

Machine Learning with Python

5 Days

Introduction to Data Science

Goal: Get an introduction to Data Science in this Module and see how Data Science helps to analyze large and unstructured data with different tools.

Objectives: At the end of this Module, you should be able to:

- Define Data Science
- Discuss the era of Data Science
- Describe the Role of a Data Scientist
- Illustrate the Life cycle of Data Science
- List the Tools used in Data Science
- State what role Big Data and Hadoop, Python, R and Machine Learning play in Data Science

Topics:

- What is Data Science?
- What does Data Science involve?
- Era of Data Science
- Business Intelligence vs Data Science
- Life cycle of Data Science
- Tools of Data Science
- Introduction to Python

Data Extraction, Wrangling, & Visualization

Goal: Discuss the different sources available to extract data, arrange the data in structured form, analyze the data, and represent the data in a graphical format.

Objectives: At the end of this Module, you should be able to:

- Discuss Data Acquisition techniques
- List the different types of Data
- Evaluate Input Data

- Explain the Data Wrangling techniques
- Discuss Data Exploration

Topics:

- Data Analysis Pipeline
- What is Data Extraction
- Types of Data
- Raw and Processed Data
- Data Wrangling
- Exploratory Data Analysis
- Visualization of Data

Hands-On/Demo:

- Loading different types of dataset in Python
- Arranging the data
- Plotting the graphs

Introduction to Machine Learning with Python

Goal: In this module, you will learn the concept of Machine Learning and its types.

Objective: At the end of this module, you should be able to:

- Essential Python Revision
- Necessary Machine Learning Python libraries
- Define Machine Learning
- Discuss Machine Learning Use cases
- List the categories of Machine Learning
- Illustrate Supervised Learning Algorithms
- Identify and recognize machine learning algorithms around us
- Understand the various elements of machine learning algorithm like parameters, hyper parameters, loss function and optimization.

Topics:

- Python Revision (numpy, Pandas, scikit learn, matplotlib)
- What is Machine Learning?
- Machine Learning Use-Cases
- Machine Learning Process Flow
- Machine Learning Categories
- Linear regression
- Gradient descent

Hands On:

- Linear Regression – Using Boston Dataset

Supervised Learning - I

Goal: In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Objective: At the end of this module, you should be able to:

- Understand What is Supervised Learning?
- Illustrate Logistic Regression
- Define Classification
- Explain different Types of Classifiers such as Decision Tree and Random Forest

Topics:

- What is Classification and its use cases?
- What is Decision Tree?
- Algorithm for Decision Tree Induction
- Creating a Perfect Decision Tree
- Confusion Matrix
- What is Random Forest?

Hands On:

- Implementation of Logistic regression, Decision tree, Random forest

Dimensionality Reduction

Goal: In this module you will learn about impact of dimensions within data. You will be taught to perform factor analysis using PCA and compress dimensions. Also, you will be developing LDA model.

Objective: At the end of this module, you should be able to:

- Define the importance of Dimensions
- Explore PCA and its implementation
- Discuss LDA and its implementation

Topics:

- Introduction to Dimensionality
- Why Dimensionality Reduction
- PCA
- Factor Analysis
- Scaling dimensional model
- LDA

Hands On:

- PCA
- Scaling

Supervised Learning - II

Goal: In this module, you will learn Supervised Learning Techniques and their implementation, for example, Decision Trees, Random Forest Classifier etc.

Objective: At the end of this module, you should be able to: • Understand What is Naïve Bayes Classifier

- How Naïve Bayes Classifier works?
- Understand Support Vector Machine
- Illustrate How Support Vector Machine works?
- Hyperparameter optimization

Topics:

- What is Naïve Bayes?
- How Naïve Bayes works?
- Implementing Naïve Bayes Classifier
- What is Support Vector Machine?
- Illustrate how Support Vector Machine works?
- Hyperparameter optimization
- Grid Search vs Random Search
- Implementation of Support Vector Machine for Classification

Hands On:

- Implementation of Naïve Bayes, SVM

Unsupervised Learning

Goal: In this module, you will learn about Unsupervised Learning and the various types of clustering that can be used to analyze the data.

Objective: At the end of this module, you should be able to:

- Define Unsupervised Learning
- Discuss the following Cluster Analysis
 - o K - means Clustering
 - o C - means Clustering
 - o Hierarchical Clustering

Topics:

- What is Clustering & its Use Cases?
- What is K-means Clustering?
- How K-means algorithm works?
- How to do optimal clustering
- What is C-means Clustering?

- What is Hierarchical Clustering?
- How Hierarchical Clustering works?

Hands On:

- Implementing K-means Clustering
- Implementing Hierarchical Clustering

Association Rules Mining and Recommendation Systems

Goal: In this module, you will learn Association rules and their extension towards recommendation engines with Apriori algorithm.

Objective: At the end of this module, you should be able to:

- Define Association Rules
- Learn the backend of recommendation engines and develop your own using python

Topics