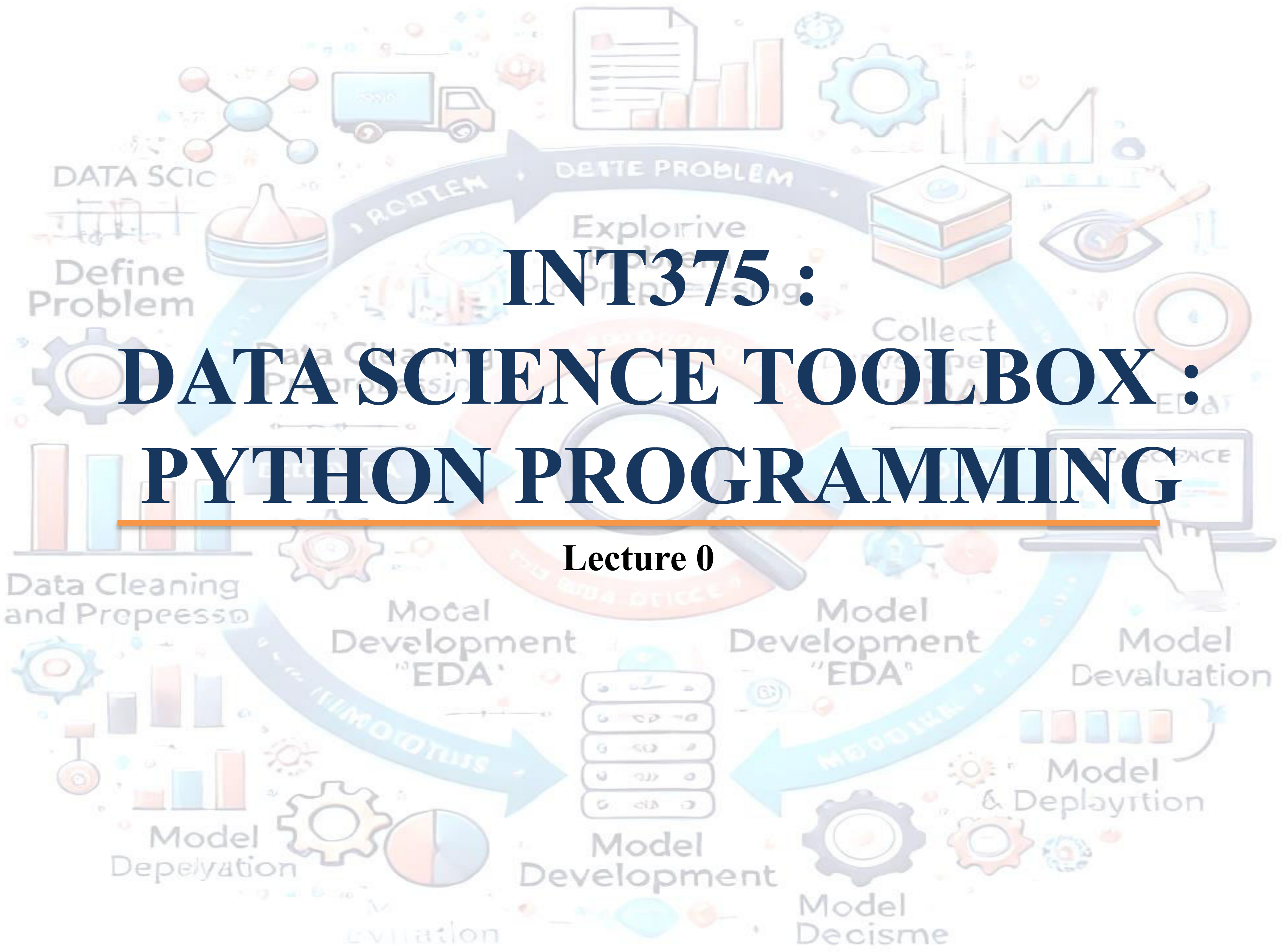
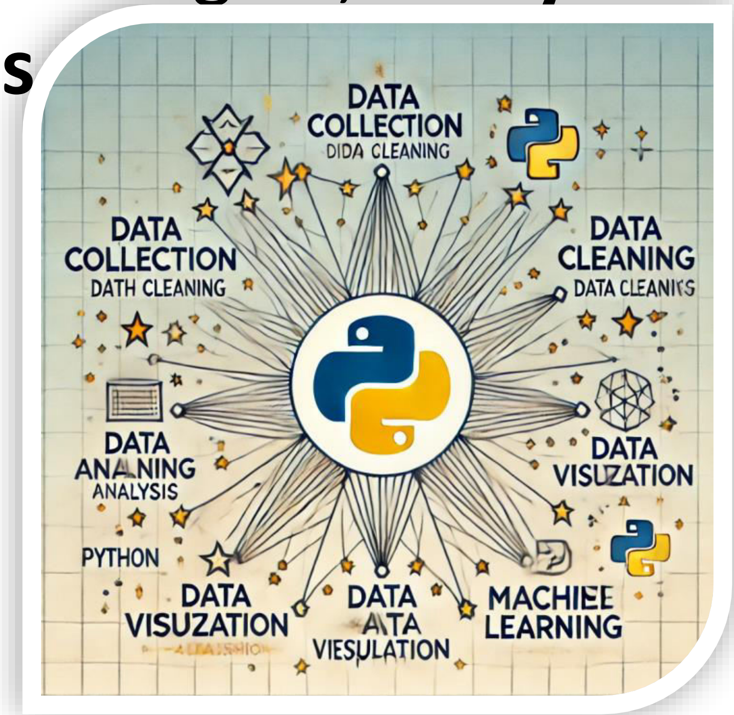


