DataStreams with Advanced Analytics", John Wiley& sons, 2012.
2. Glenn J. Myatt, "Making Sense of Data", John Wiley & Sons, 2007.
3. Pete Warden, "Big Data Glossary", O'Reilly, 2011.
4. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, Reprinted 2008.


## WEB SOURCES:

- https://www.geeksforgeeks.org/what-is-big-data-analytics/
- https://www.ibm.com/topics/big-data-analytics
- https://course.fast.ai/

**24CMDS32** **NATURAL LANGUAGE PROCESSING**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

- Natural Language Processing (NLP) is an interdisciplinary field that combines linguistics, computer science, and artificial intelligence to enable computers to understand, interpret, and generate human language.
- NLP will provide an overview of the key techniques and algorithms used in NLP,
- It also includes text preprocessing, language models, and advanced methods for understanding and generating natural language.

### UNIT I - INTRODUCTION TO NLP 10

Overview of NLP**:** Applications, challenges, and history-Text Processing Basics: Tokenization, stopword removal, stemming, and lemmatization-Text Preprocessing Tools: NLTK, SpaCy, and other Python libraries

### UNIT II - LINGUISTIC FOUNDATIONS FOR NLP 13

Syntax and Parsing: Parts of speech, syntactic trees, and dependency parsing- Semantics and Pragmatics:Word meanings, ambiguity, word senses-Named Entity Recognition (NER): Techniques and applications-Introduction to Language Corpora: Text corpora, annotated datasets

### UNIT III - PROBABILITY AND STATISTICAL MODELS 14

Probability in NLP**:** N-grams, conditional probabilities-Hidden Markov Models (HMMs)**:** Part-of-speech tagging, sequence labeling-Naive Bayes Classifier**:** Text classification, spam detection

### UNIT IV- LANGUAGE MODELS 13

n-gram Models: Unigrams, bigrams, and higher-order n-grams-Smoothing Techniques: Laplace, Kneser-Ney Evaluation Metrics: Perplexity, cross-entropy- Word Embeddings: Word2Vec, GloVe, and FastText Contextual Word Representations: ELMo, BERT, and GPT-Evaluation of Word Representations: Similarity, analogy tasks, and out-of-vocabulary (OOV) words.

**UNIT V- MACHINE LEARNING IN NLP** **10**

Supervised Learning for Text Classification: Logistic regression, SVMs-Text Classification Applications: Sentiment analysis, spam detection-Evaluation Metrics: Precision, recall, F1-score, confusion matrix.

**Total hours: 60**

**COURSE OUTCOMES:**

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

**CO-1:** Understand the Foundations of NLP

**CO-2:** Design a traditional NLP model for an application of your choice.

**CO-3:** Analyze the text preprocessing, tokenization, part-of-speech tagging, named entity recognition, and sentiment analysis.

**CO-4:** Apply efficient machine learning algorithms to solve real world problems

**CO-5:** To understand the principles of NLP and its capabilities.

**CO-6:** Understand text preprocessing steps, including text normalization, stemming, lemmatization, stop word removal, and tokenization.

**CO-7:** To get knowledge of the fundamentals of machine learning and implement them in NLP.

**CO-8:** Analyze various applications of NLP

**TEXTBOOKS**:

1. "Speech and Language Processing" by Daniel Jurafsky and James H. Martin, 2020 - 3rd Edition
2. "Natural Language Processing with Transformers" by Lewis Tunstall, Leandro von Werra and Thomas Wolf, 2022

**REFERENCE BOOKS:**

**1.** “Natural Language Processing with Python" by Steven Bird, Ewan Klein, and Edward Loper,2009 (1st Edition)

**2.** "Pattern Recognition and Machine Learning" by Christopher Bishop,2006

**WEB SOURCES:**

- https://web.stanford.edu/class/cs224n/
- https://www.coursera.org/specializations/natural-language-processing
- https://course.fast.ai/

**24CMDS33** **APPLIED STATISTICS**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

**COURSE OBJECTIVES:**

- To understand the student's statistics plays an important role in data analysis.
- To help the students to read, classify and then interpret the statistical data.
- To acquire knowledge the students about the Statistical functions.
- To implement the concepts of statistical tools like Diagrams, Graphs, Time series, Index numbers, Statistics Quality control (SQC) etc.

