File Ref.No.260197/GA - IV - J1/2021/Admn

UNIVERSITY OF CALICUT

Abstract

General and Academic Branch IV -Faculty of Science-The Regulation, scheme and syllabus of PG Diploma in Data Science and Analytics" for University Teaching Department, with effect from 2023 Admission (Project mode course)-lapsed due to the failure in laying before the succeeding meeting of the Senate – approved by the Vice Chancellor as approved by Academic Council on 20.12.2023, exercising the powers conferred under Section 10(13) of the Calicut University Act, 1975 - To lay before the next Senate - sanctioned-Orders issued

---

G & A - IV - J

U.O.No. 5336/2024/Admn Dated, Calicut University.P.O, 23.03.2024

Read:-1) Minutes of the Board of Studies in Computer Science & Application (PG), held on 16/10/2023.

2) Item No.I.25 in the minutes of Faculty of Science held on 17/11/2023

3) Item No.II.C of the minutes of the LXXXVI meeting of Academic council held on 20/12/2023

4) Statute 3, Chapter 39 of the Calicut University First Statutes, 1977

5) Orders of Vice Chancellor in the file of even No. dated 23/03/2024

ORDER

1. The Board of Studies in Computer Science & Application (PG), held on 16/10/2023 recommended to approve the Regulation and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department, w.e.f 2023 admission, vide paper read as(1).
2. Meeting of Faculty of Science has approved the minutes of the Board of Studies in Computer Science & Application (PG), vide paper read as (2).
3. The LXXXV meeting of Academic Council held on 20/12/2023 vide paper read as (3), has ratified the action of the Vice Chancellor in having approved the minutes of the meeting of the Board of Studies in Computer Science & Application-PG held on 16/10/2023.
4. As per Statute 3, Chapter 39 of the Calicut University First Statutes, 1977, "All Regulations made or repealed by the Academic Council shall be laid before the Senate during its next succeeding meeting as laid down in section 39 of the Act. The Senate shall have the power to cancel or amend any Regulation in the manner laid down in clause (c) of sub- section (2) of section 19 of the Act. If any Regulation or an amending or repeal thereof is not so laid before the Senate, the Regulation or amendment shall lapse or the Regulation repealed shall revive as the case may be, after the next succeeding meeting of the Senate"
5. The Regulation Scheme and Syllabus approved by the Academic Council on 20/12/2023 was not laid before the succeeding meeting of the Senate held on 21/12/2023, since it was on the following day of the meeting of the Academic Council. Hence as per Statute 3, Chapter 39 of the Calicut University First Statutes, 1977, the Regulations, scheme and syllabus of PG Diploma in Data Science and Analytics" for University Teaching Department, with effect from 2023 Admission (Project mode course) approved by the Academic Council on 20/12/2023 lapsed on 21/12/2023
6. Considering the matter in detail, Vice Chancellor, vide paper read as (5) has accorded sanction for the following by exercising the powers conferred under Section 10(13) of the Calicut

University Act, 1975

- To approve the Regulations, scheme and syllabus of PG Diploma in Data Science and Analytics" for University Teaching Department, with effect from 2023 Admission (Project mode course) approved by the Academic Council held on 20/12/2023 as such.
- To report the matter to the next meeting of the Academic Council
- To lay the Regulations, scheme and syllabus of PG Diploma in Data Science and Analytics" for University Teaching Department, with effect from 2023 Admission (Project mode course) before the next meeting of the Senate scheduled to be held on 26.03.2024

7. Orders are issued accordingly.

Ajayakumar T.K

Assistant Registrar

To

Head of the Department, 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/CHMK Library/Information Centres/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 2023 Admission onwards)

UNDER THE FACULTY OF SCIENCE

Board Of Studies in Computer Science and Applications (PG)
University Of Calicut, Kerala 673 635

REGULATIONS FOR THE PROJECT MODE PROGRAMME

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

With effect from 2023 Admission

1. Introduction

We are awash in 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.

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. Bachelors 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. with Mathematics at 10 + 2 Level or at Graduation Level or any equivalent degree or any other University / Institution, recognized by this University as equivalent thereto, with a minimum aggregate of 55% marks or equivalent grade.

