



M.S. Data Science

2 Years (4 Semesters)

Dept. Of Computer Science

SVU College of CM & CS, SRI VENKATESWARA UNIVERSITY: TIRUPATI

An Industry Collaborative Programme

WITH EXAFLUENCE INC., USA

SYLLABUS & REGULATIONS

For Approval W.E.F AY 2025 – 2026

Contents

SNO	CONTENT	PAGE NO.
1	Preamble	2
2	Duration of the Program	3
3	Minimum Qualifications for Admission	3
4	Admission Procedure	3
5	Structure of the Programme	4
6	The Credit System	5
7	Minimum Instruction Days	6
8	Medium of Instruction and Examination	6
9	Credits required for Award of M.S. Data Science Degree	6
10	Evaluation of Academic Performance	7
11	University Examinations	8
12	Attendance Requirements	9
13	Academic Requirements	9
14	Condition of Promotion	11
15	Award of Degree	11
16	Award Grades and Grade points	11
17	Award of Rank	13
18	Transitory Regulations	13
19	Institutional Framework and Methodology	14
20	Fee & Remuneration	15
21	Remuneration / Honorarium and Allowances	16
22	Amendments of Regulation	16
23	Careers	16
24	Internship & Industrial Training	16
25	Important Instructions	17
26	Website Information	18
27	Visiting Faculty	18
28	Industry Connectivity	18
29	Communication	18
30	University's Vision and Mission	19
31	Department's Vision and Mission	19
32	Program Education Objective (PEO)	27
33	Program Specific outcomes (PSO)	27
34	Program Outcome (PO)	28
35	Unit Wise Details	29
36	Model Question Paper	86

Department of Computer Science
SVU COLLEGE OF COMMERCE MANAGEMENT & COMPUTER SCIENCE
SRI VENKATESWARA UNIVERSITY, TIRUPATI

M.S. DATA SCIENCE

An Industry Collaborative Programme in Hybrid Mode (online and offline)

EXAFLUENCE INC having its register office at 3759, US Route 1South, Suite 203, New Jersey, South Brunswick, United States of America- 08852 with overseas development centres in Bangalore, Chennai, Dubai, Canada and UK and ODC at #201, Ecco Greens- DTP, Upadhyaya Nagar, Tirupati 517507 Bangalore, Chennai in India and in Dubai entered a memorandum of understanding on 12th day of August 2022, with **SRI VENKATESWARA UNIVERSITY, TIRUPATI**, Andhra Pradesh, one of the oldest and renewed state-owned University in Andhra Pradesh.

Sri Venkateswara University was established in 1954, in Tirupati to cater the educational needs and aspirations of people of the Rayalaseema Region of Andhra Pradesh. After completing 68 years of excellence in teaching, research, extension and outreach activities, the University is committed to cater the needs of higher education offering a full range of post-graduate programs in Arts, Sciences, Law, Management, Education, Physical Education, Engineering and Pharmacy disciplines. From a humble beginning of one College with six departments, the University has now grown into the second largest University in Andhra Pradesh having five constituent Colleges Viz. College of Arts, College of Sciences, College of Commerce, Management & Computer Science, College of Pharmacy, and College of Engineering accommodating 52 departments offering 88 programs.

Exafluence head quartered in New Jersey USA with overseas development centres in Bangalore, Chennai, Tirupati in India, Dubai, Canada and UK, Exafluence has presence across global the intent is to enroll students in worldwide and bring in international students to SV University. Exafluence is promoted by alumni of SV University Mr. RaviKiran Dharmavaram and COO TN Sreenivas, and it's associated with MongoDB, AWS, Microsoft, Google etc., and with rich experience in data science, the students can get benefited with latest technologies and work in Real Time environment and connect with technical leaders spread across global.

1. PREAMBLE

Data plays a crucial role in driving innovation and strategic decisions across industries such as social media, healthcare, and finance, leading to a growing need for skilled professionals who can extract insights and facilitate data-driven decision-making.

The MS Data Science program delivers a rigorous, industry-aligned curriculum shaped by global experts and stakeholders. It builds strong foundations in data mining, AI, machine learning, analytics, and visualization, preparing students for real-world challenges. Developed in collaboration with Exafluence Education, Inc., USA, the program ensures academic excellence and global relevance in today's data-driven landscape.

The curriculum integrates foundational and advanced data science topics through a mix of core and elective courses. Core subjects build proficiency in essential areas such as programming (e.g., Python), data visualization, statistical analysis, natural language processing, AI/ML, time series analysis, soft computing, and more. Elective courses, offered in the second and third semesters, allow students to specialize in areas of their interest, providing a personalized learning experience.

Aligned with the vision of the **National Education Policy (NEP) 2020**, the program embraces a multidisciplinary approach, offering flexibility through electives, continuous evaluation, and project-based learning. The outcome-based education model ensures that students graduate with practical skills, analytical thinking, and the capacity to innovate in a data-centric environment.

Students can take open elective courses on Massive Open Online Courses (MOOC) platforms such as SWAYAM, NPTEL, etc., and the credit will be transferred to the course bucket as equivalent to classroom-based courses.

Practical application is at the heart of the program. Students engage in real-world projects and internships that enable them to apply theoretical knowledge to solve complex data challenges and deliver actionable insights to organizations. The inclusion of a dedicated Research Methodology course fosters a strong research mindset, empowering students to contribute meaningfully to academic and industry advancements in data science.

Graduates of this program would be competent to shoulder different job roles, such as data scientists, data analysts, machine learning engineers, data engineers, and data consultants. Opportunities span across sectors, including finance, healthcare, e-commerce, marketing, technology, and research. The program also serves as a strong foundation for those interested in pursuing doctoral studies.

2. Duration Of the Program:

- The professional postgraduate programme leading to the degree of Master of Science in Data Science will extend 2 academic years/over a period of 4 semesters.
- Each academic year being divided into two consecutive (one odd + one even) semesters
- Each semester will normally have duration of 90 working days.
- Minimum duration of the full-time M.S. Data Science Professional Programme is two consecutive academic years i.e. four semesters and maximum period is four academic years, i.e. eight semesters

3. Minimum Qualifications for Admission:

Candidates for admission into the first semester of the M.S. DATA SCIENCE Professional programme must have a

- B.Sc. with Mathematics as main subject along with any two of the following subjects: Physics, Chemistry, Statistics, Electronics, Computer Science
- B.Tech. CSE/ IT/ Electronics
- BCA

of academic input of any university recognized by the Universities Grants Commission.

4. Admission Procedure

- 4.1** The first academic year admission into M.S. Data Science is through an online entrance examination which consisted of 50 multiple choices objective type questions for 50 Marks. There is no negative marking for the wrong answers.
- 4.2** The selection process for the qualified students in the entrance examination
- 4.3** The short-listed candidates are called for counselling by Directorate of Admissions (DOA), SV University, for certificate verification, and first year course fee payment (through online portal <https://pay.svupayonline.edu.in/>.)
- 4.4** Admission of Foreign / NRI Students for M.S. DATA SCIENCE Professional Programme:
- Students should have earned their qualifying degree from a University / Institute recognized by the Association of Indian Universities / similar Indian body.

- Students nominated by Foreign Governments may be accepted to M.S. DATA SCIENCE Professional programme, without any further test / interview, if their request for admission is routed through MHRD / any agency of the Government of India.

- 4.5** The candidate at the time of admission should produce a Transfer certificate (T.C.), Degree certificate, Marks statement or equivalent credit statements, Migration certificate and any other certificate as stipulated by the University.
- 4.6** Candidates should abide by the conditions, rules and regulations stipulated by the University / Exafluence Inc. and Laws of the University applicable from time to time.
- 4.7** For further information the candidate may contact the ACADEMIC INCHARGE - Exafluence Education Campus Centre, Next to Distance Education Centre Bldg.SV University, Tirupati – 517502.email ID: admissions@svudatascience.in

Note: This MS programme is a self- Supporting course and is **not** applicable for benefits under **government scholarships**.

Note: Admission will be given as per rules in vogue from time to time.

5. Structure of the Programme:

The M.S. DATA SCIENCE Professional programme has a curriculum, with syllabus consisting of:

- **Core courses**, which give a broad base in the main field of study in the academic programme concerned.
- **Elective courses** chosen by the student in consultation with the faculty adviser
Elective course is a course which can be chosen from a pool of papers. It may be:
 - Supportive to the discipline of study
 - Providing an expanded scope
 - Enabling an exposure to some other discipline/domain
 - Nurturing student's proficiency/skill.

An elective may be "Generic Elective" focusing on those courses which add generic proficiency to the students. These electives shall be "Discipline centric".

- Second type elective may be open elective and shall be offered for other disciplines.

- **The Foundation Courses** may be of two kinds: Compulsory Foundation and Elective foundation. “Compulsory Foundation” courses are the courses based upon the content that leads to Knowledge enhancement. They are mandatory for all disciplines. Elective Foundation courses are value-based and are aimed at man-making education.
- **Skill enhancement Course** and Multidisciplinary Course/Project are introduced as per Action Plan of National Education Policy-NEP-2020.
 - Laboratory, project work and software design courses where special emphasis is laid on the application of knowledge to real-life problems.
 - The program will also include involve lectures /tutorials/ lab work, Technical Seminars, Group discussions and practical Industry exposure training, as prescribed by the Board of Studies in the curriculum & syllabus and approved by the Academic Council.
- **Audit Course**
 - The main aim of Audit Course is too aware the Contemporary Societal issues. This course is under self-study.
 - In addition, the students may have to choose audit courses, in consultation with the faculty adviser.
- **MOOCS and e-Learning:**

Discipline centric elective course through MOOCs (Massive Open Online Course) platform. Students can register for the course offered by authorized Institutions/Agencies online with the approval of the course coordinator. The certificate issued by the Institutions/Agencies after successful completion of the course will be considered for the award of the Grade to that course in open electives only.

6. The Credit System:

- Each course has a certain number of credits assigned to it, depending upon whether it is a lecture or tutorial, or practical work and the number of periods assigned per week. The practical courses shall include laboratory work and project work.

- **The credits are assigned according to the following pattern:**

- 1 credit for each lecture period/week,
- 1 credit for each tutorial period/week,
- 1 credit for each practical session of 2 periods/week &
- 2 credits for each project session of 3 periods / week.

7. Minimum Instruction Days:

- The normal duration of M.S. DATA SCIENCE Course is 4 semesters.
- Semesters I, II, III, IV shall consist of a minimum of 90 instruction days (based on six instruction days per week) excluding the days allotted for internal assessment, semester end examinations and preparation holidays.
- 4th semester shall consist of a minimum of 15 weeks for undertaking major project work either in the College or any of the recognized Public Service Sectors/ Government Sectors, National Laboratories and Industrial Houses or any other organizations approved by the Department course coordinator & Exafluence Inc., USA.

8. Medium of Instruction and Examination:

The medium of instruction will be English, and the examination will be in English only.

9. Credits Required for Award of M.S. DATA SCIENCE Degree:

A student earns credits by passing courses every semester. A student, who has registered the M.S. DATA SCIENCE Professional programme, must acquire 90 credits to become eligible for the award of the degree.

- It is mandatory for a student to complete successfully all the core courses pertaining to his/her specialization of study.
- A student may choose Generic Electives from the list of elective courses offered from his/her specialization of study.
- Further, a student may select from a list of Elective courses from other Departments as Open Electives to “suit the required” number of credits, such that the total credits is at least 90.
- There should be a register maintained by the course coordinator & Head of the Department indicating for each student, the course (s) registered by the student within the department, so that “Generic Electives” opted by the student are

indicated.

- In the case of Open Elective, the course coordinator & Head of the Department should prepare a statement/register indicating the courses chosen/ opted by the students of the department in other departments.
- The Course coordinator (or) Head of the Department should send the list of registered papers (opted by the student) to the principal with a copy to the controller of examinations immediately within a week commencement of each semester.
- A copy of the courses registered by the students in each semester approved by Dept. Course coordinator shall be sent to the Academic Branch as well as Examination Branch.
- The list of students registered for MOOCs shall be furnished by dept. course coordinator giving details of the programme with a copy to the Principal and Controller of Examinations.
- A model of Registers to be maintained by the course coordinator & Head of the Department. It is mandatory on the part of the course coordinator & Head of the Department to maintain Register for each Course separately.

10.Evaluation of Academic Performance:

The performance of the students in each semester shall be evaluated paper wise.

The distribution of marks between sessional work (internal assessment) and University Examination is as follows:

Paper Category	Sessional Marks	University Examination Marks
Theory	30	70
Practical	30 Designing of the Program (10) + Program Execution (10) + Viva-Voce (10)	70 Record (20) + Designing of the Program (20) + Program Execution (20) + Viva Voce (10)
Minor Project Work	30 Presentation (15) + Viva	70 Documentation (10) + Presentation

	Voce (15)	(20) + Execution (20) + Viva Voce (20)
Major Project Work	100 Project Documentation (50) + Presentation (30) + Viva Voce (30)	200 National / International Conference Publication Proceedings (External): 50 Marks (Paper based on project should be submitted to conference and published in the form of proceedings) or National / International Journal Publication (External): 50 Marks (Paper based on project should be submitted to the journal and should be published) 1. Project Documentation: 50 Marks 2. Presentation: 50 Marks 3. Viva Voce: 50 Marks

11. University Examinations:

- 11.1 For each theory subject, there shall be a comprehensive University Examination of three hours' duration. Setting of Question papers shall be done by external examiners from the panels recommended by the Board of Studies.
- 11.2 The semester end university examination in each course and paper is of three hours duration with a maximum of 70 marks. The Question paper consists of two parts; A & B. Part A consists of 10 short answer questions and 5 to be answered (5 X 4 = 20 Marks). Part B consists of 10 questions with choice from each unit (5 X 10 = 50 Marks).
- 11.3 (a) For each practical subject, the University Examination shall be conducted by one Internal examiner and one external examiner/two examiners.
(b) Viva -Voce examination in practical subject shall be conducted by Examiners (One Internal and One External Examiner/Two Examiners)
- 11.4 Viva-Voce Examination in Minor Project Work shall be conducted by One Internal Examiner and One External Examiner / Two Examiners
- 11.5 Viva-voce examination in Major Project Work shall be conducted by a committee consisting of Two External Examiners and One Internal Examiner / Three Examiners

- 11.6 (a) The Examiners for Practical examinations and Viva-voce examinations (Stated in 11.3) shall be appointed from among the panels submitted by the Chairperson/Chairman of the Board of Studies.
- (b) The Examiners for Project report evaluation for Minor/Major Project Works and Viva Voce examination for Minor/Major Project Works (Stated in 11.4, 11.5) shall be appointed from among the panels submitted by the Chairperson/Chairman of the Board of Studies.
- 11.7 All students enrolled in MOOCS and seeking academic credit or certificate must pay the prescribed examination fee as notified by the prescribed MOOC's Platform
- 11.8 Regulation on Supplementary Examination for MOOCs

Students who fail or are absent in a MOOC (Massive Open Online Course) approved for credit transfer by the University shall be permitted to appear for the University-conducted supplementary examination in the respective course, subject to fulfillment of eligibility criteria and payment of the prescribed examination fee. The supplementary examination shall be conducted in accordance with the University's examination schedule and evaluation procedures.

12. Attendance Requirements:

- 12.1 Regular course of study, in a semester, means a minimum average attendance of 75% in all the papers computed by totalling the number of periods of lectures, tutorials, and practical's and project as the case may be, held in every paper as the denominator and the total number of periods attended by the student in all papers, as the numerator.
- 12.2 However, a student has to put in a minimum attendance of 50% in each subject, in addition to the condition laid down in clause 12.1.
- 12.3 No consideration whatsoever in attendance will be shown to any student for late admission due to any reason.
- 12.4 Condonation of shortage of attendance may be recommended on medical grounds only provided a student puts in at least 62.5% attendance, in all the papers put together as calculated in clause 12.1 above, along with a minimum of 50% attendance in each subject as stated in clause 12.2.
- 12.5 A student, who could not satisfy these requirements of attendance as given in clauses above, in any semester, shall have to repeat that semester in the subsequent academic years with the written permission of the university.
- 12.6 A student shall not be permitted to study any semester for more than 2 times during the course of study.

- 12.7** A certificate of satisfactory attendance must be submitted by the student from the organization where he/she undertakes Project Work. For the above purpose, the candidate is not expected to avail more than 12 working days of leave of absence.
- 12.8** Further, a student is required to complete the course of study satisfying the attendance requirements in all the 4 semesters, within a period of first 8 semesters from the time of admission, failing which he/she shall forfeit his /her seat.
- 12.9** A student, who has not satisfied the minimum attendance requirements in any semester, may repeat that semester after obtaining written permission from the university, canceling the previous record of attendance and sessional marks of that semester. However, this facility may be availed by any student not more than twice during the entire course of study and the entire course of study shall be within the first eight semesters as stipulated in clause 12.8.

