(Approved Under Section 2(f) and 12(B) of UGC act of 1956)

Shri Mata Vaishno Devi University

School of Computer Science and Engineering
Faculty of Engineering, SMVDU
Kakryal, Katra 182320

Ref No: SMVDU/SoCSE/2026/639 Date: 27.1.2026

19th Meeting of the Board of Studies (BoS)

The 19th Meeting of the Board of Studies (BoS) of the School of Computer Science & Engineering (SoCSE), Shri Mata Vaishno Devi University (SMVDU), was held on 19th January 2026 at 12:00 Noon in online mode.

The details of the online meeting are as follows:
Mode: Online (via Google Meet)
Venue: School of Computer Science & Engineering (SoCSE)
Google Meet Link:- https://meet.google.com/rhb-rger-anv
Date: 19th January 2026
Time: 12:00 Noon

The following members attended the meeting:

1. Prof. Vijay Laxmi, Dept. of Computer Science and Engineering, MNIT Jaipur, External Member
2. Dr. Sunanda, Chairman & Head, SoCSE
3. Mr. Manoj Kumar Verma, Associate Head, SoCSE
4. Prof. (Dr.) Manoj Kumar Gupta
5. Prof. (Dr.) Ajay Kaul
6. Prof. (Dr.) Baijnath Kaushik
7. Dr. Pooja Sharma
8. Dr. Sakshi Arora
9. Dr. Naveen Kumar Gondhi
10. Dr Rohit Tanwar
11. Dr. Archana Purwar
12. Dr. Sonika Gupta
13. Mr. Sanjay Kumar Sharma
14. Dr. Deo Prakash
15. Mr. Sudesh Kumar
16. Dr. Vipal Kumar Sharma
17. Dr. Ravi Datta Sharma (Coordinator M.Sc. Data Science Program)
18. Dr. Ravi Datta
19. Mr. Abhigya

The agenda items were placed before the members of the Board of Studies and the following discussions/recommendations were made:

<u>Agenda Item No. 1</u>: To Review and Approve the Updated Course Structure for M.Sc. in Data Science *(With effect from Entry Batch 2026)*

The proposed course structure for the M.Sc. in Data Science programme, was placed before the Board for consideration and discussion.

The key changes in the **L–T–P structure and credits** are as follows:

Sr. No.	Course Code	Semester	Course Name	Current L-T-P	Current Credits	Proposed L-T-P	Proposed Credits	Applicable from Entry Batch
1	DSL DC605	I	Research Methodology	3-0-0	3	4-0-0	4	2026
2	DSM MA601	I	Introduction to Linux and Database Technologies	2-0-2	3	3-0-2	4	2026
			Electives					
3	DSM DE601	I	Programming in R	2-0-4	4	3-0-2	4	2026
4	DSM DE603	I	Programming in Python	2-0-4	4	3-0-2	4	2026
5	DSM DE605	I	Big Data Tools & Technologies-I	2-0-4	4	3-0-2	4	2026
6	DSM DE602	II	Big Data Tools & Technologies-II	2-0-4	4	3-0-2	4	2026
7	DSM DE604	II	Time Series	2-0-4	4	3-0-2	4	2026
8	DSM DE606	II	Cloud Computing	2-0-4	4	3-0-2	4	2026
9	DSM DE608	II	Pattern Recognition	2-0-4	4	3-0-2	4	2026
10	DSM DE701	III	Natural Language Processing	2-0-4	4	3-0-2	4	2026
11	DSM DE703	III	Big Data Tools & Technologies-III	2-0-4	4	3-0-2	4	2026
12	DSM DE705	III	Customer Behavior Analysis and Fraud Detection	2-0-4	4	3-0-2	4	2026
13	DSM DE707	III	Blockchain	2-0-4	4	3-0-2	4	2026
14	DSM DE709	III	Next Generation Sequencing Data Analysis	2-0-4	4	3-0-2	4	2026
15	DSM DE711	III	Emerging Technologies in AI	2-0-4	4	3-0-2	4	2026

The Course Structure approved by the Board; applicable to Entry Batch 2026 onwards is attached as **Annexure 19.1** for reference.

<u>Agenda Item No. 2:</u> To consider and approve the Updated Course Structure for M.Sc. in Data Science *(Entry Batch 2025)*

The proposed updations in Course Structure for the M.Sc. in Data Science programme were discussed for the Entry Batch 2025.

The key changes in the **L–T–P structure and credits** are as follows:

Sr. No.	Course Code	Semester	Course Name	Current L-T-P	Current Credits	Proposed L-T-P	Proposed Credits	Applicable from Entry Batch
			Electives					
1	DSM DE602	II	Big Data Tools & Technologies-II	2-0-4	4	3-0-2	4	2025
2	DSM DE604	II	Time Series	2-0-4	4	3-0-2	4	2025
3	DSM DE606	II	Cloud Computing	2-0-4	4	3-0-2	4	2025
4	DSM DE608	II	Pattern Recognition	2-0-4	4	3-0-2	4	2025
5	DSM DE701	III	Natural Language Processing	2-0-4	4	3-0-2	4	2025
6	DSM DE703	III	Big Data Tools & Technologies-III	2-0-4	4	3-0-2	4	2025
7	DSM DE705	III	Customer Behavior Analysis and Fraud Detection	2-0-4	4	3-0-2	4	2025
8	DSM DE707	III	Blockchain	2-0-4	4	3-0-2	4	2025
9	DSM DE709	III	Next Generation Sequencing Data Analysis	2-0-4	4	3-0-2	4	2025
10	DSM DE711	III	Emerging Technologies in AI	2-0-4	4	3-0-2	4	2025

The Course Structure as approved by the board applicable from *Entry Batch 2025* is attached as **Annexure "19.2"** for reference and record.

Agenda Item No. 3: To review and approve the Course Structure for M.Sc. in Data Science (Part-Time) *(With effect from Entry Batch 2026)*

To review and approve the proposed Course Structure for the M.Sc. in Data Science (Part-Time) programme, to be implemented with effect from the Entry Batch 2026, in order to ensure academic rigor, industry relevance, and alignment with current data science practices and regulatory guidelines. The proposal aims to:

- Provide flexibility for working professionals through a part-time learning structure;
- Ensure parity in academic standards with the full-time programme;
- Emphasize application-oriented learning through laboratories, case studies, and project work;
- Align course content with emerging trends and industry requirements in Data Science.

The Board approved the proposed course structure of the M.Sc. in Data Science (Part-Time), applicable from the Entry Batch 2026 onwards. However, the expert emphasized the need to appoint additional faculty to conduct the two M.Sc. Data Science programmes being run in parallel in full-time and part-time modes. The approved course structure is attached as Annexure "19.3".

Agenda Item No. 4: To consider and Approve the syllabus of the New Course "Generative Artificial Intelligence" as School Elective–III

The syllabus of the proposed course titled *"Generative Artificial Intelligence"*, to be offered as School Elective–III / DEC–III, was placed before the Board for consideration and discussion. The Board members reviewed the course structure in detail and, after deliberation, approved it with the following modifications, **in view of the growing importance of secure and responsible deployment of Large Language Models**:

1. The syllabus shall include topics related to the security of Large Language Models (LLMs).
2. The course shall also be offered as an elective for M.Tech (Artificial Intelligence) and M.Sc. (Data Science) programmes.

The revised and approved syllabus is attached as **Annexure "19.4"**.

The resultant updated Course Structure applicable to the **B.Tech. (Computer Science and Engineering)** programme for **Entry Batch 2023 and Entry Batch 2024 onwards**, after inclusion of the elective course **"Generative Artificial Intelligence"**, is attached as **Annexure "19.4(a)"** and **Annexure "19.4(b)"**, respectively.

Agenda Item No. 5: To consider and approve Introduction and Syllabi of New Courses under AEC / VAC / SEC Categories
(To be offered from Even Semester of Academic Session 2025–26)

The syllabi of the new courses proposed under the **AEC / VAC / SEC categories**, to be offered from the Even Semester of Academic Session 2025–26 was placed before the board:

Value Added Course Offered by School (VAC)

S. No	Course Code	Course Title	L	T	S/P	Credit
1	CSL VA104	Digital and AI Literacy	2	0	0	2

Ability Enhancement Courses Offered by School (AEC)

S. No	Course Code	Course Title	L	T	S/P	Credit
1	CSL AE104	Smart Devices Network	2	0	0	2

Skill Enhancement Courses Offered by School (SEC)

S. No	Course Code	Course Title	L	T	S/P	Credit
1	CSL SE106	Data Centre Management	2	0	0	2

The Board members thoroughly reviewed the syllabi and after deliberation, approved with the following modifications:

1. The course *"Digital and AI Literacy"* should include dedicated units focusing on Generative Artificial Intelligence, with emphasis on its concepts, tools, applications, and responsible use.

2. The course *"Smart Device Networks"* should include AI tools and their usage in IoT based smart devices.
3. The course *"Data Centre Management"* should include topics on firewalls, Domain Name System (DNS), and other essential network and security components, providing students with a comprehensive understanding of data centre operations and infrastructure management.

The syllabi of the above courses namely "Digital and AI Literacy" ; "Smart Devices Network" and "Data Centre Management" after the desired modifications is attached as Annexure **19.5(a)**, **19.5(b) , 19.5(c)** respectively.

Agenda Item No. 6: To consider and approve the List of Permitted NPTEL Courses for the Even Semester of Academic Session 2025–26

The list of permitted **NPTEL courses** to be offered to students during the **Even Semester of Academic Session 2025–26 were placed before the Board.**

The board approved the list of NPTEL courses to be offered for the Even Semester, Session 2025–26. They are attached as **Annexure "19.6"**.

Agenda Item No. 7: To consider and approve the Course Codes and Syllabus Revisions for Elective Courses of B.Tech. in "Robotics & Artificial Intelligence" Programme offered by School of Mechanical Engineering

Proposals for Consideration

1. **Approval of Course Codes for Elective Courses**
   The board approved the course codes of the elective courses offered by the School of Computer Science and Engineering for the B.Tech. in Robotics & Artificial Intelligence programme..

The course codes of the elective courses offered by SoCSE are given below:-

S. No	Course Code	Course Title	L	T	P	Credit	School Elective	Semester
1	CSL DE335	Fuzzy Logic	3	1	0	4	DEC–I	5
2	CSM DE337	Python Programming	3	0	2	4	DEC–I	5
3	CSL DE340	Advanced Artificial Intelligence	3	1	0	4	DEC–III	6
4	CSL DE342	Internet of Things for Robotics	3	1	0	4	DEC–IV	6
5	CSL DE419	Robotics Vision Control	3	1	0	4	DEC–V	7
6	CSL DE421	Reinforcement Learning	3	1	0	4	DEC–VI	7

2. **Revision of Syllabus of the Course "Machine Learning"**
   The board approved the revision of the syllabus of the course **Machine Learning** to include

content related to Artificial Intelligence. The updated syllabus along with the approved course code is attached as **Annexure 19.7(a)**.

3. **Introduction of the Course "Augmented Reality (AR) and Virtual Reality (VR)" as an Elective**
   The board approved the inclusion of the course **Augmented Reality (AR) and Virtual Reality (VR)** as an elective. The syllabus is attached as **Annexure 19.7(b)**.
4. **Rationalization of Experiments in Relevant Laboratory Courses**
   The board approved the rationalization of the number of experiments in the relevant laboratory courses. The list of proposed experiments is attached as **Annexure 19.7(c)**.

It was emphasized by the external expert that with a significant number of new programmes and courses being introduced, there is a strong need to appoint additional faculty members to effectively manage teaching, curriculum delivery, and academic quality across all programmes.

Agenda Item No. 8 : To consider and approve the Guidelines for Internship for B.Tech. Students

The proposed guidelines for internship were placed before the Board for consideration and discussion.

The approved internship guidelines are attached as **Annexure 19.8**.

Agenda Item No. 9: To Review and approve the Vision, Mission, and Programme Educational Objectives (PEOs)

To review and approve the revised **Vision** and **Mission** statements of the School, and the **Programme Educational Objectives (PEOs)** for the following programmes:

- B.Tech. (Computer Science and Engineering)
- M.Tech. (Artificial Intelligence)

The revisions have been proposed to ensure alignment with **outcome-based education, industry relevance**, **national education policies**, and **stakeholder expectations**.

The approved Vision, Mission, and proposed PEOs are attached as **Annexure 19.9**.

Agenda Item No. 19.10 : Approval of Title; Course Code and Syllabus of the Course "Business Fundamentals for Entrepreneurs"

To consider and approve the Title, Course Code and syllabus of the course titled: **Business Fundamentals for Entrepreneurs being offered to B.Tech CSE VI semester**.

(4)

The proposed syllabus of the new course was placed for discussion before the Board and is attached as **Annexure 19.10**. The same was approved by the board and is attached as Annexure 19.10.

Prof. (Dr.) Vijay Laxmi
Professor
Deptt of Computer Science & Engg.
MNIT Jaipur

Dr. Sakshi Arora
Associate Professor &
Member Secretary, BoS

Mr. Manoj Kumar Verma
Assistant Professor &
Associate Head,
Member

Prof. (Dr.) Manoj Kumar Gupta
Professor
Member

Prof. (Dr.) Ajay Kaul
Professor
Member

Prof. (Dr.) Baij Nath Kaushik
Professor
Member

Dr. Naveen Kumar Gondhi
Associate Professor
Member

Dr. Pooja Sharma
Associate Professor
Member

Dr. Archana Purwar
Associate Professor
Member

Dr. Rohit Tanwar
Associate Professor
Member

Dr. Sonika Gupta
Assistant Professor
Member

Mr. Sanjay Kumar Sharma
Assistant Professor
Member

Dr. Deo Prakash
Assistant Professor
Member

Mr. Anuj Mahajan
Assistant Professor
Member

(5)

Mr. Sudesh Kumar
Assistant Professor
Member

Dr. Vipal Kumar Sharma
Assistant Professor
Member

Dr. Ravi Datta Sharma
Assistant Professor
Special Invitee,
Member

Mr. . Abhigya
Assistant Professor
Member

Dr. Sunanda
Associate Professor & Head
(School of Computer Science & Engineering)

Copy to:

1. Chairperson
2. Prof., Vijay Laxmi, External Expert
3. Members of the BoS
4. Dr. Ravi Datta Sharma, Assistant Professor, School of Biotechnology
5. Record/office File

(Approved Under Section 2(f) and 12(B) of UGC act of 1956)

Shri Mata Vaishno Devi University

School of Computer Science and Engineering

Faculty of Engineering, SMVDU

Kakryal, Katra 182320

Annexure 19.1

Course Structure of the M.Sc. Data Science Programmew.e.f 2026 Batch

SMVDU PG Structure 2 (Level 6.5)**NEP- Coursework with Research**

Semester – I

Course Code	Course Name	Type of Course	L-T-P	Credits
DSL DC601	Data Structure & Algorithm Design	DSC-1	3-0-0	3
DSP DC601	Data Structure & Algorithm Design Lab	DSC-1	0-0-2	1
DSL DC603	Statistics and Probability for Analytics	DSC-2	3-0-0	3
DSP DC603	Statistics and Probability for Analytics Lab	DSC-2	0-0-2	1
DSL DC605	Research Methodology	DSC-3	**4-0-0**	**4**
DSL NC601	Mathematics Essential	NC-1	3-0-0	NC
DSM MA601	Introduction to Linux and Database technologies	MA-1	**3-0-2**	**4**
-	Elective-1	DSE-1	**3-0-2**	4
-	Elective-2	DSE-2	**3-0-2**	4
				Total Credit=24

Semester – I (Bridge and Non-Credit (NC) course)
Mathematics Essential is going to be compulsory for students who have no exposure to mathematics in graduation or intermediate.

Semester – I (Mandatory course, Introduction to Linux and Database technologies)

Semester – I (Electives)
Students can choose any two electives from the list.

1. Programming in R (DSM DE601)
2. Programming in Python (DSM DE603)
3. Big Data Tools &Technologies-I (DSM DE605)

Semester – II

Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC602	Machine Learning	DSC-4	3-0-0	3

DSP DC602	Machine Learning Lab	DSC-4	0-0-2	1
DSL DC604	Regression Theory & Analysis	DSC-5	3-0-0	3
DSP DC604	Regression Theory & Analysis Lab	DSC-5	0-0-2	1
DSL DC606	Data Visualization	DSC-6	3-0-0	3
DSP DC606	Data Visualization lab	DSC-6	0-0-2	1
-	Elective -1	DSE-3	**3-0-2**	4
-	Elective -2	DSE-4	**3-0-2**	4
				Total Credit=20

Semester – II (Electives)
Students can choose any two electives from the list.

1. Big Data Tools & Technologies-II (DSM DE602)
2. Time Series (DSM DE604)
3. Cloud Computing (DSM DE606)
4. Pattern Recognition (DSM DE608)

Semester – III

Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC701	Deep Learning & Neural Networks	DSC-7	3-0-0	3
DSP DC701	Deep Learning & Neural Networks Lab	DSC-7	0-0-2	1
DSL DC703	Generalized & Linear Modeling	DSC-8	3-0-0	3
DSP DC703	Generalized & Linear Modeling lab	DSC-8	0-0-2	1
DSL DC705	Graph & Social Network Analysis	DSC-9	3-0-0	3
DSP DC705	Graph & Social Network Analysis Lab	DSC-9	0-0-2	1
-	Elective-1	DSE-5	**3-0-2**	4
-	Elective-2	DSE-6	**3-0-2**	4
				Total Credit=20

Semester – III (Electives)
Students can choose any two electives from the list.

1. Natural Language Processing (DSM DE701)
2. Big Data Tools & Technologies-III (DSM DE703)
3. Customer Behavior Analysis and Fraud detection (DSM DE705)
4. Blockchain (DSM DE707)
5. Next Generation Sequencing Data Analysis (DSM DE709)
6. Emerging technologies in AI (DSM DE711)

Semester – IV

Course Code	Course Name	Type of Course	L-T-P	Credit

Annexure 19.1

DSD PR701	Research Thesis / Project /Patent	PR	0-0-40	20
				Total Credit=20

Outcomes expected of Research thesis/project/Patent **in the 2nd Year of PG Programmes**

Semester – IV (Applicable toStructure 2)

The following outcomes must be achieved by the end of the IV Semester:

i. Research Problem Identification
ii. Review of literature
iii. Research design formulation
iv. Commencement of experimentation, fieldwork, or similar tasks
v. Completion of experimentation/ fieldwork
vi. Submission of dissertation, presentation with Viva Voce
vii. Research output in the form of any one of the following –
- Prototype or product development/ patent
- Publication/Presentation of research work in National/International Conference/conference proceedings
- Publication in a reputed journal such as Scopus-indexed journals or other similar quality journals
- Book or Book Chapter in a publication by a reputed publisher.
- Development of software or Database/Analysis pipeline/ Extensive analysis of Data
- Any other scholastic work as recommended by the School Research Committee and approved by the competent authority.

Summary (Excluding contact hours of Project)

	First Semester	Second Semester	Third Semester	Fourth Semester	Total
Credits	24	20	20	20	84

Course Code : **DSL DC601**
Course Title : **Data Structure &Algorithm Design**
L-T-P/S=Credits : **3-0-0**

Unit I: Introduction and Programming strategies Introduction, Objects and ADTs with example, Constructors and destructors, Data structure, methods, Pre and post conditions, C conventions, Error handling, some programming language notes. Data structures: Arrays; lists; stacks and stack frames; Recursion -Recursive functions with example of factorial, Queue, Degeue.

Unit II: Searching, Queues and Sorting, Sequential and binary search, Trees, binary search tree, complexity. Red-Black trees, AVL trees, general n-ary trees, hash tables; Hashing and collision resolution, Priority queues and heaps, Sorting: Bubble, Heap, Quick, Bin, Radix, Dynamic algorithm Fibonacci numbers, bionomial coefficients, optimal binary search trees, matrix chain multiplication, longest common subsequence, optimal triangulation.

Unit III: Graphs, Minimum spanning tree and Djkstra's algorithm, Huffman encoding, FFT, Hard or intractable problems Eulerian or Hamiltonian paths, Travelling salesman problem.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding different types of data structures and their working principle	CO1
2	Able to describe different Data structures	CO2
3	Explain various sorting and searching algorithms.	CO3
4	Describe and solve computational problems including graphs and encoding.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGraw-Hill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.
- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Annexure 19.1

Course Code : **DSP DC601**
Course Title : **Data Structure & Algorithm Design Lab**
L-T-P/S=Credits : **0-0-1**

Unit I : Stack implementation through arrays, link list, Programs for recursion functions, Implementation of queues and leap structures, Application of binary trees in pre-order, post-order and in-order evaluation,

Unit II : A VL tree implementation, Optimal matrix multiplication

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding and implementation of different data structures	CO1
2	Understanding of Binary Trees and implementation for expression evaluation.	CO2
3	Describe and implement dynamic algorithm for Matrix Chain Multiplication.	CO3
4	Implement of Recursion Function, Heap and Queues Data Structures.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGrawHill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.

References:

- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Course Code : **DSL DC603**
Course Title : **Statistics And Probability for Analytics**
L-T-P/S=Credits : **3-0-0**

Introduction to Data analysis and Types : Data Analysis Types of Data Analysis – Descriptive, Exploratory, Predictive, Inferential, Steps in Data Analysis

Unit 1: Descriptive Statistics
Statistics: Preliminary concepts: Measures of Central Tendency: Mean, Median, Mode
Measures of Dispersion: Range, Standard deviation, Variance, Covariance, Graphical Representation of Statistics: Histograms, Bar plots, Scatter plots etc.

Unit 2: Probability: Random Experiments, Trial and Event, Sample Space, Mutually Exclusive or Disjoint Events, Mutually Exhaustive Events, Equally Probable Events, Complementary Event, Classical definition of Probability, Statistical definition of Probability, Axiomatic definition of Probability, Addition theorem, Multiplication theorem, Conditional Probability, Bayes' Theorem, Expectation, Continuous Distribution: Normal Distribution, Properties of Normal distribution

Unit 3: Correlation: Bivariate distribution Correlation, Types of Correlation, Simple Correlation Coefficient for ungrouped data, Properties and Interpretation of Correlation Coefficient, Coefficient of determination, Scatter diagram, Standard Error, Probable error of Correlation Coefficient. Rank correlation, Some examples.

Unit 4: Introduction to the Inferential Statistics: Parameter, Statistic, Null hypothesis, Alternative hypothesis, Critical region, Type1 Error, Type 11 Error, Level of significance, P-value and its applications. Test of Significance for Small samples: One sample t-test, Paired t-test, Degrees of freedom for t-test, F test for equality of Population variances, Degrees of freedom for F-test. Test of Significance for Large samples: Normal test for sample mean and population mean, Normal test for two sample means. Chi-square Test: Test of goodness of fit, Test of Independence of attributes, Degrees of freedom for Chi Square test, Coefficient of contingency, Yates' correction for continuity. Analysis of Variance: One way and Two way (only Examples) , Introduction to Model Building: Basics of Model building, Definition of a Model, Point estimation, Confidence intervals, Testing

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4
5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- S. Ross, A First Course in Probability, 10th ed. Boston, MA: Pearson, 2018, ISBN: 978-0134753119.
- L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, 1st ed. New York, NY: Springer, 2004, ISBN: 978-0387402727.
- G. Casella and R. L. Berger, Statistical Inference, 2nd ed. Boston, MA: Cengage Learning, 2001, ISBN: 978-0534243128.
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174
- Practical Statistics for Data Scientists, Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : **DSP DC603**
Course Title : **Statistics And Probability for Analytics Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Descriptive Statistics: Calculation and interpretation of Descriptive Statistics

Unit 2: Probability: Understanding the preliminary concepts of probability through case studies, Continuous Distribution: Generation of statistical distributions

Unit 3: Correlation: Calculation of correlation understanding and interpreting correlation through case studies

Unit 4: Introduction to the Inferential Statistics: Understanding inferential statics through case studies, Introduction to Model Building: Understanding Basics of Model building, Definition of a Model, Point estimation, Confidence intervals, Testing through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4
5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- Practical Statistics for Data Scientists, Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : DSL DC605
Course Title : Research Methodology
L-T-P/S=Credits : 4-0-0

Unit 1: Research: Definition, Characteristics, Objectives, Research and Scientific method Types of Research: Descriptive vs. Analytical Research, Applied vs. I, Fundamental Research, Quantitative vs. Qualitative Research, Conceptual vs. Empirical Research, Research Problem: Research Problem, Selecting the problem, Necessity of Defining the problem, Technique Involved in Defining a Problem.

