PONDICHERRY UNIVERSITY
(A CENTRAL UNIVERSITY)
SCHOOL OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

REGULATIONS, CURRICULUM & SYLLABUS
(For Affiliated Colleges)

B.Sc.(Honors) DEGREE PROGRAMME

B.Sc. Data Science (Honors with Research)

B.Sc. Data Science (Honors)

(Under the National Education Policy 2020)

Effective from the Academic Year 2023 - 2024

November 2023

Table of Contents

1. PREAMBLE

In today's rapidly evolving world, the importance of data and its analysis has never been more profound. The field of Data Science stands at the forefront of this data revolution, driving innovations, solving complex problems, and shaping the future of industries across the globe. Recognizing the profound significance of Data Science in modern society, there is an impending need to introduce an undergraduate program in Data Science.

Data Science, at its core, is the discipline that empowers individuals and organizations to harness the power of data for informed decision-making. It blends mathematics, statistics, computer science, and domain expertise to extract valuable insights, predict future trends, and optimize various processes. This field has already permeated every sector, from healthcare and finance to marketing and environmental sciences, and its impact continues to expand.

The significance of including an undergraduate program in Data Science within educational framework is multifold. Firstly, it addresses the growing demand for professionals with specialized skills in data analysis, machine learning, and artificial intelligence. Secondly, an undergraduate program in Data Science fosters interdisciplinary learning, as students gain proficiency in mathematics, statistics, programming, and data ethics. This interdisciplinary approach encourages a holistic understanding of complex problems and enhances critical thinking and problem-solving skills.

Thirdly, our commitment to Data Science education aligns with the global imperative of cultivating a workforce that can contribute to scientific research, economic growth, and societal welfare. Through this program, we aim to produce graduates who are equipped with the tools and knowledge to tackle real-world challenges and drive innovation.

Moreover, by offering an undergraduate program in Data Science, students are empowered to participate actively in the data-driven society, facilitating data literacy and digital citizenship. This, in turn, enhances their career opportunities and prepares them for a future where data will continue to play an increasingly central role.

In conclusion, introducing an undergraduate program in Data Science is a testament to the commitment to fostering a generation of professionals who can harness the power of data for positive and transformative change. This program embodies the objective to equipping students

with the knowledge, skills, and ethical values necessary to navigate the complexities of the data-driven world, ensuring a brighter future for both individuals and society at large.

The present Curriculum Framework for B.Sc (Data Science) degree is intended to facilitate the students to achieve the following.

- To provide an understanding and knowledge of the basic theory of Computer Science and Information Technology with good foundation on theory, systems and applications such as algorithms, data structures, data handling, data communication and computation
- To offer a strong foundation in Data Science, including statistical analysis, machine learning, data management, and data ethics, in line with the NEP's emphasis on a well-rounded education.
- To equip students with practical skills that are immediately applicable in the industry. It focuses on hands-on experience in data analysis, programming, and using data science tools to nurture skill development, aligning with NEP's skill-based learning approach.
- To encourage students to engage in cross-disciplinary learning, promoting a holistic understanding of data science's real-world applications and connecting it with other domains, fostering a multidisciplinary approach as outlined in the NEP.
- To encourage students to undertake projects, collaborate on research, and contribute to the development of cutting-edge data science solutions.

1. PROGRAM OUTCOMES:

By the end of the program the following outcomes will be achieved by the students:

- Students will demonstrate a solid understanding of fundamental concepts in data science, including statistics, machine learning, data processing, and data visualization.
- Students will possess practical skills in data analysis, programming, and the use of data science tools, enabling them to tackle real-world data challenges effectively.
- Students will be adept at applying data science techniques in various domains, fostering a multidisciplinary approach to problem-solving and decision-making.
- Students will have the ability to conduct data-driven research and innovation in data science, contributing to the advancement of the field.
- Students will excel in effectively communicating complex data-driven insights through reports, visualizations, and presentations.

- Students will be proficient in problem-solving and critical thinking, applying these skills to address data-related challenges creatively.
- Students will be well-prepared for careers in data-related industries and will have an entrepreneurial mindset, capable of developing data-driven business solutions.

3. DEFINITIONS

Terms used in the NEP Regulations shall have the meaning assigned to them as given below unless the context otherwise requires:

A. Credit: A credit is the number of hours of instruction required per week for the given subject in a given semester of 16-18 weeks. One credit is equivalent to 15 hours of teaching (lecture or tutorial) or 30 hours of practice or field work or community engagement and service per Semester.

B. Academic Year: Means the year starting on 1st day of July and ends on the 30th day of June succeeding year.

C. Residence time: Means the time a student spends for attending classes in the College/Institution (either Online/Offline) as a full-time student and enrolled in any Academic programme of the Institution.

D. **Semester:** Means 18 weeks (90 Working days) of teaching-learning session of which two weeks shall be set apart for examinations and evaluation.

E. Grade: Means a letter grade assigned to a student in a course for his/her performance at academic sessions as denoted in symbols of: O(Outstanding), A+(Excellent), A(Very good), B+(Good), B(Above average), C(Average), P(Pass), F(Fail) and Ab(Absent) with a numeric value of O=10, A+=9, A=8, B+=7, B=6, C=5, P=4, and F=0, Ab=0.

F. Grade Point Average (GPA): Means an average of the Grades secured by a student in all courses in a given academic session duly weighted by the number of credits associated to each of the courses.

G. Cumulative GPA (CGPA): Means the weighted average of all courses the student has taken in the entire programme of study.

H. Common courses: Means the set of courses that all students who are admitted are required to study; these courses include, Languages (English- Modern Indian languages), NEP specific courses

viz. Understanding India, Environmental sciences/Education, Health and wellbeing/Yoga, and Digital & Technological solutions.

I. Major Discipline Courses: Means the core subjects mandatory for the Computer Science discipline. These courses are common across all specializations of Computer Science.

J. Minor Discipline Courses: Means allied/elective/specialization specific subjects of Computer Science discipline. Based on the set of Minor Discipline Courses the candidate study, specialization in Computer Science will be awarded. Eg: B.Sc. (Computer Science) with minor discipline courses in Data Science will be awarded as B.Sc. Computer Science with Specialization in Data Science.

K. Credit Requirements: For a Degree/Diploma/Certificate Programme means the minimum number of credits that a student shall accumulate to achieve the status of being qualified to receive the said Degree, Diploma/Certificate as the case may be.

L. Exit option: Means the option exercised by the student, to leave the Programme at the end of any given Academic year.

M: Lateral entry: Means a student being admitted into an ongoing Programme of the University otherwise than in the 1st year of the programme.

N: Vocational Studies/Education: Means set of activities for participation in an approved project or practical or lab, practices of application of scientific theories, studio activities involving students in creative artistic activities, workshop-based activities, field-based shop-floor learning, and Community engagement services, etc. **(These courses are expected to enable students to incorporate the learned skills in daily life and start up entrepreneurship.)**

O: Skill-based learning/project: Means activities designed to understand the different socio-economic contexts, first-hand understanding of the policies, regulations, organizational structures, processes, and programmes that guide the development process.

P: Work-based internship: Means structured internships with Software Companies, Research and Higher Educational Institution Laboratories, Corporate offices, etc. which will further improve employability.

4. AWARD OF UG DEGREE/DIPLOMA/CERTIFICATE

Candidates who complete Eight semesters and earn a minimum of 160 credits will be awarded either of the following degrees after successful completion of the said requirements.

4.1 Degree and Nomenclature

Candidates who complete Eight semesters and earn a minimum of 160 credits will be awarded either of the following degrees after successful completion of the said requirements.

B.Sc. Data Science (Honors with Research) *

B.Sc. Data Science (Honors) **

* - for candidates who complete a research project work in the Eighth Semester

** - for candidates who complete 3 theory courses (MJD 21, MJD 22, and MJD 23) instead of the research project work in the Eighth Semester

4.2 Exit Options

Candidates can exercise the following exit options and obtain the said certificate or diploma or degree, if the minimum required credits are earned and other conditions are met.

Exit after 2nd Semester: Certificate in Data Science will be awarded for candidates who exit the course at the end of 2nd semester and earned a minimum of 40 credits and have completed a Summer Internship of 4 credits for 4 – 6 weeks duration, during the summer vacation post 2nd semester.

Exit after 4th Semester: Diploma in Data Science will be awarded for candidates who exit the course at the end of 4th semester and earned a minimum of 80 credits and have completed a Summer Internship of 4 credits for 4 – 6 weeks duration, during the summer vacation post 4th semester.

