

A course titled Computational Data Analysis (10 credits) was introduced in B.Tech. Semester 1 from 2018-2022 batch following Project based Learning (PBL) approach. This document includes incremental updates of three cycles, 2018-22, 2019-23 and 2020-24 batch respectively.

Iteration 1 of Computational Data Analysis

First Course File

Course Name: Computational Data Analysis

Faculty Name(s): Dr. Sonal Jain, Dr. Umesh Gupta, Dr. Kavita Choudhary

Faculty Lead: Dr. Sonal Jain

Target Start Date: Second Week of Dec-18

Course Description: This course introduces computational analysis of data based on Linear Algebra Principles and Statistics. The computational analysis will include learning and utilizing Python as a programming language. This course will lead to a technical project that will include learnings from the course duration.

Course Goals:

Able To Do:

- To enable the students with modelling engineering problems based on Linear Algebra principles and Statistics
- To enable the students to learn concepts and tools for working with data and have experience in analyzing real data using Statistics
- To enable the students to learn Python programming language and to use this in solving data related technical problems

Syllabus

Introduction to Algorithms, Hardware Overview, Python as a Tool, Installing Python and Writing a Program, Variables & Expressions, Decision Statements, How to Debug?, Control Structures: Loops & Iterations, Linear Data Structure: String, List, Tuple, Data Dictionary and Set, Python Library (Pandas, Numpy, PyPlot), Functions, Classes & Objects, Working with Files

Matrix Operations, Eliminations, Matrix Inversion, Transformation, Solution of Linear , Simultaneous Equation, Eigen Values & Eigen Vectors , Linear Transformation, Linear Combination, Vector Spaces and Subspaces

Probability, Baye's Rule, Sampling, Data Processing and Pre-processing, Random Variable, Discrete & Continuous Distribution, Hypothesis Formulation , Test of Hypothesis, ANOVA, Correlation, Curve Fitting, Regression, Time Series Analysis, Forecasting

Reference Books

1. Allen B. Downey. Think Python. Green Tea Press, Massachusetts, USA.
2. Kenneth Hoffman and Ray Kunze. Linear Algebra. PHI Learning Private Limited, 2nd Edition, 2012.
3. Gilbert Strang. Introduction to Linear Algebra. Wellesley-Cambridge Press, 4th edition, 2009.
4. Allen B. Downey. Think Stats. Green Tea Press, Massachusetts, USA.



Sonal

(Dr. Sonal Jain)

5. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc., 3rd Edition (2004).
6. Rishard A. Johnson, Miller and Freund's probability and Statistics for Engineers, PHI

Domain Specific Knowledge:

Python Programming Language, Linear Algebra, Statistical Techniques

Professional Skills:

- Collaboration, Leadership, Team-work
- Idea Generation, Professional Ethics, Time Management
- Presentation Skills, Report Generation, Desire to Learn

Expectations from the Students

- To adhere to the course guidelines/course code throughout its duration
- To work in team with a significant and well-defined contribution from each individual
- To work beyond class duration in the team on projects
- Self-learning on the topics relevant to the course sessions is required pre and post sessions.

Expectations from the faculty members

- To support in the situation where students are feeling stuck in the project
- To provide supporting notes and reading material at various stages, as per requirement
- To announce the assessment score at every evaluation stage so that each student is aware about his/her progress at each stage

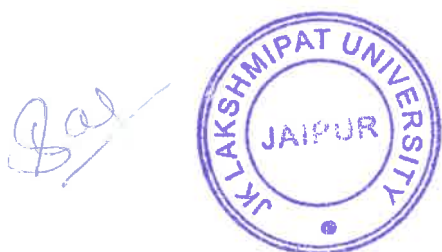
Course Feedback: Online Every Fortnight

Course Evaluation:

Component	Weightage (%)	Details	Marks (100)
Theory Examinations	20	Theory Exam 1	10
		Theory Exam 2	10
Continuous Evaluation	30	Assignment (8 Nos)	16
		Class Participation (7 Weeks)	14
Project	50	Mini Project 1	10
		Mini Project 2	10
		Final Project	30

Project Evaluation Components – 50 Marks (50%)

Attendance Marks (% Weightage)	Documentation	Time Mgmt.	Amount of Work Done	Presentation	
				Presentation Skills	Q & A
10 (20%)	10 (20%)	05(10%)	10 (20%)	10 (20%)	05(10%)



Marks = 10 * (% Attendance)

--	--	--	--	--	--

Class participation will include learning outcomes in writing, random questions from individuals in the class and surprise quizzes.