Unit 2: Research Design, and Data Collection: Research Design: Meaning, Need, Features of Good Design, Concepts, Relating to Research Design, Different Research Design, Basic Principle of Experimental Designs. Data Collection: Observation Method, Interview Method, Questionnaires, Case Study Method. Big Data and Data Science, why data science? Who is a Data Scientist? Data Science Life Cycle

Unit 3: Data Science Process: Overview, defining goals, Retrieving data, Data preparation, Data exploration, Data modeling, Presentation. Sampling, Measurement and Scaling Techniques Sampling: Steps in Sampling Design, Different Types of Sample Designs, Measurement and Scaling: Measurement in Research, Measurement Scales, Technique of Developing Measurement Tools,

Scaling, Important Scaling Techniques. Data processing: Data description, Data processing, Importance of data preparation, feature engineering and scaling of datasets, Data imbalance, Data normalization and standardization, Treatment of missing values

Unit 4: Ethics in Research and Data Science: Research ethics, Data Science ethics: Doing good data science, – Owners of the data, Valuing different aspects of privacy, Getting informed consent, The Five Cs, Diversity, Inclusion.Report Writing: Significance of Report Writing, Different Steps in Writing Report, Layouts of Research Report, Types of Reports, Oral Presentation, Mechanics and Precautions for Writing Research Report.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the Fundamentals of Research, formulation of research problems and hypothesis	CO1
2	Understanding how to collect data, ethics in data collection and preprocessing of data	CO2
3	Basic understanding of Ethical and Legal Aspects of Research	CO3
4	Making review of research, writing reports and presentation	CO4

Text & References:

- An Introduction to Data Science, Jeffrey Stanton, Syracuse University, SAGE Publications, Inc, ISBN: 978-1506377537, 2017
- A Simple Introduction to DATA SCIENCE, Lars Nielsen, Noreen Burlingame
- Introduction to Data Science, DAN POTTER, CARSTEN BINNING,ELI UPFAL
- Davy Cielen and Arno Meysman, Introducing Data Science. Simon and Schuster, 2016.
- M. Loukides, H. Mason, and D. Patil, Ethics and Data Science. O'Reilly Media, 2018.
- C. R. Kothari, Research Methodology Methods and Techniques. 3rd. ed. New Delhi: NewAge International Publishers, Reprint 2014.
- Zina O'Leary, The Essential Guide of Doing Research. New Delhi: PHI, 2005

Course Code : **DSL NC601**
Course Title : **Mathematics Essential**
L-T-P/S=Credits : **Non-credit**

Unit1: Set Theory - Number system, Sets and their operations, Relations and functions - Relations and their types, Functions and their types, Sequence and Limits - Function of One variable - Function of one variable,Graphs and Tangents,Limits for sequences, Limits for function of one variable, Limits and Continuity

Unit 2: Linear algebra, vector spaces and transformations in n-dimensions. Matrix Operations (Addition, Multiplication, Inversion), formulating and solving linear systems as matrix-vector equations, performing basic computations involving matrix algebra, orthogonal projections, and eigenvalue-eigenvector problems. Singular Value Decomposition (SVD), Principal Component Analysis (PCA)

Unit 3: Derivatives, Tangents and Critical points - Differentiability and the derivative, Computing derivatives and L'H'opital's rule Derivatives, tangents and linear approximation. Critical points: local maxima and minima, Partial Derivatives,

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental aspect of mathematics	CO1
2	Development of analytical and problem solving skills	CO2
3	Understanding of linear algebra, vector space, operation on matrices etc	CO3
4	Finding optimization : local maxima and minima, derivatives	CO4

Text & References:

- G. Strang, Linear Algebra and Its Applications, 4th ed. Boston, MA: Cengage Learning, 2006, ISBN: 978-0030105678.
- Mathematics for Machine Learning, Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Cambridge University Press, 978-1108455145
- C. D. Meyer, Matrix Analysis and Applied Linear Algebra, 1st ed. Philadelphia, PA: SIAM, 2000, ISBN: 978-0898714548.
- Stewart, Calculus: Early Transcendentals, 8th ed. Boston, MA: Cengage Learning, 2015, ISBN: 978-1285741550.
- S. Boyd and L. Vandenberghe, Convex Optimization, 1st ed. Cambridge, UK: Cambridge University Press, 2004, ISBN: 978-0521833783.
- M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning, 1st ed. Cambridge, UK: Cambridge University Press, 2020, ISBN: 978-1108455145.

Course Code : **DSM MA601**
Course Title : **Introduction To Linux and Database Technologies**
L-T-P/S=Credits : **3-0-1**

Unit 1: Operating System: Difference between Windows and Unix/Linux, Definition; Concepts; Function of Operating System; UNIX/LINUX Operating Systems Concepts; Layers of UNIX; Tree Structure of UNIX; Root File System; /bin Directory; /dev Directory; /bin Directory; /etc Directory; /lib Directory; /proc Directory; /mnt Directory; /root Directory; /sbin Directory; /tmp Directory; /var Directory; Relative Path; Absolute Path; Creation of Directory; Creating file; removing file; Listing Files and Directories copying file; renaming file; Changing File Permission; Changing Directory Permission; Changing Group; Changing Owner; Pipe; Filters; pwd command; date command; head command; tail command less command; more command; grep command; VI Editor (Creating a new File; Inserting Text in File; Deleting Text in File; Copy , Cut & Paste Text; Save File). System monitoring, process information, log files, run levels, Introduction to Process in Unix/Linux: Process; Process State (New, Running, Waiting, Ready, Termination); Introduction to Management of Memory; Swapping Virtual Memory; Paging.

Unit 2: Introduction: Relational Databases, NoSQL Databases, Multimodal Databases. Working with Relational Databases using SQL (Structured Query Language), Data Definition Language (DDL), Data Manipulation Language (DML), Basic SQL queries, Integrity constraints on tables,Data Control

Language Commands(DCL): Grant and Revoke SQL Functions, SQL querying to do operations such as identifying nulls, special characters, blank rows/columns, and run distributions, run data summaries, merge tables, get unique counts, SQL Joins, Aggregate functions and GROUP BY, Nested queries and sub queries. GROUP BY CLAUSE along basic aggregations such as SUM, COUNT, AVG RANK(), ROWNUM() &DENSE_RANK. UNION and UNION ALL CASE statement. Introduction to Advanced SQL concepts: Indexes, Sèquence, Clusters, Views, Cursors and Triggers, Embedded SQL, Introduction to SQLITE

Unit 3: Introduction to NoSQL and its capabilities, Contrast MySQL with NOSQL, MongoDB: A Database for the internet, MongoDB's key features, Diving into the MongoDB shell, Introduction to Creating and Querying with indexes, basic Administration and writing Program using MongoDB. Application Development in MongoDB and processing MongoDB Mastery: Indexing and query optimization, Text Search, MongoDB vs Apache Cassandra: Big Data and Apache Cassandra, Importance of Cassandra, Cassandra as a Distributed Database,Cassandra as a Schema Free Database.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of computer and operating system	CO1
2	Understand Linux operating system, file structure and basic commands	CO2
3	Establishment of basic understanding of databases: SQL - and NO-SQL databases	CO3
4	Ability to work with SQL commands and interaction with database	CO4
5	To identify the need and working with NO-SQL databases in Data Science	CO5

Text & References:

- Peter Baer Galvin. 2016. Operating System Concepts. BPB Publication · Stuart E. Madnick. 2001. Operating System. Tata Mac Graw Hill.
- Kenneth H. Roshan. 2007. The complete reference Unix Tata Mac Graw Hill..
- D. M. Dhamethire. , 2011. System Programming and Operating Systems. Tata Mac Graw Hill.
- Kirrgcox. 2001. Red Hat Linux by. Printice Hill India.
- Andrew S. Talenbaum. 2008. Modern Operating system. Printice Hill India.
- Sumetabha Das. 2013. Unix (Concept and Application). Tata Mac Graw Hill.
- Ramesh Bangia. 2015. Learning Unix. BPB Publication.
- Database System Concepts, Abraham Silberschatz, Henry F. Korth, and S. Sudarshan
- An Introduction to Database Systems, C J Date
- Fundamentals of Databases – Elmasri and Navathe.
- Database Management Systems – Raghu Ramakrishna, Johannes Gehrke
- MongoDB: The definitive guide by Kristina Chodorow, Michael Dirolf
- Cassandra: The Definitive Guide, Eben Hewitt
- MangoDB in Action, Kyle Banker, Peter Bakkum, Shaun Verch, Douglas Garrett, Tim Hawkins

Course Code : **DSM DE601**
Course Title : **Programming In R**
L-T-P/S=Credits : **3-0-1**

Annexure 19.1

Unit 1: An Introduction to R, Overview of R programming, applications, usage and comparative study with other softwares and introduction to R for Data Science, Setting up the R environment and packages. Setting up R environment and install packages and supporting libraries in R, Basics of R : Manage your data and workspace, Save your work, How to use R, Data structure in R, Data creation and curation and special function using R, Matrices and Lists

Unit 2: File Handling: Reading different file format using R, file handling and processing, writing output file. Graphics using R: Graphics device, Basic plot function, scatter plot, 3-D scatter plot, pair plots, Lineplot, MatplotMatpoints, Bar plot, Histogram plot, Density plot, Dot plot, Pie chart, Venn diagram, Grid graphics, Lattice, ggplot2, Interactive plotting, combine multiple plots in same graphics screen, save graphics to a file., Conditional Executions, Comparison Operators, Logical Operators, Control Structures, If statements, Ifelse statements, Loops, For loop, While loop, Apply loop family, Other loops,

Unit 3: Functions: Define and Call functions, Syntax Rules for functions, Control utilities for functions, Writing own function, Advance R, Advance R functions and Regular expressions, Object oriented programming, Building R package

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of programming in R and basic of setup environment	CO1
2	Develop Skills for Data Manipulation	CO2
3	Enable Data Visualization	CO3
4	Perform statistical analysis on various datasets	CO4
5	Application of R in machine learning and artificial intelligence	CO5

Text & References:

- A Handbook of Statistical analysis using R, Brain Everitt and Torsten Hothorn
- The art of R programming, Norman Matloff
- Data Analysis and Graphics using R, W. John Braun
- R Graphics, Paul murrell
- R for Data Science, Garrett Grolemund and Hadley Wickham
- Linear Models with R, Julian J. Faraway
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174

Course Code : **DSM DE603**
Course Title : **Programming In Python**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Python: Overview of Python, applications, usage and comparative study with other software. Basics of Python : Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), Basic Programming with Python:If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass,

Unit 2: Data Structures of Python: Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, Operations, Function and Methods, Files,

Modules, Dictionaries, Functions and Functional, Programming, Declaring and calling Functions, Declare, assign and retrieve values from Lists, Introducing Tuples, Accessing tuples

Unit 3: Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception, Exception Handling, Except clause, Try finally clause, User Defined Exceptions, Python Libraries : Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental of programming in python and basic setup	CO1
2	Develop Skills for Data Manipulation	CO2
3	Perform statistical analysis on various datasets	CO3
4	Application of python in machine learning and artificial intelligence	CO4

Text & References:

- Think Python by Allen B. Downey
- Introducing Python by Bill Lubanovic
- Hello World by Warner Sande and Carter Sande
- Learning Python , 5th Edition , Mark Lutz
- Python For Data Analysis by W Mckinney
- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media

Course Code : DSM DE605
Course Title : Big Data Tools & Technologies-I
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to Big Data Big Data and its Importance, V's of Big Data, Drivers for Big Data, Introduction to Big Data Analytics, Big Data Analytics applications. Characteristics of Data: Structured, Semi-structured and Unstructured data, data at rest, data in motion, etc, good data versus bad data, Challenges in handling large data sets,

Unit 2: Introduction to Hadoop: Hadoop Introduction, Hadoop key characteristics, Hadoop Core Components, Hadoop Ecosystem, Hadoop Services, Different Hadoop Modes, Hadoop's Parallel World – Data discovery, Hadoop Installation, Hadoop Cluster Setup, Storing Data in Hadoop, HDFS Commands (Basics, Intermediate and Advanced), Processing your data with Map Reduce, Run test example using Map reduce, Distributed Cache, Map side join vs Reduce Side Join, Running Hadoop applications, Processing Big Data: Integrating disparate data stores, Mapping data to the programming framework, Connecting and extracting data from storage, Transforming data for processing, Subdividing data in preparation for Hadoop Map Reduce. Hadoop Map Reduce: Employing Hadoop Map Reduce, Creating the components of Hadoop Map Reduce jobs, Executing / Monitoring Hadoop Map Reduce jobs,

Unit 3: Fundamental of Apache Pig: What is Pig?, Introduction to Pig Data Flow Engine, Pig and Map Reduce in Detail, When should Pig Used?, Pig and Hadoop Cluster, Pig Interpreter and Map Reduce, Pig Relations and Data Types, Debugging and Generating Example in Apache Pig, Pig philosophy and architecture, Pig installation, Grunt shell, Loading data, Exploring Pig Latin commands, Pig Transformations functions, Joins in Pig, Hands on Exercises

Annexure 19.1

Unit 4: Fundamental of Apache Hive: What is Hive?, Architecture of Hive, Hive Services, Hive Clients, How Hive Differs from Traditional RDBMS, Introduction to HiveQL, Data Types and File Formats in Hive, File Encoding , Common problems while working with Hive Hive architecture, Hive installation, Hive vs. RDBMS, HiveQL and the Hive shell, Data types and schemas, Creating tables (external vs. managed), Creating Partitions, Creating Views, UDF function in java in Hive, Using hive to create diff types of format

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of big data, characters of big data and structure	CO1
2	Establishment of platform to handle and manipulate big-data	CO2
3	Apply Hadoop, Apache pig and hive palatiform to big data analytics	CO3
4	To acquire skills for Challenges in handling large data sets	CO4

Text & References:

- Big Data and Analytics, Seema Acharya , Subhashini Chellappan
- Professional Hadoop Solution, Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich
- Programming PiG, Wiley Publication, Alan Gates
- Programming Hive, Wiley Publication, Jason Rutherglen, Dean Wamplor& Edward Capriolo

Course Code : **DSL DC602**
Course Title : **Machine Learning**
L-T-P/S=Credits : **3-0-0**

Unit 1: An Introduction to Machine Learning: What is ML, Types of ML, Applications used for ML, AI vs ML, Essential for ML and AI. Techniques of Machine Learning: Introduction to Supervised, unsupervised, semi-supervised, and reinforced machine learning techniques.

Unit 2: Unsupervised learning: Basic concepts to calculate distances: Concepts to calculate distances: euclidean distance, Manhattan distance, L1 and L2-norm, Minikowski distance etc., Discuss theory and concepts behind unsupervised learning, Clustering: Finding grouping of data. Hierarchical clustering, k-means and K-medoid clustering algorithm, Dimension Reduction: PCA, t-SNE.

Unit 3: Supervised learning: Fundamentals of Supervised machine learning: Components of supervised learning model: Data, Model: (Classification and regression), Training, Validation, Evaluating, Inference and model's complexity: over and underfitting, Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based methods (Decision tree and Random Forest), Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Key concepts: Agent, Environment, State, Action, Reward, applications, advantages and disadvantages

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Machine Learning by Tom M. Mitchell
- Introduction to Machine Learning, Ethem Alpaydın
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie , Robert Tibshirani , Jerome Friedman

Course Code : **DSP DC602**
Course Title : **Machine Learning Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Introduction: Understanding classification and regression problem. Splitting the data into Training and test data, Understanding Overfitting and complexity with case studies,

Annexure 19.1

Unit 2: Unsupervised learning: Finding grouping of data: Hierarchical clustering, k-means and K-medoid clustering algorithm, dimensionality reduction through case studies

Unit 3: Supervised learning: Working with different models: Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based methods (Decision tree and Random Forest). Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Solve a case study on real world data and apply reinforcement learning concepts.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Programming Collective Intelligence by Toby Segaran
- Machine Learning for Hackers by Drew Conway and John Myles
- Machine Learning by Tom M. Mitchell
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani , Jerome Friedman

Course Code : DSL DC606
Course Title : Data Visualization
L-T-P/S=Credits : 3-0-0

Unit 1: Introduction to data visualization: Introduction to data visualization. Explore and manipulate all graphical parameter. Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact basedteam collaboration, Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529
- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : DSP DC606
Course Title : Data Visualization Lab
L-T-P/S=Credits : 0-0-1

Unit 1: Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact based team collaboration, Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes, Case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529

- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : **DSL DC604**
Course Title : **Regression Theory & Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination, Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction, Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit

Unit 2: Transformations: Variance stabilizing transformations, Transformations to linearize the model, Methods to select a transformation, weighted least squares, Regression and random effect, Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity

Unit 3: Validation of regression models: Techniques for validation, Nonlinear least squares, transformations, Parameter estimation, Logistic and Poisson regression

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : **DSP DC604**
Course Title : **Regression Theory & Analysis Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination through case studies Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction through case studies

Unit 2: Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit through case studies

Unit 3: Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : **DSM DE602**
Course Title : **Big Data Tools & Technologies-II**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Apache Spark : Introduction to Apache Spark, Features of Apache Spark, Apache Spark Stack, Introduction to RDD's, RDD's Transformation, What is good and Bad in Map Reduce , Why to use Apache Spark Spark: A Hadoop Replacement? : Java, Scala, or Python? , Scala, Packages, Data Types, Classes, Calling Functions, Operators, Control Structures, A Quick Intro to Spark: Starting the shell, Data Sources, Testing Spark, Spark Monitor, Comparing Hadoop Map Reduce to Spark, Writing Standalone program with Spark

Unit 2: Spark SQL, Basic Concepts, Using Spark SQL with RDDs, Spark Streaming: Basic Concepts, Creating Your First Stream with Scala, Creating Your First Stream with Java

Unit 3: Fundamental of Azure databricks: Control plane and data plane, Workspace (assets, objects), Build end to end pipeline (Create cluster, exploring source data, data ingestion, data preparation, Query, creating and running databrick job), Spark integration

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamentals for big data technologies	CO1
2	Understanding use of distributing computing in big data	CO2
3	Apply and preparation of data processing workflow	CO3
4	Application to real-world data and analysis	CO4

Text & References:

- Spark in Action, Marko Bonaci, Peter Zecevic
- Spark Cookbook, Rishi Yadav
- Machine Learning for Big Data , Jason Bell
- Spark : The Definitive Guide, Matei Zaharia, Bill Chambers
- Azure Databricks Cookbook: Accelerate and scale real-time analytics solutions using the Apache Spark-based analytics service, by Phani Raj, Vinod Jaiswal, Packt Publishing Limited, ISBN: 978-1789809718

Annexure 19.1

Course Code : **DSM DE604**
Course Title : **Time Series**
L-T-P/S=Credits : **3-0-1**

Introduction: Introduction to time series

Unit 1: Autoregressive-moving average models ARIMA: Moving average models MA(q). Condition of invertability. Autoregressive models AR(p). Yull-Worker). Yull-Worker equations. Stationarity conditions. Autoregressive-moving average models ARMA (p,q).

Unit 2: Coefficient estimation in ARIMA processes: Coefficients estimation in autoregressive models. Coefficient estimation in ARMA (p) processes. Quality of adjustment of time series models. AIC information criterion. BIC information criterion. "Portmonto" statistics.

Unit 3: Forecasting in the framework of Box-Jenkins model Forecasting, trend and seasonality in Box-Jenkins model. Non-stationary time series, Non-stationary time series. Time series with non-stationary variance. Non-stationary mean. ARIMA (p,d,q) models. The use of Box-Jenkins methodology to determination of order of integration.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the principles of autoregressive-moving average models (ARIMA)	CO1
2	Estimation of coefficients in ARIMA processes	CO2
3	Explain the methodology for forecasting	CO3
4	Apply time series models to real-world datasets for forecasting and decision-making applications.	CO4

Text & References:

- Enders W. Applied Econometric Time Series. John Wiley & Sons, Inc., 1995
- Mills, T.C. The Econometric Modelling of Financial Time Series. Cambridge University Press, 1999
- Andrew C. Harvey. Time Series Models. Harvester wheatsheaf, 1993.

Course Code : **DSM DE606**
Course Title : **Cloud Computing**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to cloud computing: History and evolution, Properties, Characteristics & Disadvantages, Pros and Cons of Cloud Computing, Benefits of Cloud Computing, AWS, Azure, Fundamentals of Cloud Computing: the primary deployment and service models., data warehousing, components, and architectures of data warehousing.

Unit 2: Cloud Computing Architecture: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Cloud Deployment Model: Cloud computing platform, Private cloud deployment, Public cloud deployment, Hybrid cloud deployment

Unit 3: Cloud security: Data on virtual network, Infrastructure Security, Network-level security, Host level security, Data Security and Storage, Cloud Access: authentication, authorization, and accounting

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basic principle of cloud computing and service model	CO1
2	How cloud computing serve as cost effective and scalable platform	CO2
3	Enable students to deploy data science models and use of cloud computing instances	CO3
4	Establishment of data security protocols	CO4

Text & References:

- Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Shahed Latif, Subra Kumaraswamy, and Tim Mather
- Enterprise Cloud Computing: Technology, Architecture, Applications, Gautam Shroff,
- Cloud Computing: Concepts, Technology & Architecture, Ricardo Puttini, Thomas Erl, and Zaigham Mahmood

Course Code : DSM DE608
Course Title : Pattern Recognition
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction: The basic concepts of pattern recognition, such as frequent pattern, closed pattern, max-pattern, and association rules. Pattern mining methods (Clustering, Apriori, ECLAT, and FPgrowth),pattern evaluation: lift, chi-square, cosine, Jaccard, and Kulczynski,

Unit 2: Methods for recognition of multi-level patterns: multi-dimensional patterns, qualitative patterns, negative correlations, compressed and redundancy-aware top-k patterns, and mining long (colossal) patterns.

Unit 3: Sequential pattern recognition methods: GSP, SPADE, PrefixSpan, and CloSpan, Graph pattern recognition methods: subgraph pattern mining, such as gSpan, CloseGraph, graph indexing methods, mining top-k large structural patterns in a single large network, and graph mining applications, such as graph indexing and similarity search in graph databases.

Unit 4: Use of Dictionary learning and collaborative filtering in pattern recognition, Pattern recognition applications: spatiotemporal and trajectory patterns, pattern mining in data streams, software bug mining, pattern discovery for image analysis

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basic concept of pattern recognition, mining and evolution method	CO1
2	Finding multilevel patterns	CO2
3	Application of Sequential pattern recognition methods	CO3
4	Modern approaches to filter patterns such as dictionary learning	CO4

Text & References:

- R.Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, NY, 1992.
- Duda R O and P E Hart, Patten classification and scene analysis, John Wiley & Sons, NY 1973
- K.S.Fu, Syntactic pattern recognition and applications, Prentice Hall, NJ, 1982
- T.Pavlidis, Structural pattern recognition, Springer-Verlag, NY, 1977
- D.H.Ballad and C.M.Brown, Algorithms for computer vision, Prentice Hall, 1982

Syllabus - Third Semester

Course Code : **DSL DC701**
Course Title : **Deep Learning & Neural Networks**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to neural networks: Conceptual overview of neural network, Supervised Learning with Neural Networks, Neural Network Representation, Computing a Neural Network's Output, Vectorizing across multiple examples, Activation functions, Derivatives of activation functions, weights and bias, Shallow neural network

Unit 2: Optimization methods: Gradient based methods, Gradient descent, Learning rate, Batch gradient descent, mini batch gradient descent, stochastic gradient descent method. Gradient free methods, Exhaustive search, Genetic algorithm, Simulated annealing,

Unit 3: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network Introduction: stride, Pooling, Flatten layers, fully connected layers.

Unit 4: Generative AI models: Introduction to generative AI and its component, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Boltzmann Machines (BMs):

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning, Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach, J. Patterson, A. Gibson · Neural Networks and Deep Learning: A Textbook, Charu C. Aggarwal · Neural Networks and Deep Learning, Michael Nielsen.

Annexure 19.1

Course Code : **DSP DC701**
Course Title : **Deep Learning & Neural Networks Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Neural networks: Getting ready with your first neural network, Vectorizing across multiple examples, Calculation of Activation functions, weights and bias, Shallow neural network

Unit 2: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network: Introduction, stride, Pooling, Flatten layers, fully connected layers.

Unit 3: Generative AI models: Application of generative AI models, Case study using real world data: Building a neural network model and generating predictions from the model using tensorflow or keras.

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning by Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach by J. Patterson, A. Gibson
- Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal
- Neural Networks and Deep Learning by Michael Nielsen

Course Code : **DSL DC703**
Course Title : **Generalized & Linear Modeling**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to Generalized Linear Models: Linear model, Non linear model, GLM, Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares

Unit 2: Nonlinear Regression Models: Linear and Nonlinear regression models, transforming a linear model, Parameter estimation Logistic and Poisson Regression Models: Logistic regression, Poison regression, Overdispersion

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation, ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson
- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill
- Categorical data analysis, Ala Agresti

Course Code : **DSP DC703**
Course Title : **Generalized & Linear Modeling Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares with case studies

Unit 2: Nonlinear Regression Models: Linear and Non linear regression models, transforming a linear model, Parameter estimation with case studies, **Logistic and Poisson Regression Models:** Logistic regression, Poison regression, Overdispersion with case studies

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation with case studies, ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson
- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill

- Categorical data analysis, Ala Agresti

Course Code : **DSL DC705**
Course Title : **Graph & Social Network Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to graph theory : Use of graph theory to construct different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph, Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital.: Data and Text Mining InSocial Media: Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg, Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.
- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.

- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K. "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

Course Code : **DSP DC705**
Course Title : **Graph & Social Network Analysis Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Study Tools for graphs and network study : Case studies will be implemented by using any programming language such as R or python. Learn other tools like Cytoscape and Gephi, and igraph, Introduction to graph theory: Input graph data : Read different network format, load network data, Construct network from matrix, Adjacency matrix. Input different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph. Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital., Data and Text Mining InSocial Media, Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg, Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.

- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K. "Social Network Analysis: Methods and Applications", Cambridge University Press: 1 edition, 1994.

