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

UNIVERSITY OF CALICUT

Abstract

General and Academic IV- Faculty of Science - Regulation and syllabus of "PG Diploma in Data Science and Analytics"for Department of Computer Science, with effect from 2022 Admission- Approved, subject to ratification by the Academic Council - Orders Issued.

G & A - IV - J

U.O.No. 17790/2022/Admn Dated, Calicut University.P.O, 20.09.2022

Read:-1. UO Note No 260197/GA-IV-J1/2021/Admn dated 21.06.2022.

2. Item No. 2 in the minutes of the meeting of the Board of Studies in Computer Science & Application-PG dated 29.07.2022
3. Remarks of the Dean, Faculty of Science, dated 04.08.2022
4. Orders of the Vice Chancellor in the file of even no, dated 13.08.2022.

ORDER

1. Head of the Department, Department of Computer Science, University of Calicut, vide paper read (1) above, has forwarded the request to approve the Regulation and syllabus of "PG Diploma in Data Science and Analytics".
2. Board of Studies in Computer Science & Application (PG), held on 29.07.2022 recommended to approve the Regulation and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department, w.e.f 2022 admission, vide paper read (2) above.
3. The minutes of the Board of Studies in Computer Science & Application (PG), has been approved by the Dean, Faculty of Science, vide paper read (3) above and by the Vice Chancellor, subject to ratification by the Academic Council, vide paper read (4) above.
4. The Regulation and syllabus of "PG Diploma in Data Science and Analytics" for University Teaching Department, with effect from 2022 Admission, is therefore implemented subject to ratification by the Academic Council.
5. Orders are issued accordingly. (Syllabus appended)

Ajitha P.P

Joint 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

PG Diploma in Data Science and Analytics with effect from 2022 Admission.

Regulations and syllabus for the programme
PG DIPLOMA IN DATA SCIENCE AND ANALYTICS
w.e.f 2022 admission onwards

1. Programme description

Name of programme : PG DIPLOMA IN DATA SCIENCE AND ANALYTICS
(PGDDSA)

1.1.Introduction : Data science and data analytics have become two of the most popular buzzwords across businesses and industries, and also the most sought after and promising skills of the future. The market demand for data scientists has gone very high in recent years. This programme PG DIPLOMA IN DATA SCIENCE AND ANALYTICS(PGDDSA) (2 semesters) is intended to focus on upskilling the students so as to equip them to build a career in Data Science and related areas. This programme is designed to utilize the expertise of resource persons from Computer Science, Mathematics and Statistics from academia as well as industry.

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 90 working days inclusive of all examinations. The maximum duration permissible for completing the Programme is fixed as 3 years.

1.3 Eligibility for admission:

Masters Degree in Computer Science/ Computer Applications/ IT
OR
Graduates / post graduates in Engineering/Technology
OR
Masters Degree in Science (with mathematics/statistics at graduation level)
OR
M. Voc in Computer science related stream with, B Voc in Computer science related stream or mathematics/statistics at graduation level
of this university or Equivalent Degree of any other University / Institution, recognized by this University as equivalent thereto, with a minimum aggregate of 55% marks or equivalent grade.
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.

1.4. Selection Criteria

The selection to the programme shall be based on the marks scored by a student in

the Entrance Examination conducted by University of Calicut. The seats of this programme will be filled by strictly adhering to the community reservation as per the University norms.

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

Sl. No	Subject	Number of Questions
1.	Computer Science	25
2.	Mathematics	25
3.	Statistics	25
4.	Quantitative Aptitude and Logical Ability	25
	Total 100 questions	
Note: Each correct response will be awarded FOUR marks, and each wrong answer will be awarded Negative 1 mark.		

1.6. Attendance

The minimum attendance required 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 will be granted by the Vice-Chancellor on genuine grounds. A student who is not eligible for condonation shall repeat the semester with the subsequent batch.

1.7 Credit

A theory class of one hour per week or a practical class of two hours per week shall be counted as one credit. The minimum credits required to complete Post Graduate Diploma in Data Science and Analytics (PGDDSA) programme is 44.

2. Evaluation Scheme

2.1 The evaluation scheme for each course (paper) shall contain two parts

a) continuous assessment CA(continuous evaluation CE)

b) end semester assessment ESA(end semester evaluation ESE)

50% weight shall be given to the continuous (internal) evaluation. The remaining 50% weight shall be for the end semester examination evaluation.

2.2 Continuous Assessment (CA) or Continuous Evaluation (CE):

Evaluation of the Continuous Assessment shall be done by the facult member

in charge of the course.

Continuous Assessment (CA) of a course shall be based on components as given below.

Components of Continuous Assessment (Theory)

Sl.No	Component	Marks
1	Internal Examinations	15
2	Assignments	10
3	Seminars	5
4	Case studies/Group Discussions/viva-voce/Mini projects	10
5	Presenting papers/Invited Talk (Outside the Department)/ OR Completing MOOC in the relevant field of study. OR Publishing Papers in Conference proceedings/ publish Book Chapters/ Publishing Articles in Approved Journals	10
	Total	**50**

Seminar

Each student should select a relevant topic and prepare a seminar report, under the guidance of a faculty members as part of the internal assessment in both the semesters. Students should prepare an abstract of the topic and submit 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 a team of faculty members (Mark distribution:50% for report and 50% for presentation and discussion). The guidelines on the format of seminar report will be issued by the department.