## UNIT I DIAGRAMMATIC AND GRAPHICAL REPRESENTATION 12

Data Analysis and Characteristics - Presentation of data –Diagrammatic and Graphical Representation.-Types of Bar Diagrams - Histogram, Frequency Curve, Frequency Polygon - Ogive curves.

## UNIT II WEIGHTED AND COMBINED MEASURES 12

Basic concepts of Mean, Median, Mode - Partition Values: Quartiles, Deciles and Percentiles. Geometric Mean (G.M.), Harmonic Mean (H.M.), Weighted Mean : Weighted A.M., G.M. and H.M.. Combined Variance and Combined Standard deviation.

## UNIT-III TIME SERIES AND INDEX NUMBERS 12

Concept– Components– Additive and multiplicative models for the analysis of time series. Measurement of trend: Graphic method- Semi Average method - Method of least squares - Method of Moving Averages. Definition– Construction of weighted index numbers – Laspeyre's, Paasche's and Fisher's index numbers. Time reversal test, Factor reversal test.

## UNIT-IV CONTROL CHARTS FOR VARIABLES 12

Control charts for attributes:*np*-chart, *p*-chart and *c*-chart. Comparison between controls Charts for variables and control charts for attributes. Descriptive Statistics-Statistical Process Control (SPC) –Acceptance Sampling- problems on control chart and descriptive statistics.

## UNIT-V STATISTICAL INFERENCE AND TESTING OF HYPOTHESIS 12

Sampling Distribution: Types of sampling methods – Central Limit Theorem-Sampling distribution of a sample. Estimation: Point Estimation-Properties of Estimation-Confidence Interval for Population Mean &Proportion- problems on sampling distribution, estimation and testing of hypothesis.

**Total Hours: 60**

## COURSE OUTCOME:

**CO1:** Adapt the knowledge of Diagrams and Graphical representation.

**CO2:** Estimate the Weighted and Combined Measures from AM.GM, HM, Variance & SD.

**CO3:** Build the knowledge of time series data and index numbers.

**CO4:** Analyze the knowledge of Statistical Quality control charts.

**CO5:** Utilize the knowledge of sampling distribution and Estimation.

**CO6:** Understand various estimation technique.

**CO7:** Analyze distribution method using sampling

**CO8:** Estimate the time series

## TEXT BOOKS:


1. Biswas and Srivastava – A textbook of mathematical Statistics, Ist Edition, Narosa Publishing House, New Delhi.
2. Gupta, S.C. and V. K. Kapoor – Mathematical Statistics, Sultan Chand and sons. Hogg, R.V. and Craig,
3. Montogomery, D. C. and Runger, G.C. -Applied Statistics and Probability for Engineers, 3rd Edition reprint, Wiley India Pvt.Ltd.,2008.


## REFERENCE BOOKS:


1. S.P.Gupta, Statistical Methods. Sultan Chand & Sons, New Delhi
2. S.C. Gupta and V.K. Kapoor, Fundamentals of Applied Statistics, Sultan Chand & Sons, 3rd Edition, 2001.
3. P.R. Vittal, Business Statistics, Margham Publications, Second Edition, 2012.


## WEB SOURCE LINKS

1. https://d3bxy9euw4e147.cloudfront.net/oscmsprodcms/media/documents/IntroductoryBusinessStatistics-OP.pdf
2. http://www.ddegjust.ac.in/studymaterial/mcom/mc-106.pdf
3. https://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/SPSS_Statistical_Analyses_u sing_SPSS.pdf

**24PMDS31** **BIGDATA ANALYTICS LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

## COURSE OBJECTIVE

- To understand the applications using Map Reduce Concepts.
- To introduce programming tools PIG & HIVE in Hadoop echo system.

## LIST OF EXPERIMENTS

1. Hadoop installation.

2. Building Hadoop single node and multi-node clusters.

3. Working with Hadoop Distributed File system commands.

4. MapReduce application for word counting on Hadoop cluster.

5. MapReduce application for matrix multiplication.

6. XYZ.com is an online music website where users listen to various tracks, the data gets collected which is given below.

The data is coming in log files and looks like as shown below.

| UserId | TrackId | Shared | Radio | Skip |
|---|---|---|---|---|
| 111115 | 222 | 0 | 1 | 0 |
| 111113 | 225 | 1 | 0 | 0 |
| 111117 | 223 | 0 | 1 | 1 |
| 111115 | 225 | 1 | 0 | 0 |

Write a MapReduce program to get the following

- Number of unique listeners
- Number of times the track was shared with others
- Number of times the track was listened to on the radio
- Number of times the track was listened to in total
- Number of times the track was skipped on the radio

7. Develop a MapReduce program to find the frequency of books published eachyear and find in which year maximum number of books were published usingthe following data.