Exit after 6th Semester: UG Degree in Data Science will be awarded for candidates who exit the course at the end of 6th semester and earned a minimum of 120 credits and have completed a Summer Internship of 4 credits for 4 – 6 weeks duration, during the summer vacation post 4th semester.

Exit after	Credits and other requirements	Awards
2nd Semester	Min: 40 Credits, Internship 4 – 6 weeks duration	Certificate in Data Science
4th Semester	Min: 80 Credits, Internship 4 – 6 weeks duration	Diploma in Data Science
6th Semester	Min: 120 Credits, Internship 4 – 6 weeks duration	B.Sc. in Data Science

5. PEDAGOGICAL APPROACHES

a) Lecture Courses	Regular classroom lectures by qualified / experienced Expert Teachers • These Lectures may also include classroom discussion, demonstrations, case analysis • Use of Models, Audio-Visual contents, Documentaries, PPTs may supplement.
b) Tutorial Courses	Problem solving Exercise classes guided discussion, supplementary readings vocational training, etc.
c) Practical / Lab work	Practical Lab activity with Theoretical support Mini projects, Activity based engagement, Program executions, Data processing and presentation exercise.
d) Seminar Course	A course requiring student to design and participate in discussions, Group Discussions, Elocution and Debate, Oral Communication Paper presentations, Poster Presentation, Role play participation, Quiz competitions, Business plan preparation/presentation, etc.
e) Internship course	Courses requiring students to *Learn by Doing* in the workplace external to the educational Institutions. Internships involve working in Software Companies, Research and Higher Educational Institution Laboratories, Corporate Offices, etc. All Internships should be properly guided and inducted for focused learning.
f) Research Project	Students need to study and analyze the recent research publications from indexed/peer reviewed journals in their area of specialization. Outcome of the study and analysis need to be presented as a thesis or research report with necessary experimental results.

6. ACADEMIC AUDIT OF COURSES

Internal Quality Assurance Cell (IQAC) at every institution is expected to supervise the implementation of NEP Regulations in these programmes. Availability of required number of Classrooms, Faculty rooms, Labs, Library facilities, Computer Centre and recruitment of Faculty members, allocation of funds for running the Science Labs/Computer Centre etc., is the responsibility of University / College Administration.

7. ADMISSIONS & LATERAL ENTRY

7.1 Admissions Eligibility:

For Affiliated Colleges: The candidates for admission to this programme shall be required to have passed 10+2 / 10+3 system of examinations or equivalent with mathematics / business mathematics / equivalent as one of the subjects of study.

Students shall be admitted to this programme based on admissions criteria fixed by the University / Government of Puducherry from time to time.

7.2 Lateral Entry:

As per NEP, students have a choice of exit and entry into the programme multiple number of times. UGC specifies that about 10% of seats over and above the sanctioned strength shall be allocated to accommodate the Lateral Entry students.

Candidates seeking entry at the second, third and fourth year, should meet the necessary eligibility criteria with respect to the certificate / diploma / degree they possess, with necessary minimum credits banked in the Academic Bank of Credits (ABC). Such students who get admitted in later years, other than first year will be guided by the following clauses:

- that the University shall notify the admission process and number of vacancies open for lateral entry.
- that the Lateral entrants shall be admitted only after such transparent screening process and such procedure that the University may prescribe from time to time. University may prescribe different methods of screening for different programmes depending on the circumstances prevailing in each case.
- Lateral entry shall be permissible only in the beginning of years 2, 3, 4 of the Under Graduate / Honors programme; provided that the students seeking lateral entry shall have obtained the minimum pass marks / grades fixed by the University in their previous academic years.

8. EVALUATION (INTERNAL & END SEMESTER ASSESSMENT) AND GRADES

All Credit courses are evaluated for 100 marks. Internal Assessment component is for 25 marks and the End Semester University exam is for 75 marks for theory courses. In case of practical courses, research project work etc., Internal Assessment component is for 50 marks and the End Semester University exam is for 50 marks.

Internal Test Scheme: Principal of the College schedules the Mid-Semester Exam for all courses during 8/9th week of start of classes. Mid-Semester exam for 90 minutes' duration need to be

conducted for all these theory courses. The evaluated marks need to be uploaded to Controller of Examinations of University. The answer books of Mid-Semester exams need to be preserved until the declaration of results by the University.

8.1 INTERNAL ASSESSMENTS (for Courses upto 6th Semester)

8.1.1 Internal Assessment Marks for Theory subjects

Total Internal Assessment mark for a theory subject is 25 marks. The breakup is as follows:

Evaluation Component	Marks
A. Mid Semester Exam (one)	20
B. Percentage of Attendance	05
Total	**25**

8.1.2 Internal Assessment marks for Practical / Internships subjects

Faculty member in-charge of Lab practical shall evaluate the practical subjects for 50 marks. The breakup is as follows:

Evaluation Component	Marks
A. Mid-Semester Practical Exam (one) / Viva-voce	20
B. Practical Record / Internship Report	25
C. Percentage of Attendance	05
Total	**50**

8.1.3 Internal Assessment marks for Research Project Work

There shall be a faculty member assigned as a Project Guide for each candidate doing the Research Project. Progress of the candidate can be assessed once in a month in a project review meeting. Three project review meetings shall be conducted for Internal Assessment.

Project review committee may be constituted and the committee shall organize project review meetings and evaluate the progress and to award the Internal Assessment marks. Internal Assessment component for the Research Project is 50 Marks. The breakup is as follows:

Evaluation Component	Marks
A. Monthly Review (3 Reviews – 10 Marks each)	30
B. Project Report	10
C. Project Presentation and viva-voce	10
Total	**50**

8.1.4 Internal Assessment marks for Theory Subjects with Practical Components

Faculty member in-charge of Theory Subjects with Practical Component shall evaluate the candidates both for their performance in theory and practical. Internal Assessment marks for Theory Subjects with Practical Components is 25 marks. The breakup is as follows:

Evaluation Component	Marks
A. Mid Semester Exam (one)	15
B. Observation Note / Practical Record	05
C. Percentage of Attendance	05
Total	**25**

8.1.5 Marks for Attendance is as follows

Attendance %	Marks
Below 75%	0
75% - 80%	1
80% - 85%	2
85% - 90%	3
90% - 95%	4
95% - 100%	5

8.2 END SEMESTER ASSESSMENTS (for Courses upto 6th Semester)

Controller of Examinations (COE) of Pondicherry University schedules the End-Semester exams for all theory and practical subjects based on university calendar. For Theory courses with Practical components, End semester exams shall be conducted separately for Theory and Practical.

A detailed Exam Time Table shall be circulated at least 15 days before the start of exams, mostly during 15/16th week of the Semester. Question Papers shall be set externally based on BoS approved syllabus. All students who have a minimum of 70% attendance are eligible to attend the end-semester exams. Attendance percentage shall be calculated for each course to decide the eligibility of the candidate for writing the end-semester examination.

8.2.1 Breakup of End Semester Marks

(All End Semester Exams shall be conducted by the Pondicherry University)
The question paper shall be set as per the Bloom's Taxonomy. Various levels along with it's description and sample questions are as follows:

Knowledge: Recall or remember previously learned information.
Example: List the basic data types in Python

Comprehension: Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas.
Example: Explain how a stack data structure works.

Application: Apply knowledge and concepts to solve problems in new situations. Use learned information in a different context.
Example: Write a Python program to solve the deadlock problem.

Analysis: Break down information into parts and examine the relationships between the parts. Identify motives or causes.
Example: Analyse the efficiency of two sorting algorithms and compare their advantages and disadvantages.

Synthesis: Create a new whole by combining elements in novel ways. Use creativity to produce something original.
Example: Design a web application that can generate a time table of a school.

Distribution of questions at various levels are as indicated.