INT375 : DATA SCIENCE TOOLBOX : PYTHON PROGRAMMING

Lecture 0

Introduction

- Data Science with Python refers to the use of Python programming language and its **powerful libraries** to extract insights, **analyze data**, and **make predictions**.



Course Overview

- Course Code : INT375
- LTP – 2 0 2[Two Lectures and Two Practical's/week]
- Credits – 3

Marks Breakup

- Credits : 3
- Marks Breakup:

Activity	**Marks**
Attendance	5
Continuous Assessment	45
End-Term Practical (ETP)	50
Total	**100**

- 2 CAs, CA1-30 marks and CA2(Project)- 100 marks
- * No MTE

CA Details

- **CA1:** BYOD Practical: Scenario based questions
- **CA2:** Skill-based Assessment (Project):

  Marks based on project uploaded on Github and LinkedIn

  RUBRICS OF GAMIFICATION:

  A)Problem Statement (Objectives) and Dataset: 10 Marks

  B) Implementation(Outcome),Report and Viva:60 marks

  C)Linkedin: 10 Marks (Based on Likes (Count-50) and Comments (Count10)

  D) Github: 20 Marks on Commits, Pull, Stars

A) Problem Statement (Objectives) and Dataset *(10 Marks)*

1. Problem Statement *(5 Marks)*

2. Dataset *(5 Marks)*

B) Implementation (Outcome), Report, and Viva *(60 Marks)*

1. Implementation *(30 Marks)*

a) Data Cleaning and Visualization *(10 Marks)*

b) EDA and Statistical Analysis *(10 Marks)*

c) Creativity and Innovation *(10 Marks)*

2. Report *(10 Marks)*

a) Format *(5 Marks)*

b) Technical Writing *(5 Marks)*

3. Viva *(20 Marks)*

C) LinkedIn Engagement *(10 Marks)*

1. Likes *(6 Marks)*

2. Comments *(4 Marks)*

D) GitHub Contributions *(20 Marks)*

1. Commits *(10 Marks)*

2. Pull Requests *(5 Marks)*

3. Stars *(5 Marks)*

Course Outcomes

CO1: Understand and apply Python programming fundamentals

CO2: utilize NumPy and Pandas for efficient data manipulation, cleaning, and preparation.