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 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:

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			

2. Programme Structure

2.1 Attendance

The minimum attendance required for each course shall be 75% of the total number of classes conducted for each semester. 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 days in a semester subject to a maximum of two spells within a programme will be granted by the Vice-Chancellor. The benefit of condonation of 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 the course is defined as a minimum of one hour of lecture or a minimum of 2 hours of lab/tutorial per week for 18 weeks in a Semester. The minimum number of credits required to complete the Post Graduate Diploma in Data Science and Analytics (PGDDSA) programme is 44.

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. 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 Continuous Assessment (CE):

Evaluation of the Continuous Assessment shall be done by the faculty member in charge of the course. The proportion of the distribution of marks including CE (Continuous Evaluation) and ESE (End Semester Examination) shall be 50-50.

Continuous Evaluation (CE) of a course shall be based on periodic written internal tests, assignments, and Seminar/Case studies/Mini Project work and Viva-Voce/ Attending workshops/ Participating and presenting papers in Conferences/Publishing articles in Journals/Proceedings, etc. in respect of each course.

Components of Continuous Assessment (Theory)

Sl. No	Component	Marks
1	Internal Examinations	15
2	Assignments	10
3	Seminars	10
4	Case Studies/Group Discussion	5
5	Presenting papers/Invited Talk (Outside the Department)/Group Discussion /Completing approved MOOC in the relevant field of study OR Publishing Papers in Conference/ Book Chapters/ Publishing Articles in Approved Journals	10
	Total	**50**

For the Journal Publication/successful completion of approved MOOC will be awarded 10 marks and Conference Proceeding/Book Chapters will get 05 Marks each.

Components of Continuous Assessment (Practical/Mini Project work)

Sl. No	Component	Marks
1	Implementing the LAB Experiments	25
2	Mini Project Work/Viva-voce	25
	Total	**50**

2.8 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.

2.9 Final Project Work and Viva-voce:

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

2.9.1 Project Evaluation

(i) Continuous Assessment of the Project Work:

A Departmental committee duly constituted by the Head of the Department will review the projects periodically. There shall be two internal presentations on the work in progress. The assessment is based on presentation, interim report, and viva voce. The total mark for CA shall be divided among the two presentations. Each internal presentation shall be evaluated based on the following components:

Sl.No	Components	Marks
1	Understanding of the problem/concepts	20
2	Adhering to methodology	5
3	Quality of presentation/Demonstration (Optional)	5
4	Quantum of work/effort	20
	Total	**50**

(ii) End Semester Assessment of the Project Work

At the end of the semester, the candidate shall submit the Project report (two bound copies and one soft copy) duly approved by the guide, and co-guide for the End Semester Assessment. Evaluation for ESA should be conducted by a board of examiners appointed by the University. (Mark Distribution will be as given below).

A student shall pass the Project course if she/he secures a separate minimum of 50 % for the external and 50% for ESE and CA put together.

If a candidate fails in the evaluation of the Project, he/she has to repeat the project course along with the next batch and undergo both CA and ESE. Unlike theory/practical courses, the CA mark will not be retained. There shall not be any improvement chance for the marks obtained in the Project course.

Sl.No	Components	Marks
1	Understanding of the problem/concepts/content	15
2	Adhering to methodology	5
3	Quality of presentation/Demonstration (Optional)	10
5	Organization and content of Project report	5
6	Viva-voce	15
	Total	**50**

2.9.2 Guideline for preparing project Report

(i). Arrangement of contents:

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

1. Cover Page & Title Page
2. Plagiarism Report
3. Bonafide Certificate
4. Abstract

5. Table of Contents
6. List of Tables
7. List of Figures
8. List of Symbols, Abbreviations and Nomenclature
9. Chapters

The chapters may be broadly divided into 3 parts (i) Introductory chapter, (ii) Chapters developing the main theme of the project work (iii) 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.

10. Appendices
11. References

The tables and figures shall be introduced at appropriate places.

(ii). 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.

(iii). 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.

3. End-Semester Assessment (ESA)

All the End Semester Assessment (Theory) Examinations will be conducted by the Controller of Examination, University of Calicut and the End Semester Assessment (Practical) Examinations will be conducted internally by the Department.