Course Components	Max. Marks	End-Sem Exam Duration
A. Theory subjects: Sec A: 10 Questions of 2 Marks each (20 Marks) *(Knowledge : 3, Comprehension : 2, Application : 3, Analysis:2)* Sec B: 5 out of 7 Questions of 5 Marks each (25 Marks) *(Knowledge : 1, Comprehension : 2, Application : 1, Analysis:3)* Sec C: 2 Either/OR choice questions of 15 Marks each (30 Marks) *(Application : 1, Analysis:1)*	75 Marks	3 Hours

Questions from all units of Syllabus equally distributed.		
B. Skill Enhancement/ Practical/Internship/Project Work subjects: **Skill Enhancement / Practical Subjects:** Based on Practical Exams conducted by CoE of University **Internship / Research Project Work:** Presentation of the work / Report / Viva-voce examinations	50 Marks	3 Hours
C. Theory Subjects with Practical Components: **i. Theory Component:** Sec A: 5 Questions of 2 Marks each (10 Marks) *(Knowledge : 3, Comprehension : 2, Application : 3, Analysis:2)* Sec B: 5 out of 7 Questions of 4 Marks each (20 Marks) *(Comprehension : 2, Application : 3, Analysis:2)* Sec C: 2 Either or type questions of 10 Marks each (20 Marks) *(Analysis / Synthesis)* **Questions from all units of Syllabus equally distributed.** **ii. Practical Component:** Based on Practical Exams / Presentation / Viva-voce with external examiner appointed by the University Controller of Examinations, and schedules exclusively prepared for such practical examinations by the University Examination Section. **The examination shall be conducted for 50 Marks and reduced to 25 Marks.** **Total Marks: 75 (Theory: 50 Marks + Practical: 25 Marks)**	50 Marks 25 Marks	3 Hours 3 Hours

8.3 CONSOLIDATION OF MARKS AND PASSING MINIMUM

Controller of Examinations of the University consolidates the Internal Assessment marks uploaded by the Colleges and marks secured by students in End-Semester examinations. The total marks will be converted into letter grades. The passing minimum is 40% marks (Internal Assessment + End Semester Assessment put together) and students who secure between 40% and 49% will be awarded 'P' (Pass Grade).

8.3.1 Arrear Exam

A student who secures less than 40% marks in aggregate is declared as ***Fail*** and that student is eligible to take up supplementary examination by registering to the failed course in the following

Semester. All other candidates who failed due to shortage of attendance and those who are seeking to improve the grade shall repeat the course.

8.3.2 Letter Grades and Calculation of CGPA

Total marks secured by a student in each subject shall be converted into a letter grade. UGC Framework has suggested a Country wide uniform letter grades for all UG courses. The following table shows the seven letter grades and corresponding meaning and the grade points for calculation of CGPA.

Equivalent Letter Grade	Meaning	Grade Points for Calculation of CGPA
O	Outstanding	10
A+	Excellent	9
A	Very Good	8
B+	Good	7
B	Above Average	6
C	Average	5
P	Pass	4
F	Fail	0
Ab	Absent	0

In order to work out the above letter grades, the marks secured by a student (Total of Internal Assessment and End Semester Assessment) would be categorized for relative grading.

The range of marks for each grade would be worked as follows:

- Highest marks in the given subject: X
- Cut of marks for grading purpose: 50 marks
- Passing minimum: 40
- Number of grades (except P - Pass) (O, A+, A, B+, B, C): G = 6
- Range of marks: K = (X - 50) / G

(i) If $K \geq 5$, then the grades shall be awarded as given in the following table.

Range of Marks in %	Letter Grade Points for	Grade Points for
X to (X-K) + 1	O	10
(X-K) to (X-2K) + 1	A+	9
(X-2K) to (X-3K) + 1	A	8
(X-3K) to (X-4K) + 1	B+	7
(X-4K) to (X-5K) + 1	B	6

(X-5K) to 50	C	5
40 – 49	P	4
Below 40	F	0
Absent (Lack of Attendance)	Ab	0

(ii) If K< 5, then the grades shall be awarded as given in the following table.

Range of Marks in %	**Letter Grade Points for**	**Grade Points for**
80-100	O	10
71-79	A+	9
66-70	A	8
61-65	B+	7
56-60	B	6
50-55	C	5
40-49	P	4
Below 40	F	0
Absent (lack of attendance)	Ab	0

8.3.3 Calculation of Semester Grade Point Average and Cumulative Grade Point Average

Semester Grade Point Average (SGPA) is calculated by taking a weighted average of all grade points secured by a candidate from all subjects registered by him/her in the given Semester. The weights being the number of credits that each subject carries.

Cumulative Grade Point Average (CGPA) shall be calculated as the weighted average of credits that course carries and the value of Grade points averaged for all subjects.

8.3.4 Computation of SGPA and CGPA

The following procedure shall be followed to compute the Semester Grade Point Average (SGPA) and Cumulative Grade Point Average (CGPA):

The SGPA is the ratio of the sum of the product of the number of credits with the grade points scored by a student in all the courses taken by a student to the sum of the number of credits of all the courses undergone by a student, i.e. SGPA $(S_i) = \Sigma(C_i \times G_i) / \Sigma C_i$

where C_i is the number of credits of the i^{th} course and G_i is the grade point scored by the student in the i^{th} course.

(i) Example for Computation of SGPA where candidate has not failed in any course

Semester	Course	Credit	Letter Grade	Grade point	Credit Point (Credit x Grade)
I	Course 1	3	A	8	3 X 8 = 24
I	Course 2	4	B+	7	4 X 7 = 28
I	Course 3	3	B	6	3 X 6 = 18
I	Course 4	3	O	10	3 X 10 = 30
I	Course 5	3	C	5	3 X 5 = 15
I	Course 6	4	B	6	4 X 6 = 24
		20			139
				SGPA	**139/20=6.95**

(ii) Example for Computation of SGPA where candidate has failed in one course

Semester	Course	Credit	Letter Grade	Grade point	Credit Point (Credit x Grade)
I	Course 1	3	A	8	3 X 8 = 24
I	Course 2	4	B+	7	4 X 7 = 28
I	Course 3	3	B	6	3 X 6 = 18
I	Course 4	3	O	10	3 X 10 = 30
I	Course 5	3	C	5	3 X 5 = 15
I	Course 6	4	F	0	4 X 0 = 00
		20			115
				SGPA	**115/20=5.75**

(iii) Example for Computation of SGPA where candidate has failed in two courses

Semester	Course	Credit	Letter Grade	Grade point	Credit Point (Credit x Grade)
I	Course 1	3	A	8	3 X 8 = 24
I	Course 2	4	B+	7	4 X 7 = 28
I	Course 3	3	F	0	3 X 0 = 00
I	Course 4	3	B	6	3 X 6 = 18
I	Course 5	3	C	5	3 X 5 = 15
I	Course 6	4	F	0	4 X 0 = 00
		20			85
				SGPA	**85/20=4.25**

The CGPA shall also be calculated in similar way as shown in examples (i), (ii) and (iii) of SGPA for all subjects taken by the students in all the semesters. However, if any student fails more than once in the same subject, then while calculating CGPA, the credit and grade point related to the subject in which the student fails in multiple attempts will be restricted to one time only. The SGPA and CGPA shall be rounded off to 2 decimal points and reported in the transcripts.

In case of audit courses offered, the students may be given (P) or (F) grade without any credits. This may be indicated in the mark sheet. Audit courses will not be considered towards the calculation of CGPA.

8.3.5 Declaration of Results

Controller of Examinations (COE) of the University shall declare the results of given UG programme following the CGPA secured by students by the end of 6th Semester and 8th Semester.

8.3.6 Classification of Divisions

Range of CGPA	Result
9.0 – 10	First Class with distinction#
6.0 - 8.99	First Class
5.0 - 5.99	Second Class
4.0 - 4.99	Pass

Distinction will be awarded ONLY to those candidates who have cleared ALL subjects in the first attempt.

8.4 INTERNAL ASSESSMENT / END-SEMESTER ASSESSMENT / PASSING MINIMUM / GRADES (FOR 7TH & 8TH SEMESTERS)

Regulations to be notified in the next revision after the confirmation from University NEP committee.

