# WORKSHOP ON
# INTRODUCTION TO DATA SCIENCE AND MACHINE LEARNING

---

## Syllabus

I. **INSTRUCTOR DETAILS**:

Madhusudan Thakur, AI Researcher

II. **COURSE DURATION:**

8 hours each day (On weekends - Saturday & Sunday), depending on the pace and depth of coverage.

III. **COURSE DESCRIPTION:**

This course is designed to introduce students to the exciting field of data science. Students will learn how to collect, clean, analyze, and visualize data to extract valuable insights and make data-driven decisions. Finally training models to predict real-life usecase using AI.

IV. **PREREQUISITES**

- No prior data science experience required
- Basic knowledge of mathematics (algebra, calculus)
- Basics of Python Programming
- Basic computer literacy, and a laptop with internet running any operating system.

V. **COURSE COMPETENCIES (CC) / LEARNING OUTCOMES**

Upon completion of this course, the student will be able to:

1. Understand and Learn to apply the concept of machine learning and use it gainfully in multiple contexts based on business requirements.

2. To become familiar with the machine learning fundamentals that go into building machine learning-based solutions for making informed and insightful business decisions.

3. Learn to apply machine learning tools and techniques to aid data-driven decision-making.

## VI.    COURSE OUTLINE:

| Module | Day | Topic | Learning Objective |
|---|---|---|---|
| **Module 1** | 1 | **Introduction to Data Science**<br><br>- What is data science? Career Aspects<br>- The data science workflow<br>- Tools and technologies in data science<br>- Introduction to Python programming | Understand and Learn to apply the concept of machine learning and use it gainfully in almost all industries.<br><br>Learn Python programming required for data science. Various packages and platforms. |
| **Module 2** | 1 | **Data Collection and Cleaning**<br>- Data sources and types<br>- Data collection methods<br>- Data cleaning and preprocessing<br>- Handling missing data | Learn about data sources and methods through which data gets processed and used for further steps of machine learning. |
| **Module 3** | 1 | **Exploratory Data Analysis (EDA)**<br>- Descriptive statistics<br>- Data visualisation with Matplotlib and Seaborn<br>- Data distribution analysis<br>- Correlation and data relationships | Learn how data tells the story and insights which can be used for generating relationships between target and independent features. |

| Module 4 | 1 | **Statistics for Data Science**<br>- Probability and probability distributions<br>- Hypothesis testing<br>- Confidence intervals<br>- Statistical significance | Learn statistics required for data science. Learn how to conduct experiments and hypothesis testing. |
|---|---|---|---|
| Module 5 | 2 | **Data Wrangling and Transformation**<br>- Data manipulation with Pandas<br>- Feature engineering<br>- Data transformation techniques | Learn how to transform raw data for further steps, and learn different standard techniques. |
| Module 6 | 2 | **Data Visualization**<br>- Advanced data visualisation with libraries like Matplotlib and Seaborn<br>- Interactive data visualisation with libraries like Plotly<br>- Storytelling with data | Learn how data tells a story. Learn to create plots and charts that mean something, all in Python. |
| Module 7 | 2 | **Machine Learning Basics**<br>- Introduction to machine learning<br>- Supervised vs. unsupervised learning techniques<br>- Scikit-Learn library for machine learning<br>- Model evaluation and selection<br>-Next steps and capstone project. | Learn multiple techniques and packages to create machine learning models to predict. Learn how to evaluate multiple models through different techniques. |

**VII.**  **FINAL CAPSTONE PROJECT:**
  - Students work on a data science project applying the skills and knowledge acquired during the course. Hands-on for practice.
  - Project presentation and peer review after a month.

**VIII.**  **ASSESSMENTS:**
  - Quizzes and assignments throughout the course

**IX.**  **RESOURCES:**
  - Online tutorials and documentation resources. Books suggestions etc.
  - Jupyter notebooks for hands-on practice, or collab
  - Datasets for analysis and projects