Assignment Evaluation Rubrics: Timely Submission, Presentation of Solution, Project Presentation Rubrics: Attendance, Documentation, Time Management, Amount of Work Done, Presentation

Course Execution Style: 2 Hrs. a Day, 5 Days a week

Resources Needed: Anaconda Platform (includes Server, Jupyter Notebook), Pycharm (IDE to write Python Programs), Cloud Computation Cycles as per need for some projects.

Sample Plan of Each day

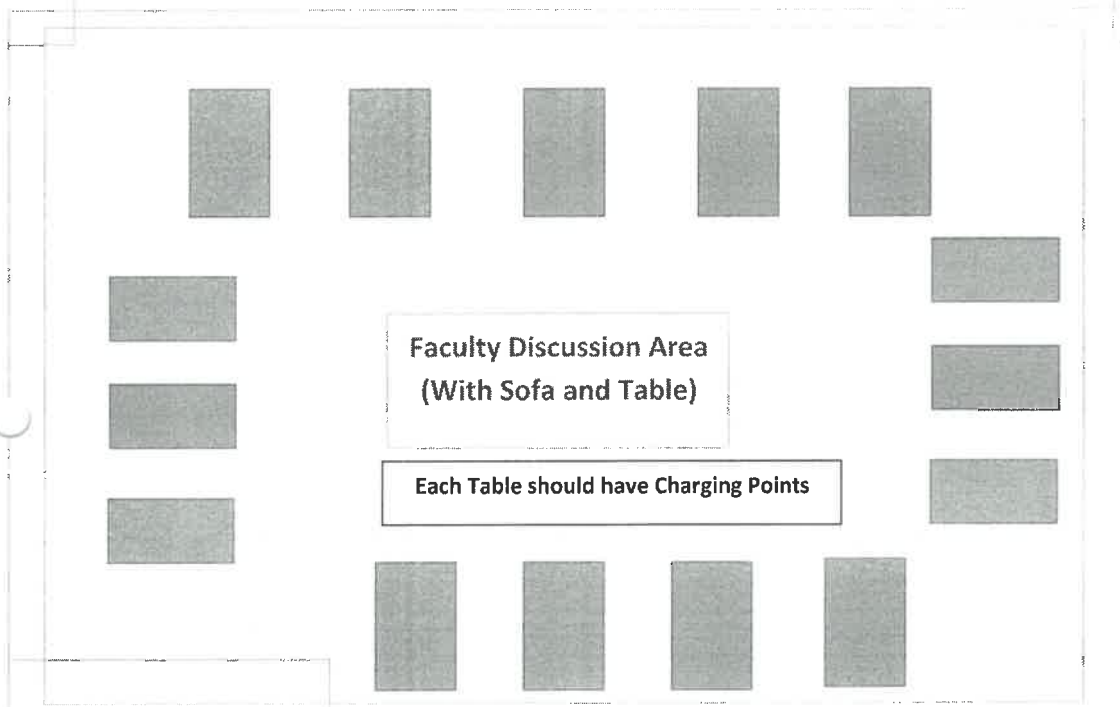
#Session 1 – Introduction

00.30	Pre-course survey and discussion
00.10	Students-faculty-Ninja introduction
00.10	Students will play tic-tac-toe game to understand Matrices
00.15	Introduction to the course
00.15	Introduction to the projects
00.40	Know your trait...

Reading Assignment: Needs and Introduction to Algorithms – Video Link: 2 Hrs

Class Setup

BOARD



List of Possible Projects

Module 1

Project 1: FIFA World Cup

Description: Team Ranking based on performance in the tournaments such as IPL, FIFA World cup, ICL etc. using Eigen value

Project 2 : Child Survival and Stable Growth Pattern

Child Survival and Stable Growth Pattern: This has been observed that one year after birth of a child remains crucial. There is high number of cases of death of a child immediately after birth. The students will study into the survival rate of child in the age group within one year by taking a data of child death within one month, three months, six months, one year and survival after one year. They will form a matrix based on these groups and apply Eigen values and Eigen vectors to find the stable growth pattern.

Project 3: Image Compression

Description: With rapid demand of storing large amount of data, the space to store this data (in the form of files) in the hard drives of the computer systems or onto the servers of big companies is getting less compared to the amount of data that is to be stored. Each image can be represented as a matrix. Apply linear algebra on this matrix and get a reduced matrix out of the original matrix. Check the Compression Ratio. Implement the Solution in Python to work with matrix of higher dimensions.