9.MINIMUM CREDITS REQUIREMENT

S.No	Component	3-year UG			4-year UG (Honors / Honors With research)		
		Credits	Courses	Cr/Course	Credits	Courses	Cr/Course
1	Major Disciplinary/ Interdisciplinary Courses	56	14	4	76	19	4
2	Minor Disciplinary/ Interdisciplinary Courses	24	6	4	32	8	4
3	Multi-Disciplinary Courses	9	3	3	9	3	3
4	Ability Enhancement Courses	8	4	2	8	4	2
5	Skill Enhancement	9	3	3	9	3	3

	Courses						
6	Value-added courses	8	4	2	8	4	2
7	Summer Internship (MJD 11)	4	1	4	4	1	4
8	Community Engagement and Service	2	1	2	2	1	2
9	Research Project/Dissertation	--	--	--	12	Project or 3 Courses[##]	
Total		**120**			**160**		

***Note:** Honors students not undertaking research will do 3 courses for 12 credits in lieu of a research project/Dissertation.*

- *MJD: Major Disciplinary (Compulsory – Hardcore Subjects)*
- *MID: Minor Disciplinary (Specialization Specific – Softcore Subjects)*
- *MLD: Multi-Disciplinary*
- *AEC: Ability Enhancement Courses*
- *SEC: Skill Enhancement Courses*
- *VAC: Value Added Courses*
- *Course Code: DS1MJ01 (DS-B.Sc Data Science , 1-Semester, MJ-Component, 1-Course Number in the respective component, E-Elective)*
-

ANNEXURE I

CURRICULUM

FIRST SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 1	DS1MJ01	Digital Logic Fundamentals	H	4	3		2
2	MID 1	DS1MI01	Foundations of Data Science-I	S	4	3		2
3	MLD 1		One course from the MLD streams 1 to 10 (Table 15)	H	3	4		
4	AEC 1	DS1AE01	English I	H	2	2		2
5	SEC 1		S.No. 1 or 2 from Table 7	S	3	2		2
6	VAC 1	DS1VA01	Understanding India	H	2	4		0
7	VAC 2		Environmental Science/ Education/Higher Order Thinking	H	2	4		0
				Total	20	30 Hours		

SECOND SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 2	DS2MJ02	Problem Solving & Programming Fundamentals	H	4	3		2
2	MID 2	DS2MI02	Foundations of Data Science-II	S	4	3		2
3	MLD 2		One course from the MLD streams 1 to 10 except the stream chosen in MLD1 (Table 15)	H	3	4		
4	AEC 2	DS2AE02	Indian Language I	H	2	2		2
5	SEC 2		S.No. 3 or 4 from Table 7	S	3	2		2
6	VAC 3		Health & Wellness/Yoga Education/Universal Human Values	H	2	2		2
7	VAC 4	DS2VA04	Digital Technologies	H	2	4		
				Total	20	30 Hours		

THIRD SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 3	DS3MJ03	Mathematical Foundations of CS	H	4	4	1	
2	MJD 4	DS3MJ04	Data Structures	H	4	3		2
3	MID 3	DS3MI03	Probability & Statistics	S	4	3		2
4	MLD 3		One course from the MLD streams 1 to 10 except the streams chosen in MLD1 and MLD2 (Table 15)	H	3	4		
5	AEC 3	DS3AE03	English II	H	2	2		2
6	SEC 3		S.No. 5 or 6 from Table 7	S	3	2		2
				Total	20	27 Hours		

FOURTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 5	DS4MJ05	Computer System Architecture	H	4	3		2
2	MJD 6	DS4MJ06	Design and Analysis of Algorithms	H	4	3		2
3	MJD 7	DS4MJ07	Object Oriented Programming	H	4	3		2
4	MID 4	DS4MI04	Applied Regression Analysis	S	4	3		2
5	AEC 4	DS4AE04	Indian Language II	H	2	2		2
6	CES 1	DS4CS01	Community Engagement and Service	H	2			6
				Total	20	30 Hours		

FIFTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 8	DS5MJ08	Operating Systems	H	4	3		2
2	MJD 9	DS5MJ09	Database Management Systems	H	4	3		2
3	MJD 10	DS5MJ10	Management Strategies & Concepts	H	4	4		
4	MID 5	DS5MI05	Artificial Intelligence	S	4	3	2	
5	MJD 11	DS5MJ11	Summer Internship	H	4			6
				Total	20	25 Hours		

SIXTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 12	DS6MJ12	Computer Networks	H	4	3		2
2	MJD 13	DS6MJ13	Software Engineering Theory and Practice	H	4	3		2
3	MJD 14	DS6MJ14	System Modelling & Simulation	H	4	3		2
4	MJD 15	DS6MJ15	Web Engineering	H	4	3	2	
5	MID 6		Any one course from Table 1	S	4	3		2
				Total	20	25 Hours		

SEVENTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 16	DS7MJ16	Software Testing and Quality Assurance	H	4	3		2
2	MJD 17	DS7MJ17	Distributed Systems	H	4	3		2
3	MJD 18	DS7MJ18	Wireless Communication Networks (5G)	H	4	3		2
4	MID 7		Any one course from Table 2	S	4	3		2
5	MID 8		Any one course from Table 3	S	4	3		2
				Total	20	25 Hours		

EIGHTH SEMESTER – B.Sc. Data Science (Honors)								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 19		Any one course from Table 4	S	4	3		2
2	MJD 20		Any one course from Table 5	S	4	3		2
3	MJD 21	DS8MJ21	Deep Learning	H	4	3		2
4	MJD 22	DS8MJ22	Time Series Analysis	H	4	3		2
5	MJD 23	DS8MJ23	Natural Language Processing	H	4	3		2
				Total	20	25 Hours		

EIGHTH SEMESTER – B.Sc. Data Science (Honors with Research)								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 19		Any one course from Table 4	S	4	3		2
2	MJD 20		Any one course from Table 5	S	4	3		2
3	MJD 21	DS8MJ24	Research Project	H	4			5
4	MJD 22	DS8MJ25	Project Report	H	4			5
5	MJD 23	DS8MJ26	Project Viva-voce	H	4			5
				Total	20	25 Hours		

Table 1: MID 6 – SIXTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MID 6	DS6MI06E1	Massive Data Management	S	4	3		2
2	MID 6	DS6MI06E2	Hadoop Eco System	S	4	3		2

Table 2: MID 7 – SEVENTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MID 7	DS7MI07E1	Big Data Analytics	S	4	3		2
2	MID 7	DS7MI07E2	Predictive Analytics	S	4	3		2

Table 3: MID 8 – SEVENTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MID 8	DS7MI08E1	Data Mining	S	4	3		2
2	MID 8	DS7MI08E2	Text and Speech Analytics	S	4	3		2

Table 4: MJD 19 – EIGHTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 19	DS8MJ19E1	Machine Learning	S	4	3		2
2	MJD 19	DS8MJ19E2	Health Care analytics	S	4	3		2

Table 5: MJD 20 – EIGHTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 20	DS8MJ20E1	Business analytics	S	4	3		2
2	MJD 20	DS8MJ20E2	Social Network Analysis	S	4	3		2

Table 6: MJD 21 / MJD 22 / MJD 23 – EIGHTH SEMESTER								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	MJD 21	DS8MJ21	Deep Learning	S	4	3		2
2	MJD 22	DS8MJ22	Time Series Analysis	S	4	3		2
3	MJD 23	DS8MJ23	Natural Language Processing	S	4	3		2

Table 7: List of Skill Enhancement Courses/ SEC 1 / SEC 2 / SEC 3 – I / II / III SEMESTERs								
S.No	Component	Course Code	Title of the Course	H/S	Credits	Hours/Week		
						L	T	P
1	SEC 1	DS1SE01E1	Python Programming	S	3	3		2
2	SEC 1	DS1SE01E2	R Programming	S	3	3		2
3	SEC 2	DS2SE02E1	Exploratory Data Analysis with Python	S	3	3		2
4	SEC 2	DS2SE02E2	Data wrangling with R	S	3	3		2
5	SEC 3	DS3SE03E1	Interactive Data Visualization	S	3	3		2
6	SEC 3	DS3SE03E2	Financial Data Analytics	S	3	3		2

Table 8: List of Major Disciplinary Courses				
S.No	Component	Course Code	Title of the Course	H/S
1.	MJD 1	DS1MJ01	Digital Logic Fundamentals	H
2.	MJD 2	DS2MJ02	Problem Solving & Programming Fundamentals	H
3.	MJD 3	DS3MJ03	Mathematical Foundations of Computer Science	H
4.	MJD 4	DS3MJ04	Data Structures	H
5.	MJD 5	DS4MJ05	Computer System Architecture	H
6.	MJD 6	DS4MJ06	Design and Analysis of Algorithms	H
7.	MJD 7	DS4MJ07	Object Oriented Programming	H
8.	MJD 8	DS5MJ08	Operating Systems	H
9.	MJD 9	DS5MJ09	Database Management Systems	H
10.	MJD 10	DS5MJ10	Management Strategies & Concepts	H
11.	MJD 11	DS5MJ11	Summer Internship	H
12.	MJD 12	DS6MJ12	Computer Networks	H
13.	MJD 13	DS6MJ13	Software Engineering Theory and Practice	H
14.	MJD 14	DS6MJ14	System Modeling & Simulation	H
15.	MJD 15	DS6MJ15	Web Engineering	H
16.	MJD 16	DS7MJ16	Software Testing and Quality Assurance	H
17.	MJD 17	DS7MJ17	Distributed Systems	H
18.	MJD 18	DS7MJ18	Wireless Communication Networks (5G)	H
19.	MJD 19		Machine Learning / Health Care Analytics	S
20.	MJD 20		Business analytics / Social Network Analysis	S