Assignments

Each student shall be required to submit a minimum of two assignments for each course.

Internal Tests

A minimum of two internal tests will be conducted for each course.

Case studies / Viva -Voce / Mini projects/Group discussions

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

Components of Continuous Assessment (Practical)

Sl.No	Component	Marks
1	LAB skill, Quality of record work	25
2	Mini Project	15
3	Viva-voce	10
	Total	50

2.3 End-Semester Assessment (ESA) or End Semester Evaluation

The End Semester Assessment examination in Practical courses shall be conducted by the Head of the Department. All the End Semester assessment examinations in theory courses and final project shall be conducted by the Controller of Examination, University of Calicut. The evaluation of the answer scripts shall be done, by the teacher offering the course and an expert (internal/external) from the approved panel, based on a well defined scheme of valuation framed by them.

2.4 Project

The Project work in the second semester 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/institution outside the campus, then a co-guide shall be selected from this Department/Institution. 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 guide from the Department before commencement of the project.

2.4.1 Project Evaluation

(i) 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 End Semester Assessment. Guidelines for the format of project report will be issued by the department. All the project reports should be plagiarism checked using Urkund/Turnitin software and the plagiarism report generated by the software should be produced while submitting the project reports.

The valuation of project shall be jointly done by, a board of examiners (internal/external) from the approved panel, based on a well-defined scheme of

valuation framed by them. Mark Distribution will be as given below.
A student shall pass in the Project course if she/he secures 40% for ESA.
If a candidate fails in the evaluation of Project(ESA), he/she has to repeat the project course along with the next batch and undergo both CA and ESA. Unlike theory/practical courses, the CA mark will not be retained.

Sl.No	Components	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,linguistic style,references etc	15
5	Viva voce based on the project work/dissertation	20
	Total	100%

(ii) Continuous Assessment of the Project:

A Departmental committee duly constituted by the Head of the Department will review the project periodically. There shall be minimum 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 and each internal presentation shall be evaluated based on the same components given above in the ESA of project work.

2.5. Pattern of question papers in ESA

Question paper for ESA (Theory) Examination shall consist of two parts as given below. The duration of the examination will be 2 hours. Questions should be uniformly distributed with the contents from all five modules in the syllabus of the course.

General Pattern of Question Paper is as given below.

First Semester PG Diploma in Data Science and Analytics Examination – 2022
Course Code: PGDDSA 01 101
Course: Mathematical Foundations of Data Science

Time: 2 Hours Total Marks: 50

Part A
(Short essay type)

Answer any 4 Questions (out of 6) , each question carries five marks.
(4 × 5marks = 20 Marks)

1.
2.
3.
4.
5.
6.

Part B
(Essay type)
Answer any three questions (out of 5). Each question carries 10 Marks.
(3 × 10 marks= 30 Marks)

1 .
2 .
3 .
4 .
5.

3.0 Grading

The indirect absolute grading system where the marks are compounded to grades based on pre-determined class intervals and letter grades based on 10-point grading system shall be followed.

Based on the percentage of marks scored (internal and external marks put together), the students are graded in each course applying the following grading system given in Table below.

PG Diploma in Data Science and Analytics with effect from 2022 Admission.

Letter Grades with Grade Points and Marks Equivalence.

Range of marks (%)	Grade point	Letter grade
85-100	8.5 - 10.0	O (Outstanding)
75-84.99	7.5 - 8.49	A+ (Excellent)
65-74.99	6.5 - 7.49	A (Very Good)
55 – 64.99	5.5 - 6.49	B+ (Good)
50 – 54.99	5.0 - 5.49	B(Above average)
45-49.99	4.50-4.99	C(Average)
40-44	4.0-4.49	D(Fails in the course if the student scores below 40% in ESA of the course)
0-39	0	F(Failed)
-	0	Incomplete
	0	Absent

Each student shall be assigned a grade point and a letter grade in each course on the basis of the percentage of marks scored in the course (internal and external marks taken together) as shown above.

- Minimum SGPA for the successful completion of a semester : 5.00.However, a student with SGPA less than 5.0 in a semester is permitted to proceed to the next semester.
- Minimum CGPA for the successful completion of a programme : 5.00
- Appearance for Continuous Evaluation and End Semester Evaluation are compulsory for a pass in each course in the programme.
- No separate minimum is required in internal evaluation for a pass in any course, but a minimum 40% in the end semester examination evaluation is required for a pass in each course.
- Minimum grade point required for passing a course is 4.0.
- Semester grade point average (SGPA)= Sum of credit points secured in a semester / Sum of credits taken in the semesters.
- Cumulative grade point average (CGPA)= Sum of credit points secured in a programme/Sum of credits taken in the programme
- Conversion Formula: Percentage of Marks = Grade point x 10
- A student is required to pass all the core courses and the stipulated minimum number of elective courses in order to complete the programme successfully.