Module 2

Project 1: Factors affecting the academic performance

To prepare a questionnaire on various factors that may affect the academic performance of students and analyzing the factors and apply proper statistical method(s) to find out the appropriate factors affecting students' academic performance.

Project 2: Relationship between players' price and their performance

Description: To select favorite sport and to collect the data of players' price and their performance in past few years. To apply appropriate statistical method(s) to establish relationship between cost and performance. Does the price of players affect their performance?

Project 3: Performance comparison between two or more machines

Description: To collect the data of measurement of product designed on different machines and to compare the performance of these machines based on means and variances in such measurements. This study may be extended to further look into the data and to find out the machines that are bringing more variance in the production.



Module 3

Project 1 : Exercises and Risk Factors

Rising health insurance costs are big concerns these days. As a result many companies encourage employees to follow a healthy life style as one measure of cutting health related costs. It is believed that regular exercise is one way of reducing the risk of having a heart related disease. The goal of this project is to find out if exercise and risk factors for heart disease have some relation.

Project 2: Team Performance Analysis

Rigorous and objective analysis of scoring statistics of teams in the game tournaments such as FIFA world cup, IPL, ICL etc.

Project 3: Ranking of Cities

Living status of cities based on their GDP, Swacchhata ranking, crime against different communities, water levels, rain conditions and other parameters

Iteration 2 of Computational Data Analysis

Faculty Name(s): Dr. Sonal Jain, Dr. Umesh Gupta, Dr. Kavita Choudhary, Dr Richa Sharma, Dr Indranath Chatterjee

During second iteration of Project following changes were incorporated:

- Evaluation scheme was changed with emphasis on Theory exams along with Projects.
- Bottom-up approach of building projects was also incorporated.
- Projects titles in second iteration was based on Sustainable development Goals.
- Posters of each project were designed during second iteration.

Sample List of Projects

Group Number	Title	SDG Goals
1	Study of Factors affecting quality of Agricultural Land	15
2	Analysis of poverty and its factors in India	2,3,5,6,8,11,16
3	Assessment of air quality in most populated cities	13,15
4	Availability of fresh water	6
5	Effect of overpopulation	3,8,12,15,16
6	Evaluation of Swacch Bharat	3,6
7	Evaluation of rural & urban development w.r.t economic development	1,2,3,4,5,8,9



8	Factors of deforestation & its impact	13,15
9	Homelessness in India	1,2,3,4,6
10	Impact of refugees	1,2,3,4,5,6,8,9,10,11,16,17
11	Startups in India	8,9
12	Status of various violences in India	10,16
13	Water quality (ground water, rivers, cannals)	6,14

Iteration 3 of Computational Data Analysis

Faculty Name(s): Dr. Sonal Jain, Dr. Umesh Gupta, Dr. Kavita Choudhary, Dr Richa Sharma

During third iteration of Project following changes were incorporated:

- Depth of content related to python was increased and students were required to Practice on Python programs more than earlier two iterations.
- Students were recommended to attend online courses hosted on platforms like coursera, cognitive.ai platforms.
- Two Teaching Assistants one for each section (Full time PhD students) were involved to help students doing programming for first time.
- Each Project report in third iteration was scanned through Turnitin and only 10% plagiarism was permitted.

Course File followed in 2020-24 batch

ES1110: Computational Data Analysis

Course Description: This course introduces computational analysis of data based on Linear Algebra Principles and Statistics. The computational analysis will include learning and utilizing Python as a programming language.

Learning Outcome

After course completion, the student will be able to

1. Write Simple Python programs using various datatypes, control structures, decision statements, libraries, functions (M1)
2. Develop Python programs using Objects, Classes and Files (M1, M2)
3. Develop Programs for analyzing and interpreting Complex situations in various domains including sustainable development by combining various Linear Algebra, Statistics and Other Problem Solving Techniques (M3)



4. Model Complex systems as Linear simultaneous equations and analyze the same using Matrix methods (M1)
5. Model Data as matrices and Find Eigen Values and Eigen Vectors and Apply the same for problem solving, e.g., ranking and performance analysis (M1)
6. Summarize and Visualize different datasets (M2)
7. Analyze and interpret different datasets using Discrete and Continuous Probability Distributions and Apply the same for problem solving, e.g., Goodness of Fit (M2)
8. Formulate and validate hypothesis with reference to different datasets (M2)
9. Apply correlation, regression, least square method and time series analysis for modeling, analysis, interpretation and forecasting (M2)

Teaching Scheme and Credits

Hrs. per Week		Credits	Duration in Weeks
In Class	Out Class		
10+2	20	10	15

Evaluation Scheme

Sr. No	Specifications	Weightage (in percentage)
01	Attendance	Nil
02	Assignment (10 LA + 5 Stats.)	15
03	Class Participation	Nil
04	Quiz (3) (1 LA+ 1 stats+ 1 python) 5 marks each	15
05	Theory Exam (Mid Term I)	Nil
06	Theory Exam (Mid Term II)	20
07	Theory Exam	Nil
08	Report-1	Nil
09	Report-2	Nil
10	Report-3	Nil
11	Project -1	Nil
12	Project -2	30
13	Project -3	Nil
14	Lab Evaluation 1	10
15	Lab Evaluation 2	10

Handwritten signature



16	Course portfolio	Nil
	Total (100)	100

Syllabus

Introduction to Algorithms, Hardware Overview, Python as a Tool, Installing Python and Writing a Program, Variables & Expressions, Decision Statements, How to Debug?, Control Structures: Loops & Iterations, Linear Data Structure: String, List, Tuple, Data Dictionary and Set, Python Library (Pandas, Numpy, PyPlot), Functions, Classes & Objects, Working with Files

Matrix Operations, Eliminations, Matrix Inversion, Transformation, Solution of Linear , Simultaneous Equation, Eigen Values & Eigen Vectors , Linear Transformation, Linear Combination, Vector Spaces and Subspaces

Probability, Baye's Rule, Sampling, Data Processing and Pre-processing, Random Variable, Discrete & Continuous Distribution, Hypothesis Formulation , Test of Hypothesis, ANOVA, Correlation, Curve Fitting, Regression

Reference Books

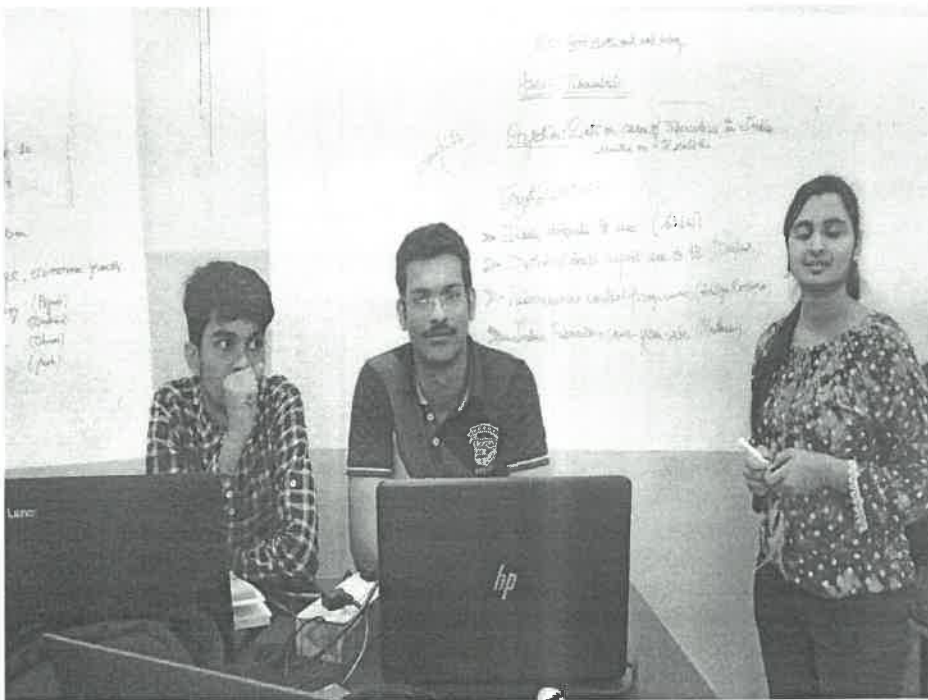
7. Allen B. Downey. Think Python. Green Tea Press, Massachusetts, USA.
8. Kenneth Hoffman and Ray Kunze. Linear Algebra. PHI Learning Private Limited, 2nd Edition, 2012.
9. Gilbert Strang. Introduction to Linear Algebra. Wellesley-Cambridge Press, 4th edition, 2009.
10. Allen B. Downey. Think Stats. Green Tea Press, Massachusetts, USA.
11. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc., 3rd Edition (2004).
12. Rishard A. Johnson, Miller and Freund's probability and Statistics for Engineers, PHI

Sample Photos



Signature





Handwritten signature or initials in blue ink.