Title Author Published year
Author
country
Language No of pages

8. Develop a MapReduce program to analyze Titanic ship data and to find the average age of the people (both male and female) who died in the tragedy. How many persons are survived in each class.

The titanic data will be.

Column 1 :PassengerId Column 2 : Survived (survived=0 &died=1)
Column 3 :Pclass Column 4 : Name
Column 5 : Sex Column 6 : Age
Column 7 :SibSp Column 8 :Parch
Column 9 : Ticket Column 10 : Fare
Column 11 :Cabin Column 12 : Embarked

9. Develop a MapReduce program to analyze Uber data set to find the days on which each basement has more trips using the following dataset.

The Uber dataset consists of four columns they are dispatching_base_number date active_vehicles trips

10. Working with managed and external tables in HIVE.

11. Write Pig Latin program to sort, group, join, project, and filter the data.

**Total Hours: 30**

## COURSE OUTCOMES

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

CO1 : Work with big data platform and explore the big data analytics techniques business applications.

CO2: Design efficient algorithms for mining the data from large volumes.

CO3: Understand the concepts of Map Reduce technologies associated with big dataanalytics.

CO4: Explore on Big Data applications Using Pig and Hive.

CO5: Understand the fundamentals of various big data analytics techniques.

## WEB SOURCES:

- https://www.geeksforgeeks.org/what-is-big-data-analytics/
- https://www.ibm.com/topics/big-data-analytics
- https://course.fast.ai/

**24PMDS32** **NATURAL LANGUAGE PROCESSING LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

## COURSE OBJECTIVE

To provide practical experience in software development using open source tools like PYTHON COLAB

## LIST OF EXPERIMENTS

**1. Text Preprocessing**

Tokenization, Stemming, Lemmatization and Stop-word removal

**2. Sentiment Analysis**

- Classifying text as positive or negative

**3. Named Entity Recognition (NER)**

- Identifying entities like names, organizations, dates, etc.

**4 . Text Classification**

- Classifying text into predefined categories

**5. Machine Translation**

- Translating text from one language to another

**6. Text Summarization**

- Summarizing long text into shorter content

**7. Question Answering**

- Given a context, find the answer to a question

**8. Word Embeddings and Similarity**

- Finding similarities between words or phrases

**9. Text Generation**

- Generating text based on a given prompt

**10. Topic Modeling**

- Discovering topics in a collection of texts

**Total Hours: 30**

## COURSE OUTCOMES

**On Completion of this course, the students can able to**

**CO1:** Create applications using PYTHON COLAB.

**CO2:** Create applications using python for the Named Entity Recognition.

**CO3:** Analyze the utilities of text classification in Python.

**CO4:** Analyze the word embeddings and similarity.

**CO5:** Perform text generation using python colab

**CO6:** Test topic modelling using python colab

**CO7:** Test the text generation, text summarization

**CO8:** Analyze different type of modelling the entities.

## WEB SOURCES:

- https://web.stanford.edu/class/cs224n/
- https://www.coursera.org/specializations/natural-language-processing
- https://course.fast.ai/

DSE 3

24DMDS32 DEEP LEARNING

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES

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

### UNIT I MACHINE LEARNING BASICS 12

Machine Learning Basics- Types of Machine Learning-Classification-Regression-Clustering-Association- Supervised Vs. Unsupervised Learning- Machine Learning Models-Applications of Machine Learning.

### UNIT II FUNDAMENTALS CONCEPTS OF NEURAL NETWORKS 12

Basic concepts — Layers – Neurons –Neural Networks in deep learning- Feed forward Neural Network (FNN) - Convolutional Neural Network (CNN) - Recurrent Neural Network (RNN).

### UNIT III INTRODUCTION TO DEEP LEARNING 12

Machine Learning Vs. Deep Learning, Representation Learning, Width vs. Depth of Neural Networks, Activation Functions: RELU, LRELU, ERELU..

### UNIT IV TRANSFER LEARNING 12

Unsupervised Training of Neural Networks, Restricted Boltzmann Machines, Auto Encoders. – Techniques- Transfer learning in deep learning.