To conduct the End Semester Assessment (Theory) and Final Project work and Viva-voce Examination, the Head of the Department/Chairman shall submit a confidential panel for the Board of Examiners as per the University norms duly approved by the Department council.

3.1 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 Question Paper

Code: Reg.No:……

Name:……..

First Semester PG Diploma in Data Science and Analytics Examination – 2023

Course Code: (eg: PGDDSA 01 101)

Course: (Eg: Mathematical Foundation of Data Science)

Time: 2 Hours 30 Min. Total Marks: 50

Part A

(Short essay type)

Answer any 5 Questions (out of 6) , each question carries two marks. (5 × 2 = 10 Marks)

1.
2.
3
4
5

6

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)

NOTE: Minimum one question from each of the five modules. Remaining questions can be from any module. There should not be more than two questions from the same module.

3.2 Grading

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.

Range of marks %	Grade points	Alphabetical grade
90–100	9	A+
80–89	8	A
70–79	7	B+
60–69	6	B
50–59	5	C
<50	0	F

A minimum of grade point 5 (Grade C) is needed for the successful completion of a Programme. Performance of a student at the end of each semester is indicated by the Grade Point Average (GPA) and is calculated by taking the weighted average of grade points of the courses successfully completed. Following formula is used for the calculation. The average will be rounded off to two decimal places.

$$\text{GPA} = \frac{\textbf{\textit{Sum of (grade points in a course multiplied by its credit)}}}{\textbf{\textit{Sum of credits of courses}}}$$

The overall performance of a student is indicated by the Cumulative Grade Point Average (CGPA) and is calculated using the same formula given above. Based on CGPA overall letter grade of the student shall be in the following way.

CGPA	Overall letter grade
8.5 and above	A+
7.5and above but less than 8.5	A
6.5 and above but less than7.5	B+
5.5 and above but less than 6.5	B
4.5 and above but less than 5.5	C

Conversion of Grades into Overall letter grade classification. The First Class with distinction: A+ and A, First-Class: B+ and B and the Second Class: C

Classification	Overall letter grade
First Class with distinction	A+ and A
First Class	B+ and B
Second Class	C

3.3 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.

4. 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 C 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.

6. Gust Faculty / Industrial Collaboration

This Programme is intended to make industrial collaborations and the Guest faculty/experts from the industry may also permitted to engage special sessions for the courses. Each course comprised 5 modules and the Guest faculty /expert from the collaborating institutes/industry may 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.

7. 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.

****

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

Programme Structure, Syllabus and Scheme of Evaluation

Semester	Theory	Practical
Semester I	5	2
Semester II	2 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 of Data Science	3	0	1	50	50	100	3
2	PGD01C02	Foundation of Data Science and Algorithm Design	3	0	1	50	50	100	3
3	PGD01C03	Machine Learning for Data Science	4	0	1	50	50	100	4
4	PGD01C04	Data Analytics and Prediction	4	0	1	50	50	100	4
5	PGD01C05	Computational Methods for Data Science	4	0	1	50	50	100	4
6	PGD01P06	Lab-I R for Data Science	0	6	2	100	-	100	3
7	PGD01P07	Lab II-Python for Data Analytics and Machine Learning	0	6	2	100	-	100	3
		Total	18	12					**24**

SEMESTER II									
NO	**Subject Code**	**Subject Title**	**Instructional Hrs/week**			**Marks**			**Credit**
			L	**P**	**T**	**CE**	**ESA**	**Total**	
	PGD02E01-04	Elective - 1	4	0	1	50	50	100	4
	PGD02E05-08	Elective - 2	4	0	1	50	50	100	4
	PGD02P03	Project work & Viva-voce	-	12	5	50	50	100	12
		Total	8	12					**20**

Elective 1:

Sl.No	Subject Code	Subject Title
1	PGD02E01	Artificial Intelligence
2	PGD02E02	Natural Language Processing
3	PGD02E03	Advanced Machine Learning
4	PGD02E04	Deep Learning for computer vision

Elective 2:

Sl.No	Subject Code	Subject Title
1	PGD02E05	Data Mining
2	PGD02E06	Big Data Analytics
3	PGD02E07	Data Visualization
4	PGD02E08	Business Analytics

Semester wise syllabus

SEMESTER- I

PGD01C01: Mathematical and Statistical Foundation of Data Science

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

- Use mathematical concepts in the field of data science.
- Employ the techniques and methods related to the area of data science in a variety of applications.
- Apply logical thinking to understand and solve the problem in context

Module -1
Sets, the principle of inclusion and exclusion, relations, equivalence relations and partition, denumerable sets, partial order relations, Mathematical Induction, Permutations and combinations.