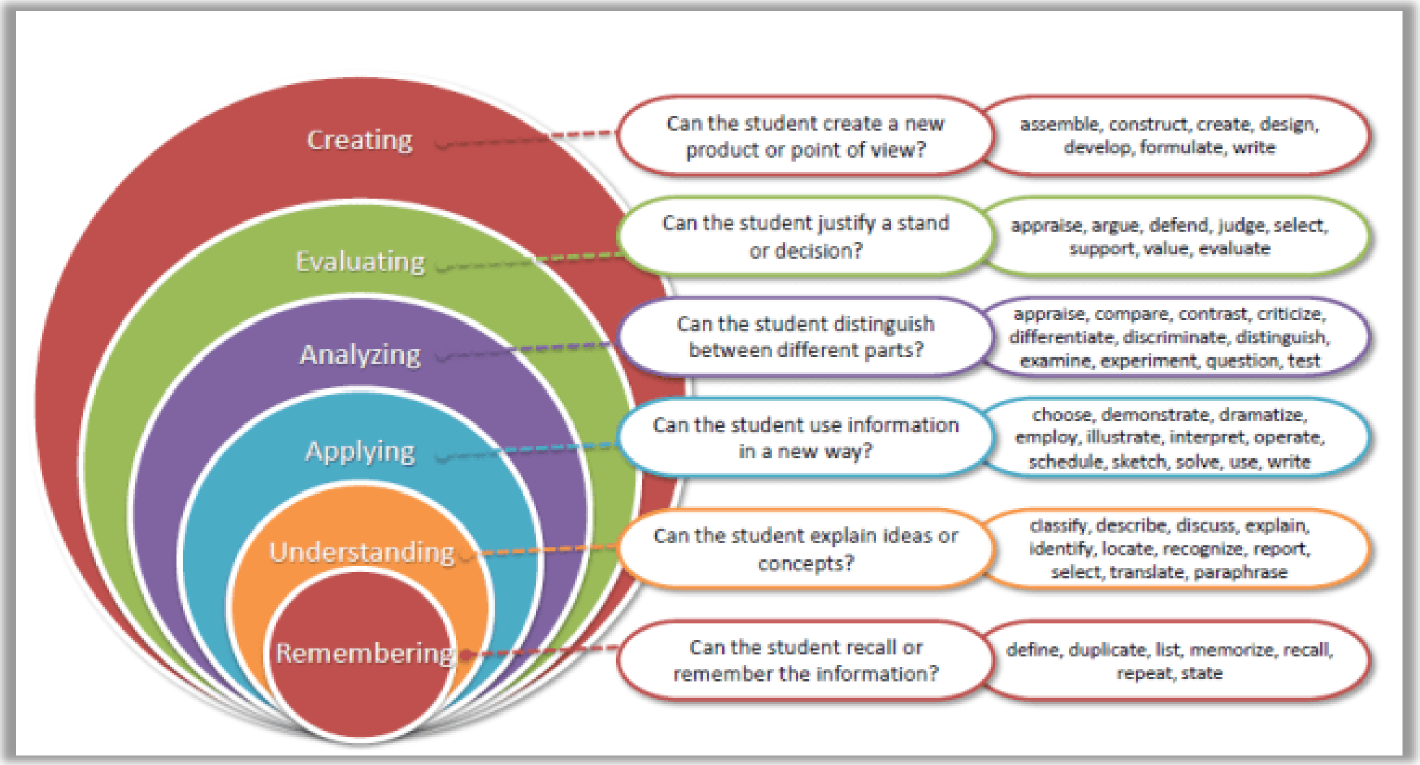
CO3: apply clear and effective data visualizations using Matplotlib and Seaborn to analyze and communicate data insights.

CO4: execute exploratory data analysis to uncover data insights using Python

CO5: perform statistical analysis and hypothesis testing using Python

CO6: associate the role of machine learning in data science

Revised Bloom’s Taxonomy



Creating – Can the student create a new product or point of view? – assemble, construct, create, design, develop, formulate, write

Evaluating – Can the student justify a stand or decision? – appraise, argue, defend, judge, select, support, value, evaluate