### UNIT V APPLICATIONS OF DEEP LEARNING 12

Image Segmentation — Object Detection — Automatic Image Captioning — Image generation with Generative Adversarial Networks — Video to Text with LSTM Models — Attention Models for Computer Vision.

**Total hours: 60**

## COURSE OUTCOMES:

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

**CO-1:** Design a learning model appropriate to the application.

**CO-2:** Design a Neural Network for an application of your choice.

**CO-3:** Analyze the foundation of neural network.

**CO-4:** Apply Efficient Convolution Algorithms to solve real world problems

**CO-5:** to understand the principles of deep learning and its capabilities.

**CO-6:** Understand Supervised and Unsupervised Learning

**CO-7:** To get knowledge of the fundamentals of neural networks and modern deep learning.

**CO-8:** Analyze various applications of Deep learning

## TEXT BOOKS:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MITPress,2016.

2. Josh Patterson and Adam Gibson, "Deep learning: A practitioner & approach", Reilly Media, First Edition, 2017.

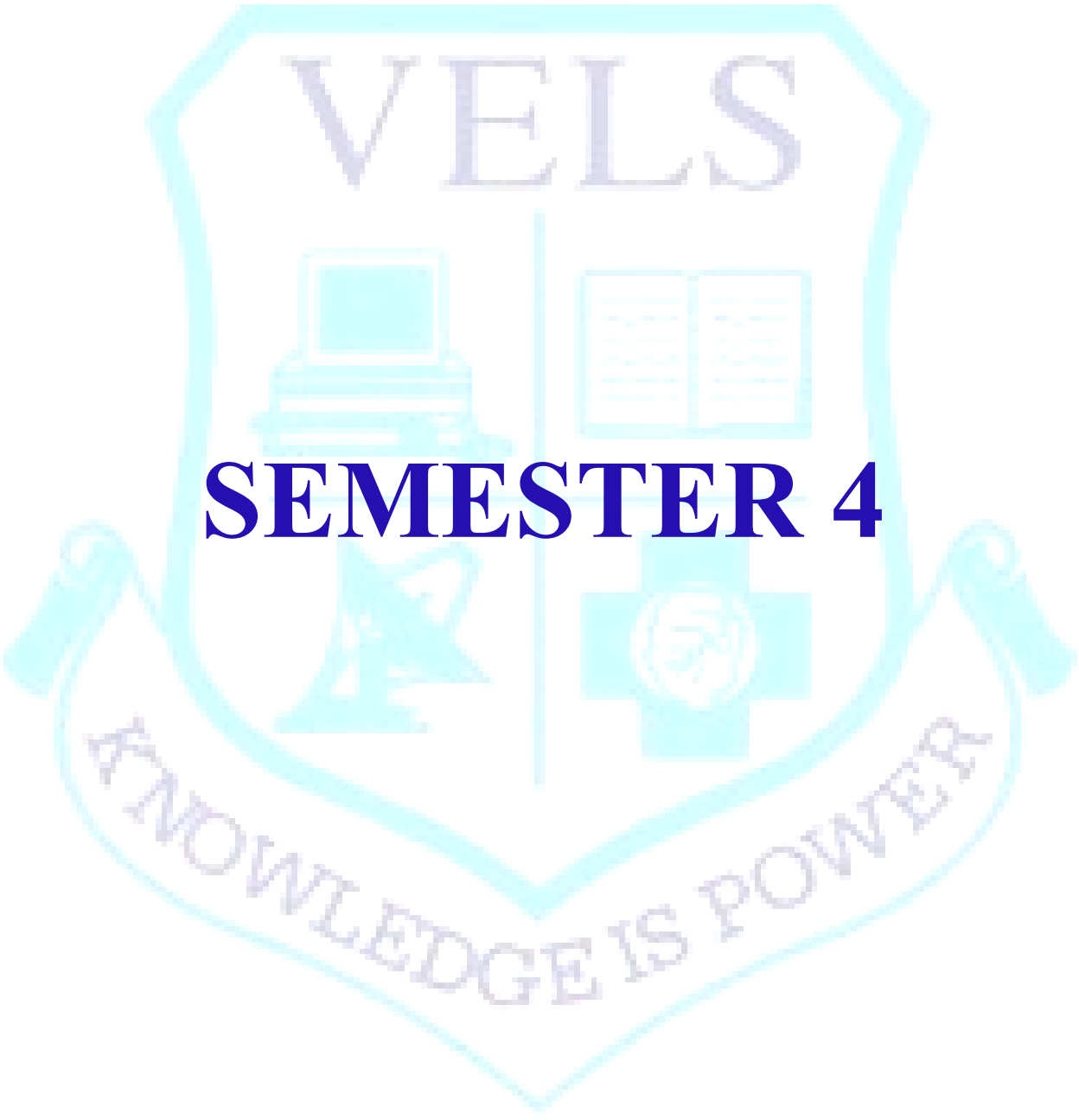
## REFERENCE BOOKS:

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:

- https://keras.io/datasets/
- http://deeplearning.net/tutorial/deeplearning.pdf
- https://arxiv.org/pdf/1404.7828v4.pdf




# SEMESTER 4

**24CMDS41** **CLOUD COMPUTING**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

**COURSE OBJECTIVES**

- To introduce the broad perceptive of cloud architecture and model
- To understand the concept of Virtualization and design of cloud Services
- To be familiar with the lead players in cloud.
- To understand the features of cloud simulator
- To apply different cloud programming model as per need.
- To learn to design the trusted cloud Computing system

**UNIT I CLOUD ARCHITECTURE AND MODEL 12**

Technologies for Network-Based System – System Models for Distributed and Cloud Computing – NIST Cloud Computing Reference Architecture. Cloud Models:- Characteristics – Cloud Services – Cloud models (IaaS, PaaS, SaaS) – Public vs Private Cloud –Cloud Solutions - Cloud ecosystem – Service management – Computing on demand.

**UNIT II VIRTUALIZATION 12**

Basics of Virtualization - Types of Virtualization - Implementation Levels of Virtualization - Virtualization Structures - Tools and Mechanisms - Virtualization of CPU, Memory, I/O Devices - Virtual Clusters and Resource management — Virtualization for Data-center Automation

**UNIT III CLOUD INFRASTRUCTURE AND IOT 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-Enabling Technologies for the Internet of Things – Innovative Applications of the Internet of Things.

**UNIT IV PROGRAMMING MODEL 12**

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

Nebula, OpenStack, Aneka, CloudSim.

**UNIT V SECURITY IN THE CLOUD 12**

Security Overview – Cloud Security Challenges and Risks – Software-as-a-Service Security – Security Governance – Risk Management – Security Monitoring – Security Architecture Design — DataSecurity — Application Security — Virtual Machine Security – Identity Management and Access Control – Autonomic Security.

**Total Hours: 60**

## TEXT BOOKS:


1. George Reese, "Cloud Application Architectures: Building Applications andInfrastructure in the Cloud" O'Reilly
2. GautamShroff,Enterprise Cloud Computing,Cambridge University Press,2011
3. James E. Smith, Ravi Nair, "Virtual Machines: Versatile Platforms for Systems and Processes", Elsevier/Morgan Kaufmann, 2005
4. John W.Rittinghouse and James F.Ransome, "Cloud Computing: Implementation, Management, and Security", CRC Press, 2010


## REFERENCE BOOKS:


1. Kai Hwang, Geoffrey C Fox, Jack G Dongarra, "Distributed and Cloud Computing,
2. From ParallelProcessing to the Internet of Things", Morgan Kaufmann Publishers,2012
3. Kumar Saurabh, "Cloud Computing – insights into New-Era Infrastructure", Wiley India,2011
4. Katarina Stanoevska-Slabeva, Thomas Wozniak, Santi Ristol, "Grid and Cloud Computing – ABusiness Perspective on Technology and Applications", Springer


## WEB SOURCES:

- *https://www.geeksforgeeks.org/cloud-computing/*
- *https://www.ibm.com/topics/cloud-computing*
- *https://www.javatpoint.com/cloud-computing*

**24CMDS41** **CLOUD COMPUTING-LAB**

| L | T | P | O | C |
|---|---|---|---|---|
| 0 | 0 | 4 | 1 | 2 |

**COURSES OBJECTIVES**:

- To understand and simulate the fundamental concepts of cloud computing and virtualization.
- To implement virtual machine management, scheduling, and load balancing techniques in a cloud environment.
- To analyze cloud performance, resource utilization, cost, and energy-efficient computing mechanisms.