Module -2
Graphs: definition, types of graphs, paths and circuits. Eulearian and Hermitian circuits. Seven bridges machine, shortest path traveling salesman problems. Planar graph. Matrix representation of graph: adjacency matrix, incidence matrix, circuit matrix, cut set matrix, path matrix, Directed Graphs, Trees, Minimum Spanning Tree of a Graph.

Module -3
Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity), Eigenvalues and eigenvectors, Matrix factorizations, Inner products, Distance measures, Projections, Notion of hyperplanes, half-planes.

Module -4
Vectors and Geometry of Space - Vectors in Space, The Dot and Cross Product of Two Vectors, Lines and Planes in Space, Distances in Space. Surfaces in Space, Cylindrical and Spherical Coordinates Vector-Valued Functions - Differentiation and Integration of Vector-Valued Functions, Arc Length and Curvature.

Module -5
Probability theory: probability spaces, conditional probability, independence – Random variables: discrete and continuous random variables, functions of random variables, generating random variables – Multivariate random variables: joint distributions, independence, generating multivariate random variables, rejection sampling – Expectation: Mean, variance and covariance, conditional expectation.

REFERENCES:

1. G. Strang "Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition", USA, 2016.
2. Bendat, J. S. and A. G. Piersol. "Random Data: Analysis and Measurement Procedures. 4th Edition". John Wiley & Sons, Inc., NY, USA, 2010
3. Montgomery, D. C. and G. C. Runger. "Applied Statistics and Probability for Engineers. 5th Edition". John Wiley & Sons, Inc., NY, USA, 2011.
4. David G. Luenberger . "Optimization by Vector Space Methods", John Wiley & Sons (NY), 1969.
5. Cathy O'Neil and Rachel Schutt . "Doing Data Science", O'Reilly Media, 2013.
6. Kenneth H. Rosen: "Discrete Mathematics and Its Applications", McGraw-Hill, 7e, 2011.
7. Joe L. Mott , Abraham Kandel & Theodore P. Baker: "Discrete Mathematics for Computer Scientists & Mathematicians", PHI, 2e, 2002.
8. Michael Mitzenmacher and Eli Upfal; "Probability and Computing, 2e", Cambridge University Press, 2017.
9. Alan Agresti, Christine A. Franklin and Bernhard Klingenberg; "Statistics: The Art and Science of Learning from Data, 4e", Pearson, 2017.

PGD01C02: Foundation of Data Science and Algorithm Design

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

- Understand the evolution of data science
- Learn the data collection and preprocessing strategies
- Develop the ability to build and assess data-based models.
- Design algorithms to perform operations with linear and non-linear data structures.
- Apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively.

Module -1

Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields, Data Collection and Data Pre-Processing.

Module -2

Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Module -3

Exploratory Data Analytics: Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map, Model Development: Simple and Multiple Regression.

Module -4

Model Evaluation using Visualization – Residual Plot – Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making. Model Evaluation: Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Underfitting and Model Selection.

Module -5:

Overview of Data Structures: Linear and Non-Linear-Algorithm Analysis: Importance of analysis of algorithms-Time and Space Complexity-Basic algorithm designing techniques (Brute Force Approach, Divide and Conquer approach, Greedy method, Dynamic Programming, Backtracking)

REFERENCES:

1. W. McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy and iPython, 2nd Ed"., O'Reilly, 2017.
2. P. Tan, M. Steinbach, AKarpatne, and V. Kumar, :Introduction to Data Mining, 2nd Ed"., Pearson Education, 2018
3. G James, D Witten, T Hastie and R Tibshirani,"An Introduction to Statistical Learning with Applications in R", Springer Texts in Statistics, Springer, 2013.
4. Cathy O'Neil and Rachel Schutt, "Doing Data Science", O'Reilly, 2015.
5. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics", EMC 2013
6. Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekaran: Fundamentals of Computer Algorithms, 2nd Edition, Universities Press, 2007.
7. V.Aho, J.E. Hoperaft, J.D.Ullman, The design and Analysis or Computer Algorithms, Addison Wesley, 1974.