13. Academic Requirements:

- 13.1** A candidate shall be declared to have passed whole examination of a semester if he/she secures a minimum aggregate of 50% along with minimum marks of 40% in the University Examination in each theory and practical paper, including project work. Aggregate for this purpose shall mean the total marks obtained in the University Examination and sessional put together in all the papers of that semester.
- 13.2** A candidate who secures a minimum aggregate of 50% in any semester as specified in clause 13.1 but fails to secure the paper minimum of 40% in the University Examination in any paper, shall be declared to have failed in that paper. The candidate may appear for the University Examination in such papers as and when conducted and pass by securing the subject minimum of 40% in each of such papers.
- 13.3** A candidate, who could not secure the minimum aggregate of 50% in any semester, as given in clause 13.1, shall be declared to have exempted in such theory papers in which he/she secured a minimum of 50% of the marks in the University Examination.
- 13.4** A candidate who could not secure the minimum aggregate of 50% in any semester, as given in clause 13.1, shall be declared to have exempted in such practical papers in which he/she secured a minimum of 40% in the University Examination and a minimum of 50% of the marks in the University Examination and sessional put together in each of such papers.

13.5 A candidate must appear for all the other papers in which he/she has not satisfied the stipulations of clauses 13.3 and 13.4 and pass compartmentally by satisfying the stipulations stated in clauses 13.3 and 13.4. Rechecking of the aggregate shall not arise when a candidate opts for compartmental passing.

13.6 Candidates who would not like to avail themselves of the facility of compartmental pass given in clauses 11.4, 11.5 and 11.6 may reappear for the whole examination of that semester, after cancelling the previous result. In such a case, clauses 13.1 to 13.5 shall apply in that order.

13.7 Candidates shall register for improvement in one or more theory papers of a semester indicating the same in the application form. However, improvement facility is not extended to practical and project work papers.

NOTE: The question of checking the aggregate shall arise only when the candidate registers for the whole examinations for the first time or when he/she registers for the whole examinations at a subsequent attempt after having cancelled the previous result.

14. Condition of Promotion:

A candidate shall be promoted to the next semester, if he/she satisfies the minimum attendance requirements of that semester of M.S. DATA SCIENCE as stipulated in clause 13.

15. Award of Degree:

The degree of M.S. DATA SCIENCE shall be conferred on a candidate who has satisfied the following conditions:

- 15.1** The candidate, after admission to the M.S. DATA SCIENCE Professional programme of the University, must have studied each of the 4 semesters at least once
- 15.2** The candidate must have satisfied minimum attendance requirements as stipulated in clause 12, and minimum academic requirements as prescribed in clause 13, in each of the 4 semesters of the M.S. DATA SCIENCE Professional programme.
- 15.3** The Programme of Study for the M.S. DATA SCIENCE shall cover normally a period of 2 academic years, comprising of 4 semesters and No student shall be permitted to complete the course of study of M.S. DATA SCIENCE earlier than 4 semesters or to take not more than 8 consecutive semesters, failing which he / she shall forfeits his / her seat (or) He/she will not be getting degree .

- 15.4** A student who has earned a minimum of 90 credits by passing in all the core courses and the minimum number of electives prescribed shall be declared to have passed the course work and shall be eligible for the award of degree.
- 15.5** A student who has earned extra credit shall be issued a separate certificate to that effect, mentioning the subject and grade.

16. Award Grades and Grade Points:

- 16.1** After a candidate has satisfied all requirements for the award of the degree as specified in clause 15, he/she shall be placed in one of the following three classifications, irrespective of whether the candidate passed compartmentally or otherwise, even after the regular period of study of four semesters.

Grade Point: It is a numerical weight allotted to each letter grade on a 10-point scale.

Letter Grade: It is an index of the performance of students in a said course. Grades are denoted by letters O, A+, A, B+, B, C, P and F.

Semester Grade Point Average (SGPA): It is a measure of performance of work done in a semester. It is ratio of total credit points secured by a student in various courses registered in a semester and the total course credits taken during that semester. It shall be expressed up to two decimal places

$$\text{SGPA (Si)} = \sum(C_i \times G_i) / \sum C_i$$

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.

The **CGPA** is also calculated in the same manner taking into account all the courses undergone by a student over all the semesters of a programme, i.e.

Cumulative Grade Point Average (CGPA): It is a measure of overall cumulative performance of a student over all semesters. The CGPA is the ratio of total credit points secured by a student in various courses in all semesters and the sum of the total credits of all courses in all the semesters. It is expressed up to two decimal places

$$\text{CGPA} = \sum(C_i \times S_i) / \sum C_i$$

Where S_i is the SGPA of the i th semester and C_i is the total number of credits in that semester. The SGPA and CGPA shall be rounded off to 2 decimal points and reported in the transcripts

Letter Grades and Grade Points

A 10-point grading system with the following letter grades is to be followed:

S.no	Grade	Grade Points
1.	O (Outstanding)	10
2.	A ⁺ (Excellent)	9
3.	A (Very Good)	8
4.	B ⁺ (Good)	7
5.	B (Above Average)	6
6.	C (Average)	5
7.	P (Pass)	4
8.	F (Fail)	0
9.	Ab. (Absent)	0

A student obtaining Grade F shall be considered failed and will be required to reappear in the examination.

For non-credit courses 'Satisfactory' or "Unsatisfactory' shall be indicated instead of the letter grade and this will not be counted for the computation of SGPA/CGPA

For this purpose, aggregate shall mean the aggregate of the marks in the University Examinations and sessional put together at all 4 semester of M.S. DATA SCIENCE.

- 16.2** A Candidate, before and even after becoming eligible for the award of the Degree, may reappear for the University Examination, as and when conducted, in any of the theory subjects, which he/she has already passed, for the purpose of improving the aggregate. However, this facility cannot be availed by a candidate who has taken the Provisional Pass Certificate. Candidates shall not be permitted to appear for University examination in Practical papers including project work for the purpose of improvement.

17.Award of Rank:

The rank shall be awarded based on the following:

- 17.1** Ranks shall be awarded for the top ten students appearing for the regular M.S. DATA SCIENCE examination.
- 17.2** Only such candidates, who satisfy the minimum academic requirements as stipulated in clause 13, by the end of their 4th semester, and become eligible for the award of the degree, are eligible for the award of rank. Candidates, who lose one or more semesters for any reason whatever, are not eligible for the award of rank.
- 17.3** For the purpose of awarding rank, aggregate of marks in University Examination and Sessional put together at all the semesters of M.S. DATA SCIENCE, secured at the first attempt only, shall be taken into account.
- 17.4** Award of Prizes, scholarships and other honors shall be according to the rank secured by a candidate, consistent with the desire of the Donor.

18.Transitory Regulations:

- 18.1** Candidates who studied the M.S. DATA SCIENCE Course under the Regulations (Semester Pattern) and syllabus from 2023-2024 but who could not satisfy the minimum attendance requirements in any semester may join the appropriate semester in the Regulations (For the full time, regular course) – 2025-27 for the remaining part of the course and be governed by the Regulations of that batch from then on.
- 18.2** University Examinations shall be conducted 2 more times, after the last regular examination, in all those papers prescribed under the Regulations & syllabus from 2024-2025
- 18.3** Candidates, who satisfied the minimum attendance requirements in any semester under the Regulations & syllabus from 2024-2025 but who are yet to pass some papers of that semester even after Two chances shall appear for the equivalent papers under Regulations from time to time, specified by the Chairman / Chairperson of the Board of Studies.

19. Institutional Framework and Methodology:

- 19.1** Exafluence Inc. designed new and innovative, industry / job-oriented, international standard course curriculum, and the methodology of training along with the practical and procedural knowledge sharing practices to be conducted in hybrid mode as per the MOU with the S V University in the department of Computer Science SVU College of CM & CS Tirupati with appropriate approvals from APSCHE, Sri Venkateswara University & Department of Computer Science.
- 19.2** Exafluence Inc. will identify Industry partners, mentors, specialist trainers, certified trainers, experienced faculty, Guest faculties from the industry (Data Sciences | Machine Learning | Artificial Intelligence | Cloud Technologies Etc.) to execute the programme.
- 19.3** Exafluence Inc. has developed educational content, curated course and curriculum to deliver, execution methodology, practical learning, Lab's setup, through its efforts for executing and conducting the programmes / courses in the SVU campus in Hybrid mode internationally in association with partner international organizations.
- 19.4** Exafluence Inc. will coordinate with SVU in enrolment of students, collection of fees, submission of student rolls, supplying the study material to the students enrolled into the programme (offline and online), assisting in industry training, models and practices of evaluation and examinations process support to SVU and all other incidental activities thereto.
- 19.4** Exafluence Inc. will coordinate with SVU and the partner national and international organizations for international student's enrollments for the Hybrid course delivery.
- 19.5** Exafluence Inc. will provide all software's and courseware to execute the programme. Exafluence Inc. along with its international and national technology partners support the labs with relevant software and technology components and lab assistants. However, most of the Data Science technology work with the open-source components.
- 19.6** Exafluence Inc. will advertise and market the programme extensively through all mediums to propagate and get the right admission through the entrance examination and admission management system.

- 19.7** Exafluence Inc. will provide the necessary industry internship, project support and job assistance to at least 50% of the students on pro rata basis.
- 19.8** University will provide the necessary infrastructure facilities for the intended course to be conducted.
- 19.9** Students will be admitted into the Industry Collaborative programme to pursue M.S. Data Science, or any other programmes offered in per view of this MOU.
- 19.10** The Programme will be closely monitored by the department “course Coordinator” who is a University Professor from the respective faculty/ department and by an Academic Board consisting of both academicians and industry professionals.
- 19.11** SVU shall conduct period-end examinations and evaluate the enrolled students in such programmes/courses under this MOU. SVU will award DEGREES to all successful students.

20. Fee & Remuneration

- 20.1** The fee payable in each year is collected in one instalment as detailed below:

Exafluence Inc. Programme	Tuition Fee Rs.	GST @ 18%	Total
II Years	3,00,000/-	54000/-	3,54,000/-
I year	150000/-	27000/-	1,77,000/-
II year	150000/-	27000/-	1,77,000/-

Note: The fee is subject to changes to be made by the University from time to time.

21. Remuneration / Honorarium and Allowances

- Honorarium for Principal 3000 * 10 months = Rs 30000/-
- Honorarium to Coordinator 3000 * 10 months = Rs 30000/-
- Honorarium for AAO 1000 * 10 months = Rs 10000/-
- Honorarium for Junior Assistant, Department of CS 850 * 10 months = Rs 8500/-

Note: Remuneration, TA & DA related to examinations and evaluation will be as per rules in vogue time to time

22. Amendments of Regulation:

The University may, from time to time, revise, amend, or change the Regulations, Scheme of Examinations and Syllabi, whenever necessary.

23. Careers

Data Scientists focus on helping their organization to use Data. They are experts in handling large quantum of data and are responsible for deriving business value. Data science has huge potential as career and with tremendous opportunities for advancement in the future. Already, demand is high, salaries are competitive, and the perks are numerous – which is why Data Scientist has been called “the most promising career”.

Opportunities for Data Scientists:

Data Architect & Administrators, Data Engineer, Data Analyst, Data Scientist, Data Specialists, Statisticians and Mathematicians, Critical Data Managers, Data Mining Engineer, Business Intelligence Analyst.

Opportunities for Data Engineers:

Database Developer, Data Architect, Data Manager, ETL Developer, Big Data Developer, Business Intelligence Developer, Analytics Consultant, Data Engineer.

24. Internship & Industrial Training:

- 24.1** The Industry will take students as “Interns” or “Trainees” for a maximum period of 4 months and a minimum period of 2 Months.
- 24.2** The Industry will select students who are ready for Industry Internship and then shortlist students for the purpose. The Programme Coordinator will make alternative arrangements for such students who are successful reaching the final semester but fail to meet the expectations of the industry.
- 24.3** The Industry may or may not pay any stipend to the students during the Internship period. If the industry is comfortable, it may plan for some incentives and other benefits when and where available.

- 24.4** The selected students will be under the direct control of the industry during the period of internship. A consultant/officer/executive will take care of the candidate and statutory requirements by a Flexi contract.
- 24.5** The Industry will render knowledge transfer on key practical areas to the students during the period of internship.
- 24.6** Each student will have a Project Guide in the Partner industry institution and or in the organization, where the student carries out his/her project work / internship.
- 24.7** The Industry will certify the project/work reports of the students which have to be submitted to the University as part of their course completion.
- 24.8** The “Interns” shall abide by all the terms and regulations of the Internship Programme as stipulated by the industry

25. Important Instructions:

- 25.1** No refund of the fee will be made once remitted. There are no fee concessions for any category of students. The fee once paid by the candidate will neither be refunded nor adjusted. Please check the eligibility before obtaining the admission.
- 25.2** All the fee paid receipts must be preserved till the candidate completes the course and obtains the degree certificate
- 25.3** All matters relating to examinations / results / marks / certificates etc will be dealt with only by the Controller of Examinations, Sri Venkateswara University, Tirupati.
- 25.4** For all legal matters the Jurisdiction is restricted to Tirupati city only.
- 25.5** Candidates are required to abide by the rules and regulations that are in force and those that will come into effect from time to time as formulated for M.S. Data Science by the Industry Collaborative Programme by Exafluence Inc and /or by the University.
- 25.6** Admission at any time may be cancelled by the coordinator/ Head/ Principal/ DOA - Industry Collaborative Programme if the candidates is found ineligible for admission into the said course or he/she is found to be guilty of indiscipline.

Website information:

www.svuniversity.edu.in & www.exfedu.in will give all the information about the M.S. Data Science Professional programme

Visiting Faculty:

Faculty visits from industry as per their specializations will be organized strictly. The Programme schedule of the visiting faculty from industry with personal details, Years of experience, teaching experience, Telephone, E-mail along with their photos should be submitted to the department course Coordinator, S.V. University at regular intervals

Industry connectivity:

Exafluence Inc. should appoint a “placement and internship officer/coordinator” who should work along with SV University for achieving 100% internships and placements only for M.S Data Science students.

Communication:

Any communication to the University by the Exafluence Inc. should be through the department course coordinator only.

M.S. Data Science (MSDS)

University's Vision and Mission:

VISION

To be a center of excellence in Education and Technology committed towards the Socio-Economic advancement of the country.

MISSION

- To impart advanced knowledge in all branches of teaching and learning.
- To transform young minds toward professional competence by inculcating values and developing skills.
- To promote research and ensure continuous value addition among students and employees.
- To strengthen association with industry, research organizations and alumni to enhance knowledge on current technologies.
- To promote next-generation technocracy and nurture entrepreneurial culture for social-economic growth.

Department's Vision and Mission:

VISION

To be a center of excellence in data science education, research, and innovation, empowering students to become globally competent professionals and ethical leaders who drive intelligent solutions through data-driven decision-making for a sustainable and inclusive future."