Course Code : **DSM DE701**
Course Title : **Natural Language Processing**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to text mining: Introduction, Challenges of Text Processing, Character-Set Dependence, Language Dependence, Corpus Dependence, Application Dependence, Exploring Text Data For Preliminary Ideas: A case study for a brief introduction, Examine multiple document corpus of text, Creation of text corpus, Text pre-processing, Stemming, Word tokenisation, Word vectorisation, Word embedding, Natural Language Processing: Sentiment Analysis, Word Frequency in text data, Wordclouds for visualizing sentiments with a case study, Wordclouds for Visualizing text data, Tidy wordclouds, Topic modelling a document

Unit 2: Text Data and Machine Learning, Clustering for text data, Clustering tweets, Regression on text data, Introduction to RTextTools, More on RTextTools, Word2Vec approach, The Doc2Vec approach, Doc2Vec approach for predicting a binary outcome, Doc2Vec approach for multi-class classification,

Unit3: Sentiment Analysis: What is sentiment, What is sentiment analysis, Scope of sentiment analysis: Document level, Sentence level, Sub-sentence level, Types of sentiment analysis, Word embeddings, Word associations or Text similarities, Building dictionaries, Summarise text, Polarity calculation, Algorithms for sentiment analysis: Rule based, Automatic Hybrid, Case study using real world data. Identify Spam-ham Emails or sms classification.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental concepts of natural language processing	CO1
2	Apply core NLP tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition (NER).	CO2
3	Analyse and preprocess text datausing vectorization techniques	CO3
4	Understand the use of deep learning models	CO4

Text & References:

- HANDBOOK OF NATURAL LANGUAGE PROCESSING Ralf Herbrich and Thore Graepel Text Mining with R by Julia Silge and David Robinson.
- Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper.
- Taming Text by Grant Ingersoll, Thomas Morton and Drew Farris.

- Foundations of Statistical Natural Language Processing by Christopher Manning and Hinrich Schütze.

Course Code : **DSM DE703**
Course Title : **Big Data Tools & Technologies-III**
L-T-P/S=Credits : **3-0-1**

Unit 1: Apache Kafka: Data Acquisition: Apache Kafka Introduction, Apache Kafka Components (Producers, consumers, consumer groups, clusters, brokers, partitions, replicas, leaders, and followers) Integration with Hadoop system, Benefits of Stream Processing and Apache Kafka use Cases

Unit 2: Apache Flink: Understanding Stream Processing and Apache Flink, capabilities and its role in the big data ecosystem, Comparison to Apache spark and storm, Flink's architecture including JobManager and TaskManager roles, execution model and dataflow, parallelism, state management, and fault tolerance, DataStream API and Transformations: implementing transformations such as map, filter, flatMap, and reduce, Windowing and Time Semantics, Stateful Stream Processing, Flink Table API and SQL: Performing relational operations on streaming data using Table API. Advanced Topics: Complex Event Processing (CEP) and machine learning with Flink.

Unit 3: ElasticSearch: Data ingestion, storage, enrichment, visualization, and analysis using elasticsearch, Components including: cluster, nodes, shards, replicas, analyzers and documents

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the architecture of distributed systems in handling large-scale and real-time data pipelines.	CO1
2	Design and implement real-time stream processing application	CO2
3	Apply and preparation of pipelines using Apache Kafka and Flink	CO3
4	Understand and implement Elasticsearch for full-text search, indexing, and analytics over large volumes of semi-structured data	CO4

Text & References:

- Apache Kafka Cookbook, Saurabh Minni, Packt Publishing Limited, ISBN: 978-1785882449
- Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications, O'Reilly Media, 2019, 978-1491974292
- Learning Apache Flink, Packt Publishing, 2017, 978-1786466228
- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, O'Reilly Media, 2015, 978-1449358549
- Elasticsearch in Action, Second Edition, Manning Publications, 9781617299858

Course Code : **DSM DE705**
Course Title : **Customer Behavior Analysis and Fraud Detection**
L-T-P/S=Credits : **3-0-1**

Unit 1: What is "customer analytics" and why do we do it?, Specific Loyalty Matrix tools & biases. Goals, Understanding customer needs, preferences, and motivations, Quantitative data (purchase history, website traffic, social media interactions), Qualitative data (customer surveys, feedback, interviews), A Brief review of what needs to be done before serious analysis can start, Sourcing business requirements, Sourcing raw data, Profiling raw data, Data quality control & remediation.

Unit 2: Association Rule Mining(Apriori algorithm), Predictive Modeling (Predicting future customer behavior using machine learning algorithms), Segmentation & Prediction: Decision Tree method, Clustering method, Using Customer Data for Personalization, Customer Lifetime Value (CLTV) Analysis, Churn Analysis

Unit 3: Define fraud, different types of fraud, and the role of data science in fraud detection, Implementation of EDA and machine learning in fraud detection

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of customer behaviour analysis	CO1
2	To be able to understand analytics to uncover patterns in customer behaviour	CO2
3	Provide skills to build customer segmentation models	CO3
4	Introduce real-time fraud detection pipelines using big data tools	CO4

Text & References:

- Behavioral Research Data Analysis with R- Yuelin Li, Jonathon Baron

Course Code : **DSM DE707**
Course Title : **Blockchain**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction: Key Innovations/Revolutions in Human History, Evolution of Economic System, Money, Ledger and Accounting, Digital Revolution & its impact across industries, Digital Money, DigiCash, E-Gold, History & Fundamentals of Blockchain, Key Elements of Blockchain, Blockchain Protocols, Benefits of Blockchain, Types of Blockchain, Centralized vs. De-centralized systems

Unit 2: Cryptography, Symmetrical vs. Asymmetrical Encryption, Secure Hash Algorithms (SHA Family), Comparison of SHA functions, Digital Signatures

Unit 3: Consensus Mechanisms – Objectives & Need, Proof of Work (PoW), Proof of Stake (PoS), Types & Functions of Node, Construct of Block, Public vs. Private Keys, Types of Ledger, Ledgers, Distributed Ledger Technology (DLT), Benefits & Use Cases of DLT, Smart Contract

Course Outcome:

S.N.	Course Outcome	CO

1	Understand the fundamental principles of blockchain technology	CO1
2	Explore how blockchain can be integrated with data science	CO2
3	Identify potential of Blockchain technology in different types of data	CO3
4	Develop simple blockchain applications	CO4

Text & References:

- Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World, 978-1101980149, 978-1260026672, Don and Alex Tapscott
- Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyon, by Chris Burniske and Jack Tatar
- The Basics of Bitcoins and Blockchains: An Introduction to Cryptocurrencies and the Technology that Powers Them (Cryptography, Derivatives Investments, Futures Trading, Digital Assets, NFT) Hardcover – Illustrated, September 15, 2018, Antony Lewis,
- The Blockchain Developer: A Practical Guide for Designing, Implementing, Publishing, Testing, and Securing Distributed Blockchain-based Projects 1st ed. Edition, Apress, 978-1484248461, 2019

Course Code : **DSM DE709**
Course Title : **Next Generation Sequencing Data Analysis**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to NGS and NGS Technologies: Introduction to sequencing technologies from a data analysts view, Concept, Applications of sequencing technologies in Whole genome assembly; Gene expression analysis; Genome annotation; Gene regulation analysis; Variation studies

Unit 2: NGS data analysis: Preprocessing: Introduction to NGS data analysis, Raw sequence files (FASTQ format), Preprocessing of raw reads: quality control (FastQC), adapter clipping, quality trimming, Introduction to read mapping (Alignment methods, Mapping heuristics), Read alignment to a reference genome (BWA, Bowtie2, TopHat), Mapping output (SAM/BAM format), Usage of important NGS toolkits (samtools, BEDtools), Mapping statistics, Visualization of mapped reads (IGV, UCSC)

Unit 3: NGS data analysis: Variant calling, DNA variant calling and Filtering DNA variants, NGS data analysis: Gene Expression Analysis, Differential Expression analysis, Quantification of expression,

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the basic principles and technology in different NGS platform	CO1
2	To be able to explain the workflow of NGS data analysis	CO2
3	To be able to perform preprocessing of the data and downstream analysis	CO3
4	Design and analyse NGS-based projects,	CO4

Text & References:

Annexure 19.1

- Exploring Personal Genomics, 1st Edition, Joel T. Dudley, Konrad J. Karczewski, ISBN-13: 9780199644490, Oxford University Press, 2013
- High-Throughput Next Generation Sequencing, Methods and Applications,Editors: Kwon, Young Min, Ricke, Steven C. (Eds.), ISBN 978-1-61779-089-8, Springer, 2011
- Next Generation Sequencing Technologies and Challenges in Sequence Assembly, El-Metwally,
  M.Sc, Sara, Ouda, Osama M., Helmy, Mohamed, ISBN 978-1-4939-0715-1, Springer 2014

(Approved Under Section 2(f) and 12(B) of UGC act of 1956)

Shri Mata Vaishno Devi University

School of Computer Science and Engineering
Faculty of Engineering, SMVDU
Kakryal, Katra 182320

Annexure 19.2

Course Structure of the M.Sc. Data Science Programme [Entry Batch 2025]

SMVDU PG Structure 2 (Level 6.5)**NEP- Coursework with Research**

Semester – I

Course Code	Course Name	Type of Course	L-T-P	Credits
DSL DC601	Data Structure & Algorithm Design	DSC-1	3-0-0	3
DSP DC601	Data Structure & Algorithm Design Lab	DSC-1	0-0-2	1
DSL DC603	Statistics and Probability for Analytics	DSC-2	3-0-0	3
DSP DC603	Statistics and Probability for Analytics Lab	DSC-2	0-0-2	1
DSL DC605	Research Methodology	DSC-3	3-0-0	3
DSL NC601	Mathematics Essential	NC-1	3-0-0	NC
DSM MA601	Introduction to Linux and Database technologies	MA-1	2-0-2	3
-	Elective-1	DSE-1	2-0-4	4
-	Elective-2	DSE-2	2-0-4	4

Total Credit=22

Semester – I (Bridge and Non-Credit (NC) course)
Mathematics Essential is going to be compulsory for students who have no exposure to mathematics in graduation or intermediate.

Semester – I (Mandatory course, Introduction to Linux and Database technologies)

Semester – I (Electives)
Students can choose any two electives from the list.

1. Programming in R (DSM DE601)
2. Programming in Python (DSM DE603)
3. Big Data Tools &Technologies-I (DSM DE605)

Semester – II

Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC602	Machine Learning	DSC-4	3-0-0	3
DSP DC602	Machine Learning Lab	DSC-4	0-0-2	1
DSL DC604	Regression Theory & Analysis	DSC-5	3-0-0	3
DSP DC604	Regression Theory & Analysis Lab	DSC-5	0-0-2	1

DSL DC606	Data Visualization	DSC-6	3-0-0	3
DSP DC606	Data Visualization lab	DSC-6	0-0-2	1
-	Elective -1	DSE-3	**3-0-2**	4
-	Elective -2	DSE-4	**3-0-2**	4
				Total Credit=20

Semester – II (Electives)
Students can choose any two electives from the list.

1. Big Data Tools & Technologies-II (DSM DE602)
2. Time Series (DSM DE604)
3. Cloud Computing (DSM DE606)
4. Pattern Recognition (DSM DE608)

Semester – III

Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC701	Deep Learning & Neural Networks	DSC-7	3-0-0	3
DSP DC701	Deep Learning & Neural Networks Lab	DSC-7	0-0-2	1
DSL DC703	Generalized & Linear Modeling	DSC-8	3-0-0	3
DSP DC703	Generalized & Linear Modeling lab	DSC-8	0-0-2	1
DSL DC705	Graph & Social Network Analysis	DSC-9	3-0-0	3
DSP DC705	Graph & Social Network Analysis Lab	DSC-9	0-0-2	1
-	Elective-1	DSE-5	**3-0-2**	4
-	Elective-2	DSE-6	**3-0-2**	4
				Total Credit=20

Semester – III (Electives)
Students can choose any two electives from the list.

1. Natural Language Processing (DSM DE701)
2. Big Data Tools & Technologies-III (DSM DE703)
3. Customer Behavior Analysis and Fraud detection (DSM DE705)
4. Blockchain (DSM DE707)
5. Next Generation Sequencing Data Analysis (DSM DE709)
6. Emerging technologies in AI (DSM DE711)

Semester – IV

Course Code	Course Name	Type of Course	L-T-P	Credit
DSD PR701	Research Thesis / Project /Patent	PR	0-0-40	20
				Total Credit=20

Outcomes expected of Research thesis/project/Patent **in the 2nd Year of PG Programmes**

Semester – IV (Applicable toStructure 2)
The following outcomes must be achieved by the end of the IV Semester:

i. Research Problem Identification
ii. Review of literature
iii. Research design formulation
iv. Commencement of experimentation, fieldwork, or similar tasks
v. Completion of experimentation/ fieldwork
vi. Submission of dissertation, presentation with Viva Voce
vii. Research output in the form of any one of the following –
- Prototype or product development/ patent
- Publication/Presentation of research work in National/International Conference/conference proceedings
- Publication in a reputed journal such as Scopus-indexed journals or other similar quality journals
- Book or Book Chapter in a publication by a reputed publisher.
- Development of software or Database/Analysis pipeline/ Extensive analysis of Data
- Any other scholastic work as recommended by the School Research Committee and approved by the competent authority.

Summary (Excluding contact hours of Project)

	First Semester	Second Semester	Third Semester	Fourth Semester	Total
Credits	22	20	20	20	82

Syllabus - First Semester

Course Code : **DSL DC601**
Course Title : **Data Structure &Algorithm Design**
L-T-P/S=Credits : **3-0-0**

Unit I: Introduction and Programming strategies Introduction, Objects and ADTs with example, Constructors and destructors, Data structure, methods, Pre and post conditions, C conventions, Error handling, some programming language notes. Data structures: Arrays; lists; stacks and stack frames; Recursion -Recursive functions with example of factorial, Queue, Degeue.

Unit II: Searching, Queues and Sorting, Sequential and binary search, Trees, binary search tree, complexity. Red-Black trees, AVL trees, general n-ary trees, hash tables; Hashing and collision resolution, Priority queues and heaps, Sorting: Bubble, Heap, Quick, Bin, Radix, Dynamic algorithm Fibonacci numbers, bionomial coefficients, optimal binary search trees, matrix chain multiplication, longest common subsequence, optimal triangulation.

Unit III: Graphs, Minimum spanning tree and Djkstra's algorithm, Huffman encoding, FFT, Hard or intractable problems Eulerian or Hamiltonian paths, Travelling salesman problem.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding different types of data structures and their working principle	CO1
2	Able to describe different Data structures	CO2
3	Explain various sorting and searching algorithms.	CO3
4	Describe and solve computational problems including graphs and encoding.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGraw-Hill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.
- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Course Code : **DSP DC601**
Course Title : **Data Structure & Algorithm Design Lab**
L-T-P/S=Credits : **0-0-1**

Unit I :Stack implementation through arrays, link list, Programs for recursion functions, Implementation of queues and leap structures, Application of binary trees in pre-order, post-order and in-order evaluation,

Unit II : A VL tree implementation, Optimal matrix multiplication

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding and implementation of different data structures	CO1
2	Understanding of Binary Trees and implementation for expression evaluation.	CO2
3	Describe and implement dynamic algorithm for Matrix Chain Multiplication.	CO3
4	Implement of Recursion Function, Heap and Queues Data Structures.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGrawHill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.

References:

- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Course Code : **DSL DC603**
Course Title : **Statistics And Probability for Analytics**
L-T-P/S=Credits : **3-0-0**

Introduction to Data analysis and Types :Data Analysis Types of Data Analysis – Descriptive, Exploratory, Predictive, Inferential, Steps in Data Analysis

Unit 1: Descriptive Statistics
Statistics: Preliminary concepts; Measures of Central Tendency: Mean, Median, Mode
Measures of Dispersion: Range, Standard deviation, Variance, Covariance, Graphical Representation of Statistics: Histograms, Bar plots, Scatter plots etc.

Unit 2: Probability: Random Experiments, Trial and Event, Sample Space, Mutually Exclusive or Disjoint Events, Mutually Exhaustive Events, Equally Probable Events, Complementary Event, Classical definition of Probability, Statistical definition of Probability, Axiomatic definition of

Probability, Addition theorem, Multiplication theorem, Conditional Probability, Bayes' Theorem. Expectation. Continuous Distribution: Normal Distribution, Properties of Normal distribution

Unit 3: Correlation: Bivariate distribution Correlation, Types of Correlation, Simple Correlation Coefficient for ungrouped data, Properties and Interpretation of Correlation Coefficient, Coefficient of determination, Scatter diagram, Standard Error, Probable error of Correlation Coefficient. Rank correlation, Some examples.

Unit 4: Introduction to the Inferential Statistics: Parameter, Statistic, Null hypothesis, Alternative hypothesis, Critical region, Type1 Error, Type 11 Error, Level of significance, P-value and its applications. Test of Significance for Small samples: One sample t-test, Paired t-test, Degrees of freedom for t-test, F test for equality of Population variances, Degrees of freedom for F-test. Test of Significance for Large samples: Normal test for sample mean and population mean, Normal test for two sample means. Chi-square Test: Test of goodness of fit, Test of Independence of attributes, Degrees of freedom for Chi Square test, Coefficient of contingency, Yates' correction for continuity. Analysis of Variance: One way and Two way (only Examples) , Introduction to Model Building: Basics of Model building, Definition of a Model. Point estimation, Confidence intervals, Testing

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4
5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- S. Ross, A First Course in Probability, 10th ed. Boston, MA: Pearson, 2018, ISBN: 978-0134753119.
- L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, 1st ed. New York, NY: Springer, 2004, ISBN: 978-0387402727.
- G. Casella and R. L. Berger, Statistical Inference, 2nd ed. Boston, MA: Cengage Learning, 2001, ISBN: 978-0534243128.
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174
- Practical Statistics for Data Scientists. Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : **DSP DC603**
Course Title : **Statistics And Probability for Analytics Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Descriptive Statistics: Calculation and interpretation of Descriptive Statistics

Unit 2: Probability: Understanding the preliminary concepts of probability through case studies, Continuous Distribution: Generation of statistical distributions

Unit 3: Correlation: Calculation of correlation understanding and interpreting correlation through case studies

Unit 4: Introduction to the Inferential Statistics: Understanding inferential statics through case studies, Introduction to Model Building: Understanding Basics of Model building, Definition of a Model, Point estimation, Confidence intervals, Testing through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4
5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- Practical Statistics for Data Scientists, Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : **DSL DC605**
Course Title : **Research Methodology**
L-T-P/S=Credits : **4-0-0**

Unit 1: Research: Definition, Characteristics, Objectives, Research and Scientific method Types of Research: Descriptive vs. Analytical Research, Applied vs. I, Fundamental Research, Quantitative vs. Qualitative Research, Conceptual vs. Empirical Research, Research Problem: Research Problem, Selecting the problem, Necessity of Defining the problem, Technique Involved in Defining a Problem.

Unit 2: Research Design, and Data Collection: Research Design: Meaning, Need, Features of Good Design, Concepts, Relating to Research Design, Different Research Design, Basic Principle of Experimental Designs. Data Collection: Observation Method, Interview Method, Questionnaires. Case Study Method. Big Data and Data Science, why data science? Who is a Data Scientist? Data Science Life Cycle

Unit 3: Data Science Process: Overview, defining goals, Retrieving data, Data preparation, Data exploration, Data modeling, Presentation. Sampling, Measurement and Scaling Techniques Sampling: Steps in Sampling Design, Different Types of Sample Designs, Measurement and Scaling: Measurement in Research, Measurement Scales, Technique of Developing Measurement Tools, Scaling, Important Scaling Techniques. Data processing: Data description, Data processing. Importance of data preparation, feature engineering and scaling of datasets, Data imbalance, Data normalization and standardization, Treatment of missing values

Unit 4: Ethics in Research and Data Science: Research ethics, Data Science ethics: Doing good data science, – Owners of the data, Valuing different aspects of privacy, Getting informed consent, The Five Cs, Diversity, Inclusion.Report Writing: Significance of Report Writing, Different Steps in Writing Report, Layouts of Research Report, Types of Reports, Oral Presentation, Mechanics and Precautions for Writing Research Report.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the Fundamentals of Research, formulation of research problems and hypothesis	CO1
2	Understanding how to collect data, ethics in data collection and preprocessing of data	CO2
3	Basic understanding of Ethical and Legal Aspects of Research	CO3
4	Making review of research, writing reports and presentation	CO4

Text & References:

- An Introduction to Data Science, Jeffrey Stanton, Syracuse University, SAGE Publications, Inc, ISBN: 978-1506377537, 2017
- A Simple Introduction to DATA SCIENCE, Lars Nielsen, Noreen Burlingame
- Introduction to Data Science, DAN POTTER, CARSTEN BINNING,ELI UPFAL
- Davy Cielen and Arno Meysman, Introducing Data Science. Simon and Schuster, 2016.
- M. Loukides, H. Mason, and D. Patil, Ethics and Data Science. O'Reilly Media, 2018.
- C. R. Kothari, Research Methodology Methods and Techniques. 3rd. ed. New Delhi: NewAge International Publishers, Reprint 2014.
- Zina O'Leary, The Essential Guide of Doing Research. New Delhi: PHI, 2005

Course Code : **DSL NC601**
Course Title : **Mathematics Essential**
L-T-P/S=Credits : **Non-credit**

Unit1: Set Theory - Number system, Sets and their operations, Relations and functions - Relations and their types, Functions and their types, Sequence and Limits - Function of One variable - Function of one variable,Graphs and Tangents,Limits for sequences, Limits for function of one variable, Limits and Continuity

Unit 2: Linear algebra, vector spaces and transformations in n-dimensions. Matrix Operations (Addition, Multiplication, Inversion), formulating and solving linear systems as matrix-vector equations, performing basic computations involving matrix algebra, orthogonal

projections, and eigenvalue-eigenvector problems. Singular Value Decomposition (SVD), Principal Component Analysis (PCA)

Unit 3: Derivatives, Tangents and Critical points - Differentiability and the derivative, Computing derivatives and L'H'opital's rule Derivatives, tangents and linear approximation, Critical points: local maxima and minima, Partial Derivatives,

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental aspect of mathematics	CO1
2	Development of analytical and problem solving skills	CO2
3	Understanding of linear algebra, vector space, operation on matrices etc	CO3
4	Finding optimization : local maxima and minima, derivatives	CO4

Text & References:

- G. Strang, Linear Algebra and Its Applications, 4th ed. Boston, MA: Cengage Learning, 2006, ISBN: 978-0030105678.
- Mathematics for Machine Learning, Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Cambridge University Press, 978-1108455145
- C. D. Meyer, Matrix Analysis and Applied Linear Algebra, 1st ed. Philadelphia, PA: SIAM, 2000, ISBN: 978-0898714548.
- Stewart, Calculus: Early Transcendentals, 8th ed. Boston, MA: Cengage Learning, 2015, ISBN: 978-1285741550.
- S. Boyd and L. Vandenberghe, Convex Optimization, 1st ed. Cambridge, UK: Cambridge University Press, 2004, ISBN: 978-0521833783.
- M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning, 1st ed. Cambridge, UK: Cambridge University Press, 2020, ISBN: 978-1108455145.

Course Code : **DSM MA601**
Course Title : **Introduction To Linux and Database Technologies**
L-T-P/S=Credits : **3-0-1**

Unit 1: Operating System: Difference between Windows and Unix/Linux, Definition; Concepts; Function of Operating System; UNIX/LINUX Operating Systems Concepts; Layers of UNIX; Tree Structure of UNIX; Root File System; /bin Directory; /dev Directory; /bin Directory; /etc Directory; /lib Directory; /proc Directory; /mnt Directory; /root Directory; /sbin Directory; /tmp Directory; /var Directory; Relative Path; Absolute Path; Creation of Directory; Creating file; removing file; Listing Files and Directories copying file; renaming file; Changing File Permission; Changing Directory Permission; Changing Group; Changing Owner; Pipe; Filters; pwd command; date command; head command; tail command less command; more command; grep command; VI Editor (Creating a new File; Inserting Text in File; Deleting Text in File; Copy , Cut & Paste Text; Save File). System monitoring, process information, log files, run levels, Introduction to Process in Unix/Linux: Process; Process State (New, Running, Waiting, Ready, Termination); Introduction to Management of Memory; Swapping Virtual Memory; Paging.