A student who secures zero grade point (F grade) in a course (for want of sufficient marks and/or attendance) is permitted to register for repeating the course when the course is offered to the next batch. The student registered for repeat course need not attend the classes if she/he has satisfied the requirements regarding attendance.
The CGPA of a student determines the academic level of the student in a programme and is the criterion for ranking the students. An overall letter grade (Cumulative Grade) for the whole programme shall be awarded to the student based on the value of CGPA using the same criterion given in the Table above for assigning letter grade for a course on the basis of the grade point. The minimum CGPA required for the successful completion of a programme is 5.0, which corresponds to 50% marks.
A student who does not complete the stipulated requirements of a course gets I Grade (Course Incomplete). However, such a student shall be permitted with the concurrence of the Department Council to complete the course at a later time along with the respective semester batch.

3.1 Grade Card

The University under its seal shall issue the semester wise grade card /consolidated grade statement and certificate on completion of the programme.

3.2 Award of Degree

The successful completion of all the courses (core and elective) and the compulsory project prescribed for the degree programme with CGPA of 5.0 shall be the minimum requirement for the award of the degree.

4.0 Supplementary Examinations for Failed Candidates

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.
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 afresh for evaluation. The Continuous Evaluation marks shall be freshly allotted in this case. Appearance for

Continuous Evaluation and End Semester Evaluation are compulsory.

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.

5.0 Industrial collaboration

This programme is intended to make industrial collaborations. Considering the special nature of this programme, requiring specific skills and specialized domain expertise, external experts from industry shall also be incorporated in running of this course. These external experts shall handle those sessions of the courses which require specific domain expertise. The external experts from the collaborating industry may handle at least two sessions(1 session = 2 hours) per module of each course.

PG DIPLOMA IN DATA SCIENCE AND ANALYTICS

Syllabus

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

SEMESTER-I

NO	Subject Code	Subject Title	Instructional Hrs/week			Marks			Credit
			L	P	T	CA	ESA	Total	
1	PGD01C01	Mathematical and Statistical Foundations of Data Science	3	0	1	50	50	100	3
2	PGD01C02	Foundations 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 techniques for Data Science	4	0	1	50	50	100	4
6	PGD01P06	Lab-I R for Data Science	0	6	2	50	50	100	3
7	PGD01P07	Lab II-Python for Data Analytics and Machine Learning	0	6	2	50	50	100	3
									24

SEMESTER II

NO	Subject Code	Subject Title	Instructional Hrs/week			Marks			Credit
			L	P	T	CA	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	-	12	5	80	120	200	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

Programme Outcomes

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

- PO-1: Apply the concepts in data science and analytics to solve real world problems.
- PO-2: Analyze research problems efficiently and to reach validated conclusions.
- PO-3: Design algorithms and develop solutions for solving complex problems to meet the specified needs of the society.
- PO-4: Create new methods by applying appropriate techniques in data analytics.
- PO-5: Develop good analytical and programming skills for a career in the highly competitive IT industry.
- PO-6: Contribute to contemporary key research issues in specialized areas of computational intelligence and data analytics.

PG Diploma in Data Science and Analytics, With Effect from 2022 Admission.

SEMESTER WISE SYLLABUS

SEMESTER- I

PGD01C01 : Mathematical and Statistical Foundations of Data Science

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

- Understand and apply the mathematical concepts in the field of data science.
- Solve the problems in the area of data science applying the mathematical techniques and methods.
- Apply logical thinking to understand and solve the problem in context
- Evaluate different methods of problem solving

Module -1

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

Module -2

Matrices, Matrix arithmetic, properties (determinants, traces, rank); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes.

Module -3

Graphs and graph models, subgraphs, types of graphs, paths, connectedness, and circuits. Euler and Hamilton paths. 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 -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. Rosen, Kenneth H., and Kamala Krithivasan. *Discrete mathematics and its applications: with combinatorics and graph theory*. Tata McGraw-Hill Education, 2012.
2. Strang, Gilbert, et al. *Introduction to linear algebra*. Wellesley-Cambridge Press, 2003.
3. Bendat, Julius S., and Allan G. Piersol. *Random data: analysis and measurement procedures*. John Wiley & Sons, 2011.
4. Montgomery, Douglas C., and George C. Runger. *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.
5. Luenberger, David G., and Yinyu Ye. *Linear and nonlinear programming*. Addison-wesley, 2016.
6. Aggarwal, Charu C., Aggarwal, and Lagerstrom-Fife. *Linear algebra and optimization for machine learning*. Springer, 2020.
7. Mitzenmacher, Michael, and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.

PGD01C02 : Foundations 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 and discover the potential of the same in different domains
- Comprehend the data collection and preprocessing strategies
- Develop the ability to build and evaluate different models.
- Analyze different data science concepts and devise new methods to solve problems in real-world contexts.
- Evaluate different approaches in data science in real world problem solving.
- Design an algorithm to solve a real world problem by employing the datascience concepts and techniques.

