DATA SCIENCE
curriculum

EDXcellence

1ST YEAR – TRAINING

Topics:-

1...Introduction to Data Science and Python

- What is Data Science?
- Why Python for Data Science?
- Python Basics (variables, data types, loops, functions)
- Installing Python and setting up the environment (Anaconda, Jupyter Notebook)

2...Python Libraries for Data Science

- Introduction to NumPy
- Introduction to Pandas
- Introduction to Matplotlib and Seaborn
- Introduction to Scikit-learn

3.... Data Manipulation with Pandas

- Pandas DataFrames and Series
- Importing and Exporting Data (CSV, Excel, SQL, etc.)
- Data Cleaning (handling missing values, duplicates, etc.)
- Data Transformation (filtering, sorting, grouping, merging)
- Data Aggregation and GroupBy Operations

1ST YEAR – TRAINING

4...Data Analysis and Exploration

- Exploratory Data Analysis (EDA)

- Descriptive Statistics (mean, median, mode, standard deviation, etc.)

- Data Visualization (line plots, bar charts, histograms, scatter plots)

- Correlation and Covariance

5... Data Visualization

- Advanced Data Visualization with Matplotlib
- Data Visualization with Seaborn
- Customizing plots (colors, labels, annotations)
- Plotting multiple plots on a single figure

6...Working with Numpy

- Numpy Arrays and Operations
- Indexing and Slicing
- Statistical Operations with Numpy
- Broadcasting and Vectorization

1ST YEAR – TRAINING

7...Data Preprocessing

- Feature Scaling (Standardization, Normalization)
- Encoding Categorical Variables
- Handling Missing Data
- Feature Selection and Extraction

8...Introduction to Machine Learning

- Supervised vs. Unsupervised Learning
- Train-Test Split
- Cross-Validation
- Model Evaluation Metrics (accuracy, precision, recall, F1-score)

9...Supervised Learning Algorithms

- Linear Regression
- Logistic Regression
- Decision Trees
- Support Vector Machines (SVM)
- k-Nearest Neighbors (k-NN)

1ST YEAR – TRAINING

10...Unsupervised Learning Algorithms

- K-Means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)

11...Model Evaluation and Improvement

- Confusion Matrix
- ROC-AUC Curve
- Hyperparameter Tuning (Grid Search, Random Search)
- Regularization (Ridge, Lasso)

12...Introduction to Natural Language Processing (NLP)

- Text Preprocessing (tokenization, stopword removal)
- Bag of Words and TF-IDF
- Sentiment Analysis
- Basic NLP models

1ST YEAR – TRAINING

13...Introduction to Deep Learning (optional)

- Basics of Neural Networks
- Introduction to TensorFlow/Keras
- Building simple neural networks for classification tasks

14...Working with Time Series Data

- Introduction to Time Series Analysis
- Time Series Decomposition
- Moving Averages
- ARIMA models

2ND YEAR – PROJECT

Projects:-

- End-to-end Data Science Project (EDA, Modeling, Evaluation)
- Kaggle Competitions
- Real-world datasets analysis

3RD YEAR – INTERNSHIP

- **Guaranteed internship opportunity to gain practical exposure.**

  Task: - Working on the internship Project.

  Here's a more engaging version of your case study description: "

- "In this project, we dive into the use of machine learning to predict housing prices, using the publicly available Surprise housing dataset. By applying models like Ridge and Lasso regression, along with advanced ensemble techniques, we uncover patterns and relationships within the data to provide accurate price predictions."

4TH YEAR – PLACEMENT

- Automated resume
- aptitude training
- mock exams
- mock interviews
- career counselling
- continuous support

- This roadmap should provide a solid foundation in Data Science using Python, covering the essential topics and preparing you for more advanced studies or real-world applications.

Top 5 Data Science Job Roles

- Data Scientist
- Data Analyst
- Machine Learning Engineer
- Data Engineer