PGD01C03: Machine Learning for Data Science

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

- Have a strong foundation for machine learning
- Understand and learn the differences between supervised and unsupervised learning
- Learn the reinforcement learning

Module -1

Introduction: Machine Learning Foundations – Overview – Design of a Learning System – Types of Machine Learning – supervised learning, unsupervised learning, reinforcement learning – Setting up your machine learning platform: training, validation and testing, over-fitting and under-fitting, different types of error calculation – Applications of Machine Learning.

Module -2

Supervised Learning: Introduction, Simple Linear Regression, Multiple Linear Regression, Polynomial Regression, Evaluating Regression Models-Model Selection, Classification, Logistic Regression-Decision Trees-Random Forests-Naive Bayes-Support Vector Machines- Evaluating Classification Models: Understand evaluation matrics like Accuracy-Precision-Recall-F1-score and ROC-AUC.

Module -3

Unsupervised Learning: Introduction, Clustering, K-Means Clustering – Hierarchical Clustering-Density-Based Clustering – Dimensionality Reduction – principal component analysis, linear discriminant analysis, canonical correlation analysis**.**

Module -4

Association Rule Mining and Reinforcement Learning: Association Rule Mining – Apriori – FP Growth-Eclat – Reinforcement Learning – Basics-Marcov Decision Processes-Upper Confidence Bound – Thompson Sampling – Q-Learning.

Module- 5

Introduction to Artificial Neural Network: Understanding the brain, perceptron, Multi-Layer perceptron as universal approximator, general architecture of artificial neural network, feed forward and backpropagation, Convolutional Networks, and Recurrent Neural Networks.

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 Mitchell, "Machine Learning", McGraw-Hill, 1997.

5. Stanford Lectures of Prof. Andrew Ng. 6. NPTEL Lectures of Prof. B.Ravindran.

PGD01C04: Data Analytics and Prediction

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

- Understand predictive modeling techniques for data analytics
- Apply data preprocessing techniques for big data
- Measure the performance of data classification and regression models
- 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, Over-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 Grolemund, “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

PGD01C05: Computational Methods for Data Science

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

- Understand different Statistical methods and their applications
- Have an idea about time-frequency analysis
- 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.

PGD01P06: R for Data Science

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

- Master the use of the R interactive environment.
- Expand R by installing R packages.
- Develop loop constructs in R.
- Use R for descriptive statistics.
- Use R for inferential statistics.

Course content:

Introduction to the statistical software R, Data objects in R, creating vectors, creating matrices, manipulating data, Accessing elements of a vector or matrix, Lists, Addition, Multiplication, Subtraction, Transpose, Inverse of matrices. Read a file. Boolean operators. R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options; Multiple plots in a single graphic window, Adjusting graphical parameters. 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:

1. Programs on data types in R.
2. Built-in functions in R.
3. Creating and manipulating vectors.
4. Creating and manipulating matrices.
5. Operations on Data Frames/Lists in R.
6. Operators in R.
7. Programs on looping constructs.
8. Customizing and Saving Graphs in R.
9. Plot functions in R.
10. Probability Distributions.
11. Correlation and Regression Models
12. Reading and Writing Different Types of Datasets
13. Data Cleaning, Measures of Variability
14. Classification and clustering models
15. Data Visualization in R

PGD01P07: Python for Data Analytics and Machine Learning

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

- To understand and write Python programs
- To Understand the OOPS concepts of Python
- To understand the concepts of File operations and Modules in Python
- Implementation of lists, dictionaries, sets and tuples as programs
- To develop web applications using Python

List of Programs :

Machine Learning

1. Numpy and Scipy
2. Data Manipulation Using Pandas
3. Data Visualization:matplotlib
4. Decision Tree Classification
5. Plot Decision Surface of Decision Tree
6. Naive Bayes Classifier
7. Plot Different SVM Classifiers
8. Support Vector Machine
9. K-Nearest Neighbors (KNN) Algorithm

Basic Python Programs

1. Python Arrays
2. Python Lists
3. Python Matrix
4. Python String
5. Python Dictionary
6. Python Tuples
7. Searching and Sorting
8. Pattern Printing
9. Python date and Time
10. File Handling in Python
11. Program to find the longest word from a sentence or text
12. Programs for creating dynamic and interactive web pages using forms.

