

A large, stylized graphic with a dark purple background. It features the words "DATA SCIENCE" in large, white, bold, sans-serif capital letters. To the left of the text are several interlocking white gears. The background is filled with various data-related icons and patterns, including line graphs, bar charts, pie charts, and network diagrams. A white, curved line resembling a stylized 'V' or a bracket is positioned to the left of the text, partially overlapping the gears and the background elements.

DATA SCIENCE

INTRODUCTION TO DATA SCIENCE

SYLLABUS & CONTENTS





This course provides an introduction to the field of data science, covering the principles and techniques used for data analysis, data visualization, and machine learning. The course is designed for students with little or no prior experience in data science, and it will equip them with the foundational knowledge and practical skills required for data-driven decision making.

- Understand the principles of data science and its role in business and society
- Collect and clean data from various sources
- Analyze and visualize data using statistical methods and tools
- Build and evaluate predictive models using machine learning algorithms
- Communicate data-driven insights to stakeholders

COURSE SYLLABUS

1. Introduction to Data Science

- What is Data Science?
- Applications of Data Science in various domains
- Data Science Lifecycle
- Tools and Technologies used in Data Science

2. Data Collection and Cleaning

- Types of Data and Data Sources
- Data Preprocessing
- Data Cleaning Techniques
- Handling Missing Data

3. Exploratory Data Analysis (EDA)

- Data Visualization Techniques
- Descriptive Statistics
- Probability Distributions
- Hypothesis Testing

4. Predictive Modeling

- Introduction to Machine Learning
- Types of Learning
- Supervised Learning
- Unsupervised Learning
- Model Selection and Evaluation

5. Regression Analysis

- Linear Regression
- Logistic Regression
- Regularization Techniques
- Model Evaluation Metrics

6. Classification Analysis

- Decision Trees
- Random Forests
- Support Vector Machines
- Neural Networks

7. Clustering and Dimensionality Reduction

- K-means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)
- t-SNE

8. Communicating Data Insights

- Storytelling with Data
- Data Visualization Best Practices
- Presenting Data Insights to Stakeholders



Content

The content for this course will be delivered through a combination of lectures, hands-on exercises, and projects. The following tools and technologies will be used:

Python programming language

Jupyter Notebook

Pandas library for data manipulation

Matplotlib and Seaborn libraries for data visualization

Scikit-learn library for machine learning

Tableau for data visualization and reporting

The weekly assignments will include programming exercises, data analysis tasks, and short quizzes. The midterm exam will cover the topics covered in the first four weeks of the course. The final project will require students to apply the concepts and techniques learned in the course to a real-world data science problem. The class participation grade will be based on attendance, participation in class discussions, and group activities.

Assessment

Weekly Assignments (40%)

Midterm Exam (20%)

Final Project (30%)

Class Participation (10%)

100% Placement Support and Interview Grooming



+91 97786 32484
+91 79073 58458



support@metailearn.com



www.metailearn.com

Metailearn, 55/523, 2nd Floor, K.K. Buildings,
Thoundayil Rd, Manorama Junction,
Panampilly Nagar, Ernakulam, Kerala 682036