MISSION

- To provide a rigorous and interdisciplinary curriculum that blends computer science, statistics, and domain knowledge to solve real-world problems using data.
- To cultivate analytical thinking, technical proficiency, and ethical responsibility among students through experiential learning, research, and industry collaborations.
- To foster innovation and entrepreneurship in the fields of artificial intelligence, machine learning, and big data technologies.
- To promote lifelong learning, professional growth, and global competencies for careers in academia, industry, and public service

SEMESTER- I

Code	Title of Paper	Category of Course	As per NEP 2020	Teaching Hours per Week			No. of Credits	Sessional Marks	SEE	Exam Duration	Max Marks
			Course Level	L	T	P					
MSP 101	Mathematical Foundation for Data Science	DSC (core course)	500-599	3	1	0	3	25	50	3 Hrs.	75
MSP 102	Statistics for Data Science	DSC (core course)	500-599	3	1	0	3	25	50	3 Hrs.	75
MSP 103	Python for Data Science	DSC (core course)	500-599	4	0	0	4	30	70	3 Hrs.	100
MSP 104	Quantum Data Science	DSC (core course)	500-599	3	1	0	3	25	50	3 Hrs.	75
MSP 105	Fundamentals of Database Systems: SQL & MongoDB	DSC Skill Oriented Course	500-599	4	0	0	4	30	70	3 Hrs.	100
MSP 106	Data Visualization & Story Telling	DSC Skill Oriented Course	500-599	4	0	0	4	30	70	3 Hrs.	100
MSP 107	Hands-on: Python for Data Science	-	500-599	0	0	4	2	15	35	3 Hrs.	50
MSP 108	Hands-on: Data Visualization & Story Telling	-	500-599	0	0	4	2	15	35	3 Hrs.	50
OOTC 109	Open Online Transdisciplinary Course - Soft Skills	-	-	3	1	0	3	-	-	-	-
IKS 110	Indian Knowledge Systems 1	IKS	200-299	2	0	0	0	-	-	2 Hrs	-
Total							28				625

Open Online Transdisciplinary Course(OOTC)

OOTC 109	Soft Skills - https://nlinecourses.nptel.ac.in/noc25_hs159/preview
-----------------	---

SEMESTER- II

Code	Title of Paper	Category of Courses	As per NEP 2020	Teaching Hours Per Week			C	IA	SEE	Exam duration	Max Marks
			Course Level	L	T	P					
MSP 201	Big Data Processing	DSC (core course)	500-599	4	0	0	4	30	70	3 Hrs.	100
MSP 202	Fundamentals of Machine Learning	DSC (core course)	500-599	4	0	0	4	30	70	3 Hrs.	100
MSP 203	Quantum Machine learning	DSC (core course)	500-599	3	1	0	3	25	50	3 Hrs.	75
MSP 204	Elective – I	DSE Skill Oriented Course	400 - 499	4	0	0	4	30	70	3 Hrs	100
MSP 205	Elective – II	DSE Skill Oriented Course	400 - 499	4	0	0	4	30	70	3 Hrs	100
MSP 206	Hands-on: Elective -I	-	400 - 499	0	0	4	2	15	35	3 Hrs.	50
MSP 207	Hands-on: Elective -II	-	400 - 499	0	0	4	2	15	35	3 Hrs.	50
MSP 208	Completion of MongoDB from MongoDB University	ABCC	300-399	3	1	0	4	-	-	-	-
IKS 209	Indian Knowledge Systems 2	IKS	200-299	2	0	0	0	-	-	2 Hrs	-
Total							27				575

ELECTIVE DETAILS

Code	Elective – I (Choose any one)	Code	Elective – II (Choose any one)
MSP 204A	MLOps	MSP 205A	Data Engineering with Python
MSP 204B	Developing Web Applications	MSP 205B	Data Engineering using Kafka and Spark Structured Streaming

SEMESTER- III

Code	Title of Paper	Category of Course	As per NEP 2020	Teaching Hours Per Week			C	IA	SEE	Exam Duration	Max Marks
			Course Level	L	T	P					
MSP 301	Artificial Intelligence and Deep Learning	DSC (core course)	500-599	4	0	0	4	30	70	3 Hrs	100
MSP 302	Cloud Computing - AWS / Azure / GCP	DSC (core course)	500-599	4	0	0	4	30	70	3 Hrs	100
MSP 303	Elective – III	DSE Skill Oriented Course	400-499	4	0	0	4	30	70	3 Hrs	100
MSP 304	Elective – IV	DSE Skill Oriented Course	400-499	4	0	0	4	30	70	3 Hrs	100
OEC 305	Open Elective(MOOCs)	OEC	400-499	3	1	0	3	-	-	-	-
MSP 306	Hands-on: Elective – III	-	400-499	0	0	4	2	15	35	3 Hrs	50
MSP 307	Hands-on: Elective - IV	-	400-499	0	0	4	2	15	35	3 Hrs	50
Total							23				500

ELECTIVE DETAILS

Code	Elective – III (Choose anyone)	Code	Elective – IV (Choose anyone)
MSP 303A	Data Science with PySpark	MSP 304A	Data Engineering using AWS Data Analytics
MSP 303B	AI and Deep Learning through TensorFlow	MSP 304B	Data Engineering with Google Big Query & Google Cloud

Open Elective (MOOCS)

OEC 305	Artificial Intelligence in Drug Discovery and Development - https://onlinecourses.nptel.ac.in/noc25_ch96/preview
OEC 305	Responsible & Safe AI Systems - https://onlinecourses.nptel.ac.in/noc25_cs118/preview

SEMESTER – IV

Code	Title of the Paper	Category of Courses	As per NEP 2020	No. of Credits	Sessional Marks	Semester End marks SEE	Exam Duration	Max Marks
			Course Level					
MSP 401	Capstone Project	DSC	400-499	12	100	200	3 Hrs.	300
Total				12				300

Total credits:

Semester	SEMESTER- I	SEMESTER- II	SEMESTER- III	SEMESTER – IV	Total
Credits	28	27	23	12	90

Program Education Objective (Peo):

- PEO1:** Build a solid academic and technical foundation in data science to meet the evolving demands of national and global industries through applied knowledge in AI, ML, big data, and analytics.
- PEO2:** Adapt and thrive in a dynamic industrial ecosystem by engaging in continuous learning, embracing technological advancements, and effectively contributing to multidisciplinary teams and innovation-driven environments.
- PEO3:** Contribute meaningfully to society by leveraging data-driven insights to solve societal problems, support decision-making in public policy, healthcare, education, and environmental sustainability.
- PEO4:** Demonstrate ethical responsibility and global professionalism, upholding data integrity, privacy, and security while working in culturally diverse and globally distributed teams.
- PEO5:** Pursue lifelong learning and leadership roles, whether in industry, academia, research, or entrepreneurship, while addressing the challenges of a globally interconnected and data-centric world.

PROGRAM-SPECIFIC OUTCOMES (Psos):

- PSO1:** Apply data science tools and techniques, including statistical analysis, machine learning, and visualization, to extract meaningful insights from structured and unstructured data.
- PSO2:** Design, develop, and deploy end-to-end data-driven web and cloud-based applications that integrate modern data engineering, AI/ML models, and real-time analytics.
- PSO3:** Demonstrate ethical responsibility, effective communication, and teamwork while solving data-centric problems in multidisciplinary and professional environments.

Program Outcome (Pos):

- PO1:** Apply knowledge of mathematics, statistics, computer science, and data science fundamentals to solve complex problems in data-driven domains.
- PO2:** Identify, formulate, and analyze complex real-world problems using data-driven approaches, supported by research literature and statistical principles.
- PO3:** Design and develop effective models, algorithms, and data pipelines that address complex data science problems, with due consideration for ethical, social, and environmental contexts.
- PO4:** Select, adapt, and apply modern tools like AI and ML algorithms, programming languages, libraries, and data science platforms for data analysis, machine learning, and visualization, with an understanding of their limitations.
- PO5:** Assess societal, legal, cultural, and ethical implications of data science solutions and ensure responsible and context-aware data usage in professional practice.
- PO6:** Understand the environmental and social impact of data-driven decisions and advocate for sustainable development through ethical data practices and responsible AI.
- PO7:** Communicate effectively with technical and non-technical stakeholders through well-structured reports, visualizations, presentations, and documentation, conveying complex data insights clearly and convincingly.
- PO8:** Adhere to ethical principles and professional standards in data acquisition, analysis, interpretation, and dissemination, with a strong commitment to data privacy and integrity.

Detailed Syllabus

Semester 1 UNIT WISE DETAILS

MSP 101-Mathematical Foundation for Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Mathematical Foundation for Data Science	Course Code	MSP 101
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	50	CIE Marks	25

Unit I: Fundamentals of Algebra and Functions – 07 Hours

Set Theory & Number Systems, Sets and their operations, Number system, Relations and Functions, Relations and their types, Functions and their types, Polynomials and Quadratic Functions Algebra of polynomials (addition, subtraction, multiplication, division), Quadratic functions: Minima, maxima, vertex, and slope, Quadratic equations, Straight Lines, Slope of a line, Parallel and perpendicular lines Representations of a line General equations of a line Straight-line fitting

Unit II: Advanced Functions and Calculus Fundamentals – 10 Hours

Functions and Transformations, Horizontal and vertical line tests, Exponential functions, Composite functions, Inverse functions, Logarithmic Functions, Properties and graphs, Exponential and logarithmic equations, Sequences, Limits, and Differentiation, Limits for sequences, Function of one variable: Graphs and tangents, Differentiability and the derivative, Tangents and critical points (local maxima and minima)

Unit III: Linear Algebra Basics – 11 Hours

Vectors and Matrices, Vectors and their operations, Matrices and matrix operations, Systems of linear equations, Linear Dependence and Transformation, Linear dependence, basis and dimension, Rank and dimension using Gaussian elimination, Rank and nullity of a matrix, Introduction to linear transformation, The null space of a matrix, Linear mapping and transformations

Unit IV: Orthogonality and Multivariable Calculus – 12 Hours

Orthogonality and Orthonormality, Gram-Schmidt method, Orthogonality and linear independence, Projections using inner products, The Gram-Schmidt process, Orthogonal transformations and rotations, Multivariable Calculus, Multivariable functions and visualization, Partial derivatives, Directional derivatives, Limits for scalar-valued multivariable functions, Continuity for multivariable functions

Unit V: Graph Theory and Algorithms – 20 Hours

Graph Representation and Traversal, Representation of graphs, Breadth-first search (BFS), Depth-first search (DFS), Complexity of BFS and DFS, Graph Theory Algorithms, Single-source shortest paths, Dijkstra's algorithm, Bellman-Ford algorithm, All-pairs shortest paths, Minimum cost spanning trees, Prim's algorithm, Kruskal's algorithm

References:

1. Textbook: Mathematics for Machine Learning" by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong
2. Reference Book: Discrete Mathematics and Its Applications" by Kenneth H. Rosen, 7th Edition

Course Outcomes [COs]

Unit	Course Outcome	Bloom's Level
I	CO1: Understand algebraic techniques to apply and analyze functions, solve quadratic equations, and interpret linear relationships in coordinate geometry.	Understand/ Apply/ Analyze
II	CO2: Analyse and interpret the behaviour of exponential, logarithmic, and composite functions, and evaluate limits and derivatives of single-variable functions.	Analyze
III	CO3: Solve systems of linear equations using matrix operations and evaluate the properties of vector spaces and linear transformations.	Evaluate
IV	CO4: Construct orthonormal bases and compute partial and directional derivatives of multivariable functions to understand gradient-based behaviour.	Understand/ Create
V	CO5: Apply graph traversal and shortest path algorithms to solve problems related to network optimization and data structures.	Apply

CO-PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Apply/ Analyze	3	2	1	1	0	0	1	1
CO2	Analyze	3	2	2	1	0	0	1	1
CO3	Evaluate	3	3	2	2	0	0	1	1
CO4	Understand/ Create	3	2	3	2	0	1	1	1
CO5	Apply	2	3	3	3	1	1	2	1

COs \ PSOs Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	1	1
CO2	3	2	1
CO3	3	2	1
CO4	3	2	1
CO5	3	3	2

MSP 102- Statistics for Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Statistics for Data Science	Course Code	MSP 102
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	50	CIE Marks	25

Unit I: Fundamentals of Data and Descriptive Statistics– 12 Hours

Types of Data, Types of data (categorical, numerical, etc.), Descriptive vs. Inferential statistics, Scales of measurement, Measures of Central Tendency, Mean, median, and mode, Quartiles and percentiles, Measures of Dispersion, Range, variance, standard deviation, Interquartile range (IQR) and five-number summary, Association Between Two Variables, Contingency tables and relative frequencies, Scatterplots, covariance, Pearson correlation coefficient, Point-biserial correlation coefficient

Unit II: Counting Principles and Probability Theory– 10 Hours

Basic Principles of Counting, Addition and multiplication rules, Factorials, Permutations and combinations, Probability Fundamentals, Definitions of probability, Events and properties of probability, Conditional Probability and Bayes' Theorem, Multiplication rule of probability, independent events, Law of total probability, Bayes' theorem

Unit III: Random Variables and Distributions– 12 Hours

Random Variables, Random experiments, and sample space, Discrete vs. continuous random variables, Probability mass function (PMF) and cumulative density function (CDF), Expectation and Variance, the Expected value of a random variable, Variance and standard deviation of a discrete random variable, Common Probability Distributions, Bernoulli trials and independent & identically distributed (i.i.d.) random variables, Binomial and Poisson distributions, Uniform and exponential distributions

Unit IV: Multiple Random Variables and Dependency– 13 Hours

Joint Distributions, two random variables and their distributions, Independence of multiple random variables, Functions of Random Variables, Visualization and transformations of multiple random variables, Expectations and Variance in Multiple Variables, expected value in different contexts (casino math, data applications), Scatter plots, spread, and variance, Covariance and correlation, Statistical inequalities

Unit V: Advanced Topics in Statistical Inference– 13 Hours

Continuous Random Variables and Models, Working with height/weight datasets, Models with multiple random variables, Dependency modelling (e.g., IPL powerplay models), Estimation and Hypothesis Testing, Bayesian estimation, Hypothesis testing, and confidence intervals

References:

1. Introductory Statistics (10th Edition) - ISBN 9780321989178, by Neil A. Weiss published by Pearson.

2. Introductory Statistics (4th Edition) - by Sheldon M. Ross.

Textbook: "Statistics and Probability for Engineering Applications" by William DeCoursey

Reference Book: "Introduction to Probability" by Joseph K. Blitzstein and Jessica Hwang

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: Describe and remember measures of central tendency and dispersion and interpret relationships between two variables using basic statistical tools.	Understand/ Remember
II	CO2: Apply counting techniques and probability rules to solve problems involving random events, including conditional probability and Bayes' Theorem.	Apply
III	CO3: Analyse random variables and their associated distributions to compute expectations and interpret real-world probabilistic models.	Analyse
IV	CO4: Evaluate joint distributions, covariance, and correlation among multiple random variables in data science contexts.	Evaluate
V	CO5: Create inferential models using estimation and hypothesis testing to draw data-driven conclusions in practical scenarios.	Create

CO-PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Remember	3	2	1	1	1	0	2	2
CO2	Apply	3	2	2	1	1	1	1	2
CO3	Analyse	3	3	2	2	1	1	1	2
CO4	Evaluate	3	3	2	2	1	1	1	2
CO5	Create	3	3	3	2	2	2	2	3

CO-PSO Mapping

CO \ PSO	PSO1	PSO2	PSO3
CO1	3	1	1
CO2	3	1	1
CO3	3	3	1
CO4	3	2	1
CO5	3	2	2

MSP 103-Python for Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Python for Data Science	Course Code	MSP 103
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: – 10 Hours

Intro to Python Programming – Features of python, Data Types-Fundamental data types, Type casting, Collection related data types, Operators, Input and output statements, Flow control- Conditional Statements & Loops, String detailing, In-built Data Structures detailing and implementing comprehension - List, Tuple, Dictionary, Set.

Unit II: – 14 Hours

Advanced Python concepts – Functions: Introduction, Types of arguments, Recursion, Types of variables in functions, lambda functions, Map, filter, reduce. Functional Programming. Modules, Packages, Object Oriented Programming, Exception Handling, File handling, Decorators, Generators, Debugging Python, Regular expressions. Asynchronous programming

Unit III: 10 Hours

Introduction to Data Science - Different Sectors Using Data Science, Data Scientist Vs Data Analyst Vs Data engineer, Data science workflow, EDA-Quantitative Technique and EDA - Graphical Technique, communicating insights and making predictions.

Unit IV: 16 Hours

Python for DS - NumPy, Pandas, Data Visualization using Matplotlib and Seaborn Machine Learning with Scikit-Learn, Natural Language Processing with Scikit Learn, Web Scraping with BeautifulSoup, APIs to get the data from web. Introduction to computer vision in Data science. Using Git and GitHub for Data science.

Unit V: 10 Hours

Algorithmic thinking and optimization for Data Science - Introduction, Sorting, Searching, and hashing in Data science, Graph algorithms, Optimization techniques, Advanced optimization and dynamic programming, Real word Applications and case study, when to use traditional algorithms vs ML models, Python implementations.