SEMESTER-II

Elective 1:

PGD02E01: Artificial Intelligence

PGD02E02: Natural Language Processing

PGD02E03: Advanced Machine Learning

PGD02E04: Deep Learning for computer vision

Elective 2:

PGD02E05: Data Mining

PGD02E06: Big Data Analytics

PGD02E07: Data Visualization

PGD02E08: Business Analytics

PGD02E01: Artificial Intelligence

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

- Explain the basic concepts of Artificial Intelligence
- Apply Constraint satisfaction problems to solve various standard problems in AI
- Compare the performance of heuristic techniques for a given problem
- 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.

PGD02E02: Natural Language Processing

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

- Identify the features and challenges of Indian languages
- Understand the morphology and parts of speech-related processing of Indian languages
- 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 Jurafskey 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.

PGD02E03: Advanced Machine Learning

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

- Understand Advanced machine learning techniques and algorithms.

Module -1

Machine Learning: Introduction, Types of machine learning, supervised learning-Basics, Overfitting the training data. Nearest Neighbor Methods, Validation: Nearest neighbor prediction, K-nearest neighbor methods, Weighted neighbor methods, the curse of dimensionality, Computational considerations, Connection to density estimation. Bayesian Classifiers, Naive Bayes classifiers, Classifiers and Error Rates.

Module -2

Linear regression: Optimization, Increasing the number of features, Overfitting and method Selection, linear classification: Characterizing a linear classifier, Training a linear classifier, Logistic regression.

Module -3

Support vector machines (SVMs), Linear SVM, Lagrangian optimization and duality, The soft margin SVM, The kernel Trick, VC dimension.

Module -4

Decision Trees: Predictor form, Training Decision trees, Decision tree classifiers, Learning Decision trees, Decision stumps. Ensemble Methods: Stacking, Bagging and Boosting. Clustering: K-means, Agglomerative, Gaussian Mixtures and EM.

Module -5

Hyperparameter Choices in Unsupervised Learning: Density Estimation and parameter choices (L2, Elbow, Links to GANs), Density based Clustering and Number of Clusters (Mean Shift, Kernel k-Means/PCA, Spectral Clustering). Hyperparameter Choices in Supervised Learning: Notions of complexity and Complexity Regularization, -Dimension reduction, early Stopping, Learning Rates, as complexity Regularization, Bayesian Approaches: complexity under Priors. Deep Learning basics: Tensorflow, fully connected networks, autodiff, stochastic gradient descent, VAE and GAN

REFERENCES:

1. Hal Daumé III, "A Course in Machine Learning", ciml.info 2012
2. Rogers and Girolami, "A First Course in Machine Learning".
3. Hastie, Tibshirani, and Friedman, "The Elements of Statistical Learning".
4. Barber, "Bayesian Reasoning and Machine Learning"
5. Tom M. Mitchell, "Machine Learning", McGrawHIll
6. Chrisopher Bishop," Pattern Recognition and Machine Learning", Springer, 2006.

PGD02E04: Deep Learning for Computer Vision

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

- Perform Feature detection
- Understand CNN and visualization using CNN

Module -1

Introduction to Computer Vision - Image Filtering - Interest Point Detection - Feature Extraction - Geometric features - SIFT, SURF, HOG, WLD, LBP. Recognition: Geometry-based – Appearance based. Applications: Object recognition - Face recognition. Implementation: Object Recognition using hand-crafted features and classifiers.

Module -2

Neural Networks – Basic concepts of artificial neurons, single and multi-layer perceptron, perceptron learning algorithm, its convergence proof, different activation functions, SoftMax cross entropy loss function.

Module -3

Convolutional Neural Networks: Building Blocks – Hyperparameter Tuning – Learning – Visualizing CNNs – Batch Normalization and Dropout – Deconvnets. Implementation: Simple Image classification using CNN.

