



UNIVERSITY OF CALICUT

Abstract

General and Academic IV - Faculty of Science - The revised Regulation, scheme and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department (Project mode course), w.e.f 2026 admission - Approved by the Vice Chancellor - Orders Issued

G & A - IV - J

U.O.No. 11896/2026/Admn

Dated, Calicut University.P.O, 22.08.2026

*Read:-*1. U.O.No. 11726/2026/Admn dated: 20.08.2026

2. Item No. 3 of the minutes of the meeting of the Board of Studies in Computer Science (PG) held on 16.07.2026

3. Remarks of the Dean, Faculty of Science, dated: 05.08.2026

ORDER

1. Vide paper read as (1), the modified Regulation, scheme and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department (Project mode course), were implemented for 2024 & 2025 admissions.
2. Considering the request of the Head, Department of Computer Science, University of Calicut, the Board of Studies in Computer Science (PG) vide paper read (2) above, recommended to approve the revised Regulation, scheme and syllabus of "PG Diploma in Data Science and Analytics" w.e.f 2026 admission.
3. The Dean, Faculty of Science, vide paper read (3), approved the recommendation of the Board of Studies in Computer Science (PG)
4. Considering the urgency, sanction has been accorded by the Vice Chancellor to implement the above recommendation under section 10(13) of Calicut University Act 1975.
5. Hence, the revised Regulation, scheme and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department (Project mode course), are implemented w.e.f 2026 admission. (the revised Regulation, scheme and syllabus appended).

Ajayakumar T.K

Assistant Registrar

To

The Head, Department of Computer Science, University of Calicut.

Copy to: PS to VC/PA to PVC/ PA to Registrar/PA to CE/JCE I/JCE V/JCE VIII/EX and EG Sections/GA I F/SUVEGA/IQAC/SF/DF/FC

Forwarded / By Order

Section Officer

UNIVERSITY OF CALICUT
DEPARTMENT OF COMPUTER SCIENCE

**OBE Regulations, Course Structure, Scheme of Evaluation
and Detailed Syllabus for the Project mode Programme**

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

(Effective from 2026 Admission onwards)

UNDER THE FACULTY OF SCIENCE

© University of Calicut

REGULATIONS FOR THE PROJECT MODE PROGRAMME

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

With effect from 2026 Admission

1. Introduction

We are surrounded by a flood of data today. In a broad range of application areas, data is being collected at an unprecedented scale and a major portion of data is unstructured. Understanding from the data is generally made by taking decisions. Decisions on the data were previously made based on guesswork, or models of reality constructed carefully, can now be made based on the data itself. Such Big Data analysis now drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, physical sciences, education, etc. Web big data analytics brings tremendous research opportunities in the areas of data analysis, social networking data analysis, business data analysis, etc. A typical big data computing life cycle consists of moving from data to information, from information to knowledge, from knowledge to intelligence, and from intelligence to decision. The promise of data-driven decision-making is now being recognized broadly.

Data is the fuel that powers AI, and large data sets make it possible for machine learning applications (machine learning is a branch of AI) to learn independently and rapidly. The abundance of data we collect supplies our AIs with the examples they need to identify differences, increase their pattern recognition capabilities, and see the fine details within the patterns. AI enables us to make sense of massive data sets, as well as unstructured data that doesn't fit neatly into database rows and columns. AI is helping organizations create new insights from data that was formerly locked away in emails, presentations, videos, and images. Databases are becoming increasingly versatile and powerful. In addition to traditional relational databases, we now have powerful graph databases that are more capable of connecting data points and uncovering relationships, as well as databases that specialize in document management. Society is in the midst of what is being called the Fourth Industrial Revolution. Data analytics and AI have tipped the balance, fundamentally changing the way we do business, treat disease, interact with technology, and communicate with each other.

Data Science addresses the problem of knowledge extraction from structured and unstructured data, including very large data sets often referred to as Big Data. Due to the government and industry's pressing need for Data Science expertise, the market demand for data scientists has soared in recent years. In this context, the Department of Computer Science, University of Calicut has decided to offer this unique Project mode Programme entitled **PG DIPLOMA IN DATA SCIENCE AND ANALYTICS (PGDDSA)** by ensuring the expertise of the faculty members of the Department of Computer Science, Mathematics and Statistics of the University and the collaborating Institutes with a focus on upskilling the students so as to equip them to build a career in Data Science and related areas. This course is offered at the 500–599 academic level.

1.1 Programme Outcomes

At the end of the Programme, the student will be able to:

- **PO-1:** Apply the knowledge of mathematics, statistics, and computer science to solve the problems related to data science and analytics.
- **PO-2:** Identify, formulate, review research literature, and analyze problems reaching validated conclusions.
- **PO-3:** Design algorithms and devise solutions for difficult problems to meet the specified needs with appropriate consideration for the society.
- **PO-4:** Create and apply appropriate techniques and use research based knowledge and methods to conduct investigations on complex problems and provide valid conclusions in the area of data science and analytics.
- **PO-5:** Unify the students to take up a career in the highly competitive IT industry with research and development skills acquired through minor and major projects.
- **PO-6:** Equip students with comprehensive knowledge and understanding of advanced theoretical fundamentals and the contemporary key research issues in specialized areas of computational intelligence and data analytics.

1.2 Duration of the Programme:

The duration of the PG **DIPLOMA IN DATA SCIENCE AND ANALYTICS** Programme shall be 1 year spread over 2 Semesters. Each semester shall have at least 18 weeks. The maximum duration permissible for completing

the Programme is fixed as 2 Years.

1.3 Number of Seats:

The Programme has 25 seats per Semester and the students will be admitted by strictly adhering to the community reservation as per the University norms.

1.4 Eligibility for admission:

i. Bachelor's Degree in Computer Science/Computer Application/Engineering/Technology or Equivalent Degree or any other University / Institution, recognized by this University as equivalent thereto, with a minimum aggregate of 55% marks or equivalent grade.

OR

B.Sc., BBA, B. Com, B.Voc., or any equivalent undergraduate degree from a recognized University/Institution with a minimum aggregate of 55% marks or equivalent grade are eligible to apply. The candidate must have studied Mathematics at the 10+2 (Higher Secondary) level or must have studied Mathematics / Business Mathematics / Statistics / Quantitative Techniques or any equivalent quantitative subject at the undergraduate level.

ii. For SEBC and Physically Challenged candidates the aggregate marks required for the Qualifying Examination is 50%. For SC/ST candidates, a minimum pass in the Qualifying Examination is sufficient.

1.5 Selection Criteria

The selection to the Programme shall be based on the Entrance Examination conducted by the University of Calicut. The admission shall be based on the marks scored by a student in the qualifying examinations/ entrance examinations as notified by the university from time to time, unless otherwise decided by the university.

The duration of the Entrance Examination will be two hours comprising multiple choice questions from Computer Science, Mathematics, and Statistics and Quantitative Aptitude Logical Ability at the undergraduate level. The pattern of the question paper shall be as follows:

Table 1. Question Paper Pattern for Entrance Examination

Table - Question Paper Pattern for Entrance Examination			
Sl. No	Subject	Number of Questions	Marks
1.	Computer Science	20	20x2 = 40
2.	Mathematics	15	15x2 = 30
3.	Statistics	15	15x2 = 30
4.	Quantitative Aptitude Logical Ability	10	10x2 = 20
	Total		120
Note: Each correct response will be awarded TWO marks, and each wrong answer will be awarded Negative 0.555 marks			

However, if the number of applications received does not exceed three times the sanctioned intake, admission shall be granted directly on the basis of the marks obtained in the qualifying undergraduate (UG) examination, without conducting an entrance examination, subject to the University's rules and the approval of the competent authority.

2. Programme Structure

2.1 Attendance

The minimum requirement of attendance during a semester shall be 75% for male students for each course and an exemption of 3% is sanctioned to female students. Those who secure the minimum attendance in a semester alone will be allowed to register for the End Semester Examination. Condonation of attendance to a maximum of 10% in a semester will be granted by the Chairman of the Academic Committee on genuine grounds, provided it is also recommended by the Department Council. The benefit of condoning attendance will be granted to the students on health grounds, for participating in University Union activities, meetings of the university bodies, and participation in extracurricular activities on production of genuine supporting documents with the recommendation of the Head of the Department concerned. A student who is not eligible for condonation shall repeat the course with the subsequent batch.