References:

1. "Python for Data Analysis" (McKinney) 3rd Edition 2022
2. "Fluent Python" (Ramalho) 2nd Edition 2022
3. "Data Science for Beginners" (Park) 1st Edition 2020
4. "Python Data Science Handbook" (VanderPlas) 2nd Edition 2023
5. "Introduction to Algorithms" (Cormen) 4th Edition 2022

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: Understand and apply Python programming fundamentals including data types, flow control, strings, and in-built data structures to build basic scripts.	Understand/ Apply
II	CO2: Analyze and implement advanced Python features like OOP, functions, modules, file handling, and asynchronous programming for modular and efficient code.	Analyze
III	CO3: Describe data science roles, workflows, and methods for exploratory data analysis and insight communication.	Understand/ Remember
IV	CO4: Apply and analyze Python libraries for data processing, machine learning, and web-based data retrieval using APIs and web scraping tools.	Apply/ Analyze
V	CO5: Evaluate algorithmic techniques and optimization strategies in solving real-world data problems using Python implementations.	Evaluate

CO-PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Apply	3	1	2	3	0	0	1	1
CO2	Analyze	3	2	3	3	0	0	1	1
CO3	Understand/ Remember	2	3	2	2	2	1	3	2
CO4	Apply/ Analyze	2	2	3	3	1	1	2	2
CO5	Evaluate	3	3	3	3	1	1	2	2

CO\PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	2	1
CO2	2	3	1
CO3	3	1	2
CO4	3	1	1
CO5	3	3	2

MSP 104-Quantum Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Quantum Data Science	Course Code	MSP 104
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	50	CIE Marks	25

Unit 1: Introduction to Quantum Mechanics and Quantum Computing

Classical vs Quantum Information: Classical bits (0 or 1) vs quantum bits (qubits) (superposition and entanglement), Introduction to quantum states: pure states and mixed states. Quantum Mechanics Fundamentals: Superposition, Entanglement: Non-locality and correlations between qubits, Measurement in quantum computing. Quantum Gates: Pauli gates: X, Y, Z, Hadamard gate: Creating superposition, CNOT gate: Entangling qubits. Quantum Circuits: Introduction to quantum circuits and the flow of quantum algorithms, Basic quantum algorithms (e.g., Deutsch-Jozsa algorithm, Bernstein-Vazirani). Overview of Quantum Computing Models: Gate-based quantum computing (e.g., IBM Qiskit, Google Cirq), Quantum annealing and adiabatic quantum computing.

Unit 2: Quantum Algorithms and Their Applications in Data Science

Grover's Search Algorithm: Introduction to Grover's algorithm for searching an unsorted database, Speedup compared to classical search algorithms (quadratic speedup). Shor's Algorithm: The problem of integer factorization, Shor's algorithm solves integer factorization, Implications for cryptography (RSA encryption). Quantum Approximate Optimization Algorithm (QAOA): Overview of QAOA and its applications in combinatorial optimization problems, Uses of QAOA in real-world optimization problems. Quantum Machine Learning Algorithms: Quantum Support Vector Machines (QSVM): role of quantum computing in improving SVM performance, Quantum k-means clustering: Using quantum algorithms for clustering tasks, Introduction to quantum neural networks (QNN). Quantum vs Classical Algorithms.

Unit 3: Introduction to Quantum Machine Learning (QML)

Overview of Quantum Machine Learning – Introduction, Importance of QML in promising for data science and AI, Differences between classical machine learning and quantum machine learning. Quantum Data Encoding- classical data encoding process into quantum states (quantum feature maps), Process of encoding large datasets into quantum states. Quantum Support Vector Machines (QSVM) - Overview of Support Vector Machines (SVM), Role of Quantum systems in enhancing the performance of SVMs. Quantum k-means Clustering: Impact of quantum computing in k-means clustering algorithm, Quantum-enhanced clustering using quantum states. Quantum Neural Networks (QNN): Basics of quantum neural networks: QNN architectures, Training QNNs with quantum circuits for classification tasks, Comparison between classical and quantum neural networks.

Unit 4: Quantum Data Structures and Quantum Databases

Quantum Data Structures: Overview - quantum arrays, quantum trees, quantum lists, difference between classical and quantum data structures. Quantum Databases: Quantum data models: quantum states as data elements, Role of quantum computers in querying and data retrieval. Quantum Search Techniques: Speedup in searching databases using quantum algorithms, Applications of Grover's algorithm to quantum databases. Quantum Memory and Storage: Theoretical models of quantum memory, Challenges and solutions in storing and accessing quantum data. Quantum Security and Privacy: Implications of quantum computing on cryptography and data privacy, Quantum-resistant encryption methods.

Unit 5: Practical Applications and Future of Quantum Data Science

Quantum Data Science Applications: Applications in finance: Portfolio optimization, fraud detection, Applications in healthcare: Drug discovery, genomic data analysis, Applications in logistics: Quantum optimization for supply chain management. Quantum Cryptography: Overview and its implications for secure communication, Quantum key distribution (QKD) and post-quantum cryptography. Hybrid Quantum-Classical Models: role of quantum computing with classical computing to solve large-scale data science problems, Examples of hybrid models in quantum machine learning. Challenges in Quantum Computing: Issues of quantum noise, error correction, and the need for stable qubits, Hardware limitations and quantum scalability. Future Directions in Quantum Data Science: Prospects for quantum computing overcoming its current challenges.

Reference Books:

1. **"Practical Quantum Computing for Developers: A Hands-on Introduction"** by Vladimir M. Dunjko and Hannes F. St. Pierre
Why: Provides hands-on guidance for implementing quantum algorithms using tools like Qiskit.
Recommended for: Unit 2 and Unit 3 (Quantum Algorithms & Quantum Machine Learning).
2. **"Quantum Computing: An Applied Introduction"** by Jack D. Hidary
Why: Focuses on practical applications and the development of quantum algorithms for real-world use.
Recommended for: Unit 5 (Practical Applications and Future of Quantum Data Science)

Course Outcomes [COs]

Unit	Course Outcome	Bloom's Level
I	CO1: Understand fundamental principles of quantum mechanics and differentiate classical vs. quantum information processing.	Understand/ Apply/ Analyze
II	CO2: Analyze and apply quantum algorithms (e.g., Grover's, Shor's) to optimization and search problems in data science.	Analyze
III	CO3: Explore quantum machine learning models like QSVM, QNN, and quantum k-means for classification and clustering tasks.	Evaluate
IV	CO4: Evaluate quantum data structures and their impact on databases, storage, retrieval, and data privacy.	Evaluate/Create
V	CO5: Evaluate real-world applications of quantum data science in finance, healthcare, logistics, and assess future directions.	Apply

CO-PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Apply/ Analyze	3	2	1	2	0	0	2	0
CO2	Analyze	3	3	2	2	0	1	2	1
CO3	Evaluate	3	3	3	3	0	1	3	2
CO4	Evaluate/Create	2	2	2	3	2	1	2	3
CO5	Apply	2	2	1	2	3	2	3	0

COs \ PSOs Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	0	3
CO2	3	2	3
CO3	3	3	3
CO4	2	2	3
CO5	3	2	3

MSP 105-Fundamentals of Database Systems: SQL & MongoDB

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Database Systems: SQL & MongoDB	Course Code	MSP 105
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: 13 Hours

Types of DB - Relational, Non-relational, ER diagrams, SQL Commands - DDL, DML, DCL, Aggregate Functions, Joins, SubQueries, Normalization, Scaling patterns, ACID Properties, Relational Query Languages and Database Design.

Unit II: 13 Hours

Relational Query Languages and Database Design - Storage Management, Indexing and Hashing, Transactions, Backup and Recovery, Query Optimization, SQL Injection, DaskSQL, CloudSQL (Athena/Big Query).

Unit III: 10 Hours

Introduction to MongoDB – Databases, Collections, Architecture and Installation, Schema Design and Data Modelling, CRUD Operations – Insert, Query, Delete, Update Documents, Bulk Write Operations, Retriable Writes & Reads.

Unit IV: 12 Hours

Indexing - Need of Index, types of Indices, Aggregation Framework: Aggregation stages, Aggregation operations, Map Reduce, MongoDB Security – Authentication, Role Base d Access Control, Encryption, Auditing.

Unit V: 12 Hours

MongoDB on the Cloud – Intro to MongoDB Atlas, Atlas account set up, Usage of Mongo Cluster – Serverless, Dedicated, Shared, Deploy free tier Cluster, MongoDB Atlas Pricing, Mongo Realm.

References:

- 1." Database Management Systems" - Abraham Silber Schatz, Henry F. Korth, S. Sudarshan , Fourth Edition
2. "MongoDB: The Definitive Guide 1st Edition" by Kristina Chodorow and Michael Dirolf 1st Edition
3. "MongoDB in Action" by Kyle Banker 2nd Edition
4. "NoSQL with MongoDB in 24 Hours, Sams Teach Yourself 1st Edition" by Brad Dayley

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: Describe the core concepts of relational and non-relational databases, ER modeling, normalization techniques, SQL commands, and ACID properties.	Understand/ Remember
II	CO2: Analyze indexing methods, storage management, transaction processing, and evaluate cloud-based SQL solutions such as DaskSQL and BigQuery for optimized performance.	Apply
III	CO3: Implement CRUD operations and design effective schemas in MongoDB using collections, documents, and data modeling strategies.	Apply
IV	CO4: Implement modular programming concepts using functions, recursion, pointers, and parameter passing techniques.	Apply
V	CO5: Demonstrate the setup and deployment of MongoDB Atlas clusters and utilize cloud-native tools and pricing models for scalable database solutions.	Understand/ Apply

CO-PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Remember	3	2	2	2	1	0	1	1
CO2	Apply	3	3	2	3	1	1	1	1
CO3	Apply	3	2	3	3	1	0	1	1
CO4	Apply	2	1	2	2	0	0	1	1
CO5	Understand/ Apply	2	2	2	3	1	1	1	1

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	2	1
CO2	3	3	1
CO3	2	3	1
CO4	2	2	1
CO5	2	3	2

MSP 106- Data Visualization & Story Telling

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Data Visualization & Story Telling	Course Code	MSP 106
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: 12 Hours

Introduction to Visualization - Visualization and its key topics, Visualization Basics, Exploratory Vs Explanatory Analysis, Decluttering and Highlighting. Multivariant Visualization Method - Geometric Representations, Icon Representation, Pixel-Oriented Representations, Types of Graphs.

Unit II: 16 Hours

Data visualization tools - Fundamentals of Data Visualization using Excel, fundamentals of Data Visualization using Tableau, fundamentals of Data Visualization using PowerBI. Plots - Area plots, Histograms, Bar Charts, Pie Charts, Box Plots, Scatter Plots.

Unit III: 12 Hours

Advanced Visualizations and Geospatial Data - Waffle Charts, Word Clouds, Seaborn and Regression Plots, Maps with Markers.

Unit IV: 10 Hours

Getting Started in Effective and Ineffective Visuals - The Human Brain and Data Visualization, Cognitive vs Perceptual Design Distinction, Tableau Desktop Vs Tableau Public. Dive into Visualization - Map Abstraction, Visual encoding of data, Visualization design, Clutter and Visual Perception, Visualization Aesthetics, Enhancing the Visualization.

Unit V: 10 Hours

Design for Understanding - Design for Purpose, Data, Relationships and Design, Static Vs Interactive Visualization, Multiple Connected Views, Language, Labelling and Scaling.
Assignment - Tell your Great Story with Data.

References:

1. "Storytelling with Data: A Data Visualization Guide for Business Professionals" by Cole Nussbaumer Knaflic 1st Edition
2. "Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures" by Claus O. Wilke 1st Edition
3. "Data Visualization: A Practical Introduction" by Kieran Healy and Kieran Joseph Healy

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: Explain the fundamentals of data visualization and differentiate between exploratory and explanatory analysis techniques.	Understand
II	CO2: Apply various data visualization tools (Excel, Tableau, Power BI) to create basic visual representations of data.	Apply/ Remember
III	CO3: Analyze and construct advanced and geospatial visualizations to derive meaningful insights.	Analyze/ Create
IV	CO4: Evaluate visualization effectiveness based on human perception, visual encoding, and design principles.	Evaluate
V	CO5: Create purpose-driven and aesthetically effective data visualizations that support storytelling and user interaction.	Create

CO\PO Mapping:

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	2	2	2	2	1	1	2	1
CO2	Apply/ Remember	2	1	2	3	1	0	2	1
CO3	Analyze/ Create	2	2	3	3	1	1	2	1
CO4	Evaluate	1	1	2	2	1	1	1	1
CO5	Create	1	2	2	2	1	1	3	1

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	3	1	1
CO2	3	1	1
CO3	3	2	1
CO4	3	1	2
CO5	3	2	2

MSP 107- Hands on Python for Data Science

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Hands on Python for Data Science	Course Code	MSP 107
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools: Python, Jupyter Notebook, VScode , NumPy, Pandas, Scikit-Learn, Matplotlib/Seaborn, Beautiful Soup and Git and GitHub.

1. Operations with different data types
2. Conditional statements
3. Functions and recursion
4. String manipulation
5. List, tuple, dictionary, and set operations
6. Lambda functions and list comprehensions
7. Functional programming (map, filter, reduce)
8. Decorators and *args/**kwargs
9. OOP concepts Exception handling (try-except blocks)
10. Basic DSA
11. Searching and sorting
12. Bit manipulation and 2D list operations
13. Exploratory Data Analysis
14. EDA with graphical techniques (histograms, box plots)
15. Data types and plotting relationships
16. NumPy operations
17. Pandas
18. Data visualization with Matplotlib/Seaborn
19. Regular expressions for pattern matching
20. Machine learning with Scikit-learn (linear regression, classification)
21. Web scraping with Beautiful Soup
22. File handling
23. Git/GitHub for version control
24. Data Structure Algorithms

MSP 108- Hands on Data Visualization & Story Telling

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Hands on Data Visualization & Story Telling	Course Code	MSP 108
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools : Excel, Tableau, PowerBI, Python(Matplotlib, Seaborn, Pywaffle)

1. Exploratory vs. Explanatory Analysis
2. Decluttering and Highlighting
3. Multivariate Visualization
4. Excel
5. Tableau
6. PowerBI
7. Advanced Plots
8. Waffle Charts & Word Clouds
9. Seaborn Regression Plots
10. Maps with Markers
11. Visual Encoding & Aesthetics
12. Labelling & Scaling
13. Storytelling with Data – Create a full dashboard with a narrative.

Open Online Transdisciplinary Course – Soft Skills

(Semester-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Soft Skills	Course Code	OOTC 109
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of SEE	-
SEE Marks	-	CIE Marks	-

Unit I: Foundations of Communication & Personality Development

Introduction to Soft Skills: Definition, Importance, and Aspects. Effective Communication: Classification (Verbal/Non-verbal), Barriers, and Strategies. Personality Development: Positive Thinking, Self-awareness, and Professional Etiquette.

Unit II: Non-Verbal & Interpersonal Communication

Non-Verbal Communication: Paralanguage, Proxemics, Haptics, and Meta-communication. Listening Skills: Types (Active, Empathetic), Techniques, and Overcoming Barriers. Cultural & Organizational Communication: Cross-cultural Nuances, Formal/Informal Channels.

Unit III: Business Writing & Professional Documentation

Principles of Business Writing: Clarity, Conciseness, and Tone. Business Letters: Formats, Styles, and Types (Inquiries, Complaints, Sales). Report Writing: Structure, Data Organization, Types (Analytical/Informational), and Style.

Unit IV: Group Dynamics & Leadership Skills

Group Communication: Meetings, Discussions, and Conflict Resolution. Leadership & Adaptability: Traits, Work Ethics, and Decision-making. Public Speaking: Debates, Speeches, and Combating Nervousness.

Unit V: Career Readiness & Advanced Professional Skills

Presentation Skills: Planning, Methods, and Delivery Techniques. Job Interviews: Résumé Preparation, Mock Interviews, and Body Language. Emotional Intelligence: Critical Thinking, Stress Management, and Applied Grammar.

References:

- 1.Butterfield, Jeff. Soft Skills for Everyone. New Delhi: Cengage Learning. 2010.
- 2.Chauhan, G.S. and Sangeeta Sharma. Soft Skills. New Delhi: Wiley. 2016.
- 3.Goleman, Daniel. Working with Emotional Intelligence. London: Banton Books. 1998.
- 4.Hall, Calvin S. et al. Theories of Personality. New Delhi: Wiley. rpt. 2011.
- 5.Holtz, Shel. Corporate Conversations. New Delhi: PHI. 2007

SEMESTER 2

MSP 201- Big Data Processing

(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Big Data Processing	Course Code	MSP 201
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: 10 Hours

Big data Basics – Big Data definition, Sources of Big Data, Big Data Characteristics-5Vs, Applications of Big Data-Big Data Benefits and Challenges, Big Data Engineering principles, Fundamentals of Hadoop - Introduction, Architecture, Ecosystem.

Unit II: 12 Hours

Data preparation, Storage and Workflow Orchestration – Types of Data, Stages of Data Preprocessing, Big Data Storage, File formats in Big data, Big Data Architecture, Workflow orchestration

Unit III: 14 Hours

Distributed Systems and batch processing - Big Data Processing Vs Traditional Distribution Systems, Batch Processing- Batch processing with MapReduce, A simple MapReduce example, Apache Spark for batch processing, Optimization techniques

Unit IV: 14 Hours

Real-Time and Streaming data Processing - Real-time processing overview-examples, real-time faces, Streaming data processing, Building Streaming pipelines with Kafka and Flink, Edge processing Basics

Unit V: 10 Hours

Advance Techniques and Future Trends – Optimization in bigdata, AI driven Data processing, Data privacy and Governance, Quantum computing potential, Emerging Analytical use cases.