Unit 2: Introduction: Relational Databases, NoSQL Databases, Multimodal Databases. Working with Relational Databases using SQL (Structured Query Language), Data Definition Language (DDL), Data Manipulation Language (DML), Basic SQL queries, Integrity constraints on tables,Data Control Language Commands(DCL): Grant and Revoke SQL Functions, SQL querying to do operations such as identifying nulls, special characters, blank rows/columns, and run distributions, run data summaries, merge tables, get unique counts, SQL Joins, Aggregate functions and GROUP BY, Nested queries and sub queries. GROUP BY CLAUSE along basic aggregations such as SUM, COUNT, AVG RANK(), ROWNUM() &DENSE_RANK. UNION and UNION ALL CASE statement. Introduction to Advanced SQL concepts: Indexes, Sequence, Clusters, Views, Cursors and Triggers, Embedded SQL, Introduction to SQLITE

Unit 3: Introduction to NoSQL and its capabilities, Contrast MySQL with NOSQL, MongoDB: A Database for the internet, MongoDB's key features, Diving into the MongoDB shell, Introduction to Creating and Querying with indexes, basic Administration and writing Program using MongoDB. Application Development in MongoDB and processing MongoDB Mastery: Indexing and query optimization, Text Search, MongoDB vs Apache Cassandra: Big Data and Apache Cassandra, Importance of Cassandra, Cassandra as a Distributed Database,Cassandra as a Schema Free Database.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of computer and operating system	CO1
2	Understand Linux operating system, file structure and basic commands	CO2
3	Establishment of basic understanding of databases: SQL - and NO-SQL databases	CO3
4	Ability to work with SQL commands and interaction with database	CO4
5	To identify the need and working with NO-SQL databases in Data Science	CO5

Text & References:

- Peter Baer Galvin. 2016. Operating System Concepts. BPB Publication · Stuart E. Madnick. 2001. Operating System. Tata Mac Graw Hill.
- Kenneth H. Roshan. 2007. The complete reference Unix Tata Mac Graw Hill..
- D. M. Dhamethire. , 2011. System Programming and Operating Systems. Tata Mac Graw Hill.
- Kirrgcox. 2001. Red Hat Linux by. Printice Hill India.
- Andrew S. Talenbaum. 2008. Modern Operating system. Printice Hill India.
- Sumetabha Das. 2013. Unix (Concept and Application). Tata Mac Graw Hill.
- Ramesh Bangia. 2015. Learning Unix. BPB Publication.
- Database System Concepts, Abraham Silberschatz, Henry F. Korth, and S. Sudarshan
- An Introduction to Database Systems, C J Date
- Fundamentals of Databases – Elmasri and Navathe.
- Database Management Systems – Raghu Ramakrishna, Johannes Gehrke
- MongoDB: The definitive guide by Kristina Chodorow, Michael Dirolf
- Cassandra: The Definitive Guide, Eben Hewitt
- MangoDB in Action, Kyle Banker, Peter Bakkum, Shaun Verch, Douglas Garrett, Tim Hawkins

Course Code : **DSM DE601**
Course Title : **Programming In R**
L-T-P/S=Credits : **3-0-1**

Unit 1: An Introduction to R, Overview of R programming, applications, usage and comparative study with other softwares and introduction to R for Data Science, Setting up the R environment and packages. Setting up R environment and install packages and supporting libraries in R, Basics of R : Manage your data and workspace, Save your work, How to use R, Data structure in R, Data creation and curation and special function using R, Matrices and Lists

Unit 2: File Handling: Reading different file format using R, file handling and processing, writing output file. Graphics using R: Graphics device, Basic plot function, scatter plot, 3-D scatter plot, pair plots, Lineplot, MatplotMatpoints, Bar plot, Histogram plot, Density plot, Dot plot, Pie chart, Venn diagram, Grid graphics, Lattice, ggplot2, Interactive plotting, combine multiple plots in same graphics screen, save graphics to a file., Conditional Executions, Comparison Operators, Logical Operators, Control Structures, If statements, Ifelse statements, Loops, For loop, While loop, Apply loop family, Other loops.

Unit 3: Functions: Define and Call functions, Syntax Rules for functions, Control utilities for functions, Writing own function, Advance R, Advance R functions and Regular expressions, Object oriented programming, Building R package

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of programming in R and basic of setup environment	CO1
2	Develop Skills for Data Manipulation	CO2
3	Enable Data Visualization	CO3
4	Perform statistical analysis on various datasets	CO4
5	Application of R in machine learning and artificial intelligence	CO5

Text & References:

- A Handbook of Statistical analysis using R, Brain Everitt and Torsten Hothorn
- The art of R programming, Norman Matloff
- Data Analysis and Graphics using R, W. John Braun
- R Graphics, Paul murrell
- R for Data Science, Garrett Grolemund and Hadley Wickham
- Linear Models with R, Julian J. Faraway
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174

Course Code : **DSM DE603**
Course Title : **Programming In Python**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Python: Overview of Python, applications, usage and comparative study with other software. Basics of Python : Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), Basic Programming with Python:If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass,

Unit 2: Data Structures of Python: Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, Operations, Function and Methods, Files, Modules, Dictionaries, Functions and Functional, Programming, Declaring and calling Functions, Declare, assign and retrieve values from Lists, Introducing Tuples, Accessing tuples

Unit 3: Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception, Exception Handling, Except clause, Try finally clause, User Defined Exceptions, Python Libraries : Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental of programming in python and basic setup	CO1
2	Develop Skills for Data Manipulation	CO2
3	Perform statistical analysis on various datasets	CO3
4	Application of python in machine learning and artificial intelligence	CO4

Text & References:

- Think Python by Allen B. Downey
- Introducing Python by Bill Lubanovic
- Hello World by Warner Sande and Carter Sande
- Learning Python , 5th Edition , Mark Lutz
- Python For Data Analysis by W Mckinney
- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media

Course Code : DSM DE605
Course Title : Big Data Tools & Technologies-I
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to Big Data Big Data and its Importance, V's of Big Data, Drivers for Big Data, Introduction to Big Data Analytics, Big Data Analytics applications. Characteristics of Data: Structured, Semi-structured and Unstructured data, data at rest, data in motion, etc, good data versus bad data, Challenges in handling large data sets,

Unit 2: Introduction to Hadoop: Hadoop Introduction, Hadoop key characteristics, Hadoop Core Components, Hadoop Ecosystem, Hadoop Services, Different Hadoop Modes, Hadoop's Parallel World – Data discovery, Hadoop Installation, Hadoop Cluster Setup, Storing Data in Hadoop, HDFS Commands (Basics, Intermediate and Advanced), Processing your data with Map Reduce, Run test example using Map reduce, Distributed Cache, Map side join vs Reduce Side Join, Running Hadoop applications, Processing Big Data: Integrating disparate data stores, Mapping data to the programming framework, Connecting and extracting data from storage, Transforming data for processing, Subdividing data in preparation for Hadoop Map Reduce. Hadoop Map Reduce: Employing Hadoop

Map Reduce, Creating the components of Hadoop Map Reduce jobs, Executing / Monitoring Hadoop Map Reduce jobs.

Unit 3: Fundamental of Apache Pig: What is Pig?, Introduction to Pig Data Flow Engine, Pig and Map Reduce in Detail, When should Pig Used?, Pig and Hadoop Cluster, Pig Interpreter and Map Reduce, Pig Relations and Data Types, Debugging and Generating Example in Apache Pig, Pig philosophy and architecture, Pig installation, Grunt shell, Loading data, Exploring Pig Latin commands, Pig Transformations functions, Joins in Pig, Hands on Exercises

Unit 4: Fundamental of Apache Hive: What is Hive?, Architecture of Hive, Hive Services, Hive Clients, How Hive Differs from Traditional RDBMS, Introduction to HiveQL, Data Types and File Formats in Hive, File Encoding , Common problems while working with Hive Hive architecture, Hive installation, Hive vs. RDBMS, HiveQL and the Hive shell, Data types and schemas, Creating tables (external vs. managed), Creating Partitions, Creating Views, UDF function in java in Hive, Using hive to create diff types of format

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of big data, characters of big data and structure	CO1
2	Establishment of platform to handle and manipulate big-data	CO2
3	Apply Hadoop, Apache pig and hive palatiform to big data analytics	CO3
4	To acquire skills for Challenges in handling large data sets	CO4

Text & References:

- Big Data and Analytics, Seema Acharya , Subhashini Chellappan
- Professional Hadoop Solution, Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich
- Programming PiG, Wiley Publication, Alan Gates
- Programming Hive, Wiley Publication, Jason Rutherglen, Dean Wamplor& Edward Capriolo

Syllabus - Second Semester

Course Code : **DSL DC602**
Course Title : **Machine Learning**
L-T-P/S=Credits : **3-0-0**

Unit 1: An Introduction to Machine Learning: What is ML, Types of ML, Applications used for ML, AI vs ML, Essential for ML and AI. Techniques of Machine Learning: Introduction to Supervised, unsupervised, semi-supervised, and reinforced machine learning techniques.

Unit 2: Unsupervised learning: Basic concepts to calculate distances: Concepts to calculate distances: euclidean distance, Manhattan distance, L1 and L2-norm, Minikowski distance etc., Discuss theory and concepts behind unsupervised learning, Clustering: Finding grouping of data. Hierarchical clustering, k-means and K-medoid clustering algorithm, Dimension Reduction: PCA, t-SNE.

Unit 3: Supervised learning: Fundamentals of Supervised machine learning: Components of supervised learning model: Data, Model: (Classification and regression), Training, Validation, Evaluating, Inference and model's complexity: over and underfitting, Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based methods (Decision tree and Random Forest), Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Key concepts: Agent, Environment, State, Action, Reward, applications, advantages and disadvantages

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Machine Learning by Tom M. Mitchell
- Introduction to Machine Learning, Ethem Alpaydın
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie , Robert Tibshirani , Jerome Friedman

Course Code : **DSP DC602**
Course Title : **Machine Learning Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Introduction: Understanding classification and regression problem. Splitting the data into Training and test data, Understanding Overfitting and complexity with case studies,

Unit 2: Unsupervised learning: Finding grouping of data: Hierarchical clustering, k-means and K-medoid clustering algorithm, dimensionality reduction through case studies

Unit 3: Supervised learning: Working with different models: Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based methods (Decision tree and Random Forest). Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Solve a case study on real world data and apply reinforcement learning concepts.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Programming Collective Intelligence by Toby Segaran
- Machine Learning for Hackers by Drew Conway and John Myles
- Machine Learning by Tom M. Mitchell
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani , Jerome Friedman

Course Code : **DSL DC606**
Course Title : **Data Visualization**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to data visualization: Introduction to data visualization. Explore and manipulate all graphical parameter. Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact basedteam collaboration, Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529
- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : DSP DC606
Course Title : Data Visualization Lab
L-T-P/S=Credits : 0-0-1

Unit 1: Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact based team collaboration, Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes, Case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529
- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : **DSL DC604**
Course Title : **Regression Theory & Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination, Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction, Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit

Unit 2: Transformations: Variance stabilizing transformations, Transformations to linearize the model, Methods to select a transformation, weighted least squares, Regression and random effect, Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity

Unit 3: Validation of regression models: Techniques for validation, Nonlinear least squares, transformations, Parameter estimation, Logistic and Poisson regression

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : **DSP DC604**
Course Title : **Regression Theory & Analysis Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination through case studies Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction through case studies

Unit 2: Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit through case studies

Unit 3: Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : DSM DE602
Course Title : Big Data Tools & Technologies-II
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to Apache Spark : Introduction to Apache Spark, Features of Apache Spark, Apache Spark Stack, Introduction to RDD's, RDD's Transformation, What is good and Bad in Map Reduce . Why to use Apache Spark Spark: A Hadoop Replacement? : Java, Scala, or Python? , Scala, Packages, Data Types, Classes, Calling Functions, Operators, Control Structures, A Quick Intro to Spark: Starting the shell, Data Sources, Testing Spark, Spark Monitor, Comparing Hadoop Map Reduce to Spark, Writing Standalone program with Spark

Unit 2: Spark SQL, Basic Concepts, Using Spark SQL with RDDs, Spark Streaming: Basic Concepts, Creating Your First Stream with Scala, Creating Your First Stream with Java

Unit 3: Fundamental of Azure databricks: Control plane and data plane, Workspace (assets, objects), Build end to end pipeline (Create cluster, exploring source data, data ingestion, data preparation, Query, creating and running databrick job), Spark integration

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamentals for big data technologies	CO1
2	Understanding use of distributing computing in big data	CO2
3	Apply and preparation of data processing workflow	CO3
4	Application to real-world data and analysis	CO4

Text & References:

- Spark in Action, Marko Bonaci, Peter Zecevic
- Spark Cookbook, Rishi Yadav

- Machine Learning for Big Data , Jason Bell
- Spark : The Definitive Guide, Matei Zaharia, Bill Chambers
- Azure Databricks Cookbook: Accelerate and scale real-time analytics solutions using the Apache Spark-based analytics service, by Phani Raj, Vinod Jaiswal, Packt Publishing Limited, ISBN: 978-1789809718

Course Code : DSM DE604
Course Title : Time Series
L-T-P/S=Credits : 3-0-1

Introduction: Introduction to time series

Unit 1: Autoregressive-moving average models ARIMA: Moving average models MA(q). Condition of invertability. Autoregressive models AR(p). Yull-Worker). Yull-Worker equations. Stationarity conditions. Autoregressive-moving average models ARMA (p,q).

Unit 2: Coefficient estimation in ARIMA processes: Coefficients estimation in autoregressive models. Coefficient estimation in ARMA (p) processes. Quality of adjustment of time series models. AIC information criterion. BIC information criterion. "Portmonto" statistics.

Unit 3: Forecasting in the framework of Box-Jenkins model Forecasting, trend and seasonality in Box-Jenkins model. Non-stationary time series, Non-stationary time series. Time series with non-stationary variance. Non-stationary mean. ARIMA (p,d,q) models. The use of Box-Jenkins methodology to determination of order of integration.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the principles of autoregressive-moving average models (ARIMA)	CO1
2	Estimation of coefficients in ARIMA processes	CO2
3	Explain the methodology for forecasting	CO3
4	Apply time series models to real-world datasets for forecasting and decision-making applications.	CO4

Text & References:

- Enders W. Applied Econometric Time Series. John Wiley & Sons, Inc., 1995
- Mills, T.C. The Econometric Modelling of Financial Time Series. Cambridge University Press, 1999
- Andrew C. Harvey. Time Series Models. Harvester wheatsheaf, 1993.

Course Code : DSM DE606
Course Title : Cloud Computing
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to cloud computing: History and evolution, Properties. Characteristics & Disadvantages, Pros and Cons of Cloud Computing, Benefits of Cloud Computing, AWS, Azure,

Fundamentals of Cloud Computing: the primary deployment and service models., data warehousing, components, and architectures of data warehousing,

Unit 2: Cloud Computing Architecture: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Cloud Deployment Model: Cloud computing platform, Private cloud deployment, Public cloud deployment, Hybrid cloud deployment

Unit 3: Cloud security: Data on virtual network, Infrastructure Security, Network-level security, Host level security, Data Security and Storage, Cloud Access: authentication, authorization, and accounting

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basic principle of cloud computing and service model	CO1
2	How cloud computing serve as cost effective and scalable platform	CO2
3	Enable students to deploy data science models and use of cloud computing instances	CO3
4	Establishment of data security protocols	CO4

Text & References:

- Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Shahed Latif, Subra Kumaraswamy, and Tim Mather
- Enterprise Cloud Computing: Technology, Architecture, Applications, Gautam Shroff,
- Cloud Computing: Concepts, Technology & Architecture, Ricardo Puttini, Thomas Erl, and Zaigham Mahmood

Course Code : **DSM DE608**
Course Title : **Pattern Recognition**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction: The basic concepts of pattern recognition, such as frequent pattern, closed pattern, max-pattern, and association rules. Pattern mining methods (Clustering, Apriori, ECLAT, and FPgrowth),pattern evaluation: lift, chi-square, cosine, Jaccard, and Kulczynski,

Unit 2: Methods for recognition of multi-level patterns: multi-dimensional patterns, qualitative patterns, negative correlations, compressed and redundancy-aware top-k patterns, and mining long (colossal) patterns.

Unit 3: Sequential pattern recognition methods: GSP, SPADE, PrefixSpan, and CloSpan, Graph pattern recognition methods: subgraph pattern mining, such as gSpan, CloseGraph, graph indexing methods, mining top-k large structural patterns in a single large network, and graph mining applications, such as graph indexing and similarity search in graph databases.

Unit 4: Use of Dictionary learning and collaborative filtering in pattern recognition, Pattern recognition applications: spatiotemporal and trajectory patterns, pattern mining in data streams, software bug mining, pattern discovery for image analysis

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basic concept of pattern recognition, mining and evolution method	CO1
2	Finding multilevel patterns	CO2
3	Application of Sequential pattern recognition methods	CO3
4	Modern approaches to filter patterns such as dictionary learning	CO4

Text & References:

- R.Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, NY, 1992.
- Duda R O and P E Hart, Patten classification and scene analysis, John Wiley & Sons, NY 1973
- K.S.Fu, Syntactic pattern recognition and applications, Prentice Hall, NJ, 1982
- T.Pavlidis, Structural pattern recognition, Springer-Verlag, NY, 1977
- D.H.Ballad and C.M.Brown, Algorithms for computer vision, Prentice Hall, 1982

Course Code : **DSL DC701**
Course Title : **Deep Learning & Neural Networks**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to neural networks: Conceptual overview of neural network, Supervised Learning with Neural Networks, Neural Network Representation, Computing a Neural Network's Output, Vectorizing across multiple examples, Activation functions, Derivatives of activation functions, weights and bias, Shallow neural network

Unit 2: Optimization methods: Gradient based methods, Gradient descent, Learning rate, Batch gradient descent, mini batch gradient descent, stochastic gradient descent method. Gradient free methods, Exhaustive search, Genetic algorithm, Simulated annealing,

Unit 3: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network Introduction: stride, Pooling, Flatten layers, fully connected layers.

Unit 4: Generative AI models: Introduction to generative AI and its component, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Boltzmann Machines (BMs):

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning, Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach, J. Patterson, A. Gibson · Neural Networks and Deep Learning: A Textbook, Charu C. Aggarwal · Neural Networks and Deep Learning, Michael Nielsen.

Course Code : **DSP DC701**
Course Title : **Deep Learning & Neural Networks Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Neural networks: Getting ready with your first neural network, Vectorizing across multiple examples, Calculation of Activation functions, weights and bias, Shallow neural network

Unit 2: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network: Introduction, stride, Pooling, Flatten layers, fully connected layers.

Unit 3: Generative AI models: Application of generative AI models, Case study using real world data: Building a neural network model and generating predictions from the model using tensorflow or keras.

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning by Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach by J. Patterson, A. Gibson
- Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal
- Neural Networks and Deep Learning by Michael Nielsen

Course Code : **DSL DC703**
Course Title : **Generalized & Linear Modeling**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to Generalized Linear Models: Linear model, Non linear model, GLM, Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares

Unit 2: Nonlinear Regression Models: Linear and Nonlinear regression models, transforming a linear model, Parameter estimation Logistic and Poisson Regression Models: Logistic regression, Poison regression, Overdispersion

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation, ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson
- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill
- Categorical data analysis, Ala Agresti

Course Code : **DSP DC703**
Course Title : **Generalized & Linear Modeling Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares with case studies

Unit 2: Nonlinear Regression Models: Linear and Non linear regression models, transforming a linear model, Parameter estimation with case studies, **Logistic and Poisson Regression Models:** Logistic regression, Poison regression, Overdispersion with case studies

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation with case studies. ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson
- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill
- Categorical data analysis, Ala Agresti

Course Code : **DSL DC705**
Course Title : **Graph & Social Network Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to graph theory : Use of graph theory to construct different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph, Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital.: Data and Text Mining InSocial Media: Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg, Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.

- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K, "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

Course Code : **DSP DC705**
Course Title : **Graph & Social Network Analysis Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Study Tools for graphs and network study : Case studies will be implemented by using any programming language such as R or python. Learn other tools like Cytoscape and Gephi, and igraph, Introduction to graph theory: Input graph data : Read different network format, load network data, Construct network from matrix, Adjacency matrix. Input different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph. Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital., Data and Text Mining InSocial Media, Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg, Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.
- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K, "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

Course Code : **DSM DE701**
Course Title : **Natural Language Processing**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to text mining: Introduction, Challenges of Text Processing, Character-Set Dependence, Language Dependence, Corpus Dependence, Application Dependence, Exploring Text Data For Preliminary Ideas: A case study for a brief introduction, Examine multiple document corpus of text, Creation of text corpus, Text pre-processing, Stemming, Word tokenisation, Word vectorisation, Word embedding, Natural Language Processing: Sentiment Analysis, Word Frequency in text data, Wordclouds for visualizing sentiments with a case study, Wordclouds for Visualizing text data, Tidy wordclouds, Topic modelling a document

Unit 2: Text Data and Machine Learning, Clustering for text data, Clustering tweets, Regression on text data, Introduction to RTextTools, More on RTextTools, Word2Vec approach, The Doc2Vec approach, Doc2Vec approach for predicting a binary outcome, Doc2Vec approach for multi-class classification,

Unit3: Sentiment Analysis: What is sentiment, What is sentiment analysis, Scope of sentiment analysis: Document level, Sentence level, Sub-sentence level, Types of sentiment analysis, Word embeddings, Word associations or Text similarities, Building dictionaries, Summarise text, Polarity calculation, Algorithms for sentiment analysis: Rule based, Automatic Hybrid, Case study using real world data. Identify Spam-ham Emails or sms classification.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental concepts of natural language processing	CO1
2	Apply core NLP tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition (NER).	CO2
3	Analyse and preprocess text datausing vectorization techniques	CO3
4	Understand the use of deep learning models	CO4

Text & References:

- HANDBOOK OF NATURAL LANGUAGE PROCESSING Ralf Herbrich and Thore Graepel Text Mining with R by Julia Silge and David Robinson.
- Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper.
- Taming Text by Grant Ingersoll, Thomas Morton and Drew Farris.
- Foundations of Statistical Natural Language Processing by Christopher Manning and Hinrich Schütze.

Course Code : **DSM DE703**
Course Title : **Big Data Tools & Technologies-III**
L-T-P/S=Credits : **3-0-1**

Unit 1: Apache Kafka: Data Acquisition: Apache Kafka Introduction, Apache Kafka Components (Producers, consumers, consumer groups, clusters, brokers, partitions, replicas, leaders, and followers) Integration with Hadoop system, Benefits of Stream Processing and Apache Kafka use Cases

Unit 2: Apache Flink: Understanding Stream Processing and Apache Flink, capabilities and its role in the big data ecosystem, Comparison to Apache spark and storm, Flink's architecture including JobManager and TaskManager roles, execution model and dataflow, parallelism, state management, and fault tolerance, DataStream API and Transformations: implementing transformations such as map, filter, flatMap, and reduce, Windowing and Time Semantics, Stateful Stream Processing, Flink Table API and SQL: Performing relational operations on streaming data using Table API. Advanced Topics: Complex Event Processing (CEP) and machine learning with Flink.

Unit 3: ElasticSearch: Data ingestion, storage, enrichment, visualization, and analysis using elasticsearch, Components including: cluster, nodes, shards, replicas, analyzers and documents

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the architecture of distributed systems in handling large-scale and real-time data pipelines.	CO1
2	Design and implement real-time stream processing application	CO2
3	Apply and preparation of pipelines using Apache Kafka and Flink	CO3
4	Understand and implement Elasticsearch for full-text search, indexing, and analytics over large volumes of semi-structured data	CO4

Text & References:

- Apache Kafka Cookbook, Saurabh Minni, Packt Publishing Limited, ISBN: 978-1785882449
- Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications, O'Reilly Media, 2019, 978-1491974292
- Learning Apache Flink, Packt Publishing, 2017, 978-1786466228

- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, O'Reilly Media, 2015, 978-1449358549
- Elasticsearch in Action, Second Edition, Manning Publications, 9781617299858

Course Code : **DSM DE705**
Course Title : **Customer Behavior Analysis and Fraud Detection**
L-T-P/S=Credits : **3-0-1**

Unit 1: What is "customer analytics" and why do we do it?, Specific Loyalty Matrix tools & biases. Goals, Understanding customer needs, preferences, and motivations, Quantitative data (purchase history, website traffic, social media interactions), Qualitative data (customer surveys, feedback, interviews), A Brief review of what needs to be done before serious analysis can start, Sourcing business requirements, Sourcing raw data, Profiling raw data, Data quality control & remediation.