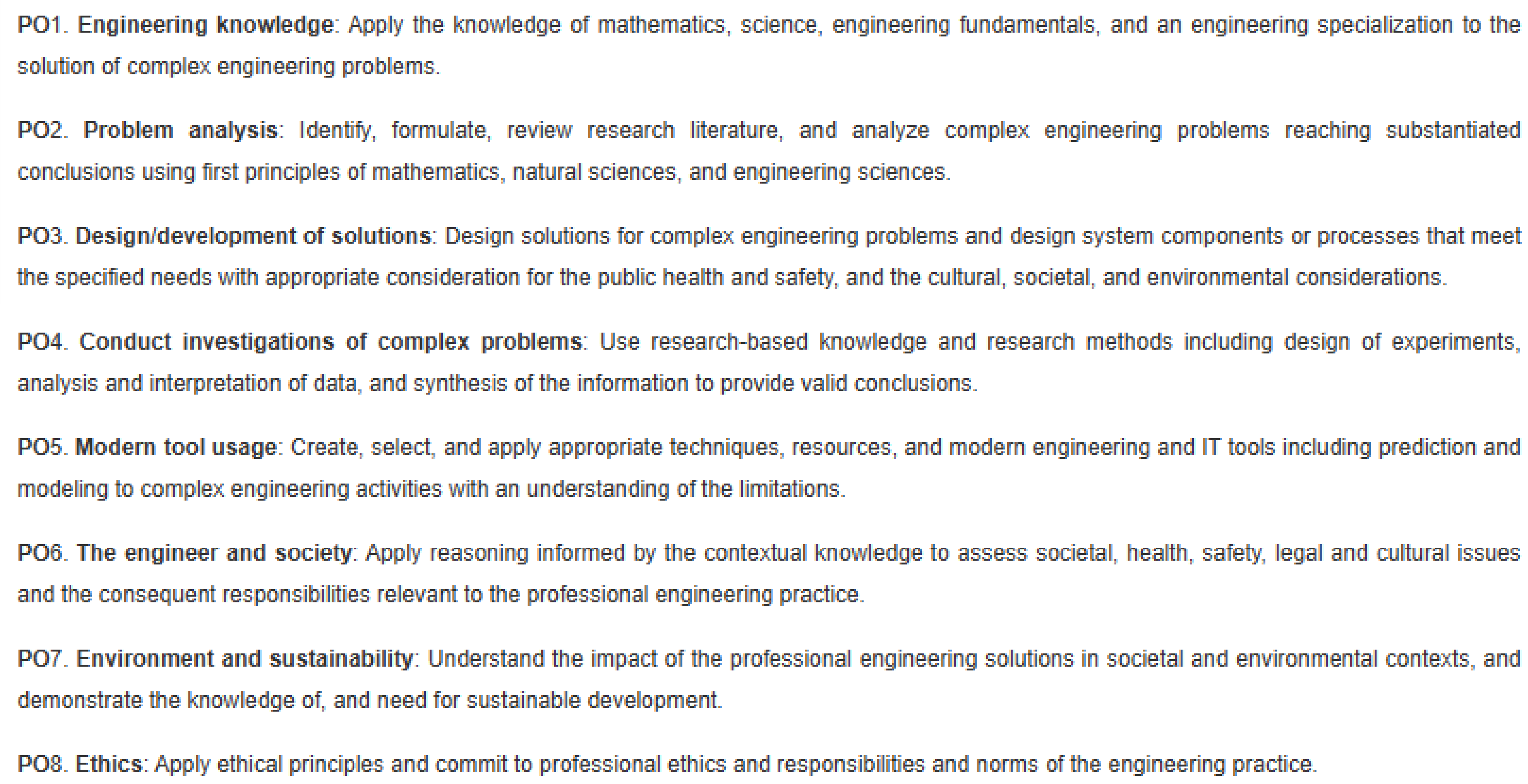
Analyzing – Can the student distinguish between different parts? – appraise, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test

Applying – Can the student use information in a new way? – choose, demonstrate, dramatize, employ, illustrate, interpret, operate, schedule, sketch, solve, use, write

Understanding – Can the student explain ideas or concepts? – classify, describe, discuss, explain, identify, locate, recognize, report, select, translate, paraphrase

Remembering – Can the student recall or remember the information? – define, duplicate, list, memorize, recall, repeat, state

Program Outcomes (POs)