2.2 Credit

One credit of a course is defined as a minimum of one hour of lecture or two hours of lab/tutorial per week for 18 weeks in a semester, with 15 instructional weeks

and 3 weeks allocated for examinations, including internal examinations, evaluation, and other academic activities. For practical-only courses, one credit is defined as a minimum of one hour per week. The minimum number of credits required to complete the Post Graduate Diploma in Data Science and Analytics (PGDDSA) programme is 48.

2.3 Seminar

Each student should select a relevant topic and prepare a seminar report, under the guidance of a faculty member as part of the internal assessment in both the semesters. Students should prepare an abstract of the topic and distribute it to faculty members in charge at least two weeks ahead of the seminar. Presentation shall be for a minimum of 30 minutes duration. The presentation and seminar report will be evaluated by an expert team consisting of at least two faculty members (Mark distribution: 50% for report and 50% for presentation and discussion).

2.4 Assignments

Each student shall be required to submit a minimum of two assignments for each course. The details such as the number of assignments, mark distribution and the weightage for each assignment will be announced by the faculty in charge of the course at the beginning of the semester.

2.5 Internal Tests

A minimum of two internal tests will be conducted for each course. The details such as the number of tests, mark distribution, and weightage for each test will be announced by the faculty in charge of the course at the beginning of the semester.

2.6 Case studies / Viva -Voce / Lab assignments

The faculty in charge of the course shall design the evaluation pattern based on one or more of these components which will be announced at the beginning of the semester.

2.7 Assessment And Evaluation

The examination has two modes, Continuous evaluation and End-semester Evaluation.

2.7.1 Continuous Assessment (CE):

1. 50% weight shall be given to the continuous (internal) evaluation: - Each semester has continuous evaluation with a particular criterion.

2. The teacher who is conducting the course should evaluate the student continuously during the semester in the specific format, prescribed by the BoS concerned. In the case of theory and practical, a dissertation/project separate format for continuous evaluation is prescribed.
3. The course teacher(s) shall maintain the academic record of each student registered for the course, which shall be forwarded to the Academic Committee through the Head of the Department/School.
4. The continuous assessment shall be based on a predetermined transparent system involving periodic written tests, viva-voce, seminars and attendance in respect of theory courses and based on written tests, viva-voce and lab skill/records in respect of practical courses as detailed below:

Table 2.Components of Continuous Assessment (Theory)

Sl.No	Component	Marks	Total
1.	Attendance*	5	50
2.	Assignment /Term Paper	10	
3.	Test Paper	25	
4.	Case Studies/Group Discussion/Viva-Voce	10	

(95% & above: 5 marks, 90 to 94.99%: 4 marks, 85to 89.99%: 3 marks, 80 to 84.99%:2 marks, 75 to 79.99%: 1 mark, below 75%: Nil)

Table 3.Components of Continuous Assessment (Practical)

Sl.No	Component	Marks	Total
1.	Lab Skill / Quality of Records	10	50
2.	Practical Test	30	
3.	Viva - Voce	10	

5. The details of executing the continuous assessment shall be decided by the concerned Departmental Council. To ensure transparency of the evaluation process, original answer scripts of the test paper shall be returned to the students after evaluation within a week of the conduct of the tests. Any dispute regarding the continuous assessment shall be taken up with the concerned teacher within 48 hours. The continuous assessment marks awarded to the students in each course in a semester shall be notified on the notice board at least one week before the

PG Diploma in Data Science and Analytics, With Effect from 2026 Admission.
commencement of the end semester examination.

6. No separate minimum is required for continuous assessment for a pass.

2.7.2. End Semester Evaluation (ESE)

50% weight shall be given to the end-of-semester examination: Each Semester has an end-of-semester examination after the completion of the syllabus prescribed for the semester.

2.7.3 Question paper setting and Evaluation

1. The End Semester examination in theory courses is to be conducted with the question papers set by a committee. The Department council or the Committee constituted by the Department Council is in charge of setting question papers for the end-semester examination. At least 3 sets of OBE model question papers set by the committee are scrutinized by the Chairman of the Department Council. The chairman of the question paper setting committee may be the Head of the Department or a faculty nominated by the Department council.
2. The evaluation of the answer scripts shall be done by the teacher offering the course and an expert, based on a well-defined scheme of valuation framed by them. If any course has no expert other than the faculty who is offering the course, the Chairman has the power to nominate an expert from the other Departments in the University or a qualified person in any other reputed institution as a special case. The remuneration for the external expert may be given as per the University rules framed by the Controller of Examinations.
3. If there is more than 15% difference in marks on double valuation, a third valuation shall be conducted by another faculty/expert. In that case, the average of the third valuation marks and the highest marks of the double valuations will be taken as the final marks. The answer scripts should be kept under the safe custody of the Chairman for a period of one year from the date of publishing the results.
4. Alternatively, the department council may opt for evaluation of the answer sheets by the teacher offering the course alone (single evaluation). In this case, the answer sheet must be shown to the respective student before the results are communicated to the Controller of Examinations. If the students are not satisfied with the result, they could submit the grievance to the chairman / HoD concerned. The chairman can take appropriate decisions including revaluation of the answer script by a faculty other than the one evaluated earlier.
5. Each course demands a minimum of 40% in the end-semester examination.
6. The End Semester practical examination shall be conducted and evaluated by

two or more examiners nominated by the Department Council.

7. The end semester examination answer scripts evaluation shall be done in a centralized valuation camp, to be held in the respective Departments/ Schools immediately after the examination, under the supervision of the Department Council /control of the Academic Committee if necessary. The chairman of the valuation camp should upload/convey the marks of all theory, practical and dissertation/projects to the Controller of Examination immediately after all the evaluation procedures and exam board approval. The Controller of Examination should publish the results online. It is desirable to have the semester results announced within 10 days of the last examination of the semester.
8. The students can apply for the Save a Year (SAY) Final Semester examination by applying within 7 days from the publication of results.
9. The SAY exams shall be conducted within 7 days from the final date of application for the SAY examination.
10. The Academic Committee is empowered to lay down the procedure for the conduct of examinations from time to time.
11. Moderation of marks may be applied during the process of result finalization by the Chairman constituted for the programme. The Examination Board, chaired by Head of the Department or a faculty nominated by the Department council and comprising all examiners involved in the valuation, shall review the consolidated marks and may recommend moderation where necessary to ensure fairness and uniformity in evaluation. The allotment and extent of moderation shall be as specified by the University of Calicut from time to time in accordance with the University regulations.

2.7.4 Practical Examinations:

Practical Examinations will be conducted internally by the Department. The details of the ESE for each practical course are specified as part of the syllabus for the course.

Table 4. End Semester Evaluation (50%)

Theory	Marks		Practical	Marks
Written Examination with double internal valuation	50		Experiments	30
			Record	10
			Viva	10

2.8 Final Project Work and Viva-voce

The Project work should be carried out over the period of 16 weeks in the Second

semester in an Industry / R&D organization / Department/Institution. If the project is carried out in an Industry / R&D organization outside the campus, then a co-guide shall be selected from the Department/ Institution concerned. Every student should do the Project individually and no grouping is allowed.

All the candidates are required to get the approval of their synopsis and the guide before commencement of the project from the Department / Institution and the matter may be intimated to the University at the beginning of the semester by the Department / Institution. The project will be reviewed periodically every month by the Department / Institutional. The continuous assessment marks (CA) will be based on the periodic progress and progress report.

2.8.1 Project Evaluation

The valuation scheme for the Project/ Dissertation: The valuation shall be jointly done by the supervisor of the project in the department and an internal expert from the approved panel, based on a well-defined scheme of valuation framed by them. If internal experts are not available in the department, the chairman has the power to nominate an expert from other departments in the university or any other reputed institution as a special case. The following break-up is suggested for the valuation. The Department Council/Board of Studies may decide on an alternative break-up, if required, specific to the discipline of study.

Table 5. Criteria for Project Evaluation

Sl.No	Particulars	Weightage (%)
1.	Review of Literature and Formulation of the Research Problem /Objective	20
2.	Methods and Description of the techniques used	15
3.	Analysis and Discussion of results	30
4.	Presentation of the report, organization, linguistics style, references etc.	15
5.	Viva Voce examination based on the Project work /Dissertation	20
	Total	100

2.8.2 Guideline for preparing project Report

1. Arrangement of contents:

The sequence in which the project report material should be arranged and bound should be as follows:

- i) Cover Page & Title Page
- ii) Bonafide Certificate
- iii) Abstract
- iv) Table of Contents
- v) List of Tables
- vi) List of Figures
- vii) List of Symbols, Abbreviations and Nomenclature
- viii) Chapters

The chapters may be broadly divided into 3 parts

- Introductory chapter,
- Chapters developing the main theme of the project work
- implementation details (if any) and Conclusion.