References:

1. "Hadoop: The Definitive Guide" (White) 4th Edition 2015
2. "Big Data: Principles and Best Practices" (Marz/Warren) 1st Edition 2015
3. "Spark: The Definitive Guide" (Chambers/Zaharia) 1st Edition 2018
4. "Streaming Systems" (Akidau) 1st Edition 2018
5. "Designing Data-Intensive Applications" (Kleppmann) 1st Edition 2017

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: <i>Understand</i> the fundamental concepts, characteristics, and architecture of Big Data systems, and the Hadoop ecosystem.	Understand/ Remember
II	CO2: <i>Apply</i> data preprocessing and workflow orchestration techniques with big data storage formats like Avro, Parquet, and ORC.	Apply
III	CO3: <i>Apply</i> batch processing operations using MapReduce and Apache Spark in a distributed environment.	Apply
IV	CO4: <i>Create</i> real-time streaming pipelines using Apache Kafka and Flink to process and analyze data streams.	Create
V	CO5: <i>Evaluate</i> cutting-edge trends in big data technologies including AI-integrated systems, data governance strategies, and quantum computing advancements.	Evaluate

CO\PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Remember	3	2	2	3	1	1	1	1
CO2	Apply	3	2	2	3	1	1	1	1
CO3	Apply	3	3	3	3	1	2	1	1
CO4	Create	3	2	3	3	1	1	2	1
CO5	Evaluate	2	3	2	2	2	2	2	2

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	3	1
CO2	2	3	1
CO3	3	3	1
CO4	3	3	1
CO5	2	2	2

MSP 202-Fundamentals of Machine Learning

(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Fundamentals of Machine Learning	Course Code	MSP 202
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Introduction to Machine Learning (Foundation) – 12 Hours

Basics of Machine Learning, Definition and Importance, Applications of Machine Learning in various domains, Data Mining Process, Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment, Learning Paradigms, Supervised Learning: Classification and Regression, Unsupervised Learning: Clustering and Dimensionality Reduction

Unit II: Exploratory Data Analysis (EDA) (Foundation for Data Preparation)– 12 Hours

Understanding Data, Data types and formats, Summary statistics (mean, median, mode, variance, skewness, kurtosis), Distributions of attributes, Visualization, Seaborn and Matplotlib for Univariate, Bivariate, and Multivariate analysis, Identifying relationships between variables, Statistical Testing Hypothesis testing, p-values, correlation analysis Handling Imbalanced Data and Challenges

Unit III: Feature Engineering and Data Preprocessing (Intermediate Level)– 12 Hours

Feature Engineering, Creating new features, Feature selection and elimination, Data Preprocessing, Data Cleaning (Handling null values, duplicate data), Type Conversions, Outlier Detection and Treatment, Data Transformation (Scaling, Encoding – Label & One-hot encoding)

Unit IV: Unsupervised Learning & Dimensionality Reduction (Intermediate Level)– 12 Hours

Dimensionality Reduction, Principal Component Analysis (PCA), Factor Analysis, Clustering Methods, Hierarchical Clustering, Agglomerative Clustering, Cluster Profiling and Visualization, Applications of Clustering

Unit V: Supervised Learning (Classification) (Advanced Level)– 12 Hours

Introduction to Supervised Learning, Classification Problems, Decision Trees (Binary and Multi-class), Implementing Decision Trees using Sklearn, Model Evaluation and Performance Metrics, Confusion Matrix, Precision, Recall, F1 Score, K-fold Cross Validation, Feature Importance

References:

1. Pattern Recognition and Machine Learning by Christopher M. Bishop, 1st Edition
2. Machine Learning by Tom M. Mitchell, 1st Edition
3. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron

Course Outcomes:

Unit	Course Outcome (CO)	Bloom's Level
I	CO1: <i>Understand</i> the fundamental concepts of machine learning, types, applications, and data mining processes including various learning paradigms.	Understand
II	CO2: <i>Apply</i> exploratory data analysis techniques using statistical summaries and visualizations to uncover patterns and distributions in data.	Apply/ Analyze
III	CO3: <i>Apply</i> preprocessing techniques including scaling, encoding, and outlier handling for effective feature engineering.	Apply
IV	CO4: <i>Apply</i> dimensionality reduction and clustering algorithms to analyze unlabeled data and visualize the underlying structure.	Apply
V	CO5: <i>Evaluate</i> classification models using performance metrics like confusion matrix, precision, recall, F1-score, and cross-validation.	Evaluate/ Analyze

CO\PO Mapping

CO \ PO	Blooms' Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	3	2	2	2	1	1	1	1
CO2	Apply/ Analyze	3	3	2	3	1	1	2	1
CO3	Apply	3	2	3	2	1	1	1	1
CO4	Apply	2	2	3	3	1	1	2	1
CO5	Evaluate/ Analyze	3	3	2	2	2	1	2	2

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	3	1
CO2	2	3	1
CO3	3	2	1
CO4	3	2	1
CO5	3	2	1

MSP 203- Quantum Machine Learning

(Semester-II of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Quantum Machine Learning	Course Code	MSP 203
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	50	CIE Marks	25

Unit 1: Introduction to Quantum Computing and Quantum Information

Introduction to Quantum Mechanics: Basic principles such as superposition, entanglement, and quantum measurement. Classical vs. Quantum Information: Classical bits vs quantum bits (qubits), Introduction to quantum states: pure and mixed states, Superposition and interference. Quantum Gates and Circuits: Overview of quantum gates (Pauli gates, Hadamard gate, CNOT gate), Quantum circuits and their operations. Quantum Computing Models: Gate-based quantum computing, Quantum annealing and adiabatic quantum computing. Quantum Algorithms Overview: High-level introduction to famous quantum algorithms (e.g., Grover's algorithm, Shor's algorithm).

Unit 2: Introduction to Machine Learning and Classical Algorithms

Overview of Classical Machine Learning: Supervised learning: Classification and regression, Unsupervised learning: Clustering and dimensionality reduction, Overview of common algorithms: Linear regression, SVMs, k-means clustering, and decision trees. Optimization in Machine Learning: The concept of loss functions and optimization, Gradient descent and its variants. Neural Networks: Basics of artificial neural networks (ANNs), Training neural networks and backpropagation. Feature Engineering: Data preprocessing and feature extraction, Importance of features in classical machine learning algorithms. Challenges in Classical Machine Learning: High computational costs for large datasets, Overfitting, underfitting, and model complexity.

Unit 3: Quantum Machine Learning Algorithms

Quantum Feature Encoding: Classical data is mapping to quantum states (quantum feature maps), Benefits of encoding high-dimensional data in quantum space. Quantum Support Vector Machine (QSVM): Classical SVM vs. Quantum SVM, Quantum kernel methods and their potential advantages. Quantum k-Means Clustering: The classical k-means algorithm vs. Quantum k-means, Speedups in clustering using quantum superposition. Quantum Principal Component Analysis (Q-PCA): Quantum version of PCA for dimensionality reduction, quantum computing can potentially speed up PCA computations. Quantum Neural Networks (QNN): Introduction to Quantum Neural Networks, Quantum circuits for representing layers of neural networks, Training and optimizing QNNs with quantum algorithms.

Unit 4: Quantum Data and Quantum Circuits for Machine Learning

Quantum Data Representation: Representation of classical data in quantum systems (quantum bits, quantum states), Quantum data encoding methods: Amplitude encoding, basis encoding, and more. Quantum Circuit Design for Machine Learning: Quantum circuit architectures for machine learning algorithms, Building and simulating quantum circuits for QML tasks. Quantum Entanglement and Its Role in Machine Learning: Quantum entanglement is used to correlate data and enhance model complexity. Hybrid Quantum-Classical Systems: Introduction to hybrid models that use quantum circuits for part of the computation and classical algorithms for the rest, Example: Variational Quantum Eigensolver (VQE) and Quantum Approximate Optimization Algorithm (QAOA). Quantum Simulators and Tools for QML: Using quantum simulators like IBM Qiskit, Google Cirq, and Microsoft QDK, how to use these tools to build and test QML algorithms.

Unit 5: Practical Applications and Future of Quantum Machine Learning

Applications in Drug Discovery: Quantum machine learning can be used to simulate molecules and discover new drugs more efficiently. Finance and Optimization: Portfolio optimization and risk management using quantum algorithms, Quantum-enhanced optimization for large-scale financial problems. Quantum Machine Learning in Computer Vision: Applying quantum machine learning techniques to image recognition and computer vision. Quantum in Natural Language Processing (NLP): Exploring quantum algorithms for text processing, sentiment analysis, and language translation. Challenges and Future Directions in QML: Current challenges in quantum hardware and noise, The future of quantum machine learning with more robust quantum systems.

Reference Books:

1. "Practical Quantum Computing for Developers: A Hands-on Introduction" by Vladimir M. Dunjko and Hannes F. St. Pierre
 - Why: A practical introduction to quantum programming that includes hands-on examples and exercises using IBM Qiskit.
 - Recommended for: Unit 4 (Quantum Circuits for Machine Learning).
2. "Quantum Machine Learning with Python" by Santanu Pattanayak
 - Why: This book provides a more practical and applied approach to QML, with Python code examples and hands-on tutorials.
 - Recommended for: Unit 4 (Quantum Circuits for Machine Learning) and Unit 5 (Practical Applications).
3. "Quantum Computing for Computer Scientists" by Noson S. Yanofsky and Mirco A. Mannucci
 - Why: Offers a comprehensive look at quantum algorithms with a focus on computer science students.
 - Recommended for: Unit 1 (Quantum Computing) and Unit 3 (Quantum Machine Learning Algorithms)

Course Outcomes [COs]

Unit	Course Outcome	Bloom's Level
I	CO1: Demonstrate a foundational understanding of quantum mechanics principles such as superposition, entanglement, and quantum gates, and compare classical and quantum computing paradigms.	Understand/ Apply
II	CO2: Apply classical machine learning techniques including supervised and unsupervised learning, neural networks, and optimization methods to analyze data and solve computational problems.	Apply/ Analyze
III	CO3: Analyze and implement quantum machine learning algorithms such as QSVM, quantum k-means, Q-PCA, and QNN, and explain their advantages over classical counterparts.	Analyze / Apply
IV	CO4: Design quantum circuits and simulate quantum data representation techniques using tools like IBM Qiskit and Google Cirq for machine learning applications.	Apply/ Create
V	CO5: Evaluate the real-world applications of quantum machine learning in domains like drug discovery, finance, and natural language processing, and assess current challenges and future directions in QML.	Evaluate / Understand

CO-PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Apply/ Analyze	2	0	0	2	0	0	0	0
CO2	Analyze	3	3	3	3	0	0	0	0
CO3	Evaluate	3	2	1	3	2	0	0	2
CO4	Understand/ Create	2	2	2	2	3	0	3	2
CO5	Apply	0	3	1	1	2	3	2	3

COs\ PSOs Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	0	0
CO2	3	3	0
CO3	1	2	2
CO4	3	3	2
CO5	1	0	3

ELECTIVE-I

MSP 204 A-MLOps

(Elective-I of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	MLOps	Course Code	MSP 204 A
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Introduction to MLOps and ML Model Development – 12 Hours

Basics of MLOps: Definitions, Need, and Business Value, DataOps and Data Engineering: Concepts and Differences, Platform Automation in ML pipelines, Configuring Continuous Integration with GitHub Actions, Hands-on: Developing a simple ML model pipeline with CI

Unit II: MLOps Foundations and Tools – 12 Hours

Maturity Levels in MLOps (0 to 3), Overview of MLOps Tools: MLflow, DVC, TFX, Kubeflow, Cloud Shell Development Environments, MLOps for Containers and Edge Devices, Dockerize a model & push to container registry, Configure a basic MLOps workflow using MLflow/Kubeflow

Unit III: MLOps Lifecycle and CI/CD – 12 Hours

MLOps Lifecycle Management, DevOps Adaptation for ML (Build, Test, Integrate, Deliver), Infrastructure Requirements for Model Deployment, Build and test a CI/CD pipeline for ML models, Hands-on: Create reusable CI/CD scripts for model training and deployment

Unit IV: Advanced Model Deployment Patterns – 12 Hours

Deployment Patterns (Batch, Online, A/B, Shadow, Blue/Green), Trends: ML on Cloud/Edge/Browsers, Containers, Docker, Kubernetes (K8s), FaaS, Introduction to ML-as-a-Service (MLaaS), Deploy model on Kubernetes cluster, Case study: Edge deployment (e.g., TensorFlow Lite or ONNX)

Unit V: Cloud-specific MLOps Implementations – 12 Hours

Overview: MLOps on AWS, GCP, Azure, AWS Lambda for ML Inference, Orchestrating ML code on Azure with Azure ML Pipelines, Build CI/CD pipelines using GCP Vertex AI or Azure ML, Develop a microservice for model inference using AWS Lambda

References:

1. Introducing MLOps: How to Scale Machine Learning in the Enterprise

Authors: Mark Treveil, Nicolas Omont, Clément Stenac, Kenji Lefevre, Du Phan, Joachim Zentici, Adrien Lavoillotte, Makoto Miyazaki, Lynn Heidmann , Publisher: O'Reilly Media, Inc.

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: <i>Describe</i> the core concepts of MLOps, including Continuous Integration (CI), Continuous Deployment (CD), and DataOps practices.	Remember/ Understand
II	CO2: <i>Explain</i> the MLOps life cycle, model development pipeline, and adaptation of DevOps in ML model deployment.	Understand
III	CO3: <i>Apply</i> MLOps tools and techniques to build, test, integrate, and deploy machine learning models using GitHub Actions and cloud-based environments.	Apply
IV	CO4: <i>Analyze</i> model deployment patterns and infrastructure requirements across different environments such as cloud, edge, and containerized platforms.	Analyze
V	CO5: <i>Design and create</i> CI/CD pipelines for deploying ML models on platforms like AWS, GCP, and Azure using modern cloud-native tools.	Create

CO\PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Remember/ Understand	2	1	2	3	1	1	1	2
CO2	Understand	2	2	3	3	1	1	1	2
CO3	Apply	3	2	3	3	1	1	2	2
CO4	Analyze	2	3	3	3	1	2	2	2
CO5	Create	3	3	3	3	2	2	2	2

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	2	1
CO2	2	3	1
CO3	3	3	2
CO4	2	3	2
CO5	3	3	2

MSP 204B – Developing Web Applications

(Elective-I of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Developing Web Applications	Course Code	MSP 204B
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Internet and HTML Fundamentals - 10 hours

Internet Basics - Protocols, Servers and their Functions, Internet Clients, Network Security, Internet Development. HTML - Fundamentals/Basic HTML, HTML vs HTML5, Text formatting on Web Pages, incorporate images, creating hyperlinks, tables, and nested tables, inserting web page, complex image maps, Setting & modifying field properties, HTML Media, Validating HTML

Unit II: CSS and XML - 10 hours

Cascading Style Sheet (CSS) - Introduction CSS, CSS vs CSS3, Designing with Style Sheets, Style Sheet Syntax, ID, Class Contextual Selectors, Cascading Order, Properties, Absolute and Relative Positioning, Layering Elements using Z-Index, Animating objects. CSS Website Layout, CSS Flex, Responsive Design. Introduction to XML, Separates Structure from Formatting

Unit III: JavaScript and Web Programming -13 Hours

Basics of JavaScript - Syntax, Comments, Variables, Operators, Data Types, Functions, Objects, Events, Strings, Arrays, Dates, Control Structures, Sets, Maps, Regexp. Advanced JavaScript - JS Objects, JS classes, JS HTML DOM, JS Browser BOM, JS web APIs, JS Ajax, JS JSON, Introduction Javascript Framework

Unit IV: PHP Web Development - 14 Hours

PHP Hypertext Preprocessor (PHP) -Introduction, PHP Document, Language Fundamentals, Statements, Operators, Decision Making Statement, Loops, Server-Side Processing, Processing Forms via GET/POST, PHP functions, Arrays & Functions, String Functions, State and Persistence - Multi Page Website for state management, File Upload in PHP Basic OOP, Introduction to PHP Frameworks-Simple Laravel Application.

