



VERSION

2

PYTHON FOR
DATA SCIENTIST

**ELYSIUM
ACADEMY**

**PYTHON FOR
DATA SCIENTIST**

ELYSIUM
ACADEMY

PYTHON FOR

DATA SCIENTIST

ELYSIUM

ACADEMY

MASTER

SR. CODE

EAPL/SOLA/2022/SLAD9

COURSE CODE

EAPDS

SUB CATEGORY

APPLICATION DEVELOPMENT



TOTAL DURATION
00
HOURS



THEORY TAKEN
00
HOURS



PRACTICAL TAKEN
00
HOURS

COURSE DESCRIPTION



The Programming for Data Science with Python program offers learners the opportunity to learn the most important programming languages used by data scientists today. Get started with the fascinating field of data science and learn Python, ML, DL, and applications with the help of experienced instructors. Learners will emerge prepared to tackle real world data analysis problems. Here are some of the skills you will need to learn if you want to become a data scientist.

COURSE GOALS



The Data Science with Python course teaches you to master the concepts of Python programming. Through this Python for Data Science training, you will learn Data Analysis, Machine Learning, Data Visualization, Web Scraping, & NLP.

FUTURE SCOPE



Data Science is an expanding area with emerging tools and technologies. The demand for data science specialists appears to have room to grow over the next ten years, ensuring the profession's future. One reason for this is the industry's constantly growing opportunities and shifting data landscape.

01

CHAPTER

GETTING STARTED

O1. Introduction to Python

- a. History and Features of Python
- b. Basic Syntax
- c. Variables and Datatypes
- d. Operators
- e. Conditional, Loop & Control Statements
- f. Functions
- g. Random Modules

O2. Introduction to Data Science

- a. Data Science
- b. Data Science vs. Data Scientist
- c. Future scope of data scientist
- d. How to link python with DS

O3. Basic Terminologies of DS

- a. Data Mining
- b. Data Sets
- c. Python packages
- d. Data Collections
- e. Data Preprocessing
- f. Data Visualization
- g. Data Modelling



10
HRS



10
HRS

O4. Installation of packages

- a. Pandas
- b. Numpy
- c. Scipy
- d. Sckit
- e. NLTK

O5. Python Scientific Libraries

- a. Pandas
- b. Numpy
- c. Scipy
- d. Matplotlib
- e. Keras
- f. Tensorflow
- g. SciKit-Learn
- h. PyTorch

02

CHAPTER

DATA COLLECTION AND PREPROCESSING

O1. Introduction to Numpy

- a. Numpy
- b. Create Numpy arrays
- c. Numpy operations
- d. Numpy for statistical operations



01
HRS



06
HRS

O2. Data Preprocessing

- a. Handling Missing/Fill-na/Replace Values
- b. Drop Column/row
- c. Label Encoding
- d. One-Hot Encoding
- e. Reshaping
- f. Data operations
- g. Data frame creations
- h. Statistical functions in data operations
- i. Pivoting
- j. Merging and joining data frame
- k. Concatenate

O3. Data Normalization

- a. What is Data Normalization?
- b. Standard Scalar
- c. Min-Max Scalar
- d. Hands on standard scalar
- e. Hands on min-max scalar

O4. Data Collection

- a. How to collect input data
- b. Read CSV Data
- c. Read JSON Data
- d. Read XLS Data
- e. Read HTML Contents
- f. View Data

03

CHAPTER

VISUALIZATIONS

O1. Matplotlib Visualization

- a. Plotting with matplotlib
- b. Bar charts
- c. Pie charts
- d. Scatter plots
- e. Box plots
- f. Histogram
- g. Bubble Chart
- h. Heat maps
- i. Graph/line graph
- j. Geographical data

O2. Seaborn Visualization

- a. Plotting with seaborn
- b. Histogram with grid
- c. Distplot
- d. Pairplot
- e. Scatter plots
- f. Box plots
- g. Lmplots
- h. Histogram
- i. Challenge: Visualize titanic dataset

O3. Feature Engineering

- a. Feature Extraction
- b. Feature Selection
- c. EDA
- d. Handling outliers
- e. Feature Scaling