The main text will be divided into several chapters and each chapter may be further divided into several divisions and sub-divisions. Each chapter should be given an appropriate title. Tables and figures in a chapter should be placed in the immediate vicinity of the reference where they are cited. Footnotes should be used sparingly. They should be typed single space and placed directly underneath in the very same page, which refers to the material they annotate.

- ix) Appendices
- x) References

The tables and figures shall be introduced at appropriate places.

2. Page Dimension and binding specifications:

The dimensions of the project report should be in A4 size. The project report should be bound using a flexible cover of the thick white art paper. The cover should be printed in black letters and the text for printing should be identical. A softcopy of the report also to be submitted.

3. All the project reports submitted by the students should be plagiarism checked and the plagiarism report generated by the authorized software should be verified and signed by the Head of the Department or Faculty in-charge of the Project.

2.9 Pattern of question papers (ESA)

The question paper for the ESA (Theory) Examination shall consist of two parts as given below. The duration of the examination is 2 hours. 30 minutes.

General pattern of the Question Paper

Code:

Register No:

Name :

First Semester PG Diploma in Data Science and Analytics Examination – 2024

Course Code: (Eg: PGD 01 C01)

Course: (Eg: Mathematical and Statistical Foundation for Data Science)

Time: 2 Hours 30 Min

Total Marks: 50

Part A

(Short essay type)

Answer any 5 Questions (out of 7), each question carries two marks.

(5 × 2 = 10 Marks)

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Part B

(Essay type)

Answer any two questions (out of 4). Each question carries 20 Marks.

(2 × 20 = 40 Marks)

- 1 (a)
(b)
- 2 (a)
(b)
- 3 (a)
(b)
- 4 (a)
(b)

2.10 Grading

2.10.1 The indirect absolute grading system, in which marks are compounded to grades based on pre-determined class intervals and letter grades based on the UGC-recommended 10-point grading system, shall be used.

2.10.2 Students are graded in each course based on the percentage of marks earned (continuous assessment and end-of-semester examination marks combined) using the grading system shown in the table.

2.10.3 An alphabetical grading system shall be adopted for the assessment of a student's performance in a course. The grade is based on a ten-point scale. The following table gives the range of marks, grade points and the alphabetical grade.

Table 6. Letter Grades with Grade Points and Marks Equivalence

Range of marks (%)	Grade point	Letter Grade
90-100	9.0 - 10.0	O (Outstanding)
80-89.99	8.0 - 8.99	A + (Excellent)
70 - 79.99	7.0 - 7.99	A (Very Good)
60- 69.99	6.0 - 6.99	B+ (Good)
50 - 59.99	5.0 - 5.99	B (Above Average)
45 - 49.99	4.5 - 4.99	C (Average)
40-44.99	4.0-4.49	D (Pass)*
0-39.99	0	F (Failed) RA (Re-appear)
-	0	Abs (Absent)
-	0	I (course incomplete)

*Pass only if the student scores a minimum 40% in the end-semester examination. There is no minimum for continuous evaluation.

2.10.4 Each student shall be assigned a grade point and a letter grade in each course on the basis of the % marks scored in the course (continuous assessment and end semester examination marks taken together) as shown above. The minimum grade point required for passing a course is 4.0.

2.10.5 Performance of a student at the end of each semester is indicated by the Semester Grade Point Average (SGPA). It is a measure of performance of work done in a semester. It is the 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.

$$\text{SGPA} = \frac{\text{Sum of Credit Points Secured in a Semester}}{\text{Sum of Credits Taken in the Semester}}$$

2.10.6 The overall performance of a student is indicated by the Cumulative Grade Point Average (CGPA) and is calculated as follows .

$$\text{CGPA} = \frac{\text{Sum of Credit Points Secured in a Programme}}{\text{Sum of Credits Taken in the Programme}}$$

2.11 Grade Card

The Controller of Examination, University of Calicut is the authority to issue the semester-wise grade card and consolidated grade statement and certificates on completion of the Programme.

2.12 Supplementary Examinations for Failed Candidates

1. Candidates who have failed (F grade) in the semester examinations (except project work) can appear for the failed papers for the particular semester along with regular students. However, the Continuous Evaluation (CE) marks shall remain the same. Two such supplementary chances will be given for each semester within two years.
2. In the event of failure in Project Work the candidate shall re-register for project work, redo the project work and resubmit the project report a fresh for evaluation. The Continuous Evaluation marks shall be freshly allotted in this case.
3. Appearance for Continuous Evaluation and End Semester Evaluation are compulsory and no grade shall be awarded to a candidate if he/she is absent for CE/ESA or both.
4. A student who fails to complete the Programme/semester can repeat the full Programme /semester once, if the department council permits so. There shall be no provision for improvement of CE or ESA.

2.13 Guest Faculty / Industrial Collaboration

This Programme is intended to make industrial collaborations and the Guest faculty/experts from the industry may also be permitted to engage special sessions for the courses. Each course comprised 5 modules and the Guest faculty /expert from the collaborating institutes/industry may be permitted to engage at most twenty-five (5 modules × 5 Sessions/Module) sessions for each course. The duration of a session is two (02) hours. The Department shall provide the remuneration/TA/DA to the expert as per the University norms. To smooth conduct of the Programme, the University has to make a provision to transfer the required fund to the Department of Computer Science on request to meet this expenditure.

2.14 Grievance Redressal Mechanism

Committees will be constituted at the Department level to investigate the written complaints regarding Continuous Evaluation (CE). The Department Level Grievance Committee (DLGC) will consist of the Department Council and two student nominees enrolled for the Programme. The grievances/complaints have to be submitted to the Department concerned within two weeks of publication of the results of CE and disposed of within two weeks of receipt of complaints.

2.15 Re-admission

A candidate who discontinues the programme after taking admission shall be eligible for re-admission to the same programme in a subsequent academic year, payment of applicable fees, and compliance with the curriculum in force at the time of re-admission. The readmission is permitted from the 2nd semester onwards.

2.16 Transcript:

On successful completion of the programme, candidates shall be issued a Transcript by the Controller of Examinations on request and payment of a prescribed fee by the student. The transcript shall contain

- i) All the information that is included in the grade card
- ii) The month and year of passing each course
- iii) The final grades awarded
- iv) The CGPA and the class, if any obtained
- v) Signature or facsimile of the Controller of Examinations/
Registrar.

2.17 Medium of Instruction

The medium of instruction and examination for the programme shall be English, unless otherwise specified by the University from time to time.

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

Programme Structure , Syllabus and Scheme of Evaluation

Semesters	Theory	Practical
Semester I	5	2
Semester II	3 Electives	1 (Project Work)

SEMESTER I									
Sl.No	Subject code	Subject Title	Instructional Hrs/week			Marks			Credit
			L	P	T	CE	ESA	Total	
1	PGD01C01	Mathematical and Statistical Foundation for Data Science	4	0	1	50	50	100	4
2	PGD01C02	Data Structures and Algorithm Design	3	2	1	50	50	100	4
3	PGD01C03	Machine Learning for Data Science	4	0	1	50	50	100	4
4	PGD01C04	Introduction to Data Science	4	0	1	50	50	100	4
5	PGD01C05	Big Data Analytics	3	2	1	50	50	100	4
6	PGD01P06	Lab-I R for Data Science	0	4	2	50	50	100	4
7	PGD01P07	Lab II-Python for Data Analytics and Machine Learning	0	4	2	50	50	100	4
Total			18	12					28

SEMESTER II									
Sl No	Subject Code	Subject Title	Instructional Hrs/week			Marks			Credit
			L	P	T	CE	ESA	Total	
1	PGD02E01-03	Elective 1	3	2	1	50	50	100	4
2	PGD02E04-06	Elective 2	3	2	1	50	50	100	4
3	PGD02E07-09	Elective 3	4	0	1	50	50	100	4
4	PGD02P03	Project work and viva-voce	-	16	5	50	50	100	8
		Total	10	20					20

Elective 1:

Sl No	Subject Code	Subject Title
1	PGD02E01	Database Systems and SQL for Data Science
2	PGD02E02	Computational Methods for Data Science
3	PGD02E03	Business Analytics

Elective 2:

Sl.No	Subject Code	Subject Title
1	PGD02E04	Data Analytics and Prediction
2	PGD02E05	Deep Learning for Data Science
3	PGD02E06	Data Visualization and Interpretation

Elective 3:

Sl. No	Subject Code	Subject Title
1	PGD02E07	Artificial and Computational Intelligence
2	PGD02E08	Data Mining
3	PGD02E09	Natural Language Processing

Semester wise syllabus

SEMESTER -I

PGD01C01:Mathematical and Statistical Foundation for Data Science

- **CO1:** Students will be able to effectively use matrix algebra tools to analyze and solve systems of linear equations.
- **CO2:** Students will be able to learn concepts of linear algebra that form the foundation of data science problems.
- **CO3:** Introducing basic concepts of probability and statistics to students.
- **CO4:** Students will be able to apply statistical techniques to understand the data.