Unit V: SQL, MySQL & PHP Integration - 13 Hours

SQL & MySQL - Creating Databases & Tables, using keys, Types of Tables in MySQL, Data Types, deleting databases and tables, Inserting, Retrieving, Updating & Deleting data, User Accounts, Access Control, Export Import MySQL. Integrating PHP and MySQL - PHP interfacing with MySQL, Database Connecting, Executing SQL, Retrieving the data set & refining the fetch, PHP Ajax.

Reference Books:

1. HTML5, CSS3 and JavaScript Made Easy by Dr. T. Vasudev et al., Dreambook Publishing, 1st Edition (2021)
2. PHP, MySQL & JavaScript All-in-One For Dummies by Richard Blum, For Dummies (Wiley), 1st Edition
3. A Hand Book On Web Development by Pritma Jashnani, Notion Press, 1st Edition (2022)
4. HTML5 Black Book by DT Editorial Services, Dreamtech Press, Second Edition.

Course Outcomes:

Unit	Course Outcome (CO)	Bloom's Level
I	CO1: <i>Understand</i> the structure of web pages using HTML5 and apply validation techniques for building user-friendly front-end designs	Understand/ Apply
II	CO2: <i>Apply</i> CSS and XML to design responsive, structured, and layered web interfaces.	Apply
III	CO3: <i>Create</i> interactive and dynamic web elements using JavaScript and associated frameworks.	Create/ Analyze
IV	CO4: <i>Create</i> and implement server-side applications using PHP with form handling, session management, and basic concepts of Laravel.	Create
V	CO5: <i>Evaluate</i> MySQL databases with PHP to perform CRUD operations and enable dynamic web functionality using AJAX.	Evaluate

CO\PO Mapping

CO \ PO	Blooms' Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand/ Apply	2	2	2	2	1	1	2	1
CO2	Apply	2	2	2	2	1	1	2	1
CO3	Create/ Analyze	2	2	3	3	1	1	2	1
CO4	Create	2	2	3	3	1	1	1	1
CO5	Evaluate	2	2	3	3	1	1	2	1

CO \ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	1	3	1
CO2	1	3	1
CO3	1	3	1
CO4	2	3	1
CO5	2	3	1

ELECTIVE-II

MSP 205 A-Data Engineering with Python

(Elective I of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering with Python	Course Code	MSP 205 A
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Foundations of Data Engineering – 8 Hours

Introduction to Data Engineering: Role of Data Engineers, Data Engineering Lifecycle; File Handling and Data Ingestion: CSV, JSON, Parquet; Reading/Writing with csv, json, and pyarrow; Serialization with pickle, Message Pack; File System Operations with PathLab; Data Governance: Metadata Management with openmetadata, PII Anonymization with faker

Unit II: Data Manipulation and Scalability – 14 Hours

Pandas for Data Manipulation: Series, Data Frame, Joins, Grouping, Missing Data Handling, Memory Optimization, NumPy for Numeric Computations: Array Operations, Vectorization, numba for JIT Compilation, Integration with Pandas; Scalable Data Processing: Parallel Processing with Dask, Distributed Computing with Ray, Optimizing Python Workloads

Unit III: Advanced Data Transformation – 12 Hours

Data Transformation with dbt: SQL-Based Modelling, Testing, Documentation Data Cleaning and Validation: High-Performance Data Frames with polars, Schema Validation with pydantic, Anomaly Handling (Outlier Detection, Imputation); Asynchronous Data Ingestion: REST API Ingestion with httpx, Batch Processing with luigi

Unit IV: Databases and Workflow Automation – 13 Hours

PostgreSQL for Advanced Data Engineering: Advanced SQL with psycopg2, SQL alchemy (Window Functions, CTEs, Query Optimization), Indexing, Partitioning, Connection Pooling, ETL Pipelines; NoSQL Databases: MongoDB with pymongo for Document Modelling, Aggregation Pipelines; Data Orchestration: Airflow for DAGs, Scheduling, Monitoring; Docker for Containerized Pipelines

Unit V: Pipeline Monitoring and Quality Assurance – 13 Hours

Pipeline Monitoring and Observability: streamlit for Pipeline Dashboards, Prometheus and Grafana for Metrics Visualization; Data Quality and Testing: Great Expectations for Validation, Structured Logging, Unit and Integration Testing with pytest and hypothesis; DataOps and Governance: Feature Engineering with featuretools, Dataset Versioning with dvc, Data Lineage with openmetadata, CI/CD with GitHub Actions

References:

1. "Data Engineering with Python" by Paul Crickard (Packt) 1st Edition 2020
2. "Fundamentals of Data Engineering" by Joe Reis and Matt Housley (O'Reilly) 1st Edition 2022
3. "Designing Data-Intensive Applications" by Martin Kleppmann (O'Reilly) 1st Edition 2017

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: <i>Describe</i> the data engineering lifecycle and file handling operations in Python including text, binary, and CSV formats.	Remember/ Understand
II	CO2: <i>Apply</i> text processing techniques in Python including parsing HTML, handling JSON/XML, and using regular expressions.	Apply
III	CO3: <i>Analyze</i> structured data using pandas by transforming, reshaping, combining, and cleaning data efficiently.	Analyze
IV	CO4: <i>Apply and analyze</i> data from relational (MySQL) and document-oriented (MongoDB) databases using Python libraries.	Apply/ Analyze
V	CO5: <i>Design and develop</i> visual representations of data using PyPlot and Pandas to aid in data interpretation and communication.	Create

CO\PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Remember/ Understand	3	2	1	2	0	0	1	1
CO2	Apply	3	2	1	3	0	0	1	1
CO3	Analyze	3	3	2	3	1	0	2	1
CO4	Apply/ Analyze	3	3	2	3	1	0	2	1
CO5	Create	2	2	2	2	1	0	3	1

CO \ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	3	2	1
CO2	3	2	1
CO3	3	2	2
CO4	3	3	2
CO5	3	2	2

MSP 205 B-Data Engineering using Kafka and Spark Structured Streaming

(Elective-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering using Kafka and Spark Structured Streaming	Course Code	MSP 205 B
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Foundations of Streaming Data Engineering – 9 Hours

Introduction to Streaming Data Engineering: Role of Streaming Engineers, Batch vs. Streaming, Event-Driven Architectures; Introduction to Kafka: Topics, Partitions, Producers, Consumers, ZooKeeper, KRaft Mode, Basic Kafka Metrics; Spark Structured Streaming Basics: Features, Spark Shell, Data Frames, Streaming Queries, Spark UI Monitoring

Unit II: Kafka Core Operations – 12 Hours

Kafka Operations: CLI for Topic Management, Producing/Consuming with Java/Scala APIs, Kafka-python; Kafka Connect: Source/Sink Connectors, Configuration Basics; Schema Management: Avro Schemas, Schema Registry, Schema Evolution for Data Consistency

Unit III: Kafka Stream Processing – 12 Hours

Kafka Streams: Stream Processing, Windowing, Joins, State Stores, Aggregations; KSQL: SQL-Based Stream Processing, Querying Streams and Tables; Fault Tolerance: Replication, Offset Management, Consumer Group Rebalancing.

Unit IV: Spark Structured Streaming and Kafka Integration – 13 Hours

Spark Structured Streaming with Kafka: Reading/Writing Kafka Topics, Joins, Aggregations; Advanced Streaming: Windowing, Watermarking, Stateful Processing, Delta Lake for Transactional Consistency; Testing and Metrics: Unit Testing with PySpark and ScalaTest, Spark UI Metrics for Throughput and Latency

Unit V: Production Streaming Pipelines – 14 Hours

Deployment: Spark on YARN/Kubernetes, Kafka Cluster Scaling, Driver/Worker Configurations; Monitoring and Observability: Prometheus and Grafana for Kafka/Spark Metrics, Alerting; Streaming Data Ops: Schema Validation with Avro, CI/CD with Jenkins, Data Lineage, Delta Lake Versioning

References:

- 1."Kafka: The Definitive Guide" by Gwen Shapira, Todd Palino, Rajini Sivaram, and Krit Petty (O'Reilly) 2nd Edition 2021
- 2."Learning Spark: Lightning-Fast Data Analytics" by Jules S. Damji, Brooke Wenig, Tathagata Das, and Denny Lee (O'Reilly) 2nd Edition 2020
- 3."Spark: The Definitive Guide" by Bill Chambers and Matei Zaharia (O'Reilly) 1st Edition 2018
- 4."Streaming Systems" by Tyler Akidau, Slava Chernyak, and Reuven Lax (O'Reilly) 1st Edition 2018

Course Outcomes:

Unit	Course Outcome	Bloom's Level
I	CO1: <i>Explain</i> the architecture, components, and basic operations of Apache Kafka and Apache Spark.	Understand
II	CO2: <i>Apply</i> basic Kafka operations such as creating topics, producing and consuming messages using CLI and Kafka Connect.	Apply
III	CO3: <i>Apply</i> the integration of Kafka with Spark and HDFS using connectors and validate data flow across components.	Apply
IV	CO4: <i>Analyze</i> data transformation and streaming workflows using Spark SQL, Spark Streaming, Spark ML, and Spark configuration.	Analyze
V	CO5: <i>Evaluate</i> and manage data engineering pipelines using Spark applications with HDFS, Hive, RDDs, and Parquet data formats.	Evaluate

CO\PO Mapping

CO \ PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	2	1	1	3	0	0	1	1
CO2	Apply	2	2	2	3	0	0	1	1
CO3	Apply	2	2	3	3	1	0	2	1
CO4	Analyze	2	2	3	3	1	1	2	1
CO5	Evaluate	3	2	3	3	1	1	2	1

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	2	2	1
CO2	3	3	1
CO3	3	3	2
CO4	3	3	2
CO5	3	3	2

MSP 206 - Hands On – Elective – I (MLOps)

(Semester-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – I (MLOps)	Course Code	MSP 206
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

TOOLS: Git, GitHub Actions, Jenkins, Docker, Kubernetes, MLflow, Kubeflow, DVC, Apache Airflow, Prometheus, Grafana, FastAPI, TensorFlow, PyTorch, AWS (Lambda, SageMaker), GCP (Vertex AI, AI Platform), Azure ML, Weights & Biases, TensorFlow Serving, TorchServe

1. Setting up GitHub Actions for CI/CD
2. Creating and Managing ML Pipelines using MLflow/Kubeflow
3. Building a ML model and versioning it using DVC
4. Data preprocessing and pipeline automation with Apache Airflow
5. Containerizing ML models with Docker
6. Deploying ML models using Flask/FastAPI and Docker
7. Using Kubernetes (K8s) for scalable model deployment
8. Implementing Model Monitoring and Logging with Prometheus/Grafana
9. CI/CD workflow for ML using Jenkins or GitHub Actions
10. Model deployment on AWS Lambda (serverless deployment)

11. Using GCP Vertex AI or AI Platform for model training and deployment
12. Orchestrating ML pipelines on Azure with Azure ML Studio or CLI
13. Implementing MLOps for edge deployment using TensorFlow Lite / ONNX
14. Creating reusable MLOps microservices (model, data, inference APIs)
15. Building a complete ML project with DataOps and MLOps integration

MSP 206 - Hands On – Elective – I (Developing Web Applications)

(Semester-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – I (Developing Web Applications)	Course Code	MSP 206
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools :Text editor, Web browser, Local server (e.g., XAMPP, WAMP), CSS editor, XML editor, PHP development environment, JavaScript development environment, MySQL database, PHPMyAdmin, Version control system (e.g., Git), Debugging tools, Code validator, FTP client, Image editor (e.g., Photoshop, GIMP)

1. Text formatting on Web Pages
2. Creating hyperlinks
3. Incorporate images
4. Designing with Style Sheets
5. Absolute and Relative Positioning
6. Animating objects
7. Benefits of XML
8. Data Sharing XML
9. DTD Building Blocks
- 10.Validation
- 11.PHP functions
- 12.Server-Side Processing
- 13.Web Application Development
- 14.Variables
- 15.Functions
- 16.Events

- 17.JS web APIs
- 18.JS JSON
- 19.Creating Databases & Tables
- 20.Inserting, Retrieving, Updating & Deleting data
- 21.User Accounts
- 22.PHP interfacing with MySQL
- 23.Connecting to MySQL
- 24.Decision Making Statement
- 25.JS Browser BOM
- 26.Executing SQL

MSP 207 - Hands On – Elective – II (Data Engineering with Python)

(Semester-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – I (Data Engineering with Python)	Course Code	MSP 207
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

TOOLS: Python, Pandas, NumPy, Matplotlib, Seaborn, MySQL, MongoDB, PyMongo, BeautifulSoup, Requests, NLTK, spaCy, re (Regular Expressions), JSON, XML

1. Reading and writing text and CSV files using Python
2. Working with binary files in Python
3. Automating file handling using Python os and os.path
4. Parsing and extracting data from HTML using BeautifulSoup
5. Text processing and tokenization using NLTK or spaCy
6. Handling JSON and XML data in Python
7. Writing and testing regular expressions for text matching
8. Creating and manipulating Pandas Series and DataFrames
9. Cleaning and transforming datasets using Pandas
10. Merging, joining, and reshaping datasets in Pandas
11. Handling missing and duplicate data using Pandas
12. Performing NumPy operations on tabular numerical data
13. Connecting to and querying a MySQL database using Python
14. Performing CRUD operations on MongoDB using PyMongo
15. Creating basic and advanced plots using Matplotlib and Pandas

MSP 207 - Hands On – Elective – II (Data Engineering using Kafka and Spark Structured Streaming)

(Semester-II of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – II(Data Engineering using Kafka and Spark Structured Streaming)	Course Code	MSP 207
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

TOOLS: Kafka, Zookeeper, Kafka CLI, Kafka Connect, Spark, Spark Shell, Spark SQL, Spark Streaming, Spark MLlib, HDFS, Hive, MapReduce, Hadoop, Parquet, GHArchive, Python, Scala

1. Setting up Kafka and producing/consuming messages using CLI
2. Creating and managing Kafka topics using Kafka CLI
3. Configuring and using Kafka Connect to produce messages
4. Validating data flow through Kafka Connect and Kafka Consumer Groups
5. Using Kafka HDFS Sink Connector to write data to HDFS
6. Creating Spark RDDs and performing basic transformations
7. Working with Spark DataFrames and Datasets
8. Writing and validating data to HDFS using Spark
9. Performing SQL queries on structured data using Spark SQL
10. Processing streaming data with Spark Streaming
11. Using Spark MLlib for basic machine learning tasks
12. Working with Spark configuration and sessions
13. Master-slave configuration setup in Spark standalone cluster
14. Analyzing GHArchive parquet data with Spark
15. Loading data incrementally to target tables and validating with Spark CLI

MSP 208 - Completion of MongoDB from MongoDB University

Department	Computer Science	Course Type	DSCC
Course Title	Completion of MongoDB from MongoDB University	Course Code	MSP 208
L-T-P	3-1-0	Credits	4
Contact Hours	60 hrs	Duration of SEE	-

Unit -1 -12 HRS

MongoDB and the Document Model - Overview of the Document Model, Data Types, Managing Databases, Collections, and Documents in Atlas, Data Relationships, Embedding and Referencing

MongoDB Data Modelling Intro - Introduction to Data Modelling, Types of Data Relationships, Modelling Data Relationships, Embedding Data in Documents, Referencing Data in Documents, Scaling a Data Model, Using Atlas Tools for Schema Help

Unit -2 -12 HRS

Connecting to a MongoDB Database-Using MongoDB Connection Strings, Connecting to a MongoDB Atlas Cluster with the Shell, Connecting to a MongoDB Atlas Cluster with Compass, Connecting to a MongoDB Atlas Cluster from an Application, Troubleshooting MongoDB Atlas Connection Errors

MongoDB CRUD Operations – (CREATE, READ, UPDATE, DELETE)

Inserting Documents, Finding Documents, Finding Documents by Using Comparison Operators, Logical Operators, Querying on Array Elements, Replacing a Document Updating MongoDB Documents by Using - updateOne(), findAndModify(), updateMany() , Deleting Documents, Sorting and Limiting, Returning Specific Data , Counting Documents

Unit -3 – 12 HRS

MongoDB Aggregation - Introduction to MongoDB Aggregation, Aggregation Pipeline stages - \$match and \$group, \$sort and \$limit, \$project, \$count, and \$set, \$out Stage.