Unit 2: Association Rule Mining(Apriori algorithm), Predictive Modeling (Predicting future customer behavior using machine learning algorithms), Segmentation & Prediction: Decision Tree method, Clustering method, Using Customer Data for Personalization, Customer Lifetime Value (CLTV) Analysis, Churn Analysis

Unit 3: Define fraud, different types of fraud, and the role of data science in fraud detection, Implementation of EDA and machine learning in fraud detection

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of customer behaviour analysis	CO1
2	To be able to understand analytics to uncover patterns in customer behaviour	CO2
3	Provide skills to build customer segmentation models	CO3
4	Introduce real-time fraud detection pipelines using big data tools	CO4

Text & References:

- Behavioral Research Data Analysis with R- Yuelin Li, Jonathon Baron

Course Code : **DSM DE707**
Course Title : **Blockchain**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction: Key Innovations/Revolutions in Human History, Evolution of Economic System, Money, Ledger and Accounting, Digital Revolution & its impact across industries, Digital Money, DigiCash, E-Gold, History & Fundamentals of Blockchain, Key Elements of Blockchain, Blockchain Protocols, Benefits of Blockchain, Types of Blockchain, Centralized vs. De-centralized systems

Unit 2: Cryptography, Symmetrical vs. Asymmetrical Encryption, Secure Hash Algorithms (SHA Family), Comparison of SHA functions, Digital Signatures

Unit 3: Consensus Mechanisms – Objectives & Need, Proof of Work (PoW), Proof of Stake (PoS), Types & Functions of Node, Construct of Block, Public vs. Private Keys, Types of Ledger, Ledgers, Distributed Ledger Technology (DLT), Benefits & Use Cases of DLT, Smart Contract

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental principles of blockchain technology	CO1
2	Explore how blockchain can be integrated with data science	CO2
3	Identify potential of Blockchain technology in different types of data	CO3
4	Develop simple blockchain applications	CO4

Text & References:

- Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World, 978-1101980149, 978-1260026672, Don and Alex Tapscott
- Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyon, by Chris Burniske and Jack Tatar
- The Basics of Bitcoins and Blockchains: An Introduction to Cryptocurrencies and the Technology that Powers Them (Cryptography, Derivatives Investments, Futures Trading, Digital Assets, NFT) Hardcover – Illustrated, September 15, 2018, Antony Lewis.
- The Blockchain Developer: A Practical Guide for Designing, Implementing, Publishing, Testing, and Securing Distributed Blockchain-based Projects 1st ed. Edition, Apress, 978-1484248461, 2019

Course Code : **DSM DE709**
Course Title : **Next Generation Sequencing Data Analysis**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to NGS and NGS Technologies: Introduction to sequencing technologies from a data analysts view, Concept, Applications of sequencing technologies in Whole genome assembly; Gene expression analysis; Genome annotation; Gene regulation analysis; Variation studies

Unit 2: NGS data analysis: Preprocessing: Introduction to NGS data analysis, Raw sequence files (FASTQ format), Preprocessing of raw reads: quality control (FastQC), adapter clipping, quality trimming, Introduction to read mapping (Alignment methods, Mapping heuristics), Read alignment to a reference genome (BWA, Bowtie2, TopHat), Mapping output (SAM/BAM format), Usage of important NGS toolkits (samtools, BEDtools), Mapping statistics, Visualization of mapped reads (IGV, UCSC)

Unit 3: NGS data analysis: Variant calling, DNA variant calling and Filtering DNA variants, NGS data analysis: Gene Expression Analysis, Differential Expression analysis, Quantification of expression,

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the basic principles and technology in different NGS platform	CO1
2	To be able to explain the workflow of NGS data analysis	CO2
3	To be able to perform preprocessing of the data and downstream analysis	CO3
4	Design and analyse NGS-based projects,	CO4

Text & References:

- Exploring Personal Genomics, 1st Edition, Joel T. Dudley, Konrad J. Karczewski, ISBN-13: 9780199644490, Oxford University Press, 2013
- High-Throughput Next Generation Sequencing, Methods and Applications,Editors: Kwon, Young Min, Ricke, Steven C. (Eds.), ISBN 978-1-61779-089-8, Springer, 2011
- Next Generation Sequencing Technologies and Challenges in Sequence Assembly, El-Metwally,
  M.Sc, Sara, Ouda, Osama M., Helmy, Mohamed, ISBN 978-1-4939-0715-1, Springer 2014

(Approved Under Section 2(f) and 12(B) of UGC act of 1956)

Shri Mata Vaishno Devi University

School of Computer Science and Engineering
Faculty of Engineering, SMVDU
Kakryal, Katra 182320

Annexure 19.3

Course Structure of the Part Time M.Sc. Data Science Programme

[Entry Batch 2026]
(6 Semester, 3 Years)

Semester – I

Course Code	Course Name	Type of Course	L-T-P	Credits
DSM MA601	Introduction to Linux and Database technologies	MA-1	3-0-2	4
DSL NC601	Mathematics Essential	NC-1	3-0-0	NC
DSL DC603	Statistics and Probability for Analytics	DSC-1	3-0-0	3
DSP DC603	Statistics and Probability for Analytics Lab	DSC-1	0-0-2	1
-	Elective-1	DSE-1	3-0-2	4
				Total Credit=12

Semester – I (Bridge and Non-Credit (NC) course)
Mathematics Essential is going to be compulsory for students who have no exposure to mathematics in graduation or intermediate.

Semester – I (Electives)
Students can choose any one elective from the list.

1. Programming in R (DSM DE601)
2. Programming in Python (DSM DE603)

Semester – II

Course Code	Course Name	Type of Course	L-T-P	Credits
DSL DC602	Machine Learning	DSC-2	3-0-0	3
DSP DC602	Machine Learning Lab	DSC-2	0-0-2	1
DSL DC604	Regression Theory & Analysis	DSC-3	3-0-0	3
DSP DC604	Regression Theory & Analysis Lab	DSC-3	0-0-2	1
-	Elective-2	DSE-2	3-0-2	4
				Total Credit=12

Semester – II (Electives)
Students can choose any one elective from the list.

1. Cloud Computing (DSM DE606)
2. Pattern Recognition (DSM DE608)

Semester – III				
Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC601	Data Structure & Algorithm Design	DSC-4	3-0-0	3
DSP DC601	Data Structure & Algorithm Design Lab	DSC-4	0-0-2	1
DSL DC703	Generalized & Linear Modeling	DSC-5	3-0-0	3
DSP DC703	Generalized & Linear Modeling lab	DSC-5	0-0-2	1
-	Elective -3	DSE-3	3-0-2	4
				Total Credit=12

Semester – III(Electives)
Students can choose any one elective from the list.

1. Big Data Tools &Technologies-I (DSM DE605)
2. Customer Behavior Analysis and Fraud detection (DSM DE705)

Semester – IV				
Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC606	Data Visualization	DSC-6	3-0-0	3
DSP DC606	Data Visualization lab	DSC-6	0-0-2	1
DSL DC605	Research Methodology	DSC-7	4-0-0	4
-	Elective-4	DSE-4	3-0-2	4
				Total Credit=12

Semester – IV (Electives)
Students can choose any two electives from the list.

1. Big Data Tools & Technologies-II (DSM DE602)
2. Time Series (DSM DE604)

Semester – V				
Course Code	Course Name	Type of Course	L-T-P	Credit
DSL DC701	Deep Learning & Neural Networks	DSC-8	3-0-0	3
DSP DC701	Deep Learning & Neural Networks Lab	DSC-8	0-0-2	1
DSL DC705	Graph & Social Network Analysis	DSC-9	3-0-0	3
DSP DC705	Graph & Social Network Analysis Lab	DSC-9	0-0-2	1
-	Elective-5	DSE-5	3-0-2	4
-	Elective-6	DSE-6	3-0-2	4
				Total Credit=16

Annexure 19.3

Semester – V (Electives)
Students can choose any two electives from the list.

1. Big Data Tools & Technologies-III (DSM DE703)
2. Next Generation Sequencing Data Analysis (DSM DE709)
3. Emerging technologies in AI (DSM DE711)
4. Natural Language Processing (DSM DE701)
5. Blockchain (DSM DE707)

Semester – VI

Course Code	Course Name	Type of Course	L-T-P	Credit
DSD PR701	Research Thesis / Project /Patent	PR	0-0-40	20

Total Credit=24

Semester – VI (Electives)
Students can choose any one elective from the list to be taken by SWAYAM

1.

Outcomes expected of Research thesis/project/Patent **PG Programmes**

Semester – VI
The following outcomes must be achieved by the end of the VI Semester:

i. Research Problem Identification
ii. Review of literature
iii. Research design formulation
iv. Commencement of experimentation, fieldwork, or similar tasks
v. Completion of experimentation/ fieldwork
vi. Submission of dissertation, presentation with Viva Voce
vii. Research output in the form of any one of the following –
- Prototype or product development/ patent
- Publication/Presentation of research work in National/International Conference/conference proceedings
- Publication in a reputed journal such as Scopus-indexed journals or other similar quality journals
- Book or Book Chapter in a publication by a reputed publisher.
- Development of software or Database/Analysis pipeline/ Extensive analysis of Data
- Any other scholastic work as recommended by the School Research Committee and approved by the competent authority.

Annexure 19.3

Summary							
	First Semester	Second Semester	Third Semester	Four Semester	Five Semester	Six Semester	Total
Credits	12	12	12	12	16	20	84

Annexure 19.3

Syllabus - First Semester

Course Code : **DSM MA601**
Course Title : **Introduction To Linux and Database Technologies**
L-T-P/S=Credits : **3-0-1**

Unit 1: Operating System: Difference between Windows and Unix/Linux, Definition; Concepts; Function of Operating System; UNIX/LINUX Operating Systems Concepts; Layers of UNIX; Tree Structure of UNIX; Root File System; /bin Directory; /dev Directory; /bin Directory; /etc Directory; /lib Directory; /proc Directory; /mnt Directory; /root Directory; /sbin Directory; /tmp Directory; /var Directory; Relative Path; Absolute Path; Creation of Directory; Creating file; removing file; Listing Files and Directories copying file; renaming file; Changing File Permission; Changing Directory Permission; Changing Group; Changing Owner; Pipe; Filters; pwd command; date command; head command; tail command less command; more command; grep command; VI Editor (Creating a new File; Inserting Text in File; Deleting Text in File; Copy , Cut & Paste Text; Save File). System monitoring, process information, log files, run levels, Introduction to Process in Unix/Linux: Process; Process State (New, Running, Waiting, Ready, Termination); Introduction to Management of Memory; Swapping Virtual Memory; Paging.

Unit 2: Introduction: Relational Databases, NoSQL Databases, Multimodal Databases. Working with Relational Databases using SQL (Structured Query Language), Data Definition Language (DDL), Data Manipulation Language (DML), Basic SQL queries, Integrity constraints on tables,Data Control Language Commands(DCL): Grant and Revoke SQL Functions, SQL querying to do operations such as identifying nulls, special characters, blank rows/columns, and run distributions, run data summaries, merge tables, get unique counts, SQL Joins, Aggregate functions and GROUP BY, Nested queries and sub queries. GROUP BY CLAUSE along basic aggregations such as SUM, COUNT, AVG RANK(), ROWNUM() & DENSE_RANK. UNION and UNION ALL CASE statement. Introduction to Advanced SQL concepts: Indexes, Sequence, Clusters, Views, Cursors and Triggers, Embedded SQL, Introduction to SQLITE

Unit 3: Introduction to NoSQL and its capabilities, Contrast MySQL with NOSQL, MongoDB: A Database for the internet, MongoDB's key features, Diving into the MongoDB shell, Introduction to Creating and Querying with indexes, basic Administration and writing Program using MongoDB. Application Development in MongoDB and processing MongoDB Mastery: Indexing and query optimization, Text Search, MongoDB vs Apache Cassandra: Big Data and Apache Cassandra, Importance of Cassandra, Cassandra as a Distributed Database,Cassandra as a Schema Free Database.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of computer and operating system	CO1
2	Understand Linux operating system, file structure and basic commands	CO2
3	Establishment of basic understanding of databases: SQL - and NO-SQL databases	CO3
4	Ability to work with SQL commands and interaction with database	CO4
5	To identify the need and working with NO-SQL databases in Data Science	CO5

Text & References:

- Peter Baer Galvin. 2016. Operating System Concepts. BPB Publication · Stuart E. Madnick. 2001. Operating System. Tata Mac Graw Hill.
- Kenneth H. Roshan. 2007. The complete reference Unix Tata Mac Graw Hill..
- D. M. Dhamethire. , 2011. System Programming and Operating Systems. Tata Mac Graw Hill.
- Kirrgcox. 2001. Red Hat Linux by. Printice Hill India.
- Andrew S. Talenbaum. 2008. Modern Operating system. Printice Hill India.
- Sumetabha Das. 2013. Unix (Concept and Application). Tata Mac Graw Hill.
- Ramesh Bangia. 2015. Learning Unix. BPB Publication.
- Database System Concepts, Abraham Silberschatz, Henry F. Korth, and S. Sudarshan
- An Introduction to Database Systems, C J Date
- Fundamentals of Databases – Elmasri and Navathe.
- Database Management Systems – Raghu Ramakrishna, Johannes Gehrke
- MongoDB: The definitive guide by Kristina Chodorow, Michael Dirolf
- Cassandra: The Definitive Guide, Eben Hewitt
- MangoDB in Action, Kyle Banker, Peter Bakkum, Shaun Verch, Douglas Garrett, Tim Hawkins

Course Code : **DSL NC601**
Course Title : **Mathematics Essential**
L-T-P/S=Credits : **Non-credit**

Unit1: Set Theory - Number system, Sets and their operations, Relations and functions - Relations and their types, Functions and their types, Sequence and Limits - Function of One variable - Function of one variable,Graphs and Tangents,Limits for sequences, Limits for function of one variable, Limits and Continuity

Unit 2: Linear algebra, vector spaces and transformations in n-dimensions. Matrix Operations (Addition, Multiplication, Inversion), formulating and solving linear systems as matrix-vector equations, performing basic computations involving matrix algebra, orthogonal projections, and eigenvalue-eigenvector problems. Singular Value Decomposition (SVD), Principal Component Analysis (PCA)

Unit 3: Derivatives, Tangents and Critical points - Differentiability and the derivative, Computing derivatives and L'H'opital's rule Derivatives, tangents and linear approximation, Critical points: local maxima and minima, Partial Derivatives,

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental aspect of mathematics	CO1
2	Development of analytical and problem solving skills	CO2
3	Understanding of linear algebra, vector space, operation on matrices etc	CO3
4	Finding optimization : local maxima and minima, derivatives	CO4

Text & References:

- G. Strang, Linear Algebra and Its Applications, 4th ed. Boston, MA: Cengage Learning, 2006, ISBN: 978-0030105678.
- Mathematics for Machine Learning, Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Cambridge University Press, 978-1108455145
- C. D. Meyer, Matrix Analysis and Applied Linear Algebra, 1st ed. Philadelphia, PA: SIAM, 2000, ISBN: 978-0898714548.
- Stewart, Calculus: Early Transcendentals, 8th ed. Boston, MA: Cengage Learning, 2015, ISBN: 978-1285741550.
- S. Boyd and L. Vandenberghe, Convex Optimization, 1st ed. Cambridge, UK: Cambridge University Press, 2004, ISBN: 978-0521833783.
- M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning, 1st ed. Cambridge, UK: Cambridge University Press, 2020, ISBN: 978-1108455145.

Course Code : **DSL DC603**
Course Title : **Statistics And Probability for Analytics**
L-T-P/S=Credits : **3-0-0**

Introduction to Data analysis and Types : Data Analysis Types of Data Analysis – Descriptive, Exploratory, Predictive, Inferential, Steps in Data Analysis

Unit 1: Descriptive Statistics
Statistics: Preliminary concepts: Measures of Central Tendency: Mean, Median, Mode
Measures of Dispersion: Range, Standard deviation, Variance, Covariance, Graphical Representation of Statistics: Histograms, Bar plots, Scatter plots etc.

Unit 2: Probability: Random Experiments, Trial and Event, Sample Space, Mutually Exclusive or Disjoint Events, Mutually Exhaustive Events, Equally Probable Events, Complementary Event, Classical definition of Probability, Statistical definition of Probability, Axiomatic definition of Probability, Addition theorem, Multiplication theorem, Conditional Probability, Bayes' Theorem. Expectation. Continuous Distribution: Normal Distribution, Properties of Normal distribution

Unit 3: Correlation: Bivariate distribution Correlation, Types of Correlation, Simple Correlation Coefficient for ungrouped data, Properties and Interpretation of Correlation Coefficient, Coefficient of determination, Scatter diagram, Standard Error, Probable error of Correlation Coefficient. Rank correlation, Some examples.

Unit 4: Introduction to the Inferential Statistics: Parameter, Statistic, Null hypothesis, Alternative hypothesis, Critical region, Typel Error, Type 11 Error, Level of significance, P-value and its applications. Test of Significance for Small samples: One sample t-test, Paired t-test, Degrees of freedom for t-test, F test for equality of Population variances, Degrees of freedom for F-test. Test of Significance for Large samples: Normal test for sample mean and population mean, Normal test for two sample means. Chi-square Test: Test of goodness of fit, Test of Independence of attributes. Degrees of freedom for Chi Square test, Coefficient of contingency, Yates' correction for continuity. Analysis of Variance: One way and Two way (only Examples) , Introduction to Model Building: Basics of Model building, Definition of a Model. Point estimation, Confidence intervals, Testing

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4
5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- S. Ross, A First Course in Probability, 10th ed. Boston, MA: Pearson, 2018, ISBN: 978-0134753119.
- L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, 1st ed. New York, NY: Springer, 2004, ISBN: 978-0387402727.
- G. Casella and R. L. Berger, Statistical Inference, 2nd ed. Boston, MA: Cengage Learning, 2001, ISBN: 978-0534243128.
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174
- Practical Statistics for Data Scientists, Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : **DSP DC603**
Course Title : **Statistics And Probability for Analytics Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Descriptive Statistics: Calculation and interpretation of Descriptive Statistics

Unit 2: Probability: Understanding the preliminary concepts of probability through case studies. Continuous Distribution: Generation of statistical distributions

Unit 3: Correlation: Calculation of correlation understanding and interpreting correlation through case studies

Unit 4: Introduction to the Inferential Statistics: Understanding inferential statics through case studies, Introduction to Model Building: Understanding Basics of Model building. Definition of a Model. Point estimation, Confidence intervals, Testing through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basics of statistics and their advancement for big data analysis	CO1
2	Description of data points distribution to understand their range, central tendencies (mean, median), and dispersion (variance, standard deviation).	CO2
3	Understanding various graphics plots such as bar, histogram and pi-chart etc.	CO3
4	Understanding probability, correlation and their application	CO4

5	Identification of statistical errors and methods for a data set	CO5

Text & References:

- A first course in Probability, Sheldon Ross
- An introduction to Probability and Statistics, Vijay K. Rohatgi and A. K. Md. Ehsanes Saleh
- Biostatistics: A foundation for analysis in the Health Sciences, W.W Daniel. Publisher: John Wiley and Sons.
- Biostatistics, P.N Arora and P.K Malhan. Publisher: Himalaya Publishing House.
- Practical Statistics for Data Scientists, Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 978-1492072942

Course Code : DSM DE601
Course Title : Programming In R
L-T-P/S=Credits : 3-0-1

Unit 1: An Introduction to R, Overview of R programming, applications, usage and comparative study with other softwares and introduction to R for Data Science, Setting up the R environment and packages. Setting up R environment and install packages and supporting libraries in R, Basics of R : Manage your data and workspace, Save your work, How to use R, Data structure in R, Data creation and curation and special function using R, Matrices and Lists

Unit 2: File Handling: Reading different file format using R, file handling and processing, writing output file. Graphics using R: Graphics device, Basic plot function, scatter plot, 3-D scatter plot, pair plots, Lineplot, MatplotMatpoints, Bar plot, Histogram plot, Density plot, Dot plot, Pie chart, Venn diagram, Grid graphics, Lattice, ggplot2, Interactive plotting, combine multiple plots in same graphics screen, save graphics to a file., Conditional Executions, Comparison Operators, Logical Operators, Control Structures, If statements, Ifelse statements, Loops, For loop, While loop, Apply loop family, Other loops.

Unit 3: Functions: Define and Call functions, Syntax Rules for functions, Control utilities for functions, Writing own function, Advance R, Advance R functions and Regular expressions, Object oriented programming, Building R package

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of programming in R and basic of setup environment	CO1
2	Develop Skills for Data Manipulation	CO2
3	Enable Data Visualization	CO3
4	Perform statistical analysis on various datasets	CO4
5	Application of R in machine learning and artificial intelligence	CO5

Text & References:

- A Handbook of Statistical analysis using R, Brain Everitt and Torsten Hothorn
- The art of R programming, Norman Matloff
- Data Analysis and Graphics using R, W. John Braun
- R Graphics, Paul murrell

- R for Data Science, Garrett Grolemund and Hadley Wickham
- Linear Models with R, Julian J. Faraway
- An Introduction to Statistical Learning: with Applications in R,Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani,Springer, 978-1071614174

Course Code : **DSM DE603**
Course Title : **Programming In Python**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Python: Overview of Python, applications, usage and comparative study with other software. Basics of Python : Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), Basic Programming with Python:If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass,

Unit 2: Data Structures of Python: Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, Operations, Function and Methods, Files, Modules, Dictionaries, Functions and Functional, Programming, Declaring and calling Functions, Declare, assign and retrieve values from Lists, Introducing Tuples, Accessing tuples

Unit 3: Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception, Exception Handling, Except clause, Try finally clause, User Defined Exceptions, Python Libraries : Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental of programming in python and basic setup	CO1
2	Develop Skills for Data Manipulation	CO2
3	Perform statistical analysis on various datasets	CO3
4	Application of python in machine learning and artificial intelligence	CO4

Text & References:

- Think Python by Allen B. Downey
- Introducing Python by Bill Lubanovic
- Hello World by Warner Sande and Carter Sande
- Learning Python , 5th Edition , Mark Lutz
- Python For Data Analysis by W Mckinney
- Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media

Syllabus - Second Semester

Course Code : **DSL DC602**
Course Title : **Machine Learning**
L-T-P/S=Credits : **3-0-0**

Annexure 19.3

Unit 1: An Introduction to Machine Learning: What is ML, Types of ML, Applications used for ML, AI vs ML, Essential for ML and AI. Techniques of Machine Learning: Introduction to Supervised, unsupervised, semi-supervised, and reinforced machine learning techniques.

Unit 2: Unsupervised learning: Basic concepts to calculate distances: Concepts to calculate distances: euclidean distance, Manhattan distance, L1 and L2-norm, Minikowski distance etc., Discuss theory and concepts behind unsupervised learning. Clustering: Finding grouping of data. Hierarchical clustering, k-means and K-medoid clustering algorithm, Dimension Reduction: PCA, t-SNE.

Unit 3: Supervised learning: Fundamentals of Supervised machine learning: Components of supervised learning model: Data, Model: (Classification and regression), Training, Validation, Evaluating, Inference and model's complexity: over and underfitting, Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based methods (Decision tree and Random Forest), Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Key concepts: Agent, Environment, State, Action, Reward, applications, advantages and disadvantages

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Machine Learning by Tom M. Mitchell
- Introduction to Machine Learning, Ethem Alpaydın
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie , Robert Tibshirani , Jerome Friedman

Course Code : DSP DC602
Course Title : Machine Learning Lab
L-T-P/S=Credits : 0-0-1

Unit 1: Introduction: Understanding classification and regression problem. Splitting the data into Training and test data, Understanding Overfitting and complexity with case studies,

Unit 2: Unsupervised learning: Finding grouping of data: Hierarchical clustering, k-means and K-medoid clustering algorithm, dimensionality reduction through case studies

Unit 3: Supervised learning: Working with different models: Binary Classification, Logistic Regression (Cost Function, regularization, etc), KNN, Naive Bayes decision, SVM, Tree based

methods (Decision tree and Random Forest). Ensemble based classifiers (Bagging, Boosting, Stacking and cascading)

Unit 4: Reinforcement learning: Solve a case study on real world data and apply reinforcement learning concepts.

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamental concept of machine learning	CO1
2	Understanding basic concept of supervised and unsupervised learning and algorithms	CO2
3	Apply clustering algorithm to find patterns in the data	CO3
4	Understand and apply machine learning algorithms	CO4
5	Practical implementation of using tools in real world data	CO5

Text & References:

- Programming Collective Intelligence by Toby Segaran
- Machine Learning for Hackers by Drew Conway and John Myles
- Machine Learning by Tom M. Mitchell
- Pattern Recognition and Machine Learning by Christopher M. Bishop (Author)
- Machine Learning Yearning by Andrew NG
- The Elements of Statistical Learning by Trevor Hastie, Robert Tibshirani , Jerome Friedman

Course Code : **DSL DC604**
Course Title : **Regression Theory & Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination. Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction, Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit

Unit 2: Transformations: Variance stabilizing transformations, Transformations to linearize the model, Methods to select a transformation, weighted least squares, Regression and random effect, Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity

Unit 3: Validation of regression models: Techniques for validation, Nonlinear least squares, transformations, Parameter estimation, Logistic and Poisson regression

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : DSP DC604
Course Title : Regression Theory & Analysis Lab
L-T-P/S=Credits : 0-0-1

Unit 1: Simple linear regression: Estimation of the Parameters, Hypothesis Testing on the Slope and Intercept, Interval Estimation in Simple Linear Regression, Prediction of New Observations, Coefficient of Determination through case studies Multiple regression: Estimation of Parameters, Hypothesis Testing, Confidence Intervals, Prediction through case studies

Unit 2: Model Adequacy testing: Residual Analysis, PRESS statistic, Lack of Fit through case studies

Unit 3: Multicollinearity: Sources, Effects, Diagnostics, Methods of dealing with Multicollinearity through case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding regression analysis for predictive model development	CO1
2	To be able to handle linear and multiple regression	CO2
3	Identification of patterns in Data	CO3
4	Establishment of relationship between variables	CO4

Text & References:

- Introduction to Linear Regression Analysis, by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining
- Introduction to Regression Analysis, M. Golberg and H.A Cho
- Applied Regression Analysis, Norman R. Draper and Harry Smith

Course Code : DSM DE606
Course Title : Cloud Computing
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to cloud computing: History and evolution, Properties, Characteristics & Disadvantages, Pros and Cons of Cloud Computing, Benefits of Cloud Computing, AWS, Azure, Fundamentals of Cloud Computing: the primary deployment and service models., data warehousing, components, and architectures of data warehousing,

Unit 2: Cloud Computing Architecture: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Cloud Deployment Model: Cloud computing platform, Private cloud deployment, Public cloud deployment, Hybrid cloud deployment

Unit 3: Cloud security: Data on virtual network, Infrastructure Security, Network-level security, Host level security, Data Security and Storage, Cloud Access: authentication, authorization, and accounting

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the basic principle of cloud computing and service model	CO1
2	How cloud computing serve as cost effective and scalable platform	CO2
3	Enable students to deploy data science models and use of cloud computing instances	CO3
4	Establishment of data security protocols	CO4

Text & References:

- Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Shahed Latif, Subra Kumaraswamy, and Tim Mather
- Enterprise Cloud Computing: Technology, Architecture, Applications, Gautam Shroff,
- Cloud Computing: Concepts, Technology & Architecture, Ricardo Puttini, Thomas Erl, and Zaigham Mahmood

Course Code : DSM DE608
Course Title : Pattern Recognition
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction: The basic concepts of pattern recognition, such as frequent pattern, closed pattern, max-pattern, and association rules. Pattern mining methods (Clustering, Apriori, ECLAT, and FPgrowth),pattern evaluation: lift, chi-square, cosine, Jaccard, and Kulczynski,

Unit 2: Methods for recognition of multi-level patterns: multi-dimensional patterns, qualitative patterns, negative correlations, compressed and redundancy-aware top-k patterns, and mining long (colossal) patterns.