Table 9: List of Minor Disciplinary Courses				
S.No	Component	Course Code	Title of the Course	H/S
1.	MID 1	DS1MI01	Foundations of Data Science - I	S
2.	MID 2	DS2MI02	Foundations of Data Science - II	S
3.	MID 3	DS3MI03	Probability & Statistics	S
4.	MID 4	DS4MI04	Applied Regression Analysis	S
5.	MID 5	DS5MI05	Artificial Intelligence	S
6.	MID 6		Massive Data Management / Hadoop Eco System	S
7.	MID 7		Big Data Analytics / Predictive Analytics	S
8.	MID 8		Data Mining / Text and speech analytics	S

Table 10: List of Multi-disciplinary Courses				
S.No	**Component**	**Course Code**	**Title of the Course**	**H/S**
1.	**MLD 1**	**DS1ML01**	Natural Sciences	**H**
2.	**MLD 2**	**DS2ML02**	Physical Sciences	**H**
3.	**MLD 3**	**DS3ML03**	Humanities & Social Sciences	**H**

Table 11: List of Ability Enhancement Courses				
S.No	**Component**	**Course Code**	**Title of the Course**	**H/S**
1.	**AEC 1**	**DS1AE01**	English I	**H**
2.	**AEC 2**	**DS2AE02**	Indian Language I	**H**
3.	**AEC 3**	**DS3AE03**	English II	**H**
4.	**AEC 4**	**DS4AE04**	Indian Language II	**H**

Table 12: List of Skill Enhancement Courses				
S.No	**Component**	**Course Code**	**Title of the Course**	**H/S**
1	**SEC 1**	**DS1SE01E1**	**Python Programming**	**S**
2	**SEC 1**	**DS1SE01E2**	**R Programming**	**S**
3	**SEC 2**	**DS2SE02E1**	**Exploratory Data Analysis with Python**	**S**
4	**SEC 2**	**DS2SE02E2**	**Data wrangling with R**	**S**
5	**SEC 3**	**DS3SE03E1**	**Interactive Data Visualization**	**S**
6	**SEC 3**	**DS3SE03E2**	**Financial Data Analytics**	**S**

Table 13: List of Value-Added Courses				
S.No	**Component**	**Course Code**	**Title of the Course**	**H/S**
1.	**VAC 1**	**DS1VA01**	Understanding India	**H**
2.	**VAC 2**		Environmental Science/ Education / Higher Order Thinking	**H**
3.	**VAC 3**		Health & Wellness / Yoga Education / Universal Human Values	**H**
4.	**VAC 4**	**DS2VA04**	Digital Technologies	**H**

Table 14: Project (WP/ Internship)				
S.No	**Component**	**Course Code**	**Title of the Course**	**H/S**
1.	**CES 1**	**DS4CS01**	**Community Engagement and Service**	**H**

*Table 15: MLD 1 / MLD 2 / MLD 3 in Sem 1 / Sem 2 / Sem 3				
S.No	**Streams**	**Course Code**	**Title of the Course**	**H/S**
1.	**Natural Science**		**Biology**	**H**
2.			**Botany**	**H**
3.			**Zoology**	**H**
4.			**Biotechnology**	**H**
5.			**Biochemistry**	**H**
6.	**Physical Sciences**		**Chemistry**	**H**
7.			**Physics**	**H**
8.			**Biophysics**	**H**
9.			**Astronomy**	**H**
10.			**Astrophysics**	**H**
11.			**Earth and Environmental Sciences**	**H**
12.	**Mathematics and Statistics**		**STATA**	**H**
13.			**SPSS**	**H**
14.			**Tally**	**H**
15.	**Computer Science & Applications**	**DS1SE01E1**	**Python Programming**	**H**
16.		**DS1SE01E2**	**R Programming**	**H**
17.		**DS2SE02E1**	**Exploratory Data Analysis with Python**	**H**

*Courses will be announced after the approval of the respective boards.

SYLLABUS
SEMESTER I

Year	I	Course Code: DS1MJ01	Credits	4
Sem.	I	Course Title: Digital Logic Fundamentals	Hours	75
Course Prerequisites, if any	Nil			
Internal Assessment Marks: 25	End Semester Marks: 75	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	• Understanding the postulates of Boolean algebra and to minimize combinational functions. • Gaining knowledge to design and analyze combinational and sequential circuits. • Learning techniques for the design of digital circuits			

Unit No.	Course Content	Hours
	Theory Component	
Unit I	**Digital Systems and Binary Numbers** Digital Systems - Binary Numbers - Number-Base Conversions - Octal and Hexadecimal Numbers - Complements of Numbers - Signed Binary Numbers - Binary Codes - Binary Storage and Registers - Binary Logic - Axiomatic Definition of Boolean Algebra - Basic Theorems and Properties of Boolean Algebra - Boolean Functions Canonical and Standard Forms - Other Logic Operations - Digital Logic Gates - Integrated Circuits	9
Unit II	**Gate-Level Minimization** Introduction - The Map Method - Four-Variable K-Map - Product-of-Sums Simplification - Don't-Care Conditions - NAND and NOR Implementation - Other Two-Level Implementations - Exclusive-OR Function - Hardware Description Language	9
Unit III	**Combinational Logic** Introduction - Combinational Circuits - Analysis Procedure - Design Procedure - Binary Adder–Subtractor - Decimal Adder - Binary Multiplier - Magnitude Comparator – Decoders – Encoders – Multiplexers - HDL Models of Combinational Circuits.	9
Unit IV	**Synchronous Sequential Logic** Introduction - Sequential Circuits - Storage Elements: Latches - Storage Elements: Flip-Flops - Analysis of Clocked Sequential Circuits - Synthesizable HDL Models of Sequential Circuits - State Reduction and Assignment - Design Procedure	9
Unit V	**Registers and Counters** Registers - Shift Registers - Ripple Counters - Synchronous Counters - Other Counters - HDL for Registers and Counters	9
	Practical Component	
Exercises	1. Binary to Decimal and vice-versa in Python 2. Decimal to Hexadecimal and Vice-Versa in Python 3. Digital Logic Gates in Python 4. Simplification of Boolean Functions in Python 5. Combinational Logic Circuits in Python i. Code Converters ii. Arithmetic (Adders, Subtractors, Multipliers, Comparators) iii. Data Handling (Multiplexers, Demultiplexers, Encoders &	30

	Decoders) 6. Combinational Logic Circuit Design in Python 7. Binary Adder-Subtractor Simulation in Python 8. Decimal Adder Simulation in Python 9. Binary Multiplier Simulation in Python 10. Sequential Circuit Storage Elements: Flip-Flop Simulation in Python (Many more programs can be included related to programming the Digital logic in Python)	
Recommended Learning Resources		
Print Resources	1. M. Morris Mano , Michael D. Ciletti,, Digital design With an Introduction to the Verilog HDL, Pearson, Fifth Edition, 2013, ISBN-13: 978-0-13-277420-8, ISBN-10: 0-13-277420-8. 2. M. Rafiquzzaman, Fundamentals of Digital Logic and Microcomputer Design, John Wiley & Sons, Inc., Fifth Edition, 2005.	

Year	I	Course Code: DS1MI01	Credits	4
Sem.	I	Course Title : Foundations of Data Science - I	Hours	75
Course Prerequisites, if any	Nil			
Internal Assessment Marks : 25	End Semester Marks : 75	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	**Students will be able to** • Demonstrate an understanding of the need for Data Science, the Data Science Process, and its application in Business Intelligence. • Apply data analysis tools in exploring and understanding datasets. • Apply probability and statistics concepts to analyze random variables, probability distributions, and sample statistics for hypothesis testing • Demonstrate proficiency in linear algebraic operations, matrix decomposition, and their application in representing relationships between data. • Apply database querying and management in a Data Science context.			