MongoDB Indexes - Single Field Index, Multikey Index, Compound Indexes, Deleting MongoDB Indexes

Unit -4 – 14 HRS

Connecting to MongoDB in Python-Using MongoDB Python Client Libraries, Connecting to an Atlas Cluster in Python Applications, Troubleshooting a MongoDB Connection in Python Applications

MongoDB CRUD Operations in Python - MongoDB Documents in Python, Inserting a Document, Querying a MongoDB Collection, Updating Documents, Deleting Documents in Python Applications, Creating MongoDB Transactions in Python Applications

MongoDB Aggregation in Python - Using MongoDB Aggregation Stages with Python: \$sort and \$project, \$match and \$group, \$sort and \$project, Building a MongoDB Aggregation Pipeline in Python Applications

Unit -5 – 10 HRS

MongoDB Atlas Search - Using Relevance-Based Search and Search Indexes, Creating a Search Index with Dynamic Mapping & Static Field Mapping, Using \$search and Compound Operators, Group Search Results by Using Facets

MongoDB Transactions – ACID Transactions and usage, MongoDB Compass, PyMongo Arrow, MongoDB VS Code Extension

SEMESTER 3

MSP 301 - Artificial Intelligence and Deep Learning

(Semester-III of MS Programme)

Department	Computer Science	Course Type	MSP
Course Title	Artificial Intelligence and Deep Learning	Course Code	MSP 301
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: – 14 Hours

Intro to AI - Basics of Artificial Intelligence, Applications of AI, Types of AI, AI agents, Search Algorithms- Logical representation of Knowledge –Intro to Machine learning- Machine learning Algorithms- Artificial Neural Networks, Fuzzy Logic, Introduction to Genetic Algorithm, Soft Computing.

Unit II: – 16 Hours

Intro to Deep Learning - Fundamentals of Deep Learning, Architecture of Deep Learning, Types of Deep Learning Networks – FFNN, RNN, CNN, Restricted Boltzmann Machine, Autoencoders, Deep Learning Applications, Limitations, Advantages, Disadvantages. Processing Sequences using RNNs & CNNs -

Introduction to CNNs and their Architecture, Convolution, Pooling, Padding, and Strides, Popular CNN Architectures: LeNet, AlexNet, VGG, ResNet, Transfer Learning with Pre-trained Models, Data Augmentation and Regularization, Image Classification using CNNs

Unit III: – 10 Hours

Recurrent Neural Networks - RNNs predict the future, the problems they face, like limited short-term memory, and solutions to these problems- LSTM (Long Short-Term Memory) and GRU cells, NLP Concepts – NLP Techniques.

Unit IV: – 10 Hours

Introduction to AI Networks with Keras - Biological neurons, multilayer perceptrons, and back-propagation, multilayer perceptron using Keras, and visualize the runs and graphs using Tensor board, Training Deep Neural Networks - Reusing pre-trained layers, using faster optimizers, and avoiding overfitting by regularization

Unit V: – 10 Hours Introduction to Gen AI - History and Evolution of Generative Models - Difference between Discriminative and Generative Models - Applications of Generative AI in Text, Image, Audio, and Video- Ethical Considerations in Generative AI - Introduction to GANs and Adversarial Learning- Types of GANs: DCGAN, CycleGAN, StyleGAN- Introduction to Self-Attention and Transformers.

References:

1. Artificial Intelligence: A Modern Approach – Stuart Russell & Peter Norvig ,Publisher: Pearson, Edition: 4th Edition (2020)
2. Deep Learning – Ian Goodfellow, Yoshua Bengio, Aaron Courville, Publisher: MIT Press, Edition: 1st Edition (2016)
3. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Aurélien Géron, Publisher: O'Reilly Media, Edition: 2nd Edition (2019)
4. Deep Learning with Python – François Chollet, Publisher: Manning Publications , Edition: 2nd Edition (2021)
5. Generative Deep Learning – David Foster, Publisher: O'Reilly Media Edition: 2nd Edition (2022)

Course Outcomes:

Unit	Course Outcome (CO)	Bloom's Level
I	CO1: <i>Understand</i> the foundational concepts of Artificial Intelligence, Machine Learning, and Soft Computing.	Understand
II	CO2: <i>Apply</i> deep learning architectures like CNNs and RNNs for image and sequence data processing.	Apply
III	CO3: <i>Analyze</i> the use of LSTM, GRU, and NLP techniques in solving real-world sequence prediction problems.	Analyze
IV	CO4: <i>Create</i> and train deep neural networks using Keras and Tensor Board for performance tuning and regularization.	Create
V	CO5: <i>Evaluate</i> generative models like GANs and Transformers and assess their ethical implications in modern AI applications.	Evaluate

CO\PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	3	2	2	2	1	1	1	1
CO2	Apply	2	2	3	3	1	1	1	0
CO3	Analyze	2	3	3	3	2	1	2	2
CO4	Create	2	2	3	3	1	1	2	2
CO5	Evaluate	2	3	3	3	2	2	2	3

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	3	2	2
CO2	3	2	1
CO3	3	2	1
CO4	3	3	1
CO5	3	2	2

MSP 302 - Cloud Computing - AWS / Azure / GCP

(Semester-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Cloud Computing - AWS / Azure / GCP	Course Code	MSP 302
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit-I: – 3 Hours

Basics of Cloud Computing, AWS Overview, Identity and Access Management (IAM)

Unit II:– 18 Hours

Introduction to AWS Compute Services, AWS Elastic Block Store (EBS) & Storage Management, AWS Load Balancing, Auto Scaling in AWS, Monitoring & Observability with CloudWatch

Unit-III :– 14 Hours

AWS Storage Services – Amazon S3, AWS Application Services, AWS Networking – Virtual Private Cloud (VPC), AWS Routing & Secure File Transfer, AWS Compute & Serverless Technologies

Unit-IV:– 15 Hours

Introduction to Google Cloud Platform (GCP), GCP Core Building Blocks, GCP Services – Databases & Data Analytics, GCP AI & Machine Learning Services, Developer Tools & Enterprise Services

Unit-V : – 10 Hours

Introduction to Microsoft Azure, Azure Storage Services, Compute Services in Azure, Azure Networking Services, Azure Application Services

References:

1. “Cloud Computing for Dummies” by Judith S. Hurwitz, Fern Halper, Robin Bloor, Marcia Kaufman
2. “AWS: The Complete Beginner’s Guide” by Stephen Barcon

Course Outcomes:

Unit	Course Outcome (CO)	Bloom's Level
I	CO1: <i>Understand</i> the fundamentals of cloud computing and the basic services offered by AWS including IAM.	Understand
II	CO2: <i>Apply</i> AWS compute, storage, and monitoring services to set up and manage scalable infrastructure.	Apply
III	CO3: <i>Analyze</i> AWS networking, file transfer, and serverless deployment services to support efficient cloud operations.	Analyze
IV	CO4: <i>Evaluate</i> and compare cloud services provided by GCP, focusing on data analytics and machine learning capabilities.	Evaluate
V	CO5: <i>Demonstrate</i> the integration of Microsoft Azure's storage, compute, and application services in building cloud solutions.	Apply

CO\PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	2	2	2	3	1	1	1	0
CO2	Apply	2	2	3	3	1	1	1	0
CO3	Analyze	2	3	3	0	1	1	2	2
CO4	Evaluate	2	3	3	3	2	2	2	2
CO5	Apply	2	2	3	3	1	1	2	2

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	1	3	1
CO2	2	3	1
CO3	2	3	1
CO4	3	2	1
CO5	2	3	1

ELECTIVES - III

MSP 303 A- Data Science with PySpark

(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Science with PySpark	Course Code	MSP 303 A
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Foundations of Data science with PySpark – 10 Hours

Role of Distributed systems in Data Science, Pyspark Fundamentals- Architecture, RDD Basics, Data frames ,Basic data operations- loading data, transformations, actions

Unit II: Data preprocessing and exploration with Pyspark – 12 Hours

Data cleaning , Data transformation, EDA with Pyspark: Summary statistics, Distributions, Anomaly Detection, Visualization Advanced data preparation techniques.

Unit III: Machine learning with Pyspark MLlib – 16 Hours

Introduction to MLlib, Pipelines, Vector Assembler, Supervised Learning: Regression- Linear, Random Forest Regressor, GBT Regressor, Classification: Logistic Regression, Decision Tree Classifier, GBT Classifier, Linear SVC, Naive Bayes, Unsupervised Learning: Clustering- Kmeans, Model Evaluation: Cross-validation, precision/recall, ROC curves, Dimensionality reduction- PCA, Market basket analysis with clustering.

Unit IV: Advance Pyspark Analytics -14 Hours

Graph Analytics- Graph frames, Page Rank algorithm, NLP Spark NLP: Tokenization, Lemmatization, Sentiment Analysis, Time Series Analysis, Recommended Engine- Collaborative vs content-based filtering, ALS, Evaluation

Unit V: Data Science Workflows and Deployment with PySpark – 8 Hours

Workflow Orchestration- Pyspark with Airflow, Running scripts on clusters, Automating ML pipelines, Visualization and Reporting: Advanced plotting - Plotly, Dashboards: Exporting to PowerBI, Model deployment, Optimization

Reference Books

1. "Learning PySpark" (Drabas/Lee) 1st Edition 2017
2. "Spark: The Definitive Guide" (Chambers/Zaharia) 1st Edition 2018
3. "Advanced Analytics with Spark" (Sandy Ryza) 2nd Edition 2017
4. "Advanced Analytics with Spark" (Sandy Ryza)", "Natural Language Processing with Spark NLP" (Thomas) 1st Edition 2020
5. "Data Pipelines with Apache Airflow" (Harensrak/de Ruiter) 1st Edition 2021

Course Outcome:

Unit	Course Outcome (CO)	Bloom's Level
I	CO1: Understand distributed computing concepts and the	Understand

	architecture of PySpark for scalable data analysis.	
II	CO2: <i>Understand and apply</i> data cleaning, transformation, and exploratory analysis techniques using PySpark DataFrames and functions.	Understand/ Apply
III	CO3: <i>Apply</i> machine learning models using PySpark MLlib for solving classification, regression, and clustering problems.	Apply
IV	CO4: <i>Analyze</i> graph data, natural language content, time-series data, and recommender systems using specialized PySpark libraries.	Analyze
V	CO5: <i>Create</i> and deploy complete data science workflows integrating PySpark, Airflow for orchestration, and Power BI for visualization.	Create

CO\PO Mapping

CO\PO	Bloom's Level	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Apply	3	2	2	3	1	1	1	1
CO2	Understand/ Apply	3	3	3	3	0	1	2	2
CO3	Analyze	3	3	3	3	2	2	2	2
CO4	Create	3	3	3	3	2	2	2	1
CO5	Apply	3	3	3	3	2	2	3	2

CO\ PSO Mapping

COs \ PSOs	PSO1	PSO2	PSO3
CO1	3	2	1
CO2	3	2	1
CO3	3	2	1
CO4	3	2	1
CO5	3	3	2

303 B-AI and Deep Learning through TensorFlow

(Elective-III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	AI and Deep Learning through TensorFlow	Course Code	MSP 303 B
L-T-P	4-0-0	Credits	4
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	70	CIE Marks	30

Unit I: Intro to AI – 10 Hours

Basics of Artificial Intelligence, Applications of AI, Types of AI, AI agents, Artificial Neural Networks, Fuzzy Logic, Introduction to Genetic Algorithm, Soft Computing.

Unit II: Intro to Deep Learning – 12 Hours

Fundamentals of Deep Learning, Architecture of Deep Learning, Types of Deep Learning Networks – FFNN, RNN, CNN, Restricted Boltzmann Machine, Autoencoders, Deep Learning Applications, Limitations, Advantages, Disadvantages.

Unit III: TensorFlow – 12 Hours

What is TensorFlow, Installation Through pip, Installation Through conda, Architecture of TensorFlow, Advantage & Disadvantage of TensorFlow, Basics of TensorFlow – Representation of a Tensor, Types of Tensors, TensorFlow Perception – Single Layer, Hidden Layer, Multilayer Perceptions.

Unit IV: ANN in TensorFlow – 14 Hours

Intro, Implementation of NN, Classification of NN, Linear Regression, CNN in TensorFlow – Intro, Working of CNN, Training of CNN, MNIST Dataset in CNN, RNN in TensorFlow – Intro, Working of RNN, RNN Time Series, LSTM RNN in TensorFlow.

Unit V: Case Studies – 12 Hours

Case 1: CNN for Digit Recognition

Case 2: CNN for Breast Cancer Detection

Case 3: CNN for Predicting the Bank Customer Satisfaction

Case 4: CNN for Credit Card Fraud Detection

Case 5: RNN - LSTM for IMDB Review Classification

Case 6: Google Stock Price Prediction with RNN and LSTM

References:

1. "Hands-on Machine learning with Scikit-Learn, Keras, and TensorFlow" by Aurelion Geron, Third Edition
2. "Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and More with TensorFlow 2 and the Keras API" by Amita Kapoor, Antonio Gulli, and Sujit Pal. , Second Edition

Course Outcomes:

CO	Course Outcome	Bloom's Taxonomy Verbs
CO1	Understand the fundamentals of Artificial Intelligence (AI), AI	Understand,

	agents, artificial neural networks, and applications of AI.	
CO2	Apply Deep Learning techniques, including FFNN, RNN, CNN, Restricted Boltzmann Machines, and Autoencoders, to solve real-world problems.	Apply, Analyze
CO3	Understand the fundamentals of TensorFlow, its architecture, and how it can be used for various machine learning tasks.	Understand
CO4	Implement Artificial Neural Networks (ANN), CNN, and RNN in TensorFlow for different applications like image recognition and time series forecasting.	Apply, Analyze
CO5	Design and deploy case studies for real-world applications using CNN and RNN, such as digit recognition, breast cancer detection, and stock price prediction.	Create, Apply

CO\PO Mapping

CO	Bloom's Verbs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand,	3	3	2	2	0	0	1	1
CO2	Apply, Analyze	3	3	3	3	1	0	2	2
CO3	Understand	3	3	2	2	0	1	1	1
CO4	Apply, Analyze	3	3	3	3	1	2	2	2
CO5	Create, Apply	3	3	3	3	1	2	2	2

CO\ PSO Mapping

CO	PSO1	PSO2	PSO3
CO1	2	1	1
CO2	3	3	2
CO3	3	2	1
CO4	3	3	2
CO5	3	3	3

ELECTIVE-IV

MSP 304 A-Data Engineering using AWS Data Analytics

(Elective-IV of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering using AWS Data Analytics	Course Code	MSP 304 A
L-T-P	4-0-0	Credits	4
Contact	60 Hrs	Duration of SEE	3 Hrs

Hours			
SEE Marks	70	CIE Marks	30

Unit I – Data Engineering Fundamentals - 8 Hours

Types of Data ,Data Engineering Process ,Responsibilities of a Data Engineer ,Types of Data Sources Data Engineering ,Tools and Skills ,Data Engineering vs Data Science, Roles and Responsibilities of a Data Engineer.

Unit II – AWS Data Engineering Concepts- 10 Hours

Introduction to AWS Data Engineering ,The AWS Data Engineer's Toolkit ,AWS Data Engineering Tools , Data Management Architectures for Analytics ,Data Cataloguing, Security and Governance

Unit III – AWS Technical Essentials - 14 Hours

AWS Infrastructure & Free Tier ,IAM, EC2, EBS, Security Groups ,AMI, Load Balancing, Auto Scaling ,CloudWatch, S3 Buckets ,AWS SFTP – Server Creation, User Setup, File Transfer ,AWS EKS ,AWS Lambda Function

Unit IV – Data Pipeline Engineering in AWS - 14 Hours

Identifying and Enabling Data Consumers ,Loading Data into a Data Mart ,Orchestrating the Data Pipeline ,Architecting Data Engineering Pipelines ,Ingesting Batch and Streaming Data ,Transforming Data to Optimize for Analytics

Unit V – Advanced AWS Data Engineering & Analytics - 14 Hours

Ad Hoc Queries with Amazon Athena Visualizing Data with Amazon QuickSight, Enabling AI and ML , Ingesting Data via REST API and SFTP , Ingesting Data into AWS RDS (PostgreSQL) , Incremental Data Loading

References:

Data Engineering with AWS by Gareth Eagar Packt Publishing.