## LIST OF EXPERIMENTS

1. Program to Simulate Cloud Environment
2. Program to Create Virtual Machines
3. Program to Create Cloudlets
4. Program for VM Allocation Policy
5. Program for Cloudlet Scheduling
6. Program to Simulate Load Balancing
7. Program to Implement Round Robin Scheduling
8. Program to Calculate Response Time & Turnaround Time
9. Program to Simulate Resource Utilization
10. Program for Cost Calculation
11. Program to Simulate VM Migration
12. Program for Energy-Aware Cloud Computing

**Total Hours: 30**

## COURSE OUTCOMES:

**On Completion of this course, the students can able to**

**CO1:** Understand and simulate a cloud computing environment and its architecture.

**CO2:** Create and manage virtual machines and cloudlets in a cloud environment.

**CO3:** Implement and analyze virtual machine allocation policies.

**CO4:** Apply cloudlet scheduling algorithms for efficient task execution.

**CO5:** Simulate and evaluate load balancing techniques including Round Robin scheduling.

**CO6:** Compute performance metrics such as response time and turnaround time.

**CO7:** Analyze resource utilization and cost calculation in cloud computing.

**CO8:** Implement VM migration and energy-aware cloud computing techniques for optimization.

**WEB SOURCES:**

- *https://www.geeksforgeeks.org/cloud-computing/*
- *https://www.ibm.com/topics/cloud-computing*
- *https://www.javatpoint.com/cloud-computing*

DSE-4

**24DMDS41** **CYBER SECURITY**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVE

- To introduce students to the fundamental concepts and evolving landscape of cyber security, including cyber threats, cyber warfare, and critical infrastructure protection.
- To provide in-depth knowledge about various types of cyber crimes, cyber law frameworks, and enforcement mechanisms both in India and globally.
- To equip students with an understanding of data privacy, data protection regulations, and effective cyber security management strategies including compliance and governance.

## UNIT I OVERVIEW OF CYBER SECURITY 9

Cyber security increasing threat landscape, Cyber security terminologies- Cyberspace, attack, attack vector, attack surface, threat, risk, vulnerability, exploit, exploitation, hacker., Non-state actors, Cyber terrorism, Protection of end user machine, Critical IT and National Critical Infrastructure, Cyberwarfare, Case Studies

## UNIT II CYBER CRIMES 9

Cyber crimes targeting Computer systems and Mobiles- data diddling attacks, spyware, logic bombs, DoS, DDoS, APTs, virus, Trojans, ransomware, data breach., Online scams and frauds- email scams, Phishing, Vishing, Smishing, Online job fraud, Online sextortion, Debit/ credit card fraud, Online payment fraud, Cyberbullying, website defacement, Cybersquatting, Pharming, Cyber espionage, Cryptojacking, Darknet- illegal trades, drug trafficking, human trafficking., Social Media Scams & Frauds- impersonation, identity theft, job scams, misinformation, fake newscyber crime against persons - cyber grooming, child pornography, cyber stalking., Social Engineering attacks, Cyber Police stations, Crime reporting procedure, Case studies.

## UNIT III CYBER LAW 9

Cyber crime and legal landscape around the world, IT Act,2000 and its amendments. Limitations of

IT Act, 2000. Cyber crime and punishments, Cyber Laws and Legal and ethical aspects related to new technologies- AI/ML, IoT, Blockchain, Darknet and Social media, Cyber Laws of other countries, Case Studies.

**UNIT IV DATA PRIVACY AND DATA SECURITY 9**

Defining data, meta-data, big data, nonpersonal data. Data protection, Data privacy and data security, Personal Data Protection Bill and its compliance, Data protection principles, Big data security issues and challenges, Data protection regulations of other countries- General Data Protection Regulations(GDPR),2016 Personal Information Protection and Electronic Documents Act (PIPEDA)., Social media- data privacy and security issues.

**UNIT V CYBER SECURITY MANAGEMENT, COMPLIANCE AND GOVERNANCE 9**

Cyber security Plan- cyber security policy, cyber crises management plan., Business continuity, Risk assessment, Types of security controls and their goals, Cyber security audit and compliance, National cyber security policy and strategy.

**COURSE OUTCOMES:**

**CO1:** To explain the key terminologies, threat vectors, and actors in the cyber security landscape.

**CO2**: To analyze various forms of cybercrimes and demonstrate awareness of online frauds, scams, and social engineering techniques.

**CO3:** To apply the relevant cyber laws and legal provisions applicable to different types of cyber offenses including AI, IoT, Blockchain and social media.