01
HRS
30
MINS



07
HRS
00
MINS

O4. File Handling

- a. open() mode
- b. read() mode
- c. write() mode
- d. append() mode

04

CHAPTER

MACHINE LEARNING

O1. Machine Learning

- a. AI vs ML vs DL
- b. Application of machine learning
- c. How do machine learns
- d. Types of Machine learning
- e. Supervised Learning
- f. Un-supervised Learning

O2. Naive Bayes

- a. What is Naïve Bayes?
- b. Types of Naïve Bayes
- c. Problem: IRIS Classification
- d. Data Processing
- e. Train and create model
- f. Performance Estimation
- g. Analyse and create confusion matrix



05
HRS



10
HRS

O3. Simple Linear Regression

- a. Ordinary Least Square and Regression Errors
- b. Data Processing
- c. Train and Test Model
- d. Test the model and Predict Y Values
- e. R-Squared and its Importance
- f. Score and Get coefficients
- g. Calculate RMSE (Root Mean Squared Error)
- h. Plot the predictions

O4. Support Vector Machine

- a. What is SVM?
- b. SVM kernel types
- c. Problem: IRIS Classification
- d. Data Processing
- e. Train and create model
- f. Performance Estimation
- g. Analyse and create confusion matrix

05

CHAPTER

ALGORITHMS

O1. Logistic Regression

- a. What is Logistic Regression?
- b. Problem: Heart Disease Prediction
- c. Build Model
- d. Performance Estimation
- e. Analyse and create confusion matrix



02
HRS



08
HRS

O2. Decision Tree

- a. What is decision tree?
- b. Decision Tree Parameters
- c. Problem: IRIS Classification
- d. Data Processing
- e. Train and create model
- f. Evaluate Model

O3. Random Forest

- a. What is random forest?
- b. Ensemble Learning
- c. Bagging and Boosting Classifiers
- d. Problem: Cardio Vascular Disease
- e. Implementation
- f. Evaluate Model

O4. K Nearest Neighbor

- a. What is KNN?
- b. KNN parameters
- c. Problem: Cardio Vascular Disease
- d. Data collection and preprocessing
- e. Implementation
- f. Evaluate Model

06

CHAPTER

EVALUATION METRICS

01. Evaluate Classification metrics

- Evaluate Accuracy
- Classification metrics
- What is Threshold and Adjusting Thresholds
- AUC ROC Curve

02. Dimension Reduction

- Why to reduce dimensions and Importance of PCA?
- Steps to calculate PCA
- Implementation of PCA
- Visualization

03. Regression

- What is Ridge regression?
- Implement Ridge Regression
- Plot Ridge Regression Line
- Lasso Regression or L1 Penalty
- Implement lasso Regression
- Plot lasso Regression Line

04. Over fitting and under fitting

- What is over fitting?
- How to avoid over fitting?
- What is under fitting?
- How to avoid under fitting?



04
HRS



06
HRS

O5. Cross Validation

- a. What is Cross Validation?
- b. How Cross Validation Works
- c. Prepare for Cross Validation
- d. Parameter and implementation of Cross Validation
- e. Understand the results of Cross Validation)
- f. Hands On - Analyse the Result

07

CHAPTER

HYPERTUNNING MODEL AND CLUSTERING

O1. Hyper Tuning for model

- a. What is Hyper parameter Tuning?
- b. Grid Search and Randomized Search Approach
- c. GridSearchCV Parameters Explained
- d. Create GridSearchCV Object
- e. Fit data to GridSearchCV
- f. Understand GridSearchCV Results

O2. Implementation of hypertuning

- a. GridSearchCV using Logistic Regression
- b. GridSearchCV using Support Vector
- c. Randomized Search using random forest
- d. Select Best Model
- e. Randomized Search
- f. Model Selection Summary



02
HRS



03
HRS

O3. Clustering

- a. What is Clustering?
- b. How the clusters are formed?
- c. Problem Understanding: Customer Segmentation
- d. Get, Visualize and Normalize the data
- e. Import KMeans and Understand Parameters
- f. Understanding KMeans++ Initialization Method
- g. Create Clusters
- h. Visualize and create different number of clusters
- i. Understand Elbow Method to Decide number of Cluster
- j. Implement Elbow Method