Course Outcomes:

CO	Course Outcome	Bloom's Taxonomy Verbs
CO1	Understand the fundamentals of data engineering, including types of data, data processes, responsibilities, and tools used in data engineering.	Understand

CO2	Apply AWS data engineering tools and services, including data cataloging, governance, and security, to manage and analyze data on AWS.	Apply
CO3	Understand and analyze essential AWS technical services such as EC2, S3, IAM, and Lambda, to support data engineering workflows and cloud architectures.	Understand, Analyze
CO4	Create and apply data pipelines on AWS, including data ingestion, transformation, and optimization for analytics.	Create, Apply
CO5	Evaluate and apply AWS data engineering solutions for querying data with Amazon Athena, visualizing data with QuickSight, and enabling AI/ML applications, along with data ingestion methods.	Evaluate, Apply

CO\PO Mapping

CO	Bloom's Verbs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	3	3	2	2	1	1	2	1
CO2	Apply	3	3	3	3	2	1	2	2
CO3	Understand, Analyze	3	3	3	2	1	2	2	2
CO4	Create, Apply	3	3	3	3	2	2	3	2
CO5	Evaluate, Apply	3	3	3	3	2	3	3	2

CO\ PSO Mapping

CO	PSO1	PSO2	PSO3
CO1	3	2	1
CO2	3	3	2
CO3	3	2	1
CO4	3	3	2
CO5	3	3	3

MSP 304-B Data Engineering with Google BigQuery & Google Cloud

(Elective-IV of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Data Engineering with Google Big Query & Google Cloud	Course Code	MSP 304 B
L-T-P	4-0-0	Credits	4
Contact	60 Hrs	Duration of SEE	3 Hrs

Hours			
SEE Marks	70	CIE Marks	30

Unit I: Introduction to Data Warehousing and Big Query – 9 Hours

Introduction to Data Warehousing: Concepts, Lifecycle, Characteristics, Architecture, Components; Introduction to Big Query: Architecture, Data Storage, Execution Engine, Big Query vs. Modern Alternatives (Spark, Snowflake); Big Query Machine Learning: Introduction to Big Query ML for Predictive Analytics.

Unit II: BigQuery Operations and Data Processing – 12 Hours

BigQuery Fundamentals: Data Types, Functions, Query Execution, Data Type Conversion; Data Processing with Dataflow: Cleaning, Transforming, Storing Data in BigQuery; Data Quality Validation: Using BigQuery SQL for Data Quality Checks; Workflow Orchestration: Scheduling Queries with Cloud Composer; Data Preparation: Cleaning and Transforming with Cloud Data prep

Unit III: Data Loading into BigQuery – 12 Hours

Loading Data into BigQuery: Basics, Error Handling, Efficient Loading Techniques; Data Ingestion Patterns: Batch and Streaming Loads, Recurring Loads; BigQuery Optimization: Partitioning and Clustering for Load and Query Performance; Data Sources: Loading from Databases, Google Drive, Federated Queries, and BigQuery Streaming API

Unit IV: Data Visualization and Querying – 13 Hours

Data Visualization: Looker Studio for Interactive Dashboards, BigQuery Integration; Advanced Querying: Joins (Theory and Hands-on), Unions, Intersections, Subqueries; Advanced Analytics: Time-Series Analysis, Geospatial Functions, Statistical Aggregations

Unit V: BigQuery and Google Cloud Integration – 14 Hours

BigQuery Query Essentials: DDL, DML, Query Optimization; Advanced BigQuery Integrations: BigQuery Storage Write API, Advanced Dataflow Optimizations for ETL/ELT, Vertex AI for ML; ETL and ELT Workflows: Design, Implementation, and Optimization in BigQuery; BigQuery Security and Auditing: Column-Level Security, Audit Logs, IAM, and Encryption

References:

1. "Google BigQuery: The Definitive Guide" by Valliappa Lakshmanan and Jordan Tigani (O'Reilly) 1st Edition 2019
2. Data Engineering with Google Cloud Platform" by Adi Wijaya (Packt) 2nd Edition 2024

Course Outcomes:

CO	Course Outcome	Bloom's Taxonomy
CO1	Understand the concept of Data Warehousing by describing its life cycle, characteristics, and architecture; explain BigQuery architecture, data	Understand

	storage, and execution while comparing it with the MapReduce processing model.	
CO2	Apply knowledge of BigQuery data types and functions to construct efficient queries; analyze data flow by improving data quality, transforming data, and visualizing it using Data Prep and scheduling queries effectively.	Apply, Analyze
CO3	Demonstrate the process of loading data into BigQuery by handling errors, designing efficient data load strategies, and implementing recurring loads within a microservice architecture.	Apply, Analyze
CO4	Use Google Sheets, BigQuery, and Data Studio to visualize and interpret data; perform advanced queries using joins, unions, and intersections; apply basic statistical functions.	Apply
CO5	Develop DDL and DML statements in BigQuery; illustrate the workflow of ETL, EL, and ELT processes; integrate BigQuery with other Google Cloud services for data pipeline management.	Create

CO\PO Mapping

CO	Bloom's Verbs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	Understand	3	3	1	2	0	0	0	1
CO2	Apply, Analyze	3	3	3	3	1	0	2	1
CO3	Apply, Analyze	3	3	3	3	0	1	2	1
CO4	Apply	3	2	3	3	1	0	2	0
CO5	Create	3	3	3	3	1	2	2	2

CO\PSO Mapping

CO	PSO1	PSO2	PSO3
CO1	2	2	1
CO2	3	2	2
CO3	2	3	2
CO4	3	2	2
CO5	3	3	2

OEC 305 - Artificial Intelligence in Drug Discovery and Development

Department	Computer Science	Course Type	DSCC
Course Title	Artificial Intelligence in Drug Discovery and Development	Course Code	OEC 305
L-T-P	3-1-0	Credits	3

Contact Hours	60 Hrs	Duration of SEE	-
SEE Marks	-	CIE Marks	-

Unit I: Foundations of Drug Discovery & AI Integration

Motivating Applications of AI in Drug Discovery, Overview of Drug Discovery Pipelines and Challenges, Conventional vs. AI-Driven Workflows, History and Evolution of AI in Drug Development, Key AI Technologies and Tools, Advantages of AI Integration.

Unit II: Core AI/ML Techniques for Drug Discovery

Introduction to Machine Learning and Neural Networks, Feature Engineering and Data Preprocessing, Evaluation Metrics for AI Models, Python Libraries for Drug Discovery (e.g., RDKit, DeepChem), Hands-on: Molecular Structure Representation and Data Handling.

Unit III: AI in Early-Stage Drug Discovery

Target Identification and Validation with Omics Data, AI for Protein Structure Prediction and Binding Site Analysis, High-Throughput Virtual Screening (Molecular Docking, QSAR), Lead Optimization Strategies, Hands-on: AI-Assisted Molecular Docking and Dynamics.

Unit IV: AI in Clinical Development & Advanced Design

ADMET Predictive Modeling and Optimization, AI in Clinical Trials (Patient Stratification, Outcome Prediction), Generative AI for De Novo Drug Design (GANs, VAEs, GNNs), Hands-on: AI-Enabled ADMET Prediction and Molecule Generation.

Unit V: Advanced Applications & Future Directions

AI in Precision Medicine and Drug Repurposing, Network Pharmacology and Multi-Target Discovery, Case Studies of AI Successes, Regulatory and Ethical Challenges, Future Trends (XAI, Federated Learning), Hands-on: End-to-End AI Workflow for Virtual Screening.

References:

1. Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more. "O'Reilly Media, Inc."
2. Brown, N. (Ed.). (2020). Artificial intelligence in drug discovery. Royal Society of Chemistry.

OEC 305 - Responsible & Safe AI Systems

Department	Computer Science	Course Type	DSCC
Course Title	Responsible & Safe AI Systems	Course Code	OEC 306
L-T-P	3-1-0	Credits	3
Contact Hours	60 Hrs	Duration of	3 Hrs

		SEE	
SEE Marks	-	CIE Marks	-

Unit I: AI Capabilities, Risks, and Responsible AI (RAI) Principles

AI Progress & Risks: Breakthroughs in AI (last 5–10 years); Imminent Risks (bias, toxicity, adversarial attacks); Long-term Risks (misuse, rogue AGI). RAI Foundations: Principles—Transparency, Accountability, Safety, Fairness, Privacy; Human-Centered Values; Interpretability.

Unit II: AI Safety & Interpretability

Attacks & Defenses: ML Poisoning (Trojans); Adversarial Attacks (Vision/NLP); Implications for Safety. Interpretability: Mechanistic Interpretability; Representation Engineering; Model Probing/Editing; Critiques of Transparency.

Unit III: Privacy, Fairness, and Regulation

Ethics & Compliance: Bias/Fairness Metrics; Explainability Tools (LIME/SHAP); Audit Mechanisms. Regulatory Landscape: GDPR, EU AI Act, DPDP (India), US Policies; Ethical Approvals; Participatory Design; Future of Work.

Unit IV: RAI in Critical Domains & AGI Governance

Domain Applications: Legal, Healthcare, Education; Policy Challenges. AGI Risks: Definition/Timelines; Instrumental Convergence (Power Seeking, Deception); Governance Frameworks.

Unit V: Practitioner Insights & Emerging Challenges

Engagement: Panel Discussions (Industry/Academia/Government); Fireside Chats. Advanced Topics: Recorded Paper Readings; Future Directions (XAI, Global Standards).

References:

1. Yampolskiy, R. V. (Ed.). (2018). Artificial Intelligence Safety and Security. CRC Press.
2. Kearns, M., & Roth, A. (2019). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford University Press.
3. Tegmark, M. (2017). Life 3.0: Being Human in the Age of Artificial Intelligence. Knopf.

MSP 306 Hands On – Elective – III (Data Science with PySpark)

(Semester III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – III (Data Science with PySpark)	Course Code	MSP 306

L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools

Google Colab

Pyspark installation on google colab or local environment

- Programming with RDD: Foreach Loop
- Using Reduce Function
- Viewing Records from MySQL
- Pyspark Joins Examples
- Word Count.
- Data Frames: Read, Write
- Explore Clean, Manipulate, Join, Aggregate.
- Model Selection and Tuning in Mllib
- Train & Test Models [Multilayer PC]
- Train & Test Models [Naive Bayes]
- Train & Test Models [Linear SVM]
- Train & Test Models [Decision Tree]
- Train & Test Models [Random Forest].
- Feature Transformers, Feature Extractors
- Regression in Pyspark Mllib
- Linear Regression
- Decision Tree Regression
- Random Forest Regression
- Gradient Boosted Tree Regression.
- K-Means & Bisecting K-Means in Mllib code
- Gaussian Mixture Modeling in Mllib
- Frequent Pattern Mining in Mllib

MSP 306 Hands On – Elective – III (AI and Deep Learning through TensorFlow)

(Semester III of MS Programme)

Department	Computer Science	Course Type	DSCC
Course Title	Hands On – Elective – III (AI and Deep Learning through TensorFlow)	Course Code	MSP 306
L-T-P	0-0-4	Credits	2

Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools: Python Programming, Jupyter Notebook, TensorFlow, Keras, PyTorch, Scikit-Learn, NumPy & Pandas, Matplotlib & Seaborn, OpenCV, NLTK, spaCY

1. Implementing a Simple AI Agent
2. Basic Artificial Neural Network
3. Fuzzy Logic Application
4. Genetic Algorithm Demo
5. Feedforward Neural Network (FFNN) Implementation
6. Recurrent Neural Network (RNN) for Sequence Prediction
7. Convolutional Neural Network (CNN) for Image Classification
8. Autoencoder Implementation
9. Tensor Operations
10. Single-Layer & Multi-Layer Perceptron in TensorFlow
11. Linear Regression with Neural Networks
12. CNN Implementation on MNIST Dataset
13. RNN for Time-Series Prediction
14. LSTM for Sequence Modelling

MSP 307 Hands On – Elective – IV (Data Engineering using AWS Data Analytics)
(Semester III of MS Programme)

Department	Computer Science	Course Type	DSCC
-------------------	-------------------------	--------------------	-------------

Course Title	Hands On – Elective – IV (Data Engineering using AWS Data Analytics)	Course Code	MSP 307
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools: AWS Account : AWS IAM, EC2, EBS, S3, AWS CLI, AWS Lambda, AWS SFTP, CloudWatch, AWS RDS (PostgreSQL), AWS Glue, AWS Glue Data Catalog, AWS Athena, Amazon QuickSight, AWS EKS, REST API

1. Exploring types of data and data sources used in data engineering
2. Setting up and managing IAM roles and policies in AWS
3. Launching and configuring EC2 instances with EBS volumes
4. Creating and using S3 buckets for data storage and management
5. Setting up AWS SFTP server and transferring files securely
6. Configuring CloudWatch for monitoring AWS resources
7. Creating and deploying Lambda functions for serverless compute
8. Working with AWS RDS (PostgreSQL) for structured data storage
9. Ingesting data using REST API and storing it in S3 or RDS
10. Building and orchestrating ETL pipelines with AWS Glue
11. Cataloguing and transforming data using AWS Glue Data Catalog
12. Loading batch and streaming data into AWS data lakes
13. Running ad hoc queries on S3 data using Amazon Athena
14. Creating dashboards and visualizations with Amazon QuickSight
15. Designing and deploying a complete data pipeline on AWS

MSP 307 Hands On – Elective – IV (Data Engineering with Google BigQuery & Google Cloud)

(Semester III of MS Programme)

Department	Computer Science	Course Type	DSCC
-------------------	-------------------------	--------------------	-------------

Course Title	Hands On – Elective – IV (Data Engineering with Google BigQuery & Google Cloud)	Course Code	MSP 307
L-T-P	0-0-4	Credits	2
Contact Hours	60 Hrs	Duration of SEE	3 Hrs
SEE Marks	35	CIE Marks	15

Tools: BigQuery, Google Cloud Platform (GCP), Google Cloud Console, Cloud Dataprep, Google Sheets, Google Data Studio, SQL, GCP IAM, GCP Storage, BigQuery DDL/DML, Python (optional), REST API (optional)

1. Exploring Data Warehouse architecture and life cycle
2. Understanding BigQuery architecture and comparing with MapReduce
3. Creating tables and querying data using BigQuery
4. Converting and transforming data types in BigQuery
5. Data cleaning and transformation using DataPrep (Cloud Dataprep)
6. Loading structured/unstructured data into BigQuery
7. Handling data load errors and optimizing performance
8. Scheduling recurring data loads in BigQuery
9. Integrating Google Sheets with BigQuery
10. Visualizing BigQuery data using Google Data Studio
11. Writing JOIN, UNION, and INTERSECT queries in BigQuery
12. Using BigQuery functions for statistical analysis
13. Writing DDL and DML statements in BigQuery
14. Building and testing an ETL/ELT pipeline in BigQuery
15. Integrating BigQuery with other Google Cloud services

QUESTION PAPER STRUCTURE FOR SEMESTER END EXAMINATIONS

The semester end university examination in each course and paper is of three hours duration with a maximum of 70 marks. The Question paper consists of two parts, A & B. Part A consists of 6 short answer questions and 5 to be answered. Part B consists of 10 questions with choice from each unit.

S V UNIVERSITY: TIRUPATI MS Data Science Course code:	
Time: 3hrs	Max. Marks: 70
PART A (5 X4 =20 MARKS) Answer any Five from the following Questions 1. a. b. c. d. e. f.	
PART B (5 X10 =50 MARKS) Answer any one question from each unit. Unit-I 2. Or 3. Unit -II	

4.

Or

5.

Unit-III

6.

Or

7.

Unit -IV

8.

Or

9.

Unit -V

10.

Or

11.



QUESTION PAPER STRUCTURE FOR OPEN ELECTIVE(MOOCs)

SUPPLEMENTARY EXAMINATION

The Open Elective (MOOCs) Supplementary university examination is of three hours duration with a maximum of 70 marks. The Question paper consists of two parts, A & B. Part A consists of 6 short answer questions and 5 to be answered. Part B consists of 10 questions with choice from each unit.

S V UNIVERSITY: TIRUPATI	
MS Data Science	
Course code:	
Time: 3hrs	Max. Marks: 70
PART A (5 X4 =20 MARKS)	
Answer any Five from the following Questions	
1.	
a.	
b.	
c.	
d.	
e.	
f.	
PART B (5 X10 =50 MARKS)	
Answer any one question from each unit.	
Unit-I	
2.	
Or	
3.	
Unit -II	

4.

Or

5.

Unit-III

6.

Or

7.

Unit -IV

8.

Or

9.

Unit -V

10.

Or

11.

QUESTION PAPER STRUCTURE FOR SEMESTER END EXAMINATIONS

The semester end university examination in each course and paper is of three hours duration with a maximum of 50 marks. The Question paper consists of two parts, A & B. Part A consists of 6 short answer questions and 5 to be answered. Part B consists of 10 questions with choice from each unit.

S V UNIVERSITY: TIRUPATI MS Data Science Course code:	
Time: 3hrs	Max. Marks: 50
PART A (5 X2 =10 MARKS) Answer any Five from the following Questions 1. a. b. c. d. e. f.	
PART B (5 X8 =40 MARKS) Answer any one question from each unit. Unit-I 2. Or 3. Unit -II 4. Or 5. Unit-III 6. Or 7. Unit -IV 8. Or 9. Unit -V 10. Or 11.	