Unit 3: Sequential pattern recognition methods: GSP, SPADE, PrefixSpan, and CloSpan, Graph pattern recognition methods: subgraph pattern mining, such as gSpan, CloseGraph, graph indexing methods, mining top-k large structural patterns in a single large network, and graph mining applications, such as graph indexing and similarity search in graph databases.

Unit 4: Use of Dictionary learning and collaborative filtering in pattern recognition, Pattern recognition applications: spatiotemporal and trajectory patterns, pattern mining in data streams, software bug mining, pattern discovery for image analysis

Course Outcome:

S.N.	Course Outcome	CO

1	Understanding the basic concept of pattern recognition, mining and evolution method	CO1
2	Finding multilevel patterns	CO2
3	Application of Sequential pattern recognition methods	CO3
4	Modern approaches to filter patterns such as dictionary learning	CO4

Text & References:

- R.Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, NY, 1992.
- Duda R O and P E Hart, Patten classification and scene analysis, John Wiley & Sons, NY 1973
- K.S.Fu, Syntactic pattern recognition and applications, Prentice Hall, NJ, 1982
- T.Pavlidis, Structural pattern recognition, Springer-Verlag, NY, 1977
- D.H.Ballad and C.M.Brown, Algorithms for computer vision, Prentice Hall, 1982

Syllabus - Third Semester

Course Code : **DSL DC601**
Course Title : **Data Structure & Algorithm Design**
L-T-P/S=Credits : **3-0-0**

Unit I: Introduction and Programming strategies Introduction, Objects and ADTs with example, Constructors and destructors, Data structure, methods, Pre and post conditions, C conventions, Error handling, some programming language notes. Data structures: Arrays; lists; stacks and stack frames; Recursion -Recursive functions with example of factorial, Queue, Degeue.

Unit II: Searching, Queues and Sorting, Sequential and binary search, Trees, binary search tree, complexity. Red-Black trees, AVL trees, general n-ary trees, hash tables; Hashing and collision resolution, Priority queues and heaps, Sorting: Bubble, Heap, Quick, Bin, Radix, Dynamic algorithm Fibonacci numbers, bionomial coefficients, optimal binary search trees, matrix chain multiplication, longest common subsequence, optimal triangulation.

Unit III: Graphs, Minimum spanning tree and Djkstra's algorithm, Huffman encoding, FFT, Hard or intractable problems Eulerian or Hamiltonian paths, Travelling salesman problem.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding different types of data structures and their working principle	CO1
2	Able to describe different Data structures	CO2
3	Explain various sorting and searching algorithms.	CO3
4	Describe and solve computational problems including graphs and encoding.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGraw-Hill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.
- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Annexure 19.3

Course Code : **DSP DC601**
Course Title : **Data Structure & Algorithm Design Lab**
L-T-P/S=Credits : **0-0-1**

Unit I : Stack implementation through arrays, link list, Programs for recursion functions, Implementation of queues and leap structures, Application of binary trees in pre-order, post-order and in-order evaluation,

Unit II : A VL tree implementation, Optimal matrix multiplication

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding and implementation of different data structures	CO1
2	Understanding of Binary Trees and implementation for expression evaluation.	CO2
3	Describe and implement dynamic algorithm for Matrix Chain Multiplication.	CO3
4	Implement of Recursion Function, Heap and Queues Data Structures.	CO4

Text & References:

Text:

- Data Structures and Algorithms, A.V. Aho, J.E. Hopcroft and J. Ullman, Addison-Wesley Publishing
- Database Design, Development and Deployment with Student CD, P. Rob and E. Semaan, McGrawHill/Irwin
- Schaum's Outline of Data Structures with C++, J.R. Hubbard, McGraw Hill Trade.

References:

- Database system concepts, A. Silberschatz, P.B. Galvin and G. Gagne, John Wiley and Sons Inc.
- Introduction to Data Structures and Application, J. Tremblay and P.G. Sorensen, McGraw Hill College Division

Course Code : **DSL DC703**
Course Title : **Generalized & Linear Modeling**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to Generalized Linear Models: Linear model, Non linear model, GLM, Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares

Unit 2: Nonlinear Regression Models: Linear and Nonlinear regression models, transforming a linear model, Parameter estimation Logistic and Poisson Regression Models: Logistic regression, Poison regression, Overdispersion

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation, ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson
- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill
- Categorical data analysis, Ala Agresti

Course Code : **DSP DC703**
Course Title : **Generalized & Linear Modeling Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Linear Regression Models: Multiple regression model, Parameter estimation, Maximum likelihood, Model adequacy checking, Weighted least squares with case studies

Unit 2: Nonlinear Regression Models: Linear and Non linear regression models, transforming a linear model, Parameter estimation with case studies, **Logistic and Poisson Regression Models:** Logistic regression, Poison regression, Overdispersion with case studies

Unit 3: The Generalized Linear Model: Exponential family, Likelihood equations, Quasi likelihood, Gamma family, The power function, Generalized Estimating Equation with case studies, ordinary least squares (OLS), model selection and evaluation: AIC, BIC, cross-validation, and likelihood ratio tests.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the theory and assumptions underlying linear regression models	CO1
2	Analyse and interpret the results of linear regression and multiple linear regression	CO2
3	Application of Nonlinear Regression Models and OLS	CO3
4	Conduct model selection and evaluation	CO4

Text & References:

- Generalised Linear Models: With Applications in Engineering and the Sciences by Raymond H. Mysers, Douglas C. Montgomery, G. Geoffrey Vining, Timothy J. Robinson

- Data Analysis Using Regression and Multilevel/ Hierarchical Models, Andew Gelman and Jennifer Hill
- Categorical data analysis, Ala Agresti

Course Code : **DSM DE605**
Course Title : **Big Data Tools & Technologies-I**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Big Data Big Data and its Importance, V's of Big Data, Drivers for Big Data, Introduction to Big Data Analytics, Big Data Analytics applications. Characteristics of Data: Structured, Semi-structured and Unstructured data, data at rest, data in motion, etc, good data versus bad data, Challenges in handling large data sets,

Unit 2: Introduction to Hadoop: Hadoop Introduction, Hadoop key characteristics, Hadoop Core Components, Hadoop Ecosystem, Hadoop Services, Different Hadoop Modes, Hadoop's Parallel World – Data discovery, Hadoop Installation, Hadoop Cluster Setup, Storing Data in Hadoop, HDFS Commands (Basics, Intermediate and Advanced), Processing your data with Map Reduce, Run test example using Map reduce, Distributed Cache, Map side join vs Reduce Side Join, Running Hadoop applications, Processing Big Data: Integrating disparate data stores, Mapping data to the programming framework, Connecting and extracting data from storage, Transforming data for processing, Subdividing data in preparation for Hadoop Map Reduce. Hadoop Map Reduce: Employing Hadoop Map Reduce, Creating the components of Hadoop Map Reduce jobs, Executing / Monitoring Hadoop Map Reduce jobs.

Unit 3: Fundamental of Apache Pig: What is Pig?, Introduction to Pig Data Flow Engine, Pig and Map Reduce in Detail, When should Pig Used?, Pig and Hadoop Cluster, Pig Interpreter and Map Reduce, Pig Relations and Data Types, Debugging and Generating Example in Apache Pig, Pig philosophy and architecture, Pig installation, Grunt shell, Loading data, Exploring Pig Latin commands, Pig Transformations functions, Joins in Pig, Hands on Exercises

Unit 4: Fundamental of Apache Hive: What is Hive?, Architecture of Hive, Hive Services, Hive Clients, How Hive Differs from Traditional RDBMS, Introduction to HiveQL, Data Types and File Formats in Hive, File Encoding , Common problems while working with Hive Hive architecture, Hive installation, Hive vs. RDBMS, HiveQL and the Hive shell, Data types and schemas, Creating tables (external vs. managed), Creating Partitions, Creating Views, UDF function in java in Hive, Using hive to create diff types of format

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the concept of big data, characters of big data and structure	CO1
2	Establishment of platform to handle and manipulate big-data	CO2
3	Apply Hadoop, Apache pig and hive palatiform to big data analytics	CO3
4	To acquire skills for Challenges in handling large data sets	CO4

Text & References:

- Big Data and Analytics, Seema Acharya , Subhashini Chellappan
- Professional Hadoop Solution, Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich

- Programming PiG, Wiley Publication, Alan Gates
- Programming Hive, Wiley Publication, Jason Rutherglen, Dean Wamplor& Edward Capriolo

Course Code : **DSM DE705**
Course Title : **Customer Behavior Analysis and Fraud Detection**
L-T-P/S=Credits : **3-0-1**

Unit 1: What is "customer analytics" and why do we do it?, Specific Loyalty Matrix tools & biases. Goals, Understanding customer needs, preferences, and motivations, Quantitative data (purchase history, website traffic, social media interactions), Qualitative data (customer surveys, feedback, interviews), A Brief review of what needs to be done before serious analysis can start, Sourcing business requirements, Sourcing raw data, Profiling raw data, Data quality control & remediation.

Unit 2: Association Rule Mining(Apriori algorithm), Predictive Modeling (Predicting future customer behavior using machine learning algorithms), Segmentation & Prediction: Decision Tree method, Clustering method, Using Customer Data for Personalization, Customer Lifetime Value (CLTV) Analysis, Churn Analysis

Unit 3: Define fraud, different types of fraud, and the role of data science in fraud detection, Implementation of EDA and machine learning in fraud detection

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the fundamentals of customer behaviour analysis	CO1
2	To be able to understand analytics to uncover patterns in customer behaviour	CO2
3	Provide skills to build customer segmentation models	CO3
4	Introduce real-time fraud detection pipelines using big data tools	CO4

Text & References:

- Behavioral Research Data Analysis with R- Yuelin Li, Jonathon Baron

Syllabus - Fourth Semester

Course Code : **DSL DC606**
Course Title : **Data Visualization**
L-T-P/S=Credits : **3-0-0**

Annexure 19.3

Unit 1: Introduction to data visualization: Introduction to data visualization. Explore and manipulate all graphical parameter. Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact based team collaboration. Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529
- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : **DSP DC606**
Course Title : **Data Visualization Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Drawing graphs such as scatterplot, stripchart, histogram, boxplot, violoin chart, bar chart, dot plot, line graph, line plot, multiplot, stacked bar chart, pie diagram, venn diagram, pair plot. Explore 3d chart, motion chart, interactive charts. Discuss selecting appropriate chart for strategy presentation. Save graphs in various formats. Discuss different ways to combine multiple figures. Discuss margin and graph size.

Unit 2: Introduction to Tableau: Creating Visual Analytics with Tableau Desktop, Connecting to your data, Building your first visualization, Creating Calculations to enhance your data, using maps to improve insight, developing an Ad Hoc Analysis Environment, Tips Tricks and Time Savers, Bringing it all together with dashboards. Installing Tableau Server, Using Tableau server to facilitate fact based team collaboration, Automating Server with Tableau's command line tools, Use case for rapid fire visual analytics, other Tools. Explore various case studies to visualize the data.

Unit 3: Introduction to PowerBI: Architecture, Data Transformation, DAX, Power BI Service, Data Visualization, Power BI Reports, Connectivity Modes, Case studies

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding principle and objective of data visualization	CO1
2	Apply techniques such as data extraction, cleaning and transformation	CO2
3	To be able to create interactive dashboards using Tableau and PowerBI	CO3
4	Integration of real time data	CO4

Text & References:

- Tableau your data, Daniel G Murray
- Streaming Data: Understanding the Real-Time Pipeline, Andrew G. Psaltis, Manning Publications, 978-1617296529
- Power BI for the Excel Analyst (1st Edition), Wyn Hopkins, Packt Publishing, 2024

Course Code : DSL DC605
Course Title : Research Methodology
L-T-P/S=Credits : 4-0-0

Unit 1: Research: Definition, Characteristics, Objectives, Research and Scientific method Types of Research: Descriptive vs. Analytical Research, Applied vs. I, Fundamental Research, Quantitative vs. Qualitative Research, Conceptual vs. Empirical Research, Research Problem: Research Problem, Selecting the problem, Necessity of Defining the problem, Technique Involved in Defining a Problem.

Unit 2: Research Design, and Data Collection: Research Design: Meaning, Need, Features of Good Design, Concepts, Relating to Research Design, Different Research Design, Basic Principle of Experimental Designs. Data Collection: Observation Method, Interview Method, Questionnaires, Case Study Method. Big Data and Data Science, why data science? Who is a Data Scientist? Data Science Life Cycle

Unit 3: Data Science Process: Overview, defining goals, Retrieving data, Data preparation, Data exploration, Data modeling, Presentation. Sampling, Measurement and Scaling Techniques Sampling: Steps in Sampling Design, Different Types of Sample Designs, Measurement and Scaling: Measurement in Research, Measurement Scales, Technique of Developing Measurement Tools, Scaling, Important Scaling Techniques. Data processing: Data description, Data processing, Importance of data preparation, feature engineering and scaling of datasets, Data imbalance, Data normalization and standardization, Treatment of missing values

Unit 4: Ethics in Research and Data Science: Research ethics, Data Science ethics: Doing good data science, – Owners of the data, Valuing different aspects of privacy, Getting informed consent, The Five Cs, Diversity, Inclusion.Report Writing: Significance of Report Writing, Different Steps in

Writing Report, Layouts of Research Report, Types of Reports, Oral Presentation, Mechanics and Precautions for Writing Research Report.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the Fundamentals of Research, formulation of research problems and hypothesis	CO1
2	Understanding how to collect data, ethics in data collection and preprocessing of data	CO2
3	Basic understanding of Ethical and Legal Aspects of Research	CO3
4	Making review of research, writing reports and presentation	CO4

Text & References:

- An Introduction to Data Science, Jeffrey Stanton, Syracuse University, SAGE Publications, Inc, ISBN: 978-1506377537, 2017
- A Simple Introduction to DATA SCIENCE, Lars Nielsen, Noreen Burlingame
- Introduction to Data Science, DAN POTTER, CARSTEN BINNING,ELI UPFAL
- Davy Cielen and Arno Meysman, Introducing Data Science. Simon and Schuster, 2016.
- M. Loukides, H. Mason, and D. Patil, Ethics and Data Science. O'Reilly Media, 2018.
- C. R. Kothari, Research Methodology Methods and Techniques. 3rd. ed. New Delhi: NewAge International Publishers, Reprint 2014.
- Zina O'Leary, The Essential Guide of Doing Research. New Delhi: PHI, 2005

Course Code : **DSM DE602**
Course Title : **Big Data Tools & Technologies-II**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to Apache Spark : Introduction to Apache Spark, Features of Apache Spark, Apache Spark Stack, Introduction to RDD's, RDD's Transformation, What is good and Bad in Map Reduce , Why to use Apache Spark Spark: A Hadoop Replacement? : Java, Scala, or Python? , Scala, Packages, Data Types, Classes, Calling Functions, Operators, Control Structures, A Quick Intro to Spark: Starting the shell, Data Sources, Testing Spark, Spark Monitor, Comparing Hadoop Map Reduce to Spark, Writing Standalone program with Spark

Unit 2: Spark SQL, Basic Concepts, Using Spark SQL with RDDs, Spark Streaming: Basic Concepts, Creating Your First Stream with Scala, Creating Your First Stream with Java

Unit 3: Fundamental of Azure databricks: Control plane and data plane, Workspace (assets, objects), Build end to end pipeline (Create cluster, exploring source data, data ingestion, data preparation, Query, creating and running databrick job), Spark integration

Course Outcome:

S.N.	Course Outcome	CO
1	Establishment of fundamentals for big data technologies	CO1
2	Understanding use of distributing computing in big data	CO2
3	Apply and preparation of data processing workflow	CO3

4	Application to real-world data and analysis	CO4

Text & References:

- Spark in Action, Marko Bonaci, Peter Zecevic
- Spark Cookbook, Rishi Yadav
- Machine Learning for Big Data , Jason Bell
- Spark : The Definitive Guide, Matei Zaharia, Bill Chambers
- Azure Databricks Cookbook: Accelerate and scale real-time analytics solutions using the Apache Spark-based analytics service, by Phani Raj, Vinod Jaiswal, Packt Publishing Limited, ISBN: 978-1789809718

Course Code : DSM DE604
Course Title : Time Series
L-T-P/S=Credits : 3-0-1

Introduction: Introduction to time series

Unit 1: Autoregressive-moving average models ARIMA: Moving average models MA(q). Condition of invertability. Autoregressive models AR(p). Yull-Worker). Yull-Worker equations. Stationarity conditions. Autoregressive-moving average models ARMA (p,q).

Unit 2: Coefficient estimation in ARIMA processes: Coefficients estimation in autoregressive models. Coefficient estimation in ARMA (p) processes. Quality of adjustment of time series models. AIC information criterion. BIC information criterion. "Portmonto" statistics.

Unit 3: Forecasting in the framework of Box-Jenkins model Forecasting, trend and seasonality in Box-Jenkins model. Non-stationary time series, Non-stationary time series. Time series with non-stationary variance. Non-stationary mean. ARIMA (p,d,q) models. The use of Box-Jenkins methodology to determination of order of integration.

Course Outcome:

S.N.	Course Outcome	CO
1	Understanding the principles of autoregressive-moving average models (ARIMA)	CO1
2	Estimation of coefficients in ARIMA processes	CO2
3	Explain the methodology for forecasting	CO3
4	Apply time series models to real-world datasets for forecasting and decision-making applications.	CO4

Text & References:

- Enders W. Applied Econometric Time Series. John Wiley & Sons, Inc., 1995
- Mills, T.C. The Econometric Modelling of Financial Time Series. Cambridge University Press, 1999
- Andrew C. Harvey. Time Series Models. Harvester wheatsheaf, 1993.

Syllabus - Fifth Semester

Course Code : **DSL DC701**
Course Title : **Deep Learning & Neural Networks**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to neural networks: Conceptual overview of neural network, Supervised Learning with Neural Networks, Neural Network Representation, Computing a Neural Network's Output, Vectorizing across multiple examples, Activation functions, Derivatives of activation functions, weights and bias, Shallow neural network

Unit 2: Optimization methods: Gradient based methods, Gradient descent, Learning rate, Batch gradient descent, mini batch gradient descent, stochastic gradient descent method. Gradient free methods, Exhaustive search, Genetic algorithm, Simulated annealing,

Unit 3: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network Introduction: stride, Pooling, Flatten layers, fully connected layers.

Unit 4: Generative AI models: Introduction to generative AI and its component, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Boltzmann Machines (BMs):

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning, Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach, J. Patterson, A. Gibson · Neural Networks and Deep Learning: A Textbook, Charu C. Aggarwal · Neural Networks and Deep Learning, Michael Nielsen.

Course Code : **DSP DC701**
Course Title : **Deep Learning & Neural Networks Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Neural networks: Getting ready with your first neural network, Vectorizing across multiple examples, Calculation of Activation functions, weights and bias, Shallow neural network

Unit 2: Deep Neural Networks: Deep L-layer neural network, Building blocks of deep neural networks, Forward and Backward Propagation, Parameters vs Hyperparameters. Convolution neural network: Introduction, stride, Pooling, Flatten layers, fully connected layers.

Unit 3: Generative AI models: Application of generative AI models, Case study using real world data: Building a neural network model and generating predictions from the model using tensorflow or keras.

Course Outcome:

S.N.	Course Outcome	CO
1	Fundamental concepts of deep learning and its importance in solving complex data science problems.	CO1
2	Provide understanding of architecture, components, operations and optimization of models	CO2
3	Describe classifiers used in neural network and deep neural network; illustrate their application on large datasets.	CO3
4	Enable practical skills in designing, training, and validating deep learning models	CO4

Text & References:

- Deep learning: adaptive computation and machine learning by Bengio, Yoshua, Courville, Aaron, Goodfellow, Ian J
- Deep Learning: A Practitioner's Approach by J. Patterson, A. Gibson
- Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal
- Neural Networks and Deep Learning by Michael Nielsen

Course Code : **DSL DC705**
Course Title : **Graph & Social Network Analysis**
L-T-P/S=Credits : **3-0-0**

Unit 1: Introduction to graph theory : Use of graph theory to construct different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph, Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital.: Data and Text Mining InSocial Media: Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg, Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.
- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K. "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

Course Code : **DSP DC705**
Course Title : **Graph & Social Network Analysis Lab**
L-T-P/S=Credits : **0-0-1**

Unit 1: Study Tools for graphs and network study : Case studies will be implemented by using any programming language such as R or python. Learn other tools like Cytoscape and Gephi, and igraph, Introduction to graph theory: Input graph data : Read different network format, load network data, Construct network from matrix, Adjacency matrix. Input different types of networks : Undirected, directed graphs, Cyclic, Directed acyclic graphs, trees, weighted graphs, bipartite graph. Study data structures of the graphs. Adjacency matrix, adjacency list. Study different network models :Small world phenomena, Scale free network, Erdös-Rényi model for random graphs, Watts and Strogatz model, Barabasi-Albert model

Unit 2: Study network properties: Degree distribution, power law, graph density, graph isomorphism, path length, diameter, clique, shortest distance, clustering coefficient, network motifs, network centralities and node ranking, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, ecentricitycentracility, subgraph centrality, matching index, Network randomization: Randomization with Erdös-Rényi model, Randomization with node degree conservation, Permutation of node labels

Unit 3: Introduction to social network: Introduction to social network data, Different data format, Paths and Connectivity-Graphs to represent social relations. Working with network data- Network Datasets-Strong and weak ties - Closure, Structural Holes, and Social Capital., Data and Text Mining InSocial Media, Data mining in nutshell, Social media, Motivations for data mining in social media, Data mining methods for social media. Social networking sites: illustrative examples. Text Mining: Keyword search, query semantics and answer ranking, Classification algorithms, Clustering algorithms., Community Detection in Social Networks: Introduction, Communities in context, core methods and algorithm for community detection, Quality functions, Agglomerative/Divisive algorithms, discuss different approached for clustering.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand and explain fundamental concepts of graph theory, network science and graph representation methods	CO1
2	Apply different network models	CO2
3	Analyse and compute key network properties	CO3
4	Investigate the behaviour of social networks	CO4

Text & References:

- Networks, Crowds and Markets by David Easley and Jon Kleinberg. Cambridge University Press, 2010
- Social and Economic Networks by Matthew O. Jackson, Princeton University Press, 2010.
- Easley and Kleinberg, "Networks, Crowds, and Markets: Reasoning about a highly connected world", Cambridge Univ. Press, 2010.
- Charu C. Aggarwal, "Social Network Data Analytics", Springer, 2011.
- Robert A. Hanneman and Mark Riddle, "Introduction to social network methods", University of California, 2005.
- Jure Leskovec, AnandRajaraman, and Jeffrey D. Ullman, "Mining of Massive Datasets", Cambridge University Press, 2 edition, 2014.
- Wasserman, S., & Faust, K, "Social Network Analysis: Methods and Applications", Cambridge University Press; 1 edition, 1994.

Course Code : **DSM DE701**
Course Title : **Natural Language Processing**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to text mining: Introduction, Challenges of Text Processing, Character-Set Dependence, Language Dependence, Corpus Dependence, Application Dependence, Exploring Text Data For Preliminary Ideas: A case study for a brief introduction, Examine multiple document corpus of text, Creation of text corpus, Text pre-processing, Stemming, Word tokenisation, Word

vectorisation, Word embedding, Natural Language Processing: Sentiment Analysis, Word Frequency in text data, Wordclouds for visualizing sentiments with a case study, Wordclouds for Visualizing text data, Tidy wordclouds, Topic modelling a document

Unit 2: Text Data and Machine Learning, Clustering for text data, Clustering tweets, Regression on text data, Introduction to RTextTools, More on RTextTools, Word2Vec approach, The Doc2Vec approach, Doc2Vec approach for predicting a binary outcome, Doc2Vec approach for multi-class classification,

Unit3: Sentiment Analysis: What is sentiment, What is sentiment analysis, Scope of sentiment analysis: Document level, Sentence level, Sub-sentence level, Types of sentiment analysis, Word embeddings, Word associations or Text similarities, Building dictionaries, Summarise text, Polarity calculation, Algorithms for sentiment analysis: Rule based, Automatic Hybrid, Case study using real world data. Identify Spam-ham Emails or sms classification.

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental concepts of natural language processing	CO1
2	Apply core NLP tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition (NER).	CO2
3	Analyse and preprocess text datausing vectorization techniques	CO3
4	Understand the use of deep learning models	CO4

Text & References:

- HANDBOOK OF NATURAL LANGUAGE PROCESSING Ralf Herbrich and Thore Graepel Text Mining with R by Julia Silge and David Robinson.
- Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper.
- Taming Text by Grant Ingersoll, Thomas Morton and Drew Farris.
- Foundations of Statistical Natural Language Processing by Christopher Manning and Hinrich Schütze.