Module -1

Sets, Relations and Functions: Sets, Operations on sets, Venn Diagrams, Multi Sets, Functions, Domain and Range, Types of Functions. Binary Relations, Equivalence Relations, Ordering Relations, Operations on Relations, Partial Orders. Statements and Notation, Connectives, Quantified Propositions, Logical Inferences, Methods of Proof of an Implication, First Order Logic and other Methods of Proof, Rules of Inference for Quantified Propositions, Proof by Mathematical Induction.

Module -2

Linear Algebra: Matrices, determinants, inverse of matrix, Transpose of a matrix. Echelon form of a matrix and Row reduced echelon form of matrix. System of Linear equations, Solving System of Linear equations, Linear Transformations-rank and nullity theorem, Consistency, and Inconsistency of linear system of equations, Linear dependence and independence. Applications in real-life problems.

Module -3

Vector Algebra: Vectors, Scalars, Addition, Scalar multiplication, dot product, vector projection, cosine similarity. Correlation coefficient, Eigen values and Eigen vectors. Principle Component analysis (PCA) – Concepts and properties. Dimensionality reduction with PCA.

Module -4

Probability, Statistics & Random Variables: Measures of central tendency and variability- data distribution, outliers, and five-point summary. Probability theory- probability spaces, axioms, conditional probability, independence, mutually exclusive events. Bayes' theorem and basics of Naïve Bayes. Random variables- discrete & continuous, functions and generation of random variables. Multivariate random variables- joint distributions, independence, rejection sampling. Expectation- mean, variance, covariance, and conditional expectation. Probability distributions- Bernoulli, Binomial, Poisson, Normal, t, F, Chi-square.

Module -5

Statistical Inference, Prediction & Advanced Models: Sampling methods and distributions, Central Limit Theorem. Estimation and confidence intervals. Hypothesis testing (means and proportions) and maximum likelihood estimation. ANOVA (single and two-factor). Correlation and regression. Time series analysis: components and moving averages. Time series models: AR, ARMA, ARIMA, SARIMA, SARIMAX. Multivariate models: VAR, VARMAX and exponential smoothing. Gaussian Mixture Models (GMM) and Expectation-Maximization (EM) algorithm.

References:

1. Kenneth H. Rosen, "Discrete Mathematics And Its Applications", 7th Ed, McGrawHill, 2012.
2. Erwin Kreyszig, "Advanced Engineering Mathematics", Wiley India, 9th Edition 2011.
3. G. Strang "Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition", USA, 2016.
4. Montgomery, D. C. and G. C. Runger. "Applied Statistics and Probability for Engineers. 5th Edition". John Wiley & Sons, Inc., NY, USA, 2011.
5. Probability and Statistics for Engineering and Sciences, 8th Edition, Jay L Devore, Cengage Learning .
6. Michael Mitzenmacher and Eli Upfal; "Probability and Computing, 2e", Cambridge University Press, 2017.

PGD01C02:Data Structures and Algorithm Design

- **C01:** Demonstrate the ability to analyse the efficiency of algorithms using asymptotic notation and characterize their run-time complexities.
- **C02:** Develop the skills to analyse recursive algorithms using recurrence relations and the Master Theorem.
- **C03:** Implement and utilize various data structures.
- **C04:** Understand and implement advanced data structures.
- **C05:** Employ algorithm design techniques such as the greedy method, divide and conquer, and dynamic programming to solve various computational problems.

Module -1

Data structures: Definition and classification. Data Abstraction & Abstract data types. Data Structure operations: Traversing, Inserting and deleting, sorting and searching. Linear Search & Binary Search- Complexity.

Linear data structure: Arrays-Records-Representation. Array operations, polynomial representation with arrays; concept of recursion, Types of recursion. Case study with the Tower of Hanoi problem.

Module -2

Stack: operations on stack. Application of stack - postfix expression evaluation, conversion of infix to postfix expression. Queue- operations and its implementations-Circular queue-dequeue - priority queues, Application of queue. Linked Lists- Operations and implementations-Doubly Linked Lists and Circular lists-Sparse matrix representation.

Module -3

Non-linear data structure: Tree- basic terminologies and properties; representation of binary tree, operations on binary tree; type of binary tree, Tree traversal, Binary search tree. Applications of tree, AVL tree, Red-Black Trees. Heap-Definition,representation,Insertion and deletion of elements , Heap sort. Graphs: basic terminologies, Properties,representation.Graph traversals - depth-first traversal - breadth -first traversal, Applications. Hashing: Overview of hashing-Hash tables- hash functions and their computations-open addressing-Linear probing- quadratic probing-double hashing algorithms and their implementations-Separate chaining-Hashing efficiency.

Module -4

Algorithm Analysis: Steps in developing algorithms, Methods of specifying an algorithm. Important Problem Types. Importance of algorithm analysis, Time and Space Complexity. Growth of Functions: Asymptotic notations. Solving recurrences: Substitution method, Recursion Tree method, Master's Theorem, problem solving using Master's theorem.

Module -5

Basic Designing Techniques: Design Principles and Strategies, Divide-and-Conquer: Merge sort, Binary Search, Strassen's Matrix Multiplication. Greedy Method: Huffman coding, Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees: Prim's Algorithm, Kruskal's Algorithm; Single Source Shortest Path algorithm.

Dynamic Programming: Floyd's Warshall's Algorithm for the All-Pairs Shortest Paths Problem, 0/1 Knapsack, The Traveling Salesperson problem. Backtracking: n-Queens problem, Branch and Bound: Assignment Problem.

References:

1. Anany Levitin: Introduction to the Design & Analysis of Algorithms, 2nd Edition, Pearson Education, 2007.
2. Thomas H Cormen, Charles E Leiserson, & Ronald L Rivest, Introduction to Algorithms, 2nd Ed. Prentice Hall of India Private Limited, New Delhi, 2001.
3. Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekaran: Fundamentals of Computer Algorithms, 2nd Edition, Universities Press, 2007.
4. Gilles Brassard & Paul Bratley, Fundamentals of Algorithmics, PrenticeHall of India, 2007.
5. Goodman S E & Hedetnieni, Introduction to the Design & analysis of Algorithms, Mc- Graw Hill, 2002
6. Alfred V.Aho, John E.Hopcroft and Jeffrey D.Ullman, Data structures and Algorithms,Pearson Education Asia,2002.
7. SartajSahni, Datastructures, Algorithms and Applications in Java, Second Edition, Universities Press(India) Pvt Ltd, 2005.
8. Robert Lafore, Datastructures and algorithms in Java, Second Edition, Sams Publishing,2003.

PGD01C03:Machine Learning for Data Science

- **C01:** Introduce students to the basic concepts and techniques of Machine Learning.
- **C02:** To gain experience of doing independent study and research in the field of Machine Learning .
- **C03:** To develop skills of using recent machine learning software tools to evaluate learning algorithms and model selection for solving practical problems.

Module -1

Introduction to Machine Learning – Machine learning basics, Types of machine learning, Applications of Machine Learning, Basic types of data in Machine learning, Data pre-processing, Predictive Vs Descriptive models, training a model, training versus testing, cross validation, overfitting & underfitting, Bias variance tradeoff, error measures, evaluating performance of a model. Introduction to feature engineering, ML tools in Python.

Module -2

Linear Regression – Problem formulation, Parameter Estimation, Bayesian Linear Regression, Multiple linear regression, fitting simple linear and multiple linear regression equations with examples, regularization techniques, case study and implementation. Logistic Regression - Interpreting Parameters in Logistic Regression, Inference for Logistic Regression, Logistic Models with Categorical Predictors, Multiple Logistic Regression, Fitting Logistic Regression Models and its implementation using real life examples.

Module -3

Support Vector Machine (SVM) Algorithm, Types of SVM, Hyperplane and Support Vectors, Working of SVM Applications of SVM; Naïve Bayes Classifier Algorithm, Bayes' Theorem, Types of Naïve Bayes Model, Working of Naïve Bayes' Classifier, Applications of Naïve Bayes Classifier.

Decision Tree:Information Theory, Entropy Based Decision Tree Construction, Avoiding Overfitting, Minimum Description Length, Handling Continuous valued attributes, missing attributes.

Module -4

Clustering:Cluster analysis concepts,Partitioning methods,Hierarchical methods for cluster analysis, Density based methods for cluster analysis, Evaluation of clustering algorithms.

Ensemble Learning: Combining Classifiers, Bagging, Random Forest, Boosting, ADABOOST, Gradient Boosting, XGBoost.

Module -5

Introduction to Artificial Neural Networks and Deep Learning Concepts
Neural networks, Perceptron, Multi-Layer Perceptron, Back propagation algorithm, Fundamentals of deep learning- Deep Feedforward Networks, Regularization for Deep Learning, Optimization for Training Deep Models, Introduction to Convolutional Networks, Sequence Modeling using Recurrent Nets.

References:

1. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2007.
2. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
3. Ethem Alpaydin, "Introduction to Machine Learning", MIT Press, Third Edition, 2014.
4. Tom M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc. Indian Edition 1997.
5. Stanford Lectures of Prof. Andrew Ng. 6. NPTEL Lectures of Prof. B.Ravindran.
6. PANG-NING TAN, MICHAEL STEINBACH, VIPIN KUMAR, Introduction To Data Mining, Pearson, 2nd Edition.
7. CHRISTOPHER J.C. BURGESS: A Tutorial on Support Vector Machines for Pattern Recognition, Kluwer Academic Publishers, Boston, pp. 1–43.

PGD01C04: Introduction to Data Science

- **C01:** Develop a strong foundation in data science concepts, application areas, and emerging trends in the field.
- **C02:** Understand fundamental statistical methods and machine learning concepts used in data science.
- **C03:** Explain the processes involved in data collection, management, and storage for data science applications.
- **C04:** Analyze the role and significance of data visualization in interpreting and presenting data effectively.
- **C05:** Distinguish between ethical and unethical practices in the use of data science and data-driven technologies.

Module -1

Introduction to Data Science: Fundamentals of Data Science, Terminologies, data science process, data science toolkit, Data Science Challenges, Types of data (Structured, Semi-Structured, Unstructured), Characteristics of data, Application Areas (Healthcare, Finance, Retail, Education, Social Media, etc).

Module -2

Data Science Teams - Defining Data Team, Roles and Responsibilities (Data Scientist, Data Engineer, Data Analyst, etc), Real World applications (Case Studies of Data Science Applications, Industry Use Cases). Data Collection and Management - Data Sources (Primary & Secondary), Data Collection Methods, Data Storage and Management, Role of Data in the Data Science Lifecycle. Data Analytics - Defining Analytics, Types of data analytics - Descriptive, Diagnostic, Predictive, Prescriptive.

Module -3

Data wrangling : Data cleaning, Data Aggregation, Sampling, Statistical descriptions of data, Measuring data similarity & dissimilarity, Handling Numeric Data, Discretization, Binarization, Normalization, Data Smoothing, Dealing with textual Data, Dealing with Images, audio and video data, Managing Categorical Attributes, Transforming Categorical to Numerical Values, Encoding techniques.

Module -4

Feature Engineering : Feature Extraction, Feature Construction, Feature Subset selection, Filter methods, Wrapper methods, Embedded methods, Feature reduction (Dimensionality Reduction). Data visualization: need for visualization, Exploratory vs Explanatory Analysis, Tables, Axis based Visualization and Statistical Plots, Data Visualization Design Process, Lessons in Data Visualization Design, Storytelling with Data, Dashboards, Power BI, Tableau.

Module -5

Ethics for Data Science, Bias and Fairness: Types of Bias, Identifying Bias, Evaluating Bias, Examples of misuse of Data, Doing Good Science, Five C's, Ethical guidelines for Data Scientist, Ethics of data scraping and storage.

References:

1. Cielen, D., Meysman, A., Introducing Data Science: Big Data, Machine

- Learning, and More, Using Python Tools, Manning Publications, 2016.
2. Cole Nussbaumer Knaflitz, Storytelling with Data, A data visualization guide for business professionals, by; Wiley, 2015.
 3. P. Tan, M. Steinbach, A. Karpatne, and V. Kumar, Introduction to Data Mining, 2nd Ed Pearson Education, 2018.
 4. Roger D Peng and Elizabeth Matsui , The Art of Data Science by, Pelican Books, 2020,
 5. Mike Loukides, Hilary Mason, DJ Patil, Ethics and Data Science, O'Reilly, 2018
 6. Jake VanderPlas, Python Data Science Handbook: Essential tools for working with data, , O'Reilly, 2016.
 7. Introducing Data Science, Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Manning Publications Co., 1st edition, 2016.

PGD01C05:Big Data Analytics

- **CO1:** Understand the fundamental concepts, characteristics, and architecture of big data technologies.
- **CO2:** Apply big data technologies and analytical techniques to design solutions for real-world problems .
- **CO3:** Develop practical knowledge of big data frameworks and tools such as Hadoop for distributed data processing.
- **CO4:** Students will be able to use R language for programming purposes .

Module -1

Introduction to Big Data – definition & importance of Big Data, four dimensions of Big Data - volume, velocity, variety, veracity, Types of Data: structured data, unstructured data, Technology Foundation for Big Data. Big Data Stack – layer 0,1,2,3 and 4 – Big Data Applications-Understanding the Basics of Virtualization- The cloud and Big Data - Big Data management – operational databases, relational databases, non-relational databases – NoSQL - key-value pair databases – document databases - columnar databases - graph databases - spatial databases.

Module -2

Big Data analysis - basic analytics - operationalized analytics - modifying business intelligence products to handle Big Data - Big Data analytics examples - Analytics solutions - text analytics - exploring unstructured data -

understanding text analytics - analysis and extraction techniques- the extracted information - text analytics tools for Big Data - custom applications for Big Data analysis - R Environment - Google Prediction API - Characteristics of a Big Data Analysis Framework.

Module -3

Hadoop – history – components – Hadoop Distributed File System –Analyzing Data with Hadoop - Application Development in Hadoop – Hadoop Streaming - getting our data into Hadoop - Map Reduce Basics – origins of MapReduce - map function – reduce function – putting them together– Map Reduce Applications – How Map Reduce Works – Map Reduce Types and Formats – Map Reduce Features.

Module -4

R Programming -Fundamentals, installation and use of software,R studio environment , Basic syntax and working with scripts , Data Types and classes. type conversion; data entry and data editing, Operators in R, Statistical analysis using Built in Functions (mean, median, variance, standard deviation, summary) , scope, logic and control statements in R , input and output operations.

Module -5

R Programming - Exploring and cleaning data for analysis, Introduction to Data Structures – Vectors, Arrays, Lists, Data Frames, and Matrices – creation, accessing, and manipulation of data. , handling missing values; Family of functions for data manipulation, Statistical modeling in R - building linear models, generalized linear models, and nonlinear models Mapping models to machine learning; evaluating and validating models. Introduction to syntax and basic graphical functions and creation of multiple plots in R .

References:

1. Hurwitz, Alan Nugent, Fern Halper and Marcia Kaufman, Big Data for Dummies.
2. Chris Elaton, Dirk Deroos, Tom Deutsch, George Lapis and Pual Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, 1st Edition,
3. Garry Turkington, Hadoop Beginner's Guide, Packt Publishing Ltd.
4. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007.

5. Tom White, Hadoop: The Definitive Guide, 3rd Edn, O'Reilly Media, 2012.
6. Dr. Bharati motwani ,Data Analytics with R New Delhi: Wiley India Pvt. Ltd.; 2019

PGD01P06: R for Data Science

- **C01:** Master the use of the R interactive environment.
- **C02:**Expand R by installing R packages.
- **C03:** Develop loop constructs in R.
- **C04:** Use R for descriptive statistics.
- **C05:** Use R for inferential statistics.

Course content:

Introduction to the statistical software R, Data objects in R: Vector, List,Array, Matrix and Factors, Creating and Manipulating Data Structures , Accessing elements of a vector or matrix, Lists, Addition, Multiplication, Subtraction, Transpose, Inverse of matrices.Input and output functions , importing and exporting different data files. Boolean operators. R-Graphics & Visualisation- Pie Chart,Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options; Multiple plots in a single graphic window, Adjusting graphical parameters. Control Structures & Looping- For loop, repeat loop, while loop, if command, if else command.

List of programs:

Course Content Set of programs to be done in the lab:

Basic Programming in R

1. Programs on data types in R.
2. Built-in functions in R.
3. Operators in R.
4. Creating and manipulating vectors.
5. Creating and manipulating matrices.
6. Operations on Data Frames/Lists/Factors in R.
7. Looping constructs.
8. Conditional statements in R.
9. Reading and Writing Different Types of Datas.
- 10.Data Cleaning and handling missing values using R

Statistical Analysis & Data Visualization in R

- 11.Measures of central tendency and variability using R

12. Visualizing correlation
13. Regression Models in R
14. Plots in R (Pie/Scatter plot/Histogram/Box plot)
15. Customizing and saving Multiple plots in a single window
16. Classification Models (Decision Trees /Naive Bayes)
17. Clustering models (K-means, Hierarchical clustering)

PGD01P07: Python for Data Analytics and Machine Learning

- **C01:** Develop Python programs using procedural and object-oriented approaches.
- **C02:** Perform data manipulation and analysis using Python libraries.
- **C03:** Visualize data effectively using plotting tools.
- **C04:** Implement basic machine learning and deep learning models.
- **C05:** Build simple data-driven applications.

Course Content:

Python Fundamentals: Python syntax and basic programming, data types and variables, control structures, strings, lists, tuples, sets, dictionaries, searching and sorting techniques, pattern printing. Functions, OOP and File Handling: Functions and modules, object-oriented programming concepts, file handling operations, date and time operations, program to find the longest word in a text. Numerical Computing and Data Analysis using Pandas: NumPy arrays and operations, matrix operations using NumPy, introduction to SciPy. Series and DataFrames, data cleaning and preprocessing, data manipulation and analysis. Data Visualization: Matplotlib basics, plotting techniques including line, bar, scatter, and histogram plots, data interpretation, Seaborn, Plotly. Machine Learning: Introduction to machine learning, decision tree classifier with visualization, Naive Bayes classifier, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), visualization of decision boundaries. Deep Learning: Introduction to neural networks, implementation of artificial neural networks (ANN), forward propagation and basic concepts, introduction to deep learning frameworks (TensorFlow/Keras), implementation of a simple ANN model. Implementation of CNN for image classification.

Sample Programs

1. Linear and binary search
2. Pattern printing (any one pattern)
3. File program to count words/lines

4. Find longest word in a file
5. Matrix operations using NumPy
6. Basic statistics (mean, median)
7. Handling missing data in Pandas
8. Simple dataset analysis using DataFrame
9. Line and bar chart using Matplotlib
10. Scatter plot for correlation
11. Histogram for data distribution
12. Foundational Data & Regression Programs (Data Import/Export, Simple Linear Regression , Multiple Linear Regression , Polynomial Regression).
13. Classification Programs (Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees & Random Forests, Naïve Bayes.
14. Unsupervised Learning Programs (K-Means Clustering, Market Basket Analysis, Anomaly Detection
15. SVM/Decision Tree/KNN classification
16. Simple ANN using Keras
17. Basic image classification (CNN)
18. Time Series Forecasting

SEMESTER -II

PGD02E01: Database Systems and SQL for Data Science

- **C01:** Introduce students to the fundamental concepts of database systems, data models and database design.
- **C02:** To develop proficiency in SQL for data manipulation, querying and analysis.
- **C03:** To develop skills in applying database systems and tools for managing and analyzing real-world data in data science applications.

Module - 1

Introduction to Database Systems: Overview of database systems, file system versus database system, database management system (DBMS) concepts and architecture, types of databases including relational, NoSQL and distributed databases, data models including hierarchical, network and relational models, three schema architecture, data independence, database users and roles, introduction to data for data science.

Module - 2

Relational Model and Database Design: Relational model concepts including tables, attributes and tuples, keys including primary key, foreign key, candidate key and composite key, integrity constraints, entity relationship (ER) model, ER diagrams and mapping to relational schema, normalization including 1NF, 2NF, 3NF and BCNF, functional dependencies, database design process for real-world applications.

Module - 3

SQL for Data Manipulation and Querying: Introduction to SQL, data definition language (DDL) ,CREATE, ALTER ,RENAME and DROP, data manipulation language (DML) including INSERT, UPDATE and DELETE, data query language (DQL) using SELECT, filtering data using WHERE, AND, OR and NOT, sorting and grouping using ORDER BY, GROUP BY and HAVING, joins including INNER JOIN, LEFT JOIN, RIGHT JOIN, and FULL OUTER JOIN, subqueries and nested queries, set operations including UNION, INTERSECT and EXCEPT, views and indexes, Program with SQL.

Module - 4

Advanced SQL and Data Analysis: Aggregate functions including COUNT, SUM, AVG, MIN and MAX, window functions and analytical queries, common table expressions (CTEs), stored procedures and functions, triggers, transactions and concurrency control, ACID properties, introduction to data warehousing, ETL process for data science, SQL for data analysis with case studies and implementation.

Module - 5

NoSQL and Modern Data Management: Limitations of relational databases, introduction to NoSQL databases, types of NoSQL databases including key-value, document, column-family and graph databases, overview of MongoDB and Apache Cassandra, big data concepts and distributed storage, introduction to Hadoop and Apache Spark, data lakes versus data warehouses, database integration with Python using libraries such as Pandas and SQL connectors, applications of databases in data science.

References:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, *Database System Concepts*, 7th Edition, McGraw-Hill, 2019.
2. Raghu Ramakrishnan, Johannes Gehrke, *Database Management Systems*, 3rd Edition, McGraw-Hill, 2002.
3. Hector Garcia-Molina, Jeffrey Ullman, Jennifer Widom, *Database Systems: The Complete Book*, 2nd Edition, Pearson, 2008.
4. C.J. Date, *An Introduction to Database Systems*, 8th Edition, Pearson, 2003.
5. Martin Kleppmann, *Designing Data-Intensive Applications*, O'Reilly, 2017.
6. Ralph Kimball, Margy Ross, *The Data Warehouse Toolkit*, 3rd Edition, Wiley, 2013.
7. Shannon Bradshaw, Eoin Brazil, Kristina Chodorow, *MongoDB: The Definitive Guide*, 3rd Edition, O'Reilly, 2019.
8. Wes McKinney, *Python for Data Analysis*, 3rd Edition, O'Reilly, 2022.

PGD02E02: Computational Methods for Data Science

- **C01:** Understand different Statistical methods and their applications.
- **C02:** Have an idea about time-frequency analysis.
- **C03:** Learn Principal Component Analysis.

Module -1

Statistical Methods and Their Applications: Basic probability concepts, Bayes' Formula, Random variables and statistical concepts, Hypothesis testing and statistical significance.

Module -2

Time-Frequency Analysis: Fourier Transforms and Wavelets: Basics of Fourier Series and the Fourier Transform, FFT Application: Radar Detection and Filtering, Radar Detection and Averaging, Time-Frequency Analysis: Windowed Fourier Transforms, Time-Frequency Analysis and Wavelets.

Module -3

Image Processing and Analysis: Basic concepts and analysis of images, Linear filtering for image denoising. Linear Algebra and Singular Value Decomposition: Basics of The Singular Value Decomposition (SVD), The SVD in broader context.

Module -4

Introduction to Principal Component Analysis (PCA), Principal Components, Diagonalization and SVD, Principal Components and Proper Orthogonal Modes, Independent Component Analysis: The concept of independent components, Image separation problem.

Module -5

Image Recognition: The SVD and Linear Discrimination Analysis, Basics of Compressed Sensing: Beyond Least-Square Fitting, Signal Reconstruction and Circumventing Nyquist, Data (Image) Reconstruction from Sparse Sampling. Dimensionality Reduction for Partial Differential Equations.

References:

1. Yeliz Karaca, Carlo Cattani "Computational methods for data analysis", O'Reilly Media, 2018.
2. Siegmund Brandt," Data Analysis: Statistical and Computational Methods for Scientists and Engineers", Springer,2017

3. D. Narayana, Sharad Ranjan, Nupur Tyagi “Basic Computational Techniques For Data Analysis” Sage,2021.

PGD02E03: Business Analytics

- **CO1:** Understand the different basic concepts / fundamentals of business statistics.
- **CO2:** Understand the importance of measures of Descriptive statistics and their implication on Business performance.
- **CO3:** Understand the concept of Probability and its usage in various business applications.
- **CO4:** Understand the practical application of Descriptive and Inferential Statistics concepts and their uses for Business Analytics.

Module -1

Descriptive Statistics: Meaning, Scope, types, functions and limitations of statistics, Measures of Central tendency – Mean, Median, Mode, Quartiles, Measures of Dispersion – Range, Inter quartile range, Mean deviation, Standard deviation, Variance, Coefficient of Variation, Skewness and Kurtosis.

Module -2

Time Series & Index Number Time series analysis: Concept, Additive and Multiplicative models, Components of time series, Trend analysis: Least Square method - Linear and Non- Linear equations, Applications in business decision- making. Index Numbers: Meaning, Types of index numbers, uses of index numbers, Construction of Price, Quantity and Volume indices: Fixed base and Chain base methods.

Module -3

Correlation & Regression Analysis -Correlation Analysis: Rank Method & Karl Pearson's Coefficient of Correlation and Properties of Correlation. Regression Analysis: Fitting of a Regression Line and Interpretation of Results, Properties of Regression Coefficients and Relationship between Regression and Correlation.

Module -4

Probability Theory & Distribution Probability: Theory of Probability, Addition and Multiplication Law, Baye's Theorem, Theoretical Distributions: Concept and application of Binomial; Poisson and Normal distributions.

Module -5

Hypothesis Testing: Null and Alternative Hypotheses; Type I and Type II errors; Testing of Hypothesis: Large Sample Tests, Small Sample test, (t, F, Z Test and Chi Square Test) Concept of Business Analytics- Meaning types and application of Business Analytics, Use of Spread Sheet to analyze data-Descriptive analytics and Predictive analytics.

References:

1. G C Beri – “Business Statistics, 3rd ed”, TATA McGrawHill.
2. Chandrasekaran & Umaparvathi-”Statistics for Managers, 1st edition, PHI Learning
3. Davis, Pecar – Business Statistics using Excel, Oxford
4. Ken Black – Business Statistics, 5th ed., Wiley India
5. Levin and Rubin – statistics for Management, 7th ed., Pearson
6. Lind, Marchal, Wathen – Statistical techniques in business and economics, 13th ed, McGrawHill
7. Newbold, Carlson, Thorne – Statistics for Business and Economics, 6th ed., Pearson
8. S. C.Gupta – Fundamentals of Statistics, Himalaya Publishing
9. Walpole – Probability and Statistics for Scientists and Engineers, 8th ed., Pearson

PGD02E04: Data Analytics and Prediction

- **CO1:**Understand predictive modeling techniques for data analytics.
- **CO2:** Apply data preprocessing techniques for big data.
- **CO3:** Measure the performance of data classification and regression models.
- **CO4:** Understand the use of Classification Trees and Rule-Based Models in big data analytics projects.

Module -1

Predictive Models, Process, Data Pre-processing, Data Transformations, FOver-Fitting and Model Tuning, Data Splitting, Resampling Techniques.

Module -2

Measuring Performance in Regression Models, The Variance-Bias Trade-off, Linear Regression for Solubility Data, Penalized Models, Nonlinear Regression

Models, Multivariate Adaptive Regression Splines, Support Vector Machines, K-Nearest Neighbors.

Module -3

Discriminant Analysis and Other Linear Classification Models, Linear Discriminant Analysis, Partial Least Squares Discriminant Analysis, Nearest Shrunken Centroids, Nonlinear Discriminant Analysis, Flexible Discriminant Analysis.

Module -4

Measuring Performance in Classification Models, Class Predictions, Class Probabilities, Evaluating Predicted Classes, Two-Class Problems, Evaluating Class Probabilities, Receiver Operating Characteristic (ROC) Curves.

Module -5

Classification Trees and Rule-Based Models, Regression Model Trees, Bagged Trees, Random Forests, Boosting, Remedies for Severe Class Imbalance, Factors That Can Affect Model Performance.

References:

1. Max Kuhn and Kjell Johnson, "Applied Predictive Modeling, 2e", Springer, 2018
2. Ankam Venkat, "Big Data Analytics", Packt Publishing Limited, Birmingham,UK, 2016
3. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", 1e, Wiley, 2015
4. Hadley Wickham, Garrett Golemund, "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1e", Shroff/O'Reilly,, 2017
5. Joel Grus, "Data Science from Scratch", Shroff, 1e, O'Reilly Media, 2015
6. James D. Miller, "Statistics for Data Science, 1e", Packt Publishing Limited, 2017
7. Thomas Rahlf, "Data Visualization with R: 100 Examples, 1e", Springer, 2017

PGD02E05: Deep Learning for Data Science

- **C01:** Introduce students to the fundamental concepts and architectures of deep learning.
- **C02:** To develop skills in designing, training and evaluating deep learning models for real-world data science problems.
- **C03:** To enable the use of modern deep learning frameworks and techniques for advanced applications such as image processing, sequence modeling and generative modeling.

Module -1

Introduction to Deep Learning: Overview of deep learning, relationship between artificial intelligence, machine learning and deep learning, biological neuron versus artificial neuron, introduction to artificial neural networks, perceptron model, limitations of perceptron, multilayer perceptron, activation functions including sigmoid, tanh and ReLU, forward propagation, loss functions, introduction to gradient descent and optimization, challenges in training deep neural networks.

Module -2

Training Deep Neural Networks: Backpropagation algorithm, computation graph and chain rule, weight initialization techniques, vanishing and exploding gradient problems, optimization algorithms including stochastic gradient descent, momentum, RMSProp and Adam, regularization techniques including L1, L2, dropout and early stopping, batch normalization, hyperparameter tuning, model evaluation and performance metrics, implementation using frameworks such as TensorFlow and PyTorch.

Module -3

Convolutional Neural Networks: Introduction to convolutional neural networks (CNNs), convolution operation, filters and feature maps, padding and stride, pooling layers, fully connected layers, popular CNN architectures including LeNet, AlexNet, ResNet and VGGNet, transfer learning and fine tuning, applications in image classification, object detection and image segmentation.

Module -4

Sequence Models and Recurrent Neural Networks: Sequential data and time series, recurrent neural networks (RNNs), vanishing gradient problem in RNNs, long short term memory (LSTM) networks, gated recurrent unit (GRU),

bidirectional RNNs, sequence to sequence models, attention mechanism and transformers, applications in natural language processing, speech recognition and time series forecasting.

Module -5

Generative Models and Advanced Topics: Introduction to generative models, autoencoders and variational autoencoders (VAE), generative adversarial networks (GANs) and GAN framework, types of GANs, applications of GANs in image synthesis and data augmentation, introduction to reinforcement learning concepts in deep learning, ethics in deep learning, challenges and future trends in deep learning for data science.

References:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016.
2. Francois Chollet, *Deep Learning with Python*, 2nd Edition, Manning Publications, 2021.
3. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd Edition, O'Reilly, 2022.
4. Lewis Tunstall, Leandro von Werra, Thomas Wolf, *Natural Language Processing with Transformers*, O'Reilly, 2022.
5. *Stanford Lectures by Andrew Ng — Deep Learning Specialization*, Coursera.

PGD02E06: Data Visualization and Interpretation

- **C01:** Understand the significance of data representation.
- **C02:** Accomplish skills to represent the facts and information.
- **C03:** Comprehend the methods for information visualization
- **C04:** Familiarize yourself with scientific visualization techniques.

Module -1

Data and types of data, Data variability, uncertainty and context. Basics of Data Visualization: Definition, Need for Visualization, how to visualize data, General types of Data Visualization, pros and Cons of Data Visualization.

Module -2

Visualization Components: Visual cues, Coordinate systems, Scales, and Context. Diagrams used for data visualization: Bar chart, Histogram, Scatter

plot, Scatter plot, Network, Streamgraph, Tree map, Gantt chart, Stripe graphic, Animated spiral graphic. Visualization based on types of data: Visualizing Categorical data, Visualizing Time series data, Visualizing Spatial data.

Module -3

Information Visualization: Definition, Objectives of Information Visualization. Visual representation of large-scale collection of non-numerical information. Design Principles of Information Visualization: Principle of Simplicity, Principle of Proximity, Principle of Similarity, Principle of Closure, Principle of Connectedness, Principle of Good Continuation, Principle of Common fate, Principle of Familiarity, Principle of Symmetry.

Module -4

Methods for Information Visualization: Cartogram, Cladogram (phylogeny), Concept Mapping, Dendrogram (classification). Graph drawing, Heat map, Hyperbolic Tree, Tree mapping Multidimensional scaling. Information visualization reference model. Case study with real world problems.

Module -5

Scientific visualization: Introduction, Methods for visualizing two-dimensional and three-dimensional data sets, volume visualization. Data Visualization using in Python matplotlib Module, pyplot, plot (), scatter, bar charts, Formatting, figure (), subplot(), text(), xlabel(), ylabel(), title(), Plotting Mathematical Functions-Presentation and visualization of data for effective communication-TABLEAU-POWER BI.

References:

1. Nathan Yau, Data Points. Wiley Big Data Series
2. Healy, Kieran, "Data Visualization: A Practical Introduction", Princeton Uty. Press.
3. Ben Bederson and Ben Shneiderman," The Craft of Information Visualization: Readings and Reflections", Morgan Kaufmann, 2003.
4. Riccardo Mazza. "Introduction to Information Visualization, Springer", 2009

PGD02E07: Artificial and Computational Intelligence

- **CO1:** Explain the basic concepts of Artificial Intelligence.
- **CO2:** Apply Constraint satisfaction problems to solve various standard problems in AI.
- **CO3:** Compare the performance of heuristic techniques for a given problem.
- **CO4:** Understand basic issues of knowledge representation techniques.

Module -1

Introduction to artificial intelligence -Definition, Overview of AI applications, Problem space and search, Propositional calculus, Predicate Calculus, Rule Based Knowledge Representation, Using Inference Rules to produce Predicate Calculus expressions, Application.

Module -2

Intelligent Agents – Agents and Environments, Good Behavior: The Concept of Rationality, The Nature of Environments, The structure of Agents. The Present and Future of AI- Agent Components, Agent Architectures.

Module -3

Heuristic search and state space search - Techniques for Heuristic Search, State Space Search Strategies for State Space Search - Hill climbing, Depth first search, Breadth first search, Best-first search, A*, Problem Reduction, AO*Heuristic in games.

Module -4

Game Playing - The Minimax search procedure, adding Alpha-beta cut-offs, Additional refinement, Iterative deepening, Planning system and its components, Understanding, Understanding as constrained satisfaction.

Module -5

Knowledge representation issues, representation and mappings, Representing simple facts in logic, Representing instances and ISA relationships, Computable functions and Predicates, Resolution, conversion to clausal form, Knowledge representation using rules, logic programming, forward versus backward reasoning, Slot and filler structures: Semantic nets, frames, conceptual dependency, scripts.

References:

1. Stuart Russell, Peter Norvig: "Artificial Intelligence: A Modern Approach ", 3rd Ed, Pearson, 2016.
2. Elaine Rich, Kevin Knight, B.Nair: "ARTIFICIAL INTELLIGENCE ", 3rd Ed, McGraw Hill, 2017.
3. Charu C. Aggarwal. "Recommended Systems. The Textbook", Springer, 2016.
4. N.P.Padhy: Artificial Intelligence and Intelligent Systems, Oxford University Press, 2009.

PGD02E08: Data Mining

- **CO1:** Understand the basic concepts of Data mining.
- **CO2:** Identify the different techniques of data preprocessing.
- **CO3:** Analyze patterns that can be discovered by classification and clustering.
- **CO4:** Understand data mining techniques of clustering.
- **CO5:** Identify complex data types based on spatial and web mining.

Module -1

Introduction, data warehousing –Multidimensional data model, OLAP operations, warehouse schema, Data warehousing Architecture, warehouse server, Metadata, OLAP engine, data warehouse Backend Process.

Module -2

Data mining - introduction - definition - data mining functionalities - major issues in data mining - data preprocessing - data cleaning - data integration and transformation- data reduction - data discretization and concept hierarchy generation. Association rule mining - efficient and scalable frequent itemset mining methods - mining various kinds of association rules - association mining to correlation analysis - constraint- based association mining.

Module -3

Classification and prediction - issues regarding classification and prediction - classification by decision tree induction - Bayesian classification - rule based classification - classification by backpropagation - support vector machines - associative classification - lazy learners - other classification methods -

prediction - accuracy and error measures - evaluating the accuracy of a classifier or predictor - ensemble methods - model selection.

Module -4

Cluster analysis - types of data in cluster analysis - a categorization of major clustering methods - partitioning methods - hierarchical methods - density-based methods - grid-based methods - model-based clustering methods - clustering high dimensional data - constraint-based cluster analysis - outlier analysis.

Module -5

Graph mining - mining object, spatial, multimedia, text and web data - multidimensional analysis and descriptive mining of complex data objects - spatial data mining - multimedia data mining - text mining - mining the World Wide Web.

References:

1. "Data Mining Techniques", A K Pujari, University press.
2. J. Han, M. Kamber, "Data Mining Concepts and Techniques", Harcourt India
3. M. Dunham, "Data Mining: introductory and Advanced Topics", Pearson Pub.
4. Pang-Ning Tan, Michael S and Vipin Kumar, "Introduction to Data Mining, 1st Edition", Pearson India, ISBN: 9332518653.

PGD02E09: Natural Language Processing

- **C01:** Identify the features and challenges of Indian languages.
- **C02:** Understand the morphology and parts of speech-related processing of Indian languages.
- **C03:** Get an overview of the Probabilistic models for natural language processing.

Module -1

Introduction: Overview - Origins and challenges of NLP- Theory of Language - Features of Indian Languages – Issues in Font –Models and Algorithms- NLP Applications.

Module -2

Morphology and Parts-of-speech: Phonology – Computational Phonology - Words and Morphemes – Segmentation – Categorization and Lemmatisation – Word Form Recognition – Valency - Agreement - Regular Expressions – Finite State Automata – Morphology- Morphological issues of Indian Languages – Transliteration.

Module -3

Probabilistic Models: Probabilistic Models of Pronunciation and Spelling – Weighted Automata – N- Grams – Corpus Analysis – Smoothing – Entropy - Parts-of-Speech – Taggers – Rule based – Hidden Markov Models – Speech Recognition.

Module -4

SYNTAX: Basic Concepts of Syntax – Parsing Techniques – General Grammar rules for Indian Languages – Context Free Grammar – Parsing with Context Free Grammars- Top-Down Parser – Earley Algorithm – Features and Unification - Lexicalised and Probabilistic Parsing.

Module -5

SEMANTICS AND PRAGMATICS: Representing Meaning – Computational Representation –Meaning Structure of Language – Semantic Analysis – Lexical Semantics – WordNet – Pragmatics –Discourse – Reference Resolution – Text Coherence – Dialogue Conversational Agents.

References:

1. Daniel Jurafsky and James H. Martin “Speech and Language Processing”, Prentice Hall, 2009.
2. Christopher D.Manning and Hinrich Schütze, “Foundation of Statistical Natural Language Processing”, MIT Press, 1999.
3. Ronald Hausser, “Foundations of Computational Linguistics”, Springer-Verlag, 1999.
4. James Allen, “Natural Language Understanding”, Benjamin/Cummings Publishing Co. 1995.

PGD02P03 Project Work & Viva -Voce

- **C01:** Demonstrates in-depth knowledge and thoughtful application through the detailed analysis of the problem chosen for the study.
- **C02:** Assess the gap by acquiring knowledge about the previous works, and its interpretation and application.
- **C03:** Demonstrates the design of the proposed methodology and its merits.
- **C04:** Organize the interim project content with proper structure and sequencing.
- **C05:** Demonstrate the academic discussion skills to emphasize, argue with clarity of purpose using evidence for the claims.