

# **Data Science Fundamentals**

## **CURRICULUM:**

### **UNIT I INTRODUCTION**

Data Science: Benefits and uses – facets of data - Data Science Process: Overview –Defining research goals – Retrieving data – data preparation - Exploratory Data analysis – build the model – presenting findings and building applications - Data Mining - Data Warehousing – Basic statistical descriptions of Data

### **UNIT II DATA MANIPULATION**

Python Shell - Jupyter Notebook - IPython Magic Commands - NumPy Arrays- Universal Functions – Aggregations – Computation on Arrays – Fancy Indexing – Sorting arrays – Structured data – Data manipulation with Pandas – Data Indexing and Selection – Handling missing data – Hierarchical indexing – Combining datasets – Aggregation and Grouping – String operations – Working with time series – High performance

### **UNIT III MACHINE LEARNING**

The modeling process - Types of machine learning - Supervised learning - Unsupervised learning - Semi-supervised learning- Classification, regression - Clustering – Outliers and Outlier Analysis

### **UNIT IV DATA VISUALIZATION**

Importing Matplotlib – Simple line plots – Simple scatter plots – visualizing errors – density and contour plots – Histograms – legends – colors – subplots – text and annotation – customization – three dimensional plotting - Geographic Data with Basemap - Visualization with Seaborn

### **UNIT V HANDLING LARGE DATA**

Problems - techniques for handling large volumes of data - programming tips for dealing with large data sets- Case studies: Predicting malicious URLs, Building a recommender system - Tools and techniques needed - Research question - Data preparation - Model building – Presentation and automation.

**Course Duration:** 45 Hours

## **Test Projects:**

The following industry use cases illustrate how data science fundamentals can be applied across various industries to extract insights, drive informed decision-making, and create value from data:

1. **Retail:** Analyzing customer purchase history to personalize marketing campaigns and recommend products.
2. **Healthcare:** Predicting patient outcomes and identifying potential health risks based on medical data.
3. **Finance:** Building models to detect fraudulent transactions and minimize financial losses.
4. **Manufacturing:** Optimizing production processes through predictive maintenance and quality control analysis.
5. **E-commerce:** Analyzing website traffic and user behavior to improve conversion rates and customer experience.
6. **Telecommunications:** Forecasting network traffic and optimizing bandwidth allocation for better service delivery.
7. **Marketing:** Segmenting customers based on demographics and behavior to target advertising more effectively.
8. **Energy:** Predicting energy consumption patterns to optimize resource allocation and reduce costs.
9. **Transportation:** Developing algorithms to optimize route planning and reduce fuel consumption in logistics.
10. **Education:** Personalizing learning experiences through adaptive learning platforms based on student performance data.
11. **Real Estate:** Predicting property prices and identifying investment opportunities based on market trends and demographic data.

12. **Entertainment:** Recommending movies, music, or other content to users based on their preferences and viewing history.
13. **Hospitality:** Analyzing guest reviews and feedback to improve service quality and customer satisfaction.
14. **Insurance:** Assessing risk factors and determining insurance premiums using predictive modeling techniques.
15. **Supply Chain Management:** Optimizing inventory levels and logistics operations to reduce costs and improve efficiency.