Module -1

Introduction: 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.

Module -5

Model Evaluation: Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Underfitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

REFERENCES:

1. McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* " O'Reilly Media, Inc.", 2012.
2. James, Gareth, et al. *An introduction to statistical learning*. springer, 2013.
3. Kroese, Dirk P., et al. *Data science and machine learning: mathematical and statistical methods*. Chapman and Hall/CRC, 2019.
4. O'Neil, Cathy, and Rachel Schutt. *Doing data science: Straight talk from the frontline.* " O'Reilly Media, Inc.", 2013.
5. Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.

PGD01C03: Machine Learning for Data Science

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

- Understand and analyse different aspects of machine learning
- Analyse different approaches in supervised and unsupervised learning
- Distinguish between the reinforcement learning and other learning techniques
- Evaluation of different algorithms and models used in machine learning
- Design algorithms based on learning methods to solve a real world problem

Module -1

Introduction: Machine Learning Foundations – Overview – Design of a Learning System – Types of Machine Learning – Supervised Learning and Unsupervised Learning – Mathematical Foundations of Machine Learning – Applications of Machine Learning.

Module -2

Supervised Learning - Simple Linear Regression – Multiple Linear Regression – Polynomial Regression – Ridge Regression – Lasso Regression – Evaluating Regression Models – Model Selection – Bagging – Ensemble Methods.

Module -3

Supervised Learning - II: Classification – Logistic Regression – Decision Tree Regression and Classification – Random Forest Regression and Classification – Support Vector Machine Regression and Classification - Evaluating Classification Models.

Module- 4

Unsupervised Learning: Clustering – K-Means Clustering – Density-Based Clustering – Dimensionality Reduction – Collaborative Filtering.

Module -5

Association Rule Learning and Reinforcement Learning: Association Rule Learning – Apriori – Eclat – Reinforcement Learning – Upper Confidence Bound – Thompson Sampling – Q-Learning.

REFERENCES:

1. Bishop, Christopher M., and Nasser M. Nasrabadi. *Pattern recognition and machine learning*. Springer, 2006.

2.Unpingco, José. *Python for probability, statistics, and machine learning*. Second Edition, Springer, 2019.

3. Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.

3. Alpaydin E. "Introduction to Machine Learning", PHI, Third Edition, 2015.

4. Mitchell, Tom M. "Key ideas in machine learning."Mc-Graw Hill, 2017.

5. Flach, Peter. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press, 2012.

6. Knox, Steven W. *Machine learning: a concise introduction*. John Wiley & Sons, 2018.

PGD01C04 : Data Analytics and Prediction

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

- Apply data preprocessing techniques for big data
- Review predictive modeling techniques in data analytics
- Analyze and evaluate data classification and regression models
- Design algorithms based on Classification Trees and Rule-Based Models in real world problems with big data

Module -1

Predictive analytics - overview and its uses. Predictive Models-different types. Data analysis - Descriptive analysis , Diagnostic analysis, Predictive analysis, Prescriptive analysis. Steps in Data analysis.

Module -2

Regression analysis-outliers, multi colinearity, Heteroscedasticity,Bias,Variance,Underfitting, Overfitting. Types of regression models. Measuring Performance in Regression Models, Penalized regression Models, Multivariate Adaptive Regression Splines. Support Vector Machines. K-Nearest Neighbour.

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

classification problems - class output, probability output. Evaluation metrics for classification model - Confusion Matrix, Precision, Recall or Sensitivity, Specificity, F1-Score, Log loss, AUC - ROC Curve. Classification Trees and Rule-Based Models, Regression Model Trees, Bagged Trees, Random Forests, Boosting, Class imbalance issues and solutions. Factors influencing Model Performance.

Module -5

Neural networks and predictive data analytics. Structure of neural network for prediction-weights, activation function, input layer, hidden layer, output layer. Working of neural networks. Applications of neural networks. Types of neural networks - Perceptron, Feed Forward Neural Networks, Multilayer Perceptron, Convolutional neural network,
Radial basis function neural networks, Recurrent neural networks, LSTM networks, Modular neural network. Improving the learning rate in neural networks. Cost function. regularization methods. Neural networks vs Deep learning

REFERENCES:

1. Kuhn, Max, and Kjell Johnson. "Applied Predictive Modeling". Springer, 2018.
2. Runkler, Thomas A. *Data analytics*. Springer, 2020.
3. Dietrich, D., B. Heller, and B. Yang. "Data Science & Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data". Wiley, 2015.
4. Wickham, Hadley, and Garrett Grolemund. *R for data science: import, tidy, transform, visualize, and model data.* O'Reilly Media, Inc. 2016.
5. Grus, Joel. *Data science from scratch: first principles with python*. O'Reilly Media, 2019.
6. Rahlf, Thomas. *Data visualisation with R: 100 examples*. Springer, 2017.
7. Aggarwal, Charu C. "Neural networks and deep learning." *Springer*,2018.
8. Mehlig, Bernhard. *Machine Learning with Neural Networks: An Introduction for Scientists and Engineers*. Cambridge University Press, 2021.