Unit No.	Course Content	Hours
	Theory Component	
Unit I	**INTRODUCTION** Need for Data Science – Data Science Process – Business Intelligence and Data Science – Prerequisites for a Data Scientist. Exploratory Data Analysis - Statistical measures, Basic tools (plots, graphs, and summary statistics) of EDA, Data Analytics Life-cycle, Preparing Data, Data Visualization, Uni-variant, Multi variant Analysis	7
Unit II	**PROBABILITY AND STATISTICS** Probability**:** Probability, Random Variables and Their Probability distribution, Multiple random variables, Sample statistics and their distribution. Statistics**:** Developing Initial Hypotheses, Identifying Potential Data Sources, and Testing hypotheses on means, proportions, and variances.	10
Unit III	**LINEAR ALGEBRA** Matrices to represent relations between data, Linear algebraic operations on matrices – Matrix Decomposition – Singular Value Decomposition – Principal Component Analysis.	10
Unit IV	**DATABASES FOR DATA SCIENCE** Structured Query Language (SQL): Data Munging, Filtering, Joins, Aggregation, Window Functions, Ordered Data, No-SQL, Document Databases, Wide-column Databases and Graphical Databases. Unstructured data: MongoDB, JSON.	8
Unit V	**DATA SCIENCE METHODOLOGY** Analytics for Data Science – Examples of Data Analytics – Data Analytics Lifecycle: Data Discovery, Data Preparation, Model Planning, Model Building.	10

Practical Component		
Exercises	1. Download, install NumPy, SciPy and pandas in Python. 2. Build a data frame using pandas from a csv file. 3. Write a program for finding the frequency, Mean, Median, Mode, Variance, and Standard Deviation of data using Python pandas data-frame. 4. Plot a graph for probability distribution using Python (Normal Distribution). 5. Create a database and establish relationships between tables. 6. Create view to extract details from two or more tables. 7. Demonstrate descriptive Statistics like mean, median, variance, and correlation for sample data 8. Demonstrate Missing value analysis using sample data. 9. Create a graph database using python. 10. Perform data analysis using SciPy .	30
Recommended Learning Resources		
Print Resources	1. Sanjeev Wagh, Manisha Bhende, Anuradha Thakare, 'Fundamentals of Data Science, CRC Press, 1st Edition, 2022. 2. Jure Leskovek, Anand Rajaraman and Jefrey Ullman., Mining of Massive Datasets. v2.1, Cambridge University Press, 2019. 3. Seema Acharya, Subhasini Chellappan, Big Data Analytics, 2nd Edition, Wiley, 2019. 4. Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", Cambridge University Press, 1st Edition, 2020. 5. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media, 1st Edition, 2015. 6. Ani Adhikari and John DeNero, 'Computational and Inferential Thinking: The Foundations of Data Science', GitBook, 2019. 7. Cathy O'Neil and Rachel Schutt, Doing Data Science, Straight Talk From The Frontline, O'Reilly, 2014. 8. Big Data and Business Analytics, Jay Liebowitz, CRC press, 2013. 9. Data mining methods,2nd edition, C. Rajan, Narosa , 2016.	

SKILL ENHANCEMENT COURSES

Year	I	**Course Code: DS1SE01E1**	**Credits**	3
Sem.	I	**Course Title : PYTHON PROGRAMMING**	**Hours**	60
Course Prerequisites, if any	Basic mathematical problem solving skills			
Internal Assessment Marks: 50	End Semester Marks: 50	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	• Understand the basics of writing Python code • Implement programs using lists, tuples and dictionaries • Understand the use of control structures • Ability to write programs using packages • Understand the file manipulation			
Unit No.	**Course Content**		**Hours**	
Theory Component				
Unit I	**Introduction, Data types** Introduction to Python – Advantages of using Python – Executing Python Programs – Python's Core data types – Numeric Types – String Fundamentals.		6	
Unit II	**Lists, Tuples, Dictionaries** Lists: list operations, list slices, list methods, list loop, mutability, aliasing, cloning lists, list parameters; Tuples: tuple assignment, tuple as return value; Dictionaries: operations and methods; advanced list processing – list comprehension.		6	
Unit III	**Control Flow, Functions, Modules** Python Statements: Assignments – Expressions – If condition – While and For Loops. Functions: Definition, Calls – Scopes – Arguments – Recursive Functions– Functional Programming tools. Classes and Object Oriented programming with Python - Modules and Packages: Purpose, using packages– Exception Handling with Python.		6	

Unit IV	**Packages** Packages: NumPy, Pandas, Scikit learn - Machine learning with Python – Cleaning up, Wrangling, Analysis, Visualization - Matplotlib package – Plotting Graphs.	6
Unit V	**File Handling** Files and exception: text files, reading and writing files, format operator; command line arguments, errors and exceptions, handling exceptions	6
Practical Component		
Exercises	1. Exchange the values of two variables 2. Finding minimum among n variables 3. Perform Simple sorting 4. Generate Students marks statement 1. Find square root, GCD, exponentiation 2. Sum the array of numbers 3. Perform linear search, binary search 4. Perform Matrix operations using NumPy 5. Perform Dataframe operations using Pandas 6. Use Matplotlib on dataset and visualise 7. Perform Word count, copy file operations	30
Recommended Learning Resources		
Print Resources	1. Mark Lutz, “Learning Python”, Fifth Edition, O’Reilly, 2013. 2. Daniel Liang, “Introduction to programming using Python”, Pearson, First edition, 2021. 3. Wes Mc Kinney, “Python for Data Analysis”, O’Reilly Media, 2012. 4. Tim Hall and J-P Stacey, “Python 3 for Absolute Beginners”, Apress, First Edition, 2009. 5. Magnus Lie Hetland, “Beginning Python: From Novice to Professional”, Apress, Second Edition, 2005.	

Year	I	**Course Code: DS1SE01E2**	**Credits**	3
Sem.	I	**Course Title : R Programming**	**Hours**	60
Course Prerequisites, if any	Basic mathematical problem solving skills			
Internal Assessment Marks : 50	End Semester Marks : 50	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	**Students will be able to** 1. Demonstrate proficiency in R programming basics, including installation, opening, saving, and editing R code, following established conventions. 2. Perform basic math operations, assign objects, and manipulate vectors in R. 3. Analyze and manipulate matrices and arrays, demonstrating skills in defining, subsetting, and performing algebraic operations on matrices. 4. Create and manipulate lists and data frames, gaining an understanding of the structure of objects and the versatility of data frames in R 5. Read and write files in R, handling R-ready datasets and external data files, showcasing competence in data file manipulation and storage.			

Unit No.	**Course Content**	**Hours**
	Theory Component	
Unit I	**Introduction** R Installation – opening – Saving and Editing – Conventions **Number, Arithmetic, assignment & Vectors** R for Basic Math – Assigning Objects – Vectors.	6
Unit II	**Matrices and Arrays** Defining a Matrix – Subsetting – Matrix Operations & Algebra – Multidimensional Arrays.	6
Unit III	**Non-numeric Values :** Logical Values – Characters - Factors **Lists and Data Frames:** Objects – Data Frames.	6
Unit IV	**Special Values , Classes, and Coercion** Some special values – Understanding Types, classes and Coercion **Basic Plotting** Using Plot with coordinate Vectors – Graphical Parameters – Adding Plots, lines, and Text – ggplot2.	6
Unit V	**Reading and Writing Files** R-Ready Data sets – Reading in External data files – Writing out Data files and Plots – Adhoc Object R/W.	6

Practical Component		
Exercises	1. Practice Installing , opening and saving files in R. 2. Create and store a vector that contains, in any configuration, the following: • A sequence of integers from 6 to 12 (inclusive). • A threefold repetition of the value 5.3 • Numbers divisible by 2. 3. Create a matrix and find the number of entries in each row which are greater than 4 4. Write a program to Add, Multiply two matrices 5. Write a program to transpose and find the inverse of a matrix. 6. Store a vector with 15 values as an object. Identify those equal to 6, those greater than or equal to 6, those less than 6 + 2, those not equal to 6 7. Identify the class of the following objects. For each object, also state whether the class is explicitly or implicitly defined. foo <- array(data=1:36,dim=c(3,3,4)) bar <- as.vector(foo) baz <- as.character(bar) qux <- as.factor(baz) quux <- bar+c(-0.1,0.1) 8. With the Weight (kg), height (cm) and Sex data of 10 students, create a plot of weight on the x-axis and height on the y-axis. Use different point characters or colors to distinguish between males and females and provide a matching legend. Label the axes and give the plot a title. **9.** Using R's built-in datasets library data frame quakes, do the following: Select only those records that correspond to a magnitude(mag) of greater than or equal to 5 and write them to a table-format file called q5.txt in an existing folder. Use a delimiting character of ! and do not include any row names. ii. Read the file back into your R workspace, naming the object q5.dframe. **10.** Demonstrate Visualization using ggplot2	30
Recommended Learning Resources		
Print Resources	**1.** Tilman M.Davies, "The Book of R: A First Course in Programming and Statistics", No Starch press, 2016. 2. Bradley C. Boehmke, " Data wrangling with R", Springer Cham, 2016. 3. Andrea de Vries, Joris Meys, "R programming for Dummies", 2nd edition, Wiley, 2016.	