**CO4:** To assess data privacy concerns and apply global data protection regulations such as GDPR and PIPEDA in real-world scenarios.

**CO5**: To design basic cybersecurity frameworks for organizations including security policies, risk assessments, and cyber crisis management plans.

**CO6:** To Analyze the structure and impact of social media platforms and assess the security risks and legal implications of online conte

**CO7:** To understand e-commerce and digital payment mechanisms, including associated frauds, security practices, and relevant legal provisions.

**CO8:** To understand about cyber security management, compliance and governance and Cyber Security policies.

**TEXTBOOKS:**

1. Cyber Security Understanding Cyber Crimes, Computer Forensics and LegalPerspectives by Sumit Belapure and Nina Godbole, Wiley India Pvt. Ltd.
2. Information Warfare and Security by Dorothy F. Denning, Addison Wesley.
3. Security in the Digital Age: Social Media Security Threats and Vulnerabilities byHenry A. Oliver, Create Space Independent Publishing Platform.

**REFERENCE BOOKS**

1. Data Privacy Principles and Practice by Natraj Venkataramanan and Ashwin Shriram,CRC Press.
2. Information Security Governance, Guidance for Information Security Managers byW. KragBrothy, 1st Edition, Wiley Publication.
3. Auditing IT Infrastructures for Compliance By Martin Weiss, Michael G. Solomon,2nd Edition, Jones Bartlett Learning.

**24DMDS42** **PREDICTIVE ANALYSIS**

| L | T | P | O | C |
|---|---|---|---|---|
| 4 | 0 | 0 | 2 | 4 |

## COURSE OBJECTIVES:

- To understand Basics of Predictive Analytics, Understand Data types and Variable types.
- To Understand Basic Modelling and work on Missing Data
- To Understand Logistic Regression and its components, Execute variable transformation and Understand Tableau visualization

## UNIT 1: DATA UNDERSTANDING AND PREPARATION 12

Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection, Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values.

## UNIT 2: MODEL DEVELOPMENT & TECHNIQUES 12

Data Partitioning, Model selection, Model Development Techniques, Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Logistic regression, Perceptron learning algorithm.

## UNIT 3: MODEL ASSESSMENT AND SELECTION 12

Bias, Variance and model complexity, Bias-variance trade off, optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters.

## UNIT 4: NEURAL NETWORKS, SUPPORT VECTOR MACHINES AND K-NEAREST NEIGHBOR 12

Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbour classifiers (Image Scene Classification).

## UNIT 5: UNSUPERVISED LEARNING AND RANDOM FORESTS 12

Association rules, Cluster analysis, Principal Components, Random forests and analysis.

**Total Hours: 60**

## COURSE OUTCOMES:

On Completion of this course, the students can able to

**CO1:** Understand how Predictive analytics can be used in the IT environment.

**CO2:** Students will grasp the meaning, benefits of Predictive analytics.

**CO3:** Students will understand analyse prediction business capabilities using Time series/Forecasting methods and Extract features

**CO4:** Understand the process of formulating business objectives.

**CO5:** Understand data selection/collection, preparation and process to successfully design, build, evaluate and implement predictive models for a various business application.

**CO6:** Compare the underlying predictive modelling techniques.

**CO7:** Select appropriate predictive modeling approaches to identify cases to progress with.

**CO8:** Understand the concepts of unsupervised learning techniques

## TEXT BOOKS

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman , *The Elements of Statistical Learning-Data Mining, Inference, and Prediction* ,Second Edition , Springer Verlag, 2009.
2. Predictive & Advanced Analytics (IBM ICE Publication), Third Edition, 2012
3. "Hands-On Predictive Analytics with R: Build Data Science Applications with R" by Marcin Kacprzyk "Machine Learning Yearning" by Andrew NG,2014

## REFERENCE BOOKS:

- Data Science for Business" by Foster Provost and Tom Fawcett, and "Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die" by Eric Siegel.
- Predictive Analytics for Dummies" by Anasse Bari, Mohamed Chaouchi, and Tommy Jung

## WEB SOURCES


1. https://medium.com/analytics-vidhya/logistic-regression-using-single-layer-perceptron-neural-network-slpnn-31757c792d5d
2. https://www.geeksforgeeks.org/ml-bias-variance-trade-off/
3. https://scikit-learn.org/stable/modules/svm.html