PGD01C05 : Computational techniques for Data Science

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

- Analyse different Statistical methods and their applications
- Summarise the potential and applications of time-frequency analysis
- Comprehend the image processing methodologies
- Analyse PCA ad apply the same to a real problem
- Design an algorithm to perform image processing with respect to a real problem

Module -1

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

Module -2

Time-Frequency Analysis - Basics of Fourier Series and the Fourier Transform, FFTs, FFT Application: Radar Detection and Filtering, Radar Detection and Averaging, Windowed Fourier Transforms, 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: L1 Norm, Signal Reconstruction, Image Reconstruction from Sparse Sampling. Basic operations in Image processing and analysis using Python - image reading, writing, converting, processing pixels with logical operators, masking, plotting histograms, filtering.

REFERENCES:

1. Kutz, J. Nathan. Data-driven modeling & scientific computation: methods for complex systems & big data. Oxford University Press, 2013.
2. Carlton, Matthew A. "Data analysis: Statistical and computational methods for scientists and engineers." Springer,2000.
3. Govaert, Gérard, ed. Data analysis. John Wiley & Sons, 2013.
4. Charbit, Maurice. Digital Signal Processing with python programming. John Wiley & Sons, 2017.
5. Gonzalez, Rafael C. Digital image processing. Pearson, 2018.
6. Dey, Sandipan. Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data. Packt Publishing Ltd, 2018.
7. Pinoli, Jean-Charles. Mathematical Foundations of Image Processing and Analysis. John Wiley & Sons, 2014.
8. Kinser, Jason M. *Image Operators: Image Processing in Python*. CRC Press, 2018.

PGD01P06 : R for Data Science

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

- Experimenting with R and its packages
- Apply R programming constructs in statistical computations and graphical representations
- Write programs for implementing machine learning algorithms
- Develop R programs for data analysis and statistical inferences

Course content:

Introduction to the statistical software R, Operators, looping constructs, Data objects in R-vector, list, matrix, array, factor, and data frame. creating data objects,reading data, manipulating data, processing data. Reading files, processing data,writing files. Boolean operators.R packages for Datascience -Dplyr, ggplot2, KernLab,DataExplorer, Caret, randomForest, Shiny, mboost, Plotly, SuperML. R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options, Multiple plots in a single graphic window, Adjusting graphical parameters.

Lab exercises

1. Programs using different data types in R.
2. Programs with Built-in functions in R.
3. Programs using Arithmetic and logical Operators in R.
4. Programs using looping constructs.
5. Programs with operations on Dataobjects
6. Programs for computations of Measures of central tendency and variability
7. implementations of Probability Distributions.
8. Finding Correlations and develop Regression Models
9. Reading and Writing Different Types of Datasets
10. Implementing Classification and clustering models
11. Data Visualization using R

REFERENCES

1. Gillespie, Colin, and Robin Lovelace. *Efficient R programming: a practical guide to smarter programming*. " O'Reilly Media, Inc.", 2016.
2. Irizarry, Rafael A. *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press, 2019.
3. Bruce, Peter, Andrew Bruce, and Peter Gedeck. *Practical statistics for data scientists: 50+ essential concepts using R and Python*. O'Reilly Media, 2020.
   Miller, Thomas W. *Modeling techniques in predictive analytics with Python and R: A guide to data science*. FT Press, 2014.

PGD01P07 : Python for Data Analytics and Machine Learning

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

- Develop Python programs for computations
- Develop python programs for data analysis making use of different libraries
- Write python programs for data visualization using different libraries
- Develop web applications using Python

Course Content:

Python - Overview and characteristics. Basics of python laguage -Datatypes, operators, programming constructs, Built-in functions. libraries - Pandas, NumPy, Scikit-Learn, SciPy, matplotlib, seaborn, PyTorch, Keras, TensorFlow. Data acquisition, Analysis and Visualisation in python. Regression, Classification and Clustering using python. Application of K-Means clustering, Support Vector Machines (SVM), KNN, Decision Trees, Naive Bayes, and PCA in python.Building artificial neural networks with Tensorflow and Keras. Classify images, data, and sentiments using deep learning.

Lab exercises:

1. Programs using different datastructures for computation
2. Data Manipulation Using Pandas
3. Reading, processing and writing of csv files
4. Computations with Numpy and Scipy
5. Data Visualization with libraries like matplotlib, seaborn
6. Statistical data analysis using python
7. Regression and classification using python
8. Implement k-means clustering with real data
9. Image recognition and classification using tensorflow/keras
10. Programs for creating dynamic and interactive web pages using forms

REFERENCES

1. Lubanovic, Bill. *Introducing Python: Modern Computing in Simple Packages.* " O'Reilly Media, Inc.", 2014.
2. Kumar, U. Dinesh, and Manaranjan Pradhan. Machine Learning using Python. Wiley, 2020.
3. McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* " O'Reilly Media, Inc.", 2012.
4. Rogel-Salazar, Jesus. *Data Science and Analytics with Python.* Chapman and Hall/CRC, 2018.
5. Brown, Martin."Python- the complete reference". McGrawHill, 2018.