SEMESTER II

Year	I	Course Code: DS2MJ02	Credits	4
Sem.	II	Course Title: Problem Solving & Programming Fundamentals	Hours	75
Course Prerequisites, if any	Nil			
Internal Assessment Marks: 25	End Semester Marks: 75	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	• Understand the basic concepts of programming languages, including syntax and semantics. • Apply programming constructs like loops, conditionals, and functions in practical scenarios. • Analyse code to identify and fix errors using debugging techniques. • Create modular programs using functions and procedures, emphasizing good programming practices.			
Unit No.	**Course Content**		**Hours**	
Theory Component				
Unit I	**Introduction to Computer Problem-Solving** The Problem-solving Aspect - Top-down Design - Implementation of Algorithms - Program Verification - The Efficiency of Algorithms - The Analysis of Algorithms		9	
Unit II	**Basic programming constructs** Basic Data types (Numerical, String) – Variables – Expressions – I/O statements – Compile and Run - Debugging.		9	
Unit III	**Decision Making – Branching & Looping** Decision making – Relational Operators - Conditional statement, Looping statement - Nested loops - Infinite loops - Switch statements.		9	
Unit IV	**Array Techniques** Array Manipulation - Different operations - one dimensional array - two-dimensional array - multi-dimensional array - Character Arrays and Strings.		9	
Unit V	**Modular solutions** Introduction to functions – Importance of design of functions – Arguments – Parameters – return values – local and global scope – Recursion.		9	
Practical Component				
Exercises	1. Program to array counting, array order reversal & find the maximum number in a set. 2. Program for removal of duplicates from an ordered array & to partition an array. 3. Program to find the k^{th} smallest element. 4. Program to exchange the values of two variables without using a third variable. 5. Program that takes a list of numbers as input and counts the total number of elements in the list. 6. Program to calculate the sum of a set of numbers entered by the user.		30	

	7. Program to compute the factorial of a given integer. 8. Program to compute the sine of an angle (in degrees) using a series expansion. 9. Program to generate the Fibonacci sequence up to a specified limit. 10. Program that takes an integer as input and reverses its digits. 11. Program that converts a number from one base to another (e.g., binary to decimal, decimal to binary).	
Recommended Learning Resources		
Print Resources	1. R. G. Dromey, "How to solve it by Computer", Pearson Education, 2007. 2. E. Balaguruswamy, "Programming In ANSI C", 4th edition, TMH Publications, 2007. 3. Yashwant Kanetkar, "Let Us C", 13th Edition, PHP, 2013. 4. Allen B. Downey, "Think Python: How to Think like a Computer Scientist", 2nd Edition, O'Reilly Publishers, 2016.	

Year	I	Course Code: DS2MI02	Credits	4
Sem.	II	Course Title : Foundations of Data Science - II	Hours	75
Course Prerequisites, if any	Foundations of Data Science I			
Internal Assessment Marks : 25	End Semester Marks : 75	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	**Students will be able to** 1. Develop skills in conveying insights from data through visual representation 2. Critically evaluate the advantages and disadvantages of real-time data science applications, demonstrating a deeper understanding and the ability to make informed judgments. 3. Practical application knowledge in processing and analyzing large-scale datasets. 4. Implement machine learning models, including Regression, Clustering, Collaborative Filtering, Association Rule Mining, Decision Trees, Naive Bayes, and Support Vector Machine 5. Apply text analytics techniques, including Information Retrieval, Natural Language Processing (NLP), and Text Mining, on textual data.			

Unit No.	Course Content	Hours
	Theory Component	
Unit I	**DATA VISUALIZATION** Introduction to Visualization, Introduction to Tableau, Dimensions, and measures, descriptive statistics, basic charts, Dashboard Design and principle, Integrate, Tableau with Google sheet.	7
Unit II	**REAL-TIME APPLICATIONS OF DATA SCIENCE** Applications of Data science – Implementation in different sectors, Advantages and Disadvantages, Exploration of Big data - Understanding and its use.	8
Unit III	**BIG DATA ANALYTICS** Terminologies – Introduction to NoSQL, Hadoop, MongoDB, JSON, Cassandra, MapReduce Programming, Hive, Pig.	10
Unit IV	**MACHINE LEARNING** Regression Model – Clustering – Collaborative Filtering – Association Rule Mining - Decision Trees, Naive Bayes, Support Vector Machine	10
Unit V	**DATA ANALYTICS ON TEXT** Major Text Mining Areas – Information Retrieval – Data Mining – Natural Language Processing NLP) – Text Analytics sub-tasks: Cleaning and Parsing, Searching, Retrieval, Text Mining, Part-of-Speech Tagging, Stemming.	10
	Practical Component	
Exercises	1. Perform visualization using Tableau. 2. Create a list of text using Tableau.	30

	3. Analyze a dataset using Tableau. 4. Install, Configure, and run Hadoop and HDFS 5. Write a program for word count/frequency using MapReduce/Python. 6. Demonstrate the use of MongoDB and Json 7. Install NLTK library and perform text processing and analysis 8. Write a program to process the text (stop words, Stemming, or Lemmatizing). 9. Build Plots on various charts using Python/Matlab/tableau 10. Practice the different ML algorithms	
Recommended Learning Resources		
Print Resources	1. Sanjeev Wagh, Manisha Bhende, Anuradha Thakare, 'Fundamentals of Data Science, CRC Press, 1st Edition, 2022. 2. Gilbert Strang, "Linear Algebra and Its Applications", New York: Academic Press, Fourth edition. 3. Seema Acharya, Subhasini Chellappan, Big Data Analytics, 2nd Edition, Wiley, 2019. 4. Suresh Kumar Mukhiya, Usman Ahmad "Hands-On Exploratory Data Analysis with Python" 1st Edition 2020. 5. Jure Leskovek, Anand Rajaraman and Jefrey Ullman., Mining of Massive Datasets. v2.1, Cambridge University Press, 2019. 6. Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", Cambridge University Press, 1st Edition, 2020. 7. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media, 1st Edition, 2015.	

SKILL ENHANCEMENT COURSES

Year	I	**Course Code: DS2SE02E1**	**Credits**	3
Sem.	II	**Course Title : Exploratory Data Analysis with Python**	**Hours**	60
Course Prerequisites, if any	Python Programming			
Internal Assessment Marks : 50	End Semester Marks : 50	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	**Students will be able to** • Perform data loading, transformation, and preliminary analysis for real-world data • Create charts and graphs to effectively communicate and interpret patterns in data during Exploratory Data Analysis. • Apply advanced statistical measures to describe and interpret datasets, including measures of central tendency and dispersion • Critically evaluate and draw meaningful conclusions from the analysis results. • Demonstrate proficiency in handling time series datasets and performing Time Series Analysis (TSA) using Python.			
Unit No.	**Course Content**		**Hours**	
Theory Component				
Unit I	**Introduction** Understanding Data Science – Significance of EDA – Making sense of Data – Comparing EDA with classical and Bayesian analysis – software tools.		6	
Unit II	**Visual aids for EDA** Line – Bar charts – Scatter Plot – Area Plot – Pie – Table – Polar chart – Histogram – Lollipop **EDA with Personal Email** Technical requirements –Loading – Transformation -Data Analysis **Data Transformation** Managing Database – Techniques – Benefits		6	
Unit III	**Descriptive Statistics** Understanding statistics – Measures of central tendency – Measures of dispersion. **Correlation** Understanding correlation – Types of analysis – Multivariate analysis using Titanic dataset.		6	