PO1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

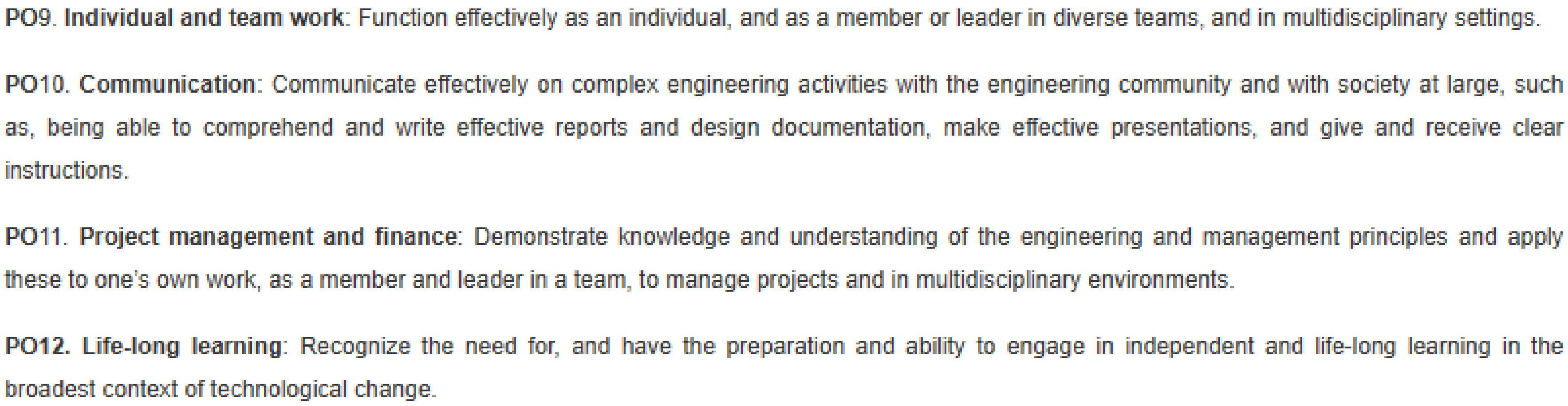
PO5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

Program Outcomes (POs)



PO9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Data Science?

Data Science

Data Science is a multidisciplinary field that combines various techniques, algorithms, processes, and tools to extract insights and knowledge from structured and unstructured data.

It involves the application of statistical analysis, machine learning, data mining, and predictive modeling to help make data-driven decisions.

Key components of Data Science:

- **Data Collection:**
  - Gathering data from various sources
- **Data Cleaning:**
  - Preparing the data
- **Exploratory Data Analysis (EDA):**
  - Analyzing the data
- **Feature Engineering:**
  - Improve model performance.
- **Modeling:**
  - Build models
- **Evaluation:**
  - Assessing model performance
- **Deployment:**
  - Implementing the model
- **Visualization:**
  - Presenting data insights using visual

PRACTICAL APPLICATIONS

- **Customer Segmentation**
  - Objective: Divide customers into distinct groups based on their behavior, demographics, or purchasing patterns.
  - Benefit: Personalized marketing, better-targeted campaigns, and improved customer retention.
- **Customer Lifetime Value (CLV) Prediction**
  - Objective: Predict the total revenue a business can expect from a customer over their entire relationship with the company.
  - Benefit: Improved resource allocation, personalized services, and better retention strategies.
- **Recommendation Systems**
  - Objective: Suggest products or services to customers based on their past behavior and preferences.
  - Benefit: Increased sales, personalized customer experiences, and improved customer satisfaction.
- **Customer Support Automation (Chatbots)**
  - Objective: Enhance customer service by providing quick, automated responses to frequently asked questions and resolving simple issues.
  - Benefit: Reduced operational costs, improved customer satisfaction, and faster issue resolution.
- **Medical Image Analysis**
  - Objective: Automate and enhance the interpretation of medical images to detect abnormalities such as tumors, fractures, and infections.
  - Benefit: Increased diagnostic accuracy, reduced workload for medical professionals, faster results, and early intervention.

Course Content

Unit 1: Introduction to Python for Data Science

Overview of Data Science, Basic Syntax and Data Types, Control Structures (if statements, loops), Functions and Modules

Unit 2: Data Manipulation with NumPy and Pandas

Introduction to NumPy: Arrays, Operations, Data Manipulation with Pandas: Series and DataFrames, Data Cleaning and Preparation, Handling Missing Data

Unit 3: Data Visualization with Matplotlib and Seaborn

Principles of Data Visualization, Creating Plots with Matplotlib, Advanced Visualization with Seaborn, Customizing Visualizations

Unit 4: Exploratory Data Analysis (EDA)

Understanding EDA and its Importance, Summary Statistics, Correlation and Covariance, Outlier Detection