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

- Summarize the concepts and applications of Artificial Intelligence
- Apply AI techniques to solve real world problems
- Develop techniques based on heuristics to solve problems
- Illustrating Soft Computing Techniques
- Analyze scope of AI in image processing steps
- Identify the challenges in NLP

Module -1

Introduction to artificial intelligence - Artificial Intelligence- Definitions, Programming Methodologies, Techniques, Intelligent Systems, Propositional calculus, Predicate Calculus, Rule Based Knowledge Representation. Unification, Resolution, Constraint Satisfaction Problem

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 -Applications of Search Techniques in Game Playing- Minimax strategy and Alpha Beta Pruning, and Planning

Module -4

Artificial Neural Networks, Fuzzy Sets & Fuzzy Logic, Rough Set Theory, Swarm Intelligence – Evolutionary Algorithms – Genetic Algorithms.

Module -5

Perception –Image Formation, Early Image-Processing Operations, Object Recognition by Appearance, Object Recognition from Structural Information. Introduction to Recommender Systems -Case study. Real time application. Natural Language Processing- Language Models, Text Classification, Information Retrieval, Information Extraction. Robotics- Robot Hardware, Robotic Perception, Robotic Software Architecture.

REFERENCES :

1. Leszek, Rutkowski. "Computational Intelligence: Methods and Techniques."Springer, 2008.
2. Aggarwal, Charu C. "An introduction to recommender systems." *Recommender systems.* Springer, 2016.
3. Russell, Stuart, and Peter Norvig "Artificial intelligence: a modern approach." Pearson,2021.
4. Rich, Elaine, and Kevin Knight. *Artificial Intelligence.* McGraw Hill, 2018.
5. Padhy, N. P. "Artificial Intelligence and Intelligent Systems". Oxford Univeristy Press, 2007.
6 Neapolitan, Richard E., and Xia Jiang. *Artificial intelligence: With an introduction to machine learning.* CRC Press, 2018.

PGD02E02: Natural Language Processing

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

- Illustrate NLP and all its aspects
- Compare the features and identify the challenges of Indian languages
- Analyze the morphology and parts of speech related processing of Indian languages
- Evaluate the Probabilistic models for natural language processing

Module -1

Overview - History-importance of NLP- Higher level capabilities of NLP- The working 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. Jurafsky, Dan. *Speech & language processing*. Pearson Education India, 2013.
2. Indurkhya, Nitin, and Fred J. Damerau. *Handbook of natural language processing*. Chapman and Hall/CRC, 2010.
3. Manning, Christopher and Schutze, Hinrich. Foundations of Statistical Natural Language Processing. MIT Press, 2000.
4. Hausser, Roland, and R. Hausser. *Foundations of computational linguistics*. Springer, 2001.
5. Vajjala, Sowmya, et al. *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly Media, 2020.
6. Kamath, Uday, John Liu, and James Whitaker. *Deep learning for NLP and speech recognition*. Springer, 2019.

PGD02E03: Advanced Machine Learning

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

- Understand all aspects of machine learning techniques.
- Applying regression to analyze variety of relationships
- Evaluate different models for classification
- Develop classifiers for given data
- Design program for clustering for real world data

Module -1

Machine Learning - Introduction and overview. Machine learning, Deep learning,Neural networks. Applications of ML in real world and challenges. Types of machine learning. supervised learning, unsupervised learning, Reinforcement learning.

Module-2

Training, testing and validation of data. Overfitting and underfitting of models. Regression Analysis. Linear regression. Simple linear regression. Multiple linear regression. Logistic regression.

Module -3

Nearest Neighbor Methods - Nearest neighbor prediction, K-nearest neighbor methods, Weighted neighbor methods, dimensionality and related issues, Computational considerations, Connection to density estimation. Bayesian Classifiers, Naive Bayes classifiers, Classifiers and Error Rates.

Module -4

Support vector machines (SVMs), Advantages of SVM, Types of SVM, Lagrangian optimization and duality, The soft margin SVM, Use of kernels, Vapnik–Chervonenkis (VC) dimension.

Module -5

Decision Trees- introduction , structure of DT, CART algorithm, working of decision trees, advantages of decision trees. Decision tree classifiers, Learning Decision trees, Decision stumps. Ensemble Methods: Stacking, Bagging and Boosting. Clustering: K-means, Agglomerative, Gaussian Mixtures and Expectation Maximization(EM).

REFERENCES:

1. Deisenroth, Marc Peter, A. Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.
2. Rogers, Simon, and Mark Girolami. *A first course in machine learning*. Chapman and Hall/CRC, 2016.
3. Unpingco, José. *Python for probability, statistics, and machine learning*. Springer, 2019.
4. Knox, Steven W. *Machine learning: a concise introduction*. John Wiley & Sons, 2018.
5. Barber, David. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
6. Bishop, Christopher M., and Nasser M. Nasrabadi. *Pattern recognition and machine learning*. springer, 2010.
7. Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2015.