Course Code : **DSM DE703**
Course Title : **Big Data Tools & Technologies-III**
L-T-P/S=Credits : **3-0-1**

Unit 1: Apache Kafka: Data Acquisition: Apache Kafka Introduction, Apache Kafka Components (Producers, consumers, consumer groups, clusters, brokers, partitions, replicas, leaders, and followers) Integration with Hadoop system, Benefits of Stream Processing and Apache Kafka use Cases

Unit 2: Apache Flink: Understanding Stream Processing and Apache Flink, capabilities and its role in the big data ecosystem, Comparison to Apache spark and storm, Flink's architecture including JobManager and TaskManager roles, execution model and dataflow, parallelism, state management, and fault tolerance, DataStream API and Transformations: implementing transformations such as map, filter, flatMap, and reduce, Windowing and Time Semantics, Stateful Stream Processing, Flink

Table API and SQL: Performing relational operations on streaming data using Table API. Advanced Topics: Complex Event Processing (CEP) and machine learning with Flink.

Unit 3: ElasticSearch: Data ingestion, storage, enrichment, visualization, and analysis using elasticsearch, Components including: cluster, nodes, shards, replicas, analyzers and documents

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the architecture of distributed systems in handling large-scale and real-time data pipelines.	CO1
2	Design and implement real-time stream processing application	CO2
3	Apply and preparation of pipelines using Apache Kafka and Flink	CO3
4	Understand and implement Elasticsearch for full-text search, indexing, and analytics over large volumes of semi-structured data	CO4

Text & References:

- Apache Kafka Cookbook, Saurabh Minni, Packt Publishing Limited, ISBN: 978-1785882449
- Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications, O'Reilly Media, 2019, 978-1491974292
- Learning Apache Flink, Packt Publishing, 2017, 978-1786466228
- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, O'Reilly Media, 2015, 978-1449358549
- Elasticsearch in Action, Second Edition, Manning Publications, 9781617299858

Annexure 19.3

Course Code : **DSM DE707**
Course Title : **Blockchain**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction: Key Innovations/Revolutions in Human History, Evolution of Economic System, Money, Ledger and Accounting, Digital Revolution & its impact across industries, Digital Money, DigiCash, E-Gold, History & Fundamentals of Blockchain, Key Elements of Blockchain, Blockchain Protocols, Benefits of Blockchain, Types of Blockchain, Centralized vs. De-centralized systems

Unit 2: Cryptography, Symmetrical vs. Asymmetrical Encryption, Secure Hash Algorithms (SHA Family), Comparison of SHA functions, Digital Signatures

Unit 3: Consensus Mechanisms – Objectives & Need, Proof of Work (PoW), Proof of Stake (PoS), Types & Functions of Node, Construct of Block, Public vs. Private Keys, Types of Ledger, Ledgers, Distributed Ledger Technology (DLT), Benefits & Use Cases of DLT, Smart Contract

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental principles of blockchain technology	CO1
2	Explore how blockchain can be integrated with data science	CO2
3	Identify potential of Blockchain technology in different types of data	CO3
4	Develop simple blockchain applications	CO4

Text & References:

- Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World, 978-1101980149, 978-1260026672, Don and Alex Tapscott
- Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyon, by Chris Burniske and Jack Tatar
- The Basics of Bitcoins and Blockchains: An Introduction to Cryptocurrencies and the Technology that Powers Them (Cryptography, Derivatives Investments, Futures Trading, Digital Assets, NFT) Hardcover – Illustrated, September 15, 2018, Antony Lewis,
- The Blockchain Developer: A Practical Guide for Designing, Implementing, Publishing, Testing, and Securing Distributed Blockchain-based Projects 1st ed. Edition, Apress, 978-1484248461, 2019

Course Code : **DSM DE709**
Course Title : **Next Generation Sequencing Data Analysis**
L-T-P/S=Credits : **3-0-1**

Unit 1: Introduction to NGS and NGS Technologies: Introduction to sequencing technologies from a data analysts view, Concept, Applications of sequencing technologies in Whole genome assembly; Gene expression analysis; Genome annotation; Gene regulation analysis; Variation studies

Unit 2: NGS data analysis: Preprocessing: Introduction to NGS data analysis, Raw sequence files (FASTQ format), Preprocessing of raw reads: quality control (FastQC), adapter clipping, quality

Dr. Ravi Datta Sharma Member | Dr. Sakshi Arora Member | Mr. Manoj Kumar Verma Member | Dr. Pooja Sharma Member

trimming, Introduction to read mapping (Alignment methods, Mapping heuristics), Read alignment to a reference genome (BWA, Bowtie2, TopHat), Mapping output (SAM/BAM format), Usage of important NGS toolkits (samtools, BEDtools), Mapping statistics, Visualization of mapped reads (IGV, UCSC)

Unit 3: NGS data analysis: Variant calling, DNA variant calling and Filtering DNA variants, NGS data analysis: Gene Expression Analysis, Differential Expression analysis, Quantification of expression,

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the basic principles and technology in different NGS platform	CO1
2	To be able to explain the workflow of NGS data analysis	CO2
3	To be able to perform preprocessing of the data and downstream analysis	CO3
4	Design and analyse NGS-based projects,	CO4

Text & References:

- Exploring Personal Genomics, 1st Edition, Joel T. Dudley, Konrad J. Karczewski, ISBN-13: 9780199644490, Oxford University Press, 2013
- High-Throughput Next Generation Sequencing, Methods and Applications.Editors: Kwon, Young Min, Ricke, Steven C. (Eds.), ISBN 978-1-61779-089-8, Springer, 2011
- Next Generation Sequencing Technologies and Challenges in Sequence Assembly, El-Metwally,
  M.Sc, Sara, Ouda, Osama M., Helmy, Mohamed, ISBN 978-1-4939-0715-1, Springer 2014

Table API and SQL: Performing relational operations on streaming data using Table API. Advanced Topics: Complex Event Processing (CEP) and machine learning with Flink.

Unit 3: ElasticSearch: Data ingestion, storage, enrichment, visualization, and analysis using elasticsearch, Components including: cluster, nodes, shards, replicas, analyzers and documents

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the architecture of distributed systems in handling large-scale and real-time data pipelines.	CO1
2	Design and implement real-time stream processing application	CO2
3	Apply and preparation of pipelines using Apache Kafka and Flink	CO3
4	Understand and implement Elasticsearch for full-text search, indexing, and analytics over large volumes of semi-structured data	CO4

Text & References:

- Apache Kafka Cookbook, Saurabh Minni, Packt Publishing Limited, ISBN: 978-1785882449
- Stream Processing with Apache Flink: Fundamentals, Implementation, and Operation of Streaming Applications, O'Reilly Media, 2019, 978-1491974292
- Learning Apache Flink, Packt Publishing, 2017, 978-1786466228
- Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine, O'Reilly Media, 2015, 978-1449358549
- Elasticsearch in Action, Second Edition, Manning Publications, 9781617299858

Course Code : DSM DE707
Course Title : Blockchain
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction: Key Innovations/Revolutions in Human History, Evolution of Economic System, Money, Ledger and Accounting, Digital Revolution & its impact across industries, Digital Money, DigiCash, E-Gold, History & Fundamentals of Blockchain, Key Elements of Blockchain, Blockchain Protocols, Benefits of Blockchain, Types of Blockchain, Centralized vs. De-centralized systems

Unit 2: Cryptography, Symmetrical vs. Asymmetrical Encryption, Secure Hash Algorithms (SHA Family), Comparison of SHA functions, Digital Signatures

Unit 3: Consensus Mechanisms – Objectives & Need, Proof of Work (PoW), Proof of Stake (PoS), Types & Functions of Node, Construct of Block, Public vs. Private Keys, Types of Ledger, Ledgers, Distributed Ledger Technology (DLT), Benefits & Use Cases of DLT, Smart Contract

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the fundamental principles of blockchain technology	CO1
2	Explore how blockchain can be integrated with data science	CO2
3	Identify potential of Blockchain technology in different types of data	CO3
4	Develop simple blockchain applications	CO4

Text & References:

- Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World, 978-1101980149, 978-1260026672, Don and Alex Tapscott
- Cryptoassets: The Innovative Investor's Guide to Bitcoin and Beyon, by Chris Burniske and Jack Tatar
- The Basics of Bitcoins and Blockchains: An Introduction to Cryptocurrencies and the Technology that Powers Them (Cryptography, Derivatives Investments, Futures Trading, Digital Assets, NFT) Hardcover – Illustrated, September 15, 2018, Antony Lewis,
- The Blockchain Developer: A Practical Guide for Designing, Implementing, Publishing, Testing, and Securing Distributed Blockchain-based Projects 1st ed. Edition, Apress, 978-1484248461, 2019

Course Code : DSM DE709
Course Title : Next Generation Sequencing Data Analysis
L-T-P/S=Credits : 3-0-1

Unit 1: Introduction to NGS and NGS Technologies: Introduction to sequencing technologies from a data analysts view, Concept, Applications of sequencing technologies in Whole genome assembly; Gene expression analysis; Genome annotation; Gene regulation analysis; Variation studies

Unit 2: NGS data analysis: Preprocessing: Introduction to NGS data analysis, Raw sequence files (FASTQ format), Preprocessing of raw reads: quality control (FastQC), adapter clipping, quality trimming, Introduction to read mapping (Alignment methods, Mapping heuristics), Read alignment to a reference genome (BWA, Bowtie2, TopHat), Mapping output (SAM/BAM format), Usage of important NGS toolkits (samtools, BEDtools), Mapping statistics, Visualization of mapped reads (IGV, UCSC)

Unit 3: NGS data analysis: Variant calling, DNA variant calling and Filtering DNA variants, NGS data analysis: Gene Expression Analysis, Differential Expression analysis, Quantification of expression,

Course Outcome:

S.N.	Course Outcome	CO
1	Understand the basic principles and technology in different NGS platform	CO1
2	To be able to explain the workflow of NGS data analysis	CO2
3	To be able to perform preprocessing of the data and downstream analysis	CO3
4	Design and analyse NGS-based projects.	CO4

Annexure 19.3

Text & References:

- Exploring Personal Genomics, 1st Edition, Joel T. Dudley, Konrad J. Karczewski, ISBN-13: 9780199644490, Oxford University Press, 2013
- High-Throughput Next Generation Sequencing, Methods and Applications,Editors: Kwon, Young Min, Ricke, Steven C. (Eds.), ISBN 978-1-61779-089-8, Springer, 2011
- Next Generation Sequencing Technologies and Challenges in Sequence Assembly, El-Metwally,
  M.Sc, Sara, Ouda, Osama M., Helmy, Mohamed, ISBN 978-1-4939-0715-1, Springer 2014

Dr. Ravi Datta Sharma
Member

Mr. Abhigya Mahajan
Member

Dr. Sakshi Arora
Member

Mr. Manoj Kumar Verma
Member

Dr. Pooja Sharma
Member

Dr. Baijnath kaushik
Member

Dr. Naveen Kr. Goudhi
Member

Dr. Sonika Gupta
Member

Mr. Sanjay Kumar Sharma
Member

Dr. Deo Prakash
Member

Mr. Sudesh Kumar
Member

Dr. Archana Purwar
Member

Dr. Rohit Tanwar
Member

Dr. Vipal Kr. Sharma
Member

Dr. Ajay Koul
Member

Dr. Manoj Kumar Gupta
Member

Mr. Anuj Mahajan
Member

Dr. Sunanda
Chairman and Head
SoCSE

26/11/26

Annexure 19.5(a)

Course Code	**: CSL VA104**
Course Title	**: Digital and AI Literacy**
L-T-P/S=Credits	**: 2-0-0**
Course Category	**: VAC**
Pre-requisite Courses (if any)	**:**
Equal Course Code (if any)	**:**
Equivalent Course Code (if any)	**:**

Detailed Syllabus

S.No.	Contents	Approx. Contact Hours
1	**Computer Fundamentals:** **Hardware and Software:** Understanding the difference between hardware (CPU, monitor, keyboard, mouse) and software (system vs. application). **Operating Systems:** Managing files and folders, personalizing desktop settings, and using common utilities in systems like Windows or Linux. **Peripherals:** Connecting and using printers, headsets, and storage devices.	30 Hours
2	**Core Software Applications (Office Productivity):** **Word Processing:** Creating, editing, formatting, and printing documents (e.g., Microsoft Word, Google Docs). **Spreadsheets:** Entering data, using basic formulas and functions, and managing workbooks (e.g., Microsoft Excel). **Presentations:** Designing slides using templates, inserting images/tables, and running slide shows (e.g., PowerPoint).	
3	**Internet and Communication:** **Web Browsing:** Navigating the World Wide Web, using search engines, and understanding URLs. **Email Management:** Creating accounts, sending/receiving attachments, and practicing "netiquette" (online etiquette). **Collaboration Tools:** Using instant messaging and cloud-based sharing for teamwork.	
4	**Digital Security and Ethics:** **Cybersecurity Basics:** Creating strong passwords, identifying phishing attempts, and understanding malware risks. **Digital Citizenship:** Learning about online privacy, social media responsibility, and ethical use of information.	
5	**Advanced or Elective Topics:** **Financial Literacy:** Using digital banking and e-governance services. **Content Creation:** Basic graphics, video editing, or blog management.	

6	**AI Literacy for Office Administration:** Introduction to Artificial Intelligence and generative AI tools for office work. Use of ChatGPT for drafting letters, emails, reports, and meeting minutes. Application of AI for reading comprehension, summarization, and language improvement. Ethical use of AI, data privacy, accuracy, and human verification in office administration.	

Suggested Books:

Sr.	Name of Book, Author, Publisher	Year of Publication/Reprint
1	Computer Fundamentals, Pradeep K. Sinha, Priti Sinha, BPB Publications	2016
2	Introduction to Computers, Peter Norton, Tata McGraw-Hill	2014
3	Microsoft Office 2019 Step by Step, Joan Lambert, Microsoft Press	2019
1	Internet and World Wide Web, A. Godbole, A. Kahate, McGraw-Hill	2012
2	Cyber Security: Understanding Cyber Crimes, N. Godbole, S. Belapure, Wiley India	2011
3	Digital Literacy, M. Madan, Laxmi Publications	2018

Course Outcomes:
(Note: Course outcomes should not more than 04 in number)

CO 1	Understand basic computer concepts including hardware, software, operating systems, and peripheral devices for effective system usage.
CO 2	Apply core office productivity tools such as word processors, spreadsheets, and presentation software to create professional digital documents.
CO 3	Use internet and communication tools effectively for web browsing, email management, and online collaboration while following netiquette.
CO 4	Demonstrate awareness of digital security, ethical practices, and digital citizenship, including safe online behavior and responsible use of digital services.

CO - PO Mapping

CO - PO Mapping for B.Tech. Four Year Full Time Degree Program														
CO's	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO1	2	-	-	-	2	-	-	-	-	-	-	1	1	-	-
CO2	-	1	2	-	3	-	-	-	1	1	-	-	2	1	-
CO3	-	-	-	-	2	1	-	2	3	2	-	1	-	2	1
CO4	-	-	-	-	1	2	2	3	1	-	1	2	-	1	2
	2	1	2	-	2	1.5	2	2.5	2.5	1.5	1	2	1.5	2	1.5

Annexure 19.5(b)

Course Code	**: CSL AE104**
Course Title	**: Smart Devices Networks**
L-T-P/S=Credits	**: 2-0-0**
Course Category	**: AEC**
Pre-requisite Courses (if any)	**: No**
Equal Course Code (if any)	**:**
Equivalent Course Code (if any)	**:**

Detailed Syllabus

Sr	Contents	Approx. Contact Hours
1	**UNIT-I** Fundamentals of IoT: Introduction, Definitions & Characteristics of IoT, IoT Architectures, Physical & Logical Design of IoT, Enabling Technologies in IoT, History of IoT, About Things in IoT, The Identifiers in IoT, About the Internet in IoT, IoT frameworks, IoT and M2M. **UNIT-II** Sensors Networks : Definition, Types of Sensors, Types of Actuators, Examples and Working, IoT Development Boards: Arduino IDE and Board Types, RaspberriPi Development Kit, RFID ,Principles and components, Wireless Sensor Networks: History and Context, The node, Connecting nodes, Networking Nodes, WSN and IoT. **UNIT-III** Wireless Technologies for IoT: WPAN Technologies for IoT: IEEE 802.15.4, Zigbee, HART, NFC, Z-Wave, BLE, Bacnet, Modbus. IP Based Protocols for IoT IPv6, 6LowPAN, RPL, REST, AMPQ, CoAP, MQTT. Edge connectivity and protocols **UNIT-IV** Data Handling& Analytics: Introduction, Bigdata, Types of data, Characteristics of Big data, Data handling Technologies, Flow of data, Data acquisition, Data Storage, Introduction to Hadoop. Introduction to data Analytics, Types of Data analytics, Local Analytics, Cloud analytics and applications **UNIT-V** Applications of IoT: Home Automation, Smart Cities, Energy, Retail Management, Logistics, Agriculture, Health and Lifestyle, Industrial IoT, Legal challenges, IoT design Ethics, IoT in Environmental Protection. **UNIT-VI** Smart device usage in packet tracer. Usage of AI tools.	30

Suggested Books:

Sr.	Name of Book, Author, Publisher	Year of Publication/Reprint
1	Hakima Chaouchi, — "The Internet of Things Connecting Objects to the Web" ISBN : 978-1- 84821-140-7, Wiley Publications	2018
2	Olivier Hersent, David Boswarthick, and Omar Elloumi, — "The Internet of Things: Key Applications and Protocols", Wiley Publications	2016
3	Vijay Madisetti and Arshdeep Bahga, — "Internet of Things (A Hands-on-Approach)", 1st Edition, VPT, 2014.	2014
4	J. Biron and J. Follett, "Foundational Elements of an IoT Solution", O'Reilly Media,	2014
5	Keysight Technologies, "The Internet of Things: Enabling Technologies and Solutions for Design and Test", Application Note, 2016. Platforms, and Use Cases", CRC Press	2016

Course Outcomes:
(Note: Course outcomes should not more than 04 in number)

CO 1	*To study fundamental concepts of IoT Smart Devices*
CO 2	*To understand roles of sensors in IoT*
CO 3	*To Learn different protocols used for smart devices*
CO 4	*To be familiar with data handling and analytics tools in IoT*

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	1	1	-	-	1	-	-	-	-	-	-	2	2	1
CO2	1	1	3	1	2	1	1	-	-	-	-	-	1	1	1
CO3	-	-	2	1	1	-	-	-	-2	1	1	1	1	2	1
CO4		2	1	2	3	1	-	1	1	-	-	-	2	2	1

Annexure 19.5(c)

Course Code : **CSL SE106**
Course Title : **Data Centre Management**
L-T-P/S=Credits : **2-0-0**
Course Category : **SEC**
Pre-requisite Courses (if any) :

Detailed Syllabus

S.No.	Contents	Approx. Contact Hours
	Unit 1 – Data Centre Hardware Systems This module covers the core hardware systems used in data centres. It includes computer system fundamentals, safety practices, and tools used for installation and maintenance. Students study servers, motherboards, processors, memory, storage devices, power supplies, graphics adapters, and peripheral devices with emphasis on installation, replacement, fault isolation, and preventive maintenance in data centre environments. **Unit 2 – Operating Systems and Server Administration** This module introduces operating system concepts relevant to data centre operations. Topics include OS architecture, boot processes, Windows and Linux server environments, operating system installation, disk partitioning, user and access management, driver and application management, firewall configuration, file systems, command-line administration, and process and memory management required for server stability and performance. **Unit 3 – Data Centre Networking and Physical Infrastructure** This module focuses on networking concepts and physical infrastructure used in data centres. It covers switching and routing basics, networking models (OSI and TCP/IP), data transmission concepts, and network components. Students learn about copper and fiber cabling, optical transceivers, structured cabling, racks, cabinets, patch panels, and cable management practices essential for scalable and reliable data centre networks. **Unit 4 – IP Networking, DNS, and Operational Networking Tasks** This module develops practical networking skills required for day-to-day data centre operations. Topics include IP addressing, subnetting, subnet mask calculation, NAT, DHCP operation, and DNS management. Students learn domain name resolution, DNS records, basic DNS configuration, monitoring, and troubleshooting. Hands-on operational tasks include network connectivity testing, DNS validation, fault diagnosis, and performance monitoring in enterprise environments. **Unit 5 – Network Security and Data Centre Safeguards** This module introduces security practices specific to data centre infrastructure. It covers security principles, firewall implementation, access control, authentication and authorization mechanisms, secure network design, encryption basics, and protection against common network and system threats. Emphasis is placed on monitoring, logging, incident handling, and adherence to security policies within data centre operations. **Unit 6 – Data Centre Operations, Management, and Best Practices** This module emphasizes data centre management and operational skills. Topics include data centre layout and zoning, power and cooling systems, environmental and asset monitoring, change and incident management, preventive maintenance, documentation standards, service continuity, escalation procedures, and compliance with operational and safety policies. The module prepares students for real-world data centre workflows and team-based operations.	30

After completing this course, students will be able to:

CO	Course Outcome
CO1	Understand basic data centre hardware components and perform simple installation and maintenance tasks safely.
CO2	Explain basic operating system concepts and perform introductory server administration tasks.
CO3	Understand basic data centre networking concepts and apply them to simple connectivity and troubleshooting tasks.
CO4	Describe essential data centre operational practices, including security awareness, monitoring, documentation, and maintenance..

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	1	–	–	1	1	–	1	–	–	–	1	1	–	1
CO2	2	1	–	–	2	–	–	–	–	–	–	1	2	1	1
CO3	2	2	–	–	2	–	–	–	1	–	–	1	1	2	1
CO4	1	–	–	–	1	2	2	2	1	1	1	2	–	1	2

ANNEXURE 19.6

(68)

श्री माता वैष्णो देवी विश्वविद्यालय

SHRI MATA VAISHNO DEVI UNIVERSITY, KATRA

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

No. SMVDU/SoCSE/2025/ 582 **Date: 04th Dec 2025**

The meeting of Academic Affairs Committee (AAC) of the School of Computer Science & Engineering (SoCSE) was held on **04th Dec 2025** at 12:00 PM in the Conference Hall. SCSE.

The following members were present during the meeting:

1. Dr. Ajay Kaul
2. Mr. Manoj Kumar Verma
3. Dr. Sonika Gupta
4. Dr. Deo Prakash
5. Dr. Baijnath Kaushik
6. Dr. Rohit Tanwar
7. Dr. Naveen Kumar Gondhi
8. Mr. Sanjay Kumar Sharma
9. Dr. Vipal Kumar
10. Mr. Anuj Mahajan
11. Dr. Manoj Kumar Gupta
12. Dr. Archana Purwar
13. Mr. Sudesh Kumar
14. Dr. Pooja Sharma

Following members could not attend the meeting:

1. Dr. Sunanda (on Leave)
2. Dr. Sakshi Arora (on Leave)

Agenda:

1. NPTEL courses to be permitted for the B.Tech CSE students for Jan-May 2026. as per the requests received from the students.

Resolution:

AAC discussed the requests received from students, and either Approved or Rejected the NPTEL courses requests, as per the attached list (**Annexure A**).
The meeting concluded with a vote of thanks.