08

CHAPTER

NLP & RECOMMENDATION SYSTEM

O1. NLP

- a. What is NLP
- b. Application of NLP
- c. Remove punctuation
- d. Tokenize
- e. Remove stop words
- f. Stem words
- g. Lemmatize
- h. Padding
- i. Part of Speech Tagging
- j. Name Entity Relationship
- k. Sentiment Analysis



03
HRS



10
HRS

O2. Build our spam detector

- a. Read and preprocess data
- b. NLP techniques
- c. Tokenization
- d. Tokens to vectors
- e. Naïve Bayes
- f. Logistic Regression
- g. Sentiment Analysis
- h. Evaluate Metrics

O3. Recommendation System

- a. What is Recommendation System?
- b. How Do Recommendation Works?
- c. Types of Recommendation

O4. Content Based Recommendation

- a. What is Content based recommendation?
- b. Advantages and drawbacks
- c. Hands on Content based recommendation code

O5. Collaborative Filtering

- a. What is Collaborative filtering?
- b. Advantages and drawbacks
- c. Kinds of collaborative filtering
- d. Hands on collaborative filtering code

O6. Hybrid Recommendation System

- a. What is hybrid recommendation?
- b. Advantages and Drawbacks

09

CHAPTER

DEEP LEARNING

01. Deep Learning

- a. What is Neuron and Artificial Neural Network?
- b. How Artificial Neural Network works?
- c. What is Keras and Tensorflow?
- d. What is a Tensor in Tensorflow?
- e. Installing Keras, backend and Tensorflow

02. Optimization

- a. Keras Model Building and Steps
- b. Layers - Overview and Parameters
- c. Activation Functions
- d. Layers - Softmax Activation Function
- e. What is a Loss Function?
- f. Cross Entropy Loss Functions
- g. Optimization - What is it?
- h. Optimization - Gradient Descent
- i. Optimization - Stochastic Gradient Descent
- j. Optimization - SGD with Momentum
- k. Optimization - SGD with Exponential Moving Average
- l. Optimization - Adagrad and RMSProp for learning rate decay
- m. Optimization - Adam



03
HRS



03
HRS

O3. Deep Neural Layers

- e. Initializers - Vanishing and Exploding Gradient Problem
- f. Layers - Initializers explained
- g. Problem Understanding: Disease Prediction
- h. Read and process the data
- i. Define the Keras Neural Network Model
- j. Compile the Keras Neural Network Model
- k. Evaluate the result

10

CHAPTER

PROJECTS

1. Project on Application of data science Part 1

- a. Problem understanding: Loan Approval Prediction
- b. Read and preprocess data
- c. Data splitting
- d. Classification – Naïve Bayes
- e. Classification – Support Vector Machine
- f. Classification - Random Forest
- g. Classification – Decision Tree
- h. Classification – Logistic Regression
- i. Train the model
- j. Evaluate performance metrics
- k. Visualization using matplotlib and seaborn



00
HRS



05
HRS

O2. Project on Application of data science Part 2

- a. Problem understanding: Zomato Restaurant Review
- b. Read and preprocess data
- c. NLP
- d. Data splitting
- e. Classification – Naïve Bayes (Multinomial)
- f. Classification – Support Vector Machine
- g. Classification - Random Forest
- h. Classification – Decision Tree
- i. Classification – Logistic Regression
- j. Train the model
- k. Evaluate performance metrics
- l. Visualization using matplotlib and seaborn

O3. Project on Application of data science Part 3

- a. Problem understanding: Product Recommendation
- b. Read and preprocess data
- c. Collaborative filtering
- d. Content based filtering
- e. Recommend the product
- f. Evaluate metrics



Placement Assistance

100%

135+ Professional Courses

Practical Sessions

90%

67+ Global Pacts

ELYSIUM
GROUP OF
COMPANIES

**ELYSIUM
ACADEMY**

**PRIVATE
LIMITED**

ELYSIUM
ACADEMY

HEADQUARTERS

#227, IInd Floor, B Block, Elysium Campus, Church Rd, Anna Nagar, Madurai, Tamil Nadu - 625020.

BRANCHES

TRICHY | TIRUNELVELI | COIMBATORE
CHENNAI | BANGALORE | UAE

+91 96777 81155 | +91 75581 84348 | +91 96777 24437

info@elysiumacademy.org | www.elysiumacademy.org