PGD02E04: Deep Learning for computer vision

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

- Develop algorithm for Feature detection
- Analyze the structure and working of CNN
- Develop program for Object Recognition using hand-crafted features and classifiers.
- Design algorithm for Image Classification using minimal neural network.
- Develop program for Image classification using CNN.
- Illustrate Image Classification using pre-trained models/Autoencoders.
- Demonstration of Video Classification / Summarization / Anomaly Detection 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 of Object Recognition using features and classifiers.

Module -2

Neural Networks – Stochastic Gradient Descent – Backpropagation – Logistic Regression – Softmax. Implementation of Image Classification using minimal neural network.

Module -3

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

Module -4

Transfer Learning – Pre-trained Models – Autoencoders. Implementation of Image Classification using pre-trained models/Autoencoders.

Module -5

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

REFERENCES :

1. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
2. Aggarwal, Charu C. "Neural networks and deep learning." *Springer*, 2018.
3. Linda, G. "Shapiro george c. stockman. computer vision." , PHI, 2022.
4. Szeliski, Richard. *Computer vision: algorithms and applications*. Springer, 2010.
5. Prince, Simon JD. *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.
6. Forsyth, David, and Jean Ponce. *Computer vision: A modern approach*. Prentice hall, 2011.

PGD02E05: Data Mining

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

- Summarize the processes in Data mining with a wide variety of techniques.
- Identify the different techniques for data preprocessing.
- Analyze patterns that can be discovered by classification and clustering.
- Understand data mining techniques for clustering and its significance.
- Comprehend complex data types in spatial and web mining.

Module -1

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

Module -2

Data mining - introduction - steps in data mining - challenges in datamining- 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 - association rules - association mining to correlation analysis - constraint- based association mining.

Module -3

Classification and prediction - issues in classification and prediction -classification methods - classification by decision tree - Bayesian classification - rule based classification - classification by backpropagation - support vector machines - associative classification - lazy learners - prediction - accuracy and error measures - evaluation metrics- evaluating the accuracy of a classifier or predictor - ensemble methods - model selection.

Module -4

Cluster analysis - types of data in cluster analysis - different 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. Pujari, Arun K. *Data mining techniques*. Universities press, 2013.

2.Dunham, Margaret H. *Data mining: Introductory and advanced topics*. Pearson Education India, 2006.

3.Shmueli, Galit, et al. *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons, 2017.

4. Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.

5.Larose, Daniel T., and Chantal D. Larose. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

PGD02E06: Big Data Analytics

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

- Illustrate Big Data with its different dimensions and significance
- Examine the big data analysis framework
- Illustrate the significant role of the HADOOP Ecosystem in dealing with very large data
- Explaining how MongoDB helps to create and deploy a highly scalable and performance-oriented database.

Module -1

Introduction to Big Data – definition & importance of Big Data, four dimensions of Big Data - volume, velocity, variety, veracity. Importance of big 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 – 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 - load data into Hadoop - Map Reduce Basics – origins of MapReduce - map function – reduce function – Map Reduce characteristics- Map Reduce formats and types- Map Reduce applications – Working of Map Reduce – Map Reduce Features.

Module -4

NoSQL databases - types - Advantages over Relational Databases - MongoDB - introduction and overview- 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.

Big data Visualization – merits of visualization- different types of analysis for visualization-visualization Techniques, interactive data visualization techniques.

REFERENCES:

1. Zikopoulos, Paul, and Chris Eaton. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw, 2017.
2. White, Tom. Hadoop: The Definitive Guide,4th edition. OReilly,2015.
3. Dean, Jared. *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. John Wiley & Sons, 2014.
4. Membrey, Peter, et al. *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing*. springer, 2010.
5. Turkington, Garry. *Hadoop Beginner's Guide*. Packt Publishing Ltd, 2013.
6. Frampton, Michael. *Big Data made easy: A working guide to the complete Hadoop toolset*. Apress, 2014.

PGD02E07: Data Visualization

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

- Illustrate the significance of data visualization
- Summarize the methods for information visualization
- Develop python programs using matplotlib for data visulaization.

Module -1

Data and types of data, Data variability, uncertainty and context. Basics of Data Visualization: Definition of Data Visualization, Need for Visualization, Methods 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, Treemap, 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

Information Visualization methods - Cartogram, Cladogram (phylogeny), Concept Mapping, Dendrogram (classification). Graph drawing, Heat map, Hyperbolic Tree, Tree mapping, Multidimensional scaling. Information visualization reference model.

Module -5

Scientific visualization: Introduction, Methods for visualizing two-dimensional and three-dimensional data sets, volume visualization. Data Visualization using Python matplotlib commands for plots - plots of basic types, plots for arrays of data and fields, plots for statistical analysis, plots for unstructured coordinates to visualize contours.