Dr. Vijay Kaul

Dr. Manoj Kumar Gupta

on Leave.
Mr. Manoj Kumar Verma

Dr. Sonika Gupta

05/12/25
Dr. Deo Prakash

5/12/25.
Dr. Baij Nath Kaushik

on Leave.
Dr. Naveen Kumar Gondhi

on Leave.
Dr. Sakshi Arora

5/12/25
Mr. Sanjay Kumar Sharma

05/12/25
Mr. Anuj Mahajan

05/12/2025
Mr. Sudesh Kumar

5/12/25
Dr. Vipal Kumar

05/12/25
Dr. Archana Purwar

Dr. Pooja Sharma

05/12/25
Dr. Rohit Tanwar
(AAC Secretary)

Sunanda.
Dr. Sunanda
(Head & AAC Chairman)

Dean (AA)

Annexure – A

(69)

Approval/Rejection status of NPTEL/SWAYM requests

S.No	Offline Course/course category (as per curriculum)	NPTEL Course Request	Duration	Request Approved/Rejected by AAC
1.	BSC 4 (Core) Discrete Structures MTL BS106	noc26-cs68 Discrete Mathematics	12 Weeks	Approved
2.	AEC	noc26-cs86 Data Analytics with Python	12 weeks	Not Approved as Data Science & Analytics is core subject ,offered in 3rd sem
3.	VAC	noc26-cs84 The Joy of Computing using Python	12 weeks	Not Approved as Python Programming is core subject offered in 2nd sem
4.	Deep Learning CSL – DCC14-CSLDC302(Core)	noc26-ee53 Deep Learning	12 weeks	Approved
5.	School Elective 4 (6th sem)	noc26_cs37 Introduction To Internet Of Things	12 weeks	Approved
6.	School elective 3 (6th sem)	noc25_cs38 Human Computer Interaction	12 weeks	Approved
7.	School elective 3 (6th sem)	Introduction to Data Warehousing and Data Mining (imb26_mg92)	12 weeks	Approved
8.	Complier Design CSLDC210 (Backlog)	noc26_cs56(Compiler Design)	12 weeks	Approved
9.	Software Engineering (CSLDC304)(Core)	cec26_cs06(Software Enginering)	12 weeks	Approved
10.	School elective 3 (6th sem)	nou26_ge22 (Android Mobile Application Development)	12 weeks	Approved
11.	Backlog (Environmental science)	noc26-hs120(Environmental science)	12 weeks	Approved
12.	Soft computing (CSLDC202)	noc26-cs44 (Introduction to soft	8 weeks	Approved

05/12/25

		computing)		
13.	AEC	noc26-cs79 (Programming , Data Structures And Algorithms Using Pyhton)	8	**Not approved**, as Data Structure is already offered as a Core subject in 2nd sem
14.	BAcklog (HPC)	Noc26-cs32 (High performance computing)	12 weeks	Approved

Minor specialization students opted for following two courses from NPTEL:

S. No	Offline Course	NPTEL Course	Duration	Request acceptd Rejected by AAC
1	Soft Computing (CSM GE401)	Soft Computing Techniques (noc26-ma35)	12 Weeks	Approved
2	Data Analytics (CSM GE306)	Data Analytics with Python (noc26-cs86)	12 Weeks	Approved

Annexure 19.7(a)

Course Code	**: CSL DC305**
Course Title	**: Machine Learning**
L-T-P/S=Credits	**: 3-0-0 =3**
Course Category	**: Departmental Core Course**
Pre-requisite Courses (if any)	**:**
Equal Course Code (if any)	**:**
Equivalent Course Code (if any)	**:**

Detailed Syllabus

S. No.	Contents	Approx. Contact Hours
1.	AI problems, foundation of AI and history of AI intelligent agents: Agents and Environments, the concept of rationality, the nature of environments, structure of agents, problem solving agents, problem formulation. Knowledge representation issues, predicate logic- logic programming, representing knowledge using rules, rules based deduction systems. Reasoning under uncertainty, review of probability, Baye's probabilistic interferences and dempster shafer theory. First order logic. Inference in first order logic, propositional vs. first order inference, unification & lifts forward chaining, Backward chaining, Resolution, Learning from observation Inductive learning	6
2.	Introduction to Machine learning: Motivation and role of machine learning in computer science and problem solving. Representation (features), linear transformations, Appreciate linear transformations and matrix vector operations in the context of data and representation. Problem formulations (classification and regression).Appreciate the probability distributions in the context of data, Prior probabilities and Bayes Rule. Introduce paradigms of Learning (primarily supervised and unsupervised. Also a brief overview of others).	2
3.	Fundamentals of Machine Learning: Statistical Decision Theory, Bias-variance Trade-off. Notion of Training, Validation and Testing; Connect to generalization and overfitting. Subset Selection, Shrinkage Methods, PCA, Linear Discriminant Analysis and Dimensionality Reduction.	4
4.	Selected Algorithms: Nearest Neighbours and KNN. Linear Regression, Multivariate Regression. Decision Tree Classifiers. Notion of Generalization and concern of Overfitting. Linear SVM, KMeans, Logistic Regression, Naïve Bayes and Ensembling and RF	6
5.	Role of Loss Functions and Optimization, (ii) Gradient Descent and Perceptron /Delta Learning, (iii) MLP, (iv) Backpropagation (v) MLP for Classification and Regression, (vi) Regularisation, Early Stopping (vii) Parameter Estimation-Maximum Likelihood estimation and Bayesian Estimation.	6
6.	Clustering Methods: Kernels (with SVM), Bayesian Methods, Generative Methods, HMM, GMM, EM, PAC learning.	6

Suggested Books:

S. No.	Name of Book, Author, Publisher	Year of Publication / Reprint
Text Books		
1	Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Mathematics for Machine Learning, Cambridge University Press.	2020
2	Tom M. Mitchell- Machine Learning- McGraw Hill Education, International Edition.	1997
Reference Books		
1	Aurelien Geron Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, Inc. 2nd Edition	2019
2	Ian Goodfellow, Yoshoua Bengio, and Aaron Courville Deep Learning MIT Press Ltd, Illustrated edition.	2016
3	Christopher M. Bishop Pattern Recognition and Machine Learning- Springer, 2nd edition.	2006

Course Outcome

S. No	Course Outcome	CO
1	Know the basics and mathematics behind various Machine Learning algorithms.	CO1
2	Capability to implement basic algorithms using basic machine learning libraries mostly in python.	CO2
3	Think analytically and suggest possible solutions to problems using Machine Learning.	CO3
4	Learning & understanding policy gradients along with its advantages and disadvantages.	CO4
5	Ensure awareness about importance of core CS principles such as algorithmic thinking and systems design in ML.	CO5

Course Code	**: CSP DC305**
Course Title	**: Machine Learning Lab**
L-T-P/S=Credits	**: 0-0-2 =1**
Course Category	**: Departmental Core Course**
Pre-requisite Courses (if any)	**:**
Equal Course Code (if any)	**:**
Equivalent Course Code (if any)	**:**

Detailed Syllabus

S. No.	Contents Lab Experiment	Approx. Contact Hours
1	Explore the **WEKA tool** and practice various pre-coded ML algorithms for regression, classification, and clustering on sample WEKA/UCI datasets.	2
2	Compute **cost function for linear regression**.	
3	Implement **gradient descent for univariate linear regression**.	2
4	Compute **cost function for multivariate linear regression**.	
5	Implement **gradient descent for multivariate linear regression**.	2
6	Given a **UCI-Iris dataset** (or any other dataset), code the classifier/clustering algorithm to predict the type (Iris-Setosa / Iris-Versicolor / Iris-Virginica) based on input features (sepal length, sepal width, petal length, petal width).	2
7	Implement and visualize **Linear Regression** using any dataset.	2
8	Implement **Logistic Regression** and plot decision boundary.	
9	Implement **K-Nearest Neighbors (KNN)** classifier and measure accuracy.	2
10	Implement **Support Vector Machine (SVM)** for classification tasks.	2
11	Perform **Naïve Bayes Classification** on text or numerical data.	2
12	Implement **Decision Trees** for classification and visualize the tree.	2
13	Implement **Random Forest** for classification and regression.	
14	Perform **K-Means Clustering** on given dataset and plot clusters.	2
15	Perform **Hierarchical Clustering** and show dendrograms.	2
16	Implement **Principal Component Analysis (PCA)** for dimensionality reduction.	2
17	Perform **Association Rule Mining** using Apriori algorithm.	
18	Implement **Gradient Descent** manually (without libraries) for a simple cost function.	2
19	Implement a simple **Neural Network** (MLP) from scratch or using a library.	

Annexure 19.7(b)

Course Code	**: CSL DE**
Course Title	**: Augmented Reality & Virtual Reality**
L-T-P/S=Credits	**: 3-0-0 =3**
Course Category	**: Department Elective Course**
Pre-requisite Courses (if any)	**:**
Equal Course Code (if any)	**:**
Equivalent Course Code (if any)	**:**

Detailed Syllabus

S. No.	Contents	Approx. Contact Hours
1	**Introduction to Virtual Reality (VR):** Defining Virtual Reality, Key elements of virtual reality experience, Virtual Reality, Telepresence, Augmented Reality and Cyberspace. Bird's-Eye View: Hardware, Software, Human Physiology and Perception.	10
2	**Input Devices:** (Trackers, Navigation, and Gesture Interfaces): Three-dimensional position trackers, navigation and manipulation, interfaces and gesture interfaces. **Output Devices:** Graphics displays, sound displays & haptic feedback.	8
3	**Modeling:** Geometric modeling, Kinematics modeling, Physical modeling, Behaviour modeling, Model management.	8
4	**Augmented Reality (AR):** Taxonomy, Technology and Features of Augmented Reality, AR Vs VR, Challenges with AR, AR systems and functionality, Augmented Reality Methods, Visualization Techniques for Augmented Reality, Enhancing interactivity in AR Environments, Evaluating AR systems AR software development : AR software, Camera parameters and camera calibration, Marker-based augmented reality, AR Toolkit.	8
5	**Interaction & Audio:** Interaction- Motor Programs and Remapping, Locomotion, Manipulation, Social Interaction. Audio-The Physics of Sound, The Physiology of Human Hearing, Auditory Perception, Auditory Rendering. Interaction- Motor Programs and Remapping, Locomotion, Manipulation, Social Interaction. Audio-The Physics of Sound, The Physiology of Human Hearing, Auditory Perception, Auditory Rendering.	8

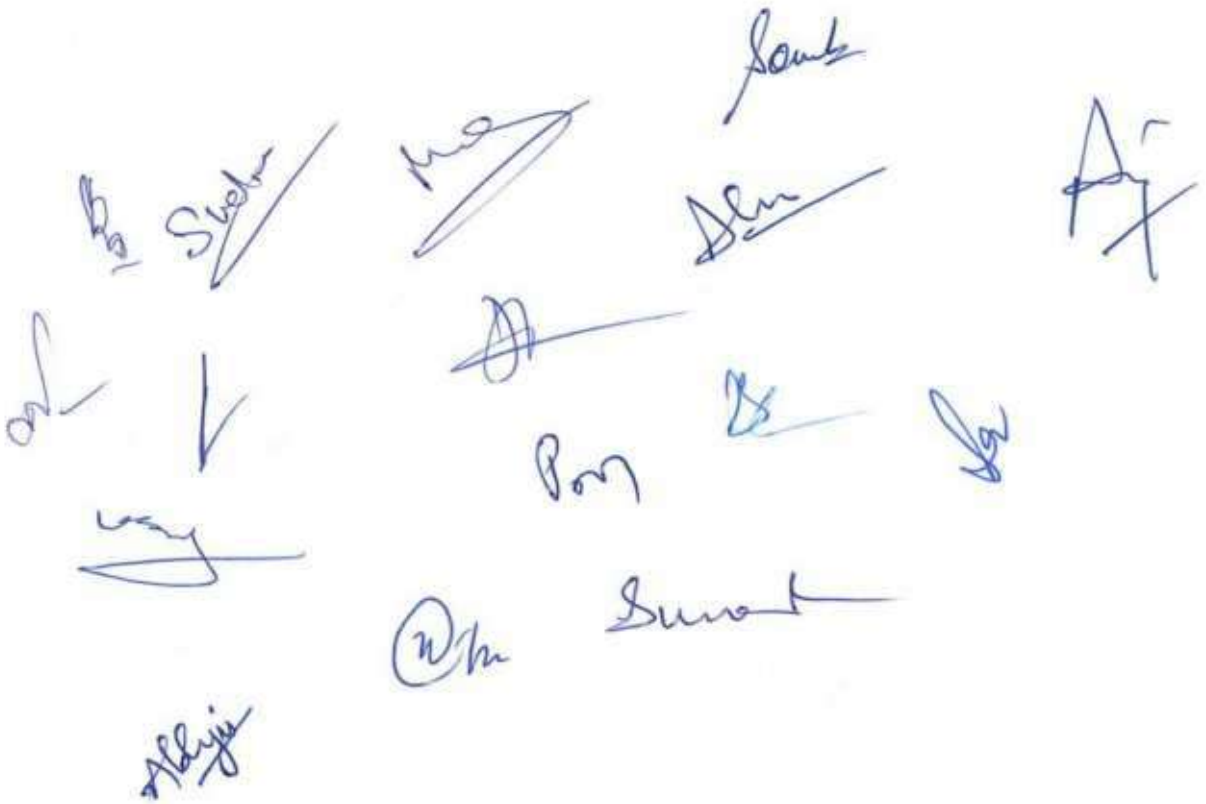
Suggested Books:

S. No.	Name of Book, Author, Publisher	Year of Publication / Reprint
Text Books		
1	Virtual Reality Technology, Second Edition, Gregory C. Burdea & Philippe Coiffet, John Wiley & Sons, Inc.	2017
2	Virtual Reality, Steven M. LaValle, Cambridge University Press.	2016
Reference Books		
1	Rajesh K. Maurya, Computer Graphics with Virtual Reality System, 3rd Edition, Wiley Publication.	2018.
2	Grigore C. Burdea, Philippe Coiffet, Virtual Reality Technology, 2nd Edition, Wiley.	2017

3	William R. Sherman and Alan B. Craig, Understanding Virtual Reality Interface, Application, and Design, 2nd Edition, Morgan Kaufmann Publishers, Elsevier.	2019
4	K.S. Hale and K. M. Stanney, Handbook on Virtual Environments, 2nd Edition, CRC Press.	2015

Course Outcome

S. No	Course Outcome	CO
1	Understand the concept of virtual reality and Communication Media.	CO1
2	Understand current virtual reality hardware and software.	CO2
3	Understand various modeling approaches.	CO3
4	Illustrate the concepts of Human Factors and Applications of VR.	CO4



Annexure 19.7(c)

Course Title : **Python Programming**
Course Code : **CSM DE337**
L-T-P/S=Credits : **3-0-2 =4**
Course Category : **Departmental Core course**
Pre-requisite Courses (if any) :
Equal Course Code (if any) :
Equivalent Course Code (if any):

Detailed Syllabus

Introduction to importance of IDEs like Spyder (Anaconda)/ PyCharm for professional programming, explore Python shell as a calculator and for inputting Python expressions directly, Data Types in. Operators in Python: comparison, arithmetic, logical, Boolean, bitwise, assignment. Python: numbers, list, tuple, strings, set, dictionary, conversion between various data types Hello World program in Python script, Python keyword and Identifiers, Indentation, Comments,

Data Types. Operators in Python: comparison, arithmetic, logical, Boolean, bitwise, assignment.

Python: numbers, list, tuple, strings, set, dictionary, conversion between various data types Input and Output in Python, if-else , for loop, while loop, break, pass, continue, global and local variables, Importing Other modules/packages and using their functions, creating random numbers/random-choice to create programs for creating Functions, functions with arguments, returning values form functions, lambda expressions, recursion, simple guessing games like Rock – Paper-Scissors. Problems On 1D/2D/3D arrays using list. Problem solving using a dictionary as a look-up table. Global and local variables, Importing other modules/packages and using their functions, creating random numbers/random-choice to Create programs for simple guessing games like Rock –Paper-Scissors. Problems on 1D/2D/3D arrays using list. Problem Solving using a dictionary as a look-up table.

Object oriented programming: Class and Object. Defining Variables and functions inside class. Creating objects, Inheritance, Multiple and Multi Level Inheritance, Function over-riding. Concept of composing objects of a different class in an object, problems on object composition

Python's de-facto GUI package like tkinter or alternative packages like: wxPython, PyQt (PySide), Pygame, Pyglet, and PyGTK.

Creating labels, buttons, entry (textbox), combo box, check button, radio button, scrolled Text (textarea), spin box,progress bar, menu bar, file dialog, tabs etc.

Creating GUI simple games like Tic-Tac-Toe Discussion about implementation of python in Machine learning Algorithms, Python with Data Sets

Suggested Books:

Sr.	Name of Book, Author, Publisher	Year of Publication/Reprint
Text Books		
1	Think Python 2nd Edition - How to Think Like a Computer Scientist, Allen B Downey, O'Reilly publication	2015
2	Learn Python 3 the Hard Way, Zed A. Shaw, Pearson publication	2022
3	Head First Programming: A Learner's Guide to Programming using the Python Language, Paul Barry David Griffiths Barry Griffiths, O'Reilly publication	2018
Reference Books		

1	Dive into Python 3, Mark Pilgrim, Apress publication	2017

Course Outcomes (theory):

CO 1	Describe the fundamental concepts of Python programming, including syntax, data types, variables, and basic input/output operations.
CO 2	Apply control structures such as conditional statements and loops to develop simple Python programs.
CO 3	Use Python's built-in data structures (lists, tuples, dictionaries, and sets) to store and manipulate data.
CO 4	Demonstrate programming and problem-solving skills that enhance employability in software development, data analysis, and related fields.

	CO - PO Mapping for B.**Tech.** Four Year Full Time Degree Program															
S No	Course Code CSL DC102	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
1	CO1	3	3	-	-	-	-	-	-	2	-	-	2	3	1	2
2	CO2	3	3	3	-	3	-	-	-	2	2	-	-	3	3	1
3	CO3	3	2	3	-	3	-	-	-	3	2	-	-	3	3	1
4	CO4	2	3	3	3	3	-	-	-	3	3	-	-	3	3	3

List of Experiments

Sr	Contents
1	Write a program to add two numbers.
2	Write a program to calculate the grade of a student.
3	Write a program to print the following pattern. * ** ***
4	Write a program to print a table of numbers.
5	Write a program to print following pattern: 0 01 012
6	Write a program to add and multiply numbers using user defined functions.
7	Write a program to concatenate strings in Python.
8	Give at least five examples of inbuilt functions of Python.
9	Give at least five examples of inbuilt Math functions of Python.
10	Write a program to calculate factorial of a number using recursion.
11	Write a program to make a List in Python and perform following operations on List: a) Length using len() function b) Print element at index 0 c) Adding an element to the list using + operator d) Appending an element to the list e) Negative indexing in list

	f) Remove the first occurrence of element a from list g) Reverse the list h) Sort list
12	Write a program to demonstrate use of Dictionaries in Python with their inbuilt functions.
13	Write a program to demonstrate use of Set in Python with their inbuilt functions.
14	Write a program to demonstrate use of Tuple in Python with their inbuilt functions.
15	Write a program to calculate Median using List.
16	Write a program to calculate Mode using List.
17	Write a program to calculate Mean using List.
18	Write a program to inherit properties of a person to a student using inheritance in Python.
19	Write a program to calculate Coefficients of given numbers.
20	Write a program to calculate Covariance of given numbers.
21	Write a program to plot a graph using the python library.
22	Write a program to make a class Vehicle and their properties.
23	Write a program to make a game Rock-Paper-Scissor.
24	Write a program to make classes for bird and animal with their properties and simulate a zoo like environment.
25	Write a program to make a GUI to take input from the user and display.

Course Outcomes (Lab):

CO 1	Describe the fundamental concepts of Python programming, including syntax, data types, variables, and basic input/output operations.
CO 2	Apply control structures such as conditional statements and loops to develop simple Python programs.
CO 3	Use Python's built-in data structures (lists, tuples, dictionaries, and sets) to store and manipulate data.
CO 4	Demonstrate programming and problem-solving skills that enhance employability in software development, data analysis, and related fields.

	CO - PO Mapping for B.**Tech.** Four Year Full Time Degree Program															
S No	Course Code	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
1	CO1	3	3	-	-	-	-	-	-	2	-	-	2	3	1	2
2	CO2	3	3	3	-	3	-	-	-	2	2	-	-	3	3	1
3	CO3	3	2	3	-	3	-	-	-	3	2	-	-	3	3	1
4	CO4	2	3	3	3	3	-	-	-	3	3	-	-	3	3	3

ANNEXURE-19.8

B.Tech. CSE 2nd /4th /6th semester

Summer Internship Guidelines

Date: 7th May 2025

The students undergoing internship after 2nd /4th /6th semester are supposed to follow these guidelines:

- Guidelines.
  - Students are supposed to provide the details of Internship like; Organization Name, Name of Supervisor, Title, Duration, etc. before the completion of final exams of 2nd /4th /6th semester respectively. Internship coordinator may ask for this information through a form or other means.
  - Duration of Internship should be of 4 weeks minimum.
  - 2nd-year students are encouraged to enroll in exploratory or introductory internships, whereas 3rd and 4th-year students should go for domain-specific and pre-placement internships.
  - Students must provide a valid certificate during assessment failure of which 'F' grade will be awarded to the student.
  - The final list of approved internships will be shared coordinator with respective students. No change in internship details will be allowed after the approved list is shared.
- Assessment - Assessment will be done after **mid semester exams** of 3rd /5th/7th semester respectively as per the following guidelines:
  - Students are required to be dressed up in *professional* attire.
  - Presentations must last no longer than 10 minutes, and will be followed by a 5- minute "question and answer" period.
  - Presentations must be made using PowerPoint slides, and must include:
    a. An overview of the internship site, including a **brief** description of its services, personnel, on-site supervisor, etc.;
    b. A **brief** description of the internship responsibilities;
    c. A **brief** description of the challenges and successes experienced during the internship;
    d. A ***very detailed*** description of the Special Project (***the majority of the presentation time should be devoted to this***);
    e. A description of what was learned from the experience and how it can be applied in the future; and
    f. A statement of future (professional and/or academic) plans.

  - Individual student needs to present their internship work to the panels comprising of faculty members. The panel will evaluate the students as per below rubrics:

Presentation	: 60 marks
Internship report (hard copy)	: 40 marks
Total	: 100 marks

Detailed Rubrics

S.No.	Parameters	Description
1	Student Dressed Professionally	Student dressed in professional attire
2	Overview Of Internship Site, Mission, Services, Personnel,etc.	Student should fully describe all the details
3	Internship Responsibilities	Student should fully describe what he/she worked on during the internship
4	Challenges Experienced	Student should describe the challenges experienced during internship
5	Successes Experienced	Student should describe the successes experienced during internship
6	Special Project: explanation and how it contributed to the mission and goals of agency	Student should describe the following: ▫ the rationale for the special project ▫ its benefit to the agency ▫ the planning involved ▫ the implementation ▫ the outcomes and evaluation
.7	What Was Learned From The Experience	Student should describe what he/she learned from the Experience
8	Future Plans (Career, etc.)	Student should clearly articulate and describe his/her future Plans that can be derived from the internship learnings.
9	Overall Descriptive and Analytical Competence	Student cover all the required Sections (above) of the presentation within the 10 minutes, and the student's descriptions of the covered sections were clearly articulated and professionally presented
10	Final Performance Evaluation by the Supervisor	Supervisor evaluation of student's performance.

Scale: Exceptional (6) ,Very Good (5),Good (4),Average (3),Below Average (2), Unsatisfactory (1)

Internship Coordinators:

Dr. Deo Prakash
(2nd Year)

Dr. Manoj Gupta
(3rd Year)

Dr. Rohit Tanwar
(4th Year)

Annexure I 9. 9a.

विज्ञान ज्योति

Shri Mata Vaishno Devi University Katra
School of Computer Science & Engineering

VISION

To provide quality professionals equipped with technical expertise in Computer Science and Engineering to the world that are competent in solving the problems in industry and society and have high moral and ethical values.

MISSION

- To have a program and value system appropriate to the pursuit of Computer Science and Engineering to meet the needs of the society.
- To nurture the students with proficiency in Computer Science and Engineering, and inculcate versatility of mind, motivation for learning and building self-reliance.

Annexure II

Shri Mata Vaishno Devi University Katra
School of Computer Science & Engineering

PROGRAM EDUCATIONAL OBJECTIVES
(**B. Tech.** Four Year Full Time Degree Program)

PEO 1: To prepare students with advanced knowledge of engineering and technology so as to keep abreast of the current trends in the field of Computer Science & Engineering.

PEO 2: To prepare students to become successful professionals in industry, government, entrepreneurial pursuit and consulting firms.

PEO 3: To prepare students having skills to function as members of multi-disciplinary teams and contribute effectively.

ANNEXURE-19.9 (B)

(74)

ANNEXURE - III 19.9(C)

Shri Mata Vaishno Devi University Katra
School of Computer Science & Engineering

PROGRAM EDUCATIONAL OBJECTIVES
(**M. Tech.** Four Year Full Time Degree Program)

PEO 1: To prepare postgraduates for successful careers in industry, research, and academia by providing them with advanced knowledge and technical expertise in the area of Artificial Intelligence.

PEO 2: To enable graduates to innovate, design, and develop intelligent systems and applications that address real-world problems.

PEO 3: To [illegible] professional ethics, effective communication, teamwork, leadership qualities, and a commitment to lifelong learning for contributing responsibly to society and the global technological community.

ANNEXURE 19.10

Course Code : **CSL AE302**

Course Title : **Business Fundamentals for Entrepreneurs**
L-T-P/S=Credits : **3-0-0 =3**
Course Category : **AEC**
Pre-requisite Courses (if any) :
Equal Course Code (if any) :
Equivalent Course Code (if any) :

Detailed Syllabus

UNIT-I
Overview of a Business, Company & Structures, Vision, Mission, Goals, Strategy & Culture. Innovation & Product Development. Manufacturing & Quality. Finance & Accounting. Leadership & Human Resources. Environment, Society & Governance

UNIT-II

Market Research, Customers & Segments, Values Based Marketing, Marketing Communication, Sales Channels & Distribution, B2B & B2C Sales, Customer Service Management, Supply Chain & Logistics, Business Risk & Management

UNIT-III
Business Information & Management, Business Types & Models, Social Business Development, Starting Up a New Technology Venture, Starting Up a New Social Venture, Business Case Discussion

Suggested Books:

Sr.	Name of Book, Author, Publisher	Year of Publication/Reprint
Text Books		
1	by James A. F. Stoner, R. Edward Freeman, Amitabh Deo Kodwani Daniel R.Gilbert, Ranjeet Nambudiri |	2018
2	Principles of Corporate Finance Richard A Brealey, Stewart C Myers, Franklin Allen,Pitabas Mohanty | McGraw Hill	
3	Human Resource Management | Gary Dessler & Biju Varkkey | Pearson Paperback	2013
4	Marketing Management | Philip Kotler, Kevin Keller, Alexander Chernev, Jagdish Sheth, GShainesh	
5	Sales and Distribution Management | Richard R. Still, Edward W. Cundiff, Norman A. P.Govoni, Sandeep Puri (Author) | Pearson Paperback	
6	Operations and Supply Chain Management | Richard B. Chase, Ravi Shankar, F. RobertJacobs | McGraw Hill	

Course Outcome

Sr	Course Outcome	CO
1	Understand the DNA of Business & Companies	CO1
2	Understand Innovation & Quality Manufacturing Processes	CO2
3	Understand Leadership, People and Financial Management, ESG	CO3
4	Understand Business Information & Risk Management	CO4
5	Understand Business Models and Social Business	CO5