Unit IV	**Grouping Datasets** Understanding groupby() – Groupby mechanics – Data aggregation – Pivot tables – Cross-tabulations. **Time series Analysis** Understanding Time series dataset – TSA with Open Power System Data.	6
Unit V	**Model Development and Evaluation** Hypothesis Testing and Regression, Model Development and Evaluation , EDA on Wine Quality Data Analysis.	6
Practical Component		
Exercises	1. Download, Install and practice opensource tools for EDA - WEKA 2. Visualize the data using various graphs 3. Perform histogram analysis using NumPy, Matplotlib, pandas. 4. Write a program to generate different charts and plots. 5. Write a program to generate pivot using groupby() method. 6. Perform Time Series analysis and test with with a predictive model 7. Write a program to identify the correlation of the features/parameters in the Titanic Dataset. 8. Perform EDA on Wine Data 9. Demonstrate different visualizations based on Exercise 7. 10. Develop and evaluate ML models on open datasets	30
Recommended Learning Resources		
Print Resources	1. Hands-On Exploratory Data Analysis with Python, Suresh Kumar Mukhiya, Usman Ahmed, 2020, PACKT Publishing 2. Exploratory Data Analysis: Uncovering Insights from Your Data, Daniel Garfield, 2023, Kindle Edition	

Year	I	Course Code: DS2SE02E2	Credits	3
Sem.	II	Course Title : Data Wrangling with R	Hours	60
Course Prerequisites, if any	Foundations of Data Science, R programming			
Internal Assessment Marks : 50	End Semester Marks : 50	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	**Students will be able to** 1. Demonstrate the ability to write and execute R code efficiently, define variables, and leverage built-in functions for data manipulation. 2. Apply data wrangling skills to various datasets, understanding the data generation process, interpreting different data types, and effectively using data to address analytical queries. 3. Utilize core functions of dplyr for efficient data manipulation, sequential operations, grouping, and joining of data frames 4. Access and integrate data from databases and web APIs using R, including making RESTful requests and processing JSON data. 5. Design and create interactive visualizations, applying principles of effective visualization with ggplot2 and additional packages like Plotly, Rbokeh, and Leaflet.			

Unit No.	Course Content	Hours
	Theory Component	
Unit I	**FUNCTIONS IN R** Programming with R- Running R Code - Comments - Defining Variables, Functions -Built-in R Functions - Loading Functions - Writing Functions - Using Conditional Statements.	6
Unit II	**DATA WRANGLING** Understanding Data - The Data Generation Process - Finding Data - Types of Data - Interpreting Data - Using Data to Answer Questions - Data Frames - Working with Data Frames -Working with CSV Data Analytics for Data Science – Examples of Data Analytics – Data Analytics Lifecycle: Data Discovery, Data Preparation, Model Planning, Model Building, Communicate Results.	6
Unit III	**MANIPULATING DATA WITH DPLYR AND TIDYR** Data Manipulation - Core dplyr Functions- Performing Sequential Operations -Analyzing Data Frames by Group - Joining Data Frames Together - dplyr in Action: Analyzing Flight Data- Reshaping Data with tidyr -From Columns to Rows: gather() - From Rows to Columns: spread() - tidyr in Action: Exploring Educational Statistics.	6
Unit IV	**ACCESSING DATABASES AND WEB APIs** Accessing a Database from R - Accessing Web APIs -RESTful Requests -Accessing Web APIs from R -Processing JSON Data -APIs in Action: Finding Cuban Food in Seattle.	6
Unit V	**INTERACTIVE DATA VISUALIZATION** Designing Data Visualizations - The Purpose of Visualization - Selecting Visual Layouts - Choosing Effective Graphical Encodings -	6

	Expressive Data Displays - Enhancing Aesthetics - Creating Visualizations with ggplot2- A Grammar of Graphics - Basic Plotting with ggplot2 - Complex Layouts and Customization - Building Maps-ggplot2 in Action: A case study. Packages: The Plotly Package - The Rbokeh Package - The Leaflet Package - Interactive Visualization in Action: Exploring Changes to the City of Seattle.	
Practical Component		
Exercises	1. Write functions to find the sum and difference of the arguments passed 2. Write a function that takes a matrix and returns a matrix that is the same as the function argument, but every odd number is doubled 3. Perform Exploratory data analysis and perform data cleaning 4. Perform operations using Dataframes 5. Practice Data transformation 6. Practice Data frame manipulation 7. Import data into R from different file formats and perform scraping-Perform web scraping 8. Download, install, practice Plotly, Rbokeh, Leaflet Package 9. Perform interactive visualization Using Seattle dataset 10. Demonstrate visualization using Plotly, Rbokeh, and Leaflet Packages	30
Recommended Learning Resources		
Print Resources	1. Michael Freeman and Joel Ross, Programming Skills for Data Science: Start Writing code to Wrangle, Analyze, and Visualize Data with R, Addison-Wesley, 2018. 2. Benjamin S. Baumer, Daniel T. Kaplan and Nicholas J. Horton, Modern Data Science with R, Chapman and Hall/CRC, 2021. 3. John Mount and Nina Zumel, Practical Data Science with R, 2nd edition, Wiley, 2019	

VALUE ADDED COURSES

Year	I	Course Code: DS2VA04	Credits	2
Sem.	II	Course Title: Digital Technologies	Hours	45
Course Prerequisites, if any	Nil			
Internal Assessment Marks: 25	End Semester Marks: 75	Duration of ESA (Theory) : 03 hrs. Duration of ESA (Practical) : 03 hrs.		
Course Outcomes	1. Knowledge about digital paradigm; 2. Realization of importance of digital technology, digital financial tools, e-commerce; 3. Know-how of communication and networks; 4. Familiarity with the e-governance and Digital India initiatives; 5. An understanding of use & applications of digital technology; 6. Basic knowledge of machine learning and big data.			
Unit No.	**Course Content**		**Hours**	
Theory Component				
Unit I	Introduction & Evolution of Digital Systems. Role & Significance of Digital Technology. Information & Communication Technology & Tools. Computer System & its working, Software and its types. Operating Systems: Types and Functions. Problem Solving: Algorithms and Flowcharts.		7	
Unit II	Communication Systems: Principles, Model & Transmission Media. Computer Networks & Internet: Concepts & Applications, WWW, Web Browsers, Search Engines, Messaging, Email, Social Networking. Computer Based Information System: Significance & Types. E-commerce & Digital Marketing: Basic Concepts, Benefits & Challenges.		7	
Unit III	Digital India & e-Governance: Initiatives, Infrastructure, Services and Empowerment. Digital Financial Tools: Unified Payment Interface, Aadhar Enabled Payment System, USSD, Credit / Debit Cards, e-Wallets, Internet Banking, NEFT/RTGS and IMPS, Online Bill Payments and PoS. Cyber Security: Threats, Significance, Challenges, Precautions, Safety Measures, & Tools, legal and ethical perspectives.		7	
Unit IV	Emerging Technologies & their applications: Overview of Cloud Computing, Big Data, Internet of Things, Virtual Reality,		7	
Unit V	Emerging Technologies & their applications: Blockchain & Cryptocurrency, Robotics, Machine Learning & Artificial Intelligence, 3-D Printing. Digital Signatures.		7	
Practical Component				
Practice	1. Operating System Installation and configuration 2. Application Software Installation and configuration 3. Hardware understanding and minor troubleshooting		10	

	4. Networking, cabling, configuration	
Recommended Learning Resources		
Print Resources	1. Pramod Kumar, Anuradha Tomar, R. Sharmila, Emerging Technologies in Computing - Theory, Practice, and Advances, Chapman and Hall / CRC, 1st Edition, 2021, eBook ISBN: 9781003121466, https://doi.org/10.1201/9781003121466. 2. V Rajaraman, Introduction to Information Technology, PHI, 3rd Edition, 2018, ISBN-10: 9387472299, ISBN-13: 978-9387472297. 3. E Balagurusamy, Fundamentals of Computers, Tata Mc GrawHill, 2nd Edition, 2011, ISBN: 9780071077880. 4. Behrouz A. Forouzan, Data Communications and Networking, McGraw Hill, 4th Edition, 2007, ISBN 978-0-07-296775-3. 5. Rajkumar Buvya, James Broberg, and Andrzej Gosciniski, Cloud Computing-Principals and Paradigms, Wiley, 2011, ISBN: 978-0-470-88799-8. 6. Stuart Russel and Peter Norvig, Artificial Intelligence - A Modern Approach, Pearson Education, 3rd Edition, 2010, ISBN- 13: 978-0-13 -604259-4. 7. Samuel Greengard, Internet of Things, The MIT Press, 2015, ISBN electronic: 9780262328937, https://doi.org/10.7551/mitpress/10277.001.0001. 8. C.S.V. Murthy, E- Commerce (Concept - Models - Strategies), Himalaya Publishing House, 2015, ISBN: 8178662760. 9. Hurwith, Nugent Halper, Kaufman, Big Data for Dummies, Wiley & Sons - Wiley, 1st Edition, 2013, ISBN-13: 978-1118504222.	