REFERENCES :

1. Vanderplas, Jacob T. *Python Data Science Handbook: Tools and Techniques for Developers.* O'Reilly, 2016.
2. Healy, Kieran. *Data visualization: a practical introduction.* Princeton University Press, 2018.
3. Chang, Winston. *R graphics cookbook: practical recipes for visualizing data.* O'Reilly Media, 2018.
4. Bederson, Benjamin B., Benjamin B. Bederson, and Ben Shneiderman, eds. *The craft of information visualization: readings and reflections.* Morgan Kaufmann, 2003.
5. Mazza, Riccardo. *Introduction to information visualization.* Springer Science & Business Media, 2009.
6. Milovanovic, Igor, Dimitry Foures, and Giuseppe Vettigli. *Python Data Visualization Cookbook.* Packt Publishing Ltd, 2015.
7. VanderPlas, Jake. *Python data science handbook: Essential tools for working with data.* " O'Reilly Media, Inc.", 2016.
8. Yau, Nathan. *Data points: visualization that means something.* John Wiley & Sons, 2013.
9. Wilke, Claus O. *Fundamentals of data visualization: a primer on making informative and compelling figures.* O'Reilly Media, 2019.

PGD02E08: Business Analytics

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

- Identify and describe complex business problems in terms of analytical models
- Apply Descriptive statistics to get basic information about variables in a dataset and to highlight potential relationships between variables.
- Discover the underlying causes of trends or systemic patterns over time applying time series analysis
- Evaluate statements by performing hypothesis tests
- Solving problems by making decisions applying business analytics

Module -1

Concept of Business Analytics- application of Business Analytics. Descriptive analytics, Predictive analytics,Prescriptive analytics. Domains of business analytics -Human resource analytics, Supply chain analytics, Customer analytics, Business process analytics,
Financial analytics. Business analytics process- business problem framing, analytics problem framing, data, methodology selection and model building, deployment. Business analytics vs Business intelligence.
Data analysis-Exploratory Data Analysis (EDA) - Data Collection , Data Cleaning, Data Preprocessing, Data Visualization. EDA techniques.

Module -2

Quantitative data analysis - 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.
Index numbers -Role of Index Numbers in data analysis, Characteristics of index numbers, Uses, Types of index numbers, Methods of Construction of index numbers - unweighted index, weighted index.

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.
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.

Module -4

Probability Theory and Distributions - Probability: Theory of Probability, Addition and Multiplication Law, Bayes' Theorem, Theoretical Probability Distributions: Concept and applications of Binomial, Poisson and Normal distributions.
Hypothesis Testing- Null and Alternative Hypotheses, Type I and Type II errors, Testing of Hypothesis. Large Sample Tests and Small Sample tests. Parametric tests and non parametric tests.t-test, z- test, ANOVA test, chi-square test.

Module -5

Decision making and simulation- Introduction, Overview of the simulation process, Simulation to model uncertainty, Monte Carlo simulation and random variables, Random number generation functions in R, simulation and decision making - case study, Data modeling using spreadsheets.

REFERENCES :

1. Black, Ken. Applied Business Statistics - Making Business Decisions. 2016, Wiley.
2. Ledolter, Johannes. *Data mining and business analytics with R.* John Wiley & Sons, 2013.
3. Levin, Richard I. *Statistics for management.* Pearson Education India, 2011.
4. Lind, Douglas A., William G. Marchal, and Samuel A. Wathen. *Statistical techniques in business & economics.* McGraw-Hill Education, 2017.
5. Camm, Jeffrey D., et al. *Essentials of business analytics.* Cengage Learning, 2016.
6. Walpole, Ronald E., et al. *Probability and statistics for engineers and scientists.* Macmillan, 2016.
7. Provost, Foster, and Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking.* " O'Reilly Media, Inc.", 2013.
8. Minelli, Michael, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses.* John Wiley & Sons, 2013.
9. Whigham, David. *Business data analysis using Excel.* Oxford University Press, 2007.

PGD02P03 Project

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

- Develop inventive and creative solutions to data research problems.
- Create and apply efficient methods to solve problems.
- Demonstrate a sound technical knowledge on the selected topic.
- Articulate discussions on the concerned area.
- Communicate the findings clearly in oral and written form.

Proposed Fee structure for the course

Sl.No	Fee details	Amount (Rs)
1	Registration fee (Application Fee)	1000
	For SC/ST students	300
2	Admission Fee	11000
3	Tuition Fee (Per Semester)	19, 800
4	Caution Deposit (Refundable)	3,000
5	Laboratory Fee (Per Semester)	3,000
6	Special Fees	1365
	-Association fee -60	
	-Sports affiliation fee-280	
	-Quasi University Fee-25	
	-University Union Fee-85	
	-Library Fee-115	
	-Stationery Fee-60	
	-Magazine Fee-115	
	-Athletic Fee-115	
	-Audio Visual Fee-35	
	Calendar Fee-40	
	Student Aid fund-15	
	Women Cell Fee-60	
	Medical Inspection Fee-20	
	Union Fee-115	
	Student Welfare Fund-225	