Unit 5: Introduction to Statistical Analysis

Descriptive and Inferential Statistics, Hypothesis Testing: Z-test, t-test, p-test, chi-squared test, variance-inflation factor(VIF), and shapiro- wilks test, Probability Distributions: Uniform Distribution, Normal Distribution, Binomial Distribution, and Poisson Distribution, - Introduction to A/B Testing

Unit 6:

Exploring the role of machine learning in data science

Introduction to Machine Learning Concepts, Supervised vs. Unsupervised Learning, Understand CRISP-DM framework using Linear Regression model, Introduction to Classification

Recent Trends

Generative AI and Its Applications: GPT-4, DALL-E ,Synthetic Data Generation

Why Star Course??

- Real-World Applications and Innovation
- Growth of Artificial Intelligence and Machine Learning
- Impact Across Industries
- High Demand for Data-Driven Insights
- Diverse Career Opportunities
- Interdisciplinary Nature
- Educational Pathways and Accessibility
- Revolutionizing Traditional Roles
- Global Job Market

Reference Books:

1. Python for Data Science, 1st edition by Mohd. Abdul Hameed, Wiley, (2021)
2. Data Science and Machine Learning using Python by Reema Thareja, Mc Graw Hills,(2022)
3. Foundational Python for Data Science, 1st edition by Kennedy Behrman, Pearson, (2022)
4. Data Science from Scratch by Joel Grus, 2nd Edition, O'Reilly, (2019).

MOOCs or Industry certification

Course Code	Course Title	Name of Mapped MOOC/Certification/Hackathon	Is Proctored	CA Benefit Count	MTT Benefit	ETE Benefit
INT357	DATA SCIENCE TOOLBOX : PYTHON PROGRAMMING	https://onlinecourses.nptel.ac.in/noc25_cs60/preview	Yes	One CA exempted	NO	NO

What are Cohorts

A group of students of a common programme who intend to attain **similar characteristics** by means of learning **similar skills** in order to target a particular career opportunity.

Purpose of Cohorts

- Student shall be able to have a goal oriented approach for his/her career
- Student identifies the goal in the very first year
- Student shall be able to follow the stage wise career progression.
- Early identification of skill set required for selected goal.

Outline Cohort’s:

- Cohort 1: Software Development (Product Based)
- Cohort 2: Data Science
- Cohort 3: Cyber Security
- Cohort 4: Full Stack Web Development
- Cohort 5: Machine Learning
- Cohort 6: Cloud Computing
- Cohort 7: Software Methodologies And Testing
- Cohort 8: Software Development (Service Based)
- Cohort 9: Entrepreneurship
- Cohort 10: Mobile Application Development
- Cohort 11: Government jobs/Higher studies

Cohort 2:   Data Science

- Companies
- Skills Required
- Skills Sources – Internal
- Skills Sources – External

Cohort 2: Data Science

- **Companies**

10-20 LPA

- Accenture
- Quick Heal
- Informatica
- IBM
- AMDOCs
- Norton

20-30 LPA

- Amazon
- Flipkart
- Deloitte
- HP

Up to 10 LPA

- TCS
- Deloitte
- Quantiphi
- Capgemini

Cohort 2: Data Science

- **Skills Required**

S1 - DATA MANAGEMENT

S2 - DATA VISUALIZATION

S3 - DATA EXPLORATION AND ANALYSIS

S4 - R LANGUAGE

S5 - PREDICTIVE ANALYTICS

S6– DATA CLASSIFICATION

S7 - DATA ANALYTICS

S8 – PYTHON LANGUAGE

S9 – RECOMMENDER SYSTEMS

S10 - DATA PREDICTION

Cohort 2: Data Science

- **Skills Sources – Internal**

MAIN COURSE

INT306: DATABASE MANAGEMENT SYSTEM[S1]

INT108: PYTHON PROGRAMMING[S8]

ELECTIVE COURSE

INT217: INTRODUCTION TO DATA MANAGEMENT [S1]

INT233: DATA VISUALIZATION[S2,S3]

INT232: R PROGRAMMING [S4]

INT234:PREDICTIVE ANALYTICS [S7,S10,S5]

INT254: FUNDAMENTALS OF MACHINE LEARNING[S6]

Next Class