Module -4

Transfer Learning – Pre-trained Models – Autoencoders. Implementation: Image Classification using pre-trained models/Autoencoders, Applications of Autoencoders, Representation learning.

Module -5

Generative Adversarial Network (GAN) – Attention Mechanism – YOLO. Application: Video Classification-Streaming CNN for action recognition - 3D convolution for temporal learning – Segmenting and captioning videos. Implementation: Video Classification / Summarization / Anomaly Detection using CNN.

REFERENCES:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", 2016
2. Charu C Aggarwal, "Neural Networks and Deep Learning", Springer, 2018.
3. Eugene Charniak," Introduction to Deep Learning", The MIT Press, 2019
4. Linda G. Shapiro and George C. Stockman, "Computer Vision", 1e, PH, 2001.
5. Richard Szeliski, "Computer Vision: Algorithms and Applications", Springer, 2010.
6. Simon J. D. Prince: "Computer Vision: Models, Learning, and Inference", 1e, Cambridge University Press, 2012.
7. David A. Forsyth and Jean Ponce, "Computer Vision: A Modern Approach", 2e, Pearson Education, 2011.
8. Ramesh Jain, Rangachar Kasturi and Brian G. Schunck, "Machine Vision, 1e", McGraw-Hill, 1995.

PGD02E05: Data Mining

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

- Understand the basic concepts of Data mining.
- Identify the different techniques of data preprocessing.
- Analyze patterns that can be discovered by classification and clustering.
- Understand data mining techniques of clustering.
- 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 section.

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.

PGD02E06: Big Data Analytics

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

- Understand basic concepts of Big Data, its dimensions and currently available other
  Databases.
- Describe big data analytics
- Illustrate the basics of the HADOOP Ecosystem
- Understand the elementary concepts of MapReduce

Module -1

Introduction to Big Data – definition & importance of Big Data, four dimensions of Big Data - volume, velocity, variety, veracity, Dypes of Data: structured data, unstructured data, Technology Foundation for Big Data. Big Data Stack – layer 0,1,2,3and 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, advanced analytics-operationalized analytics, monetization analytics- modifying business intelligence products to handle Big Data - Big Data analytics examples- Analytics solutions - text analytics - exploring unstructured data ,analysis and extraction techniques - the extracted information - text analytics tools for Big Data – New models and Approaches to support Big Data– Characteristics - 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

NoSQL databases - types - Advantages over Relational Databases - MongoDB - introduction - MongoDB philosophy - the data model - designing the database - collections - documents - data types - the _id Field - indexes - viewing available

databases and collections - opening a database - inserting data - querying for data - retrieving documents - aggregation commands - grouping results - conditional operators.

Module -5

Application of Big Data Using Pig and Hive – Data Processing Operators in Pig – Hive Services – HiveQL _Querying Data in Hive – Fundamentals of HBase and Zookeeper – Visualization – Visual data analysis Techniques, interaction techniques; Systems and applications.

REFERENCES:

1. Hurwitz, Alan Nugent, Fern Halper and Marcia Kaufman, "Big Data for Dummies", ISBN: 9781118504222.
2. Eelco Plugge, Peter Membrey and Tim Hawkins, "The Definitive Guide to MongoDB: The NoSQL Database for Cloud and Desktop Computing, 1st Edition", Apress, ISBN: 9781430230519.
3. Chris Elaton, Dirk Deroos, Tom Deutsch, George Lapis and Pual Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, 1st Edition", ISBN: B006AU BO6.
4. Garry Turkington, "Hadoop Beginner's Guide", Packt Publishing Ltd, ISBN: 1849517304.

PGD02E07: Data Visualization

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

- Understand the significance of data representation.
- Accomplish skills to represent the facts and information.
- Comprehend the methods for information visualization
- 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

PGD02E08: Business Analytics

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

- Understand the different basic concepts / fundamentals of business statistics.
- Understand the importance of measures of Descriptive statistics which includes measures of central tendency, Measures of Dispersion, Time Series Analysis, Index Number, Correlation and Regression analysis and their implication on Business performance.
- Understand the concept of Probability and its usage in various business applications.
- 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

PGD02P03 Project Work & Viva -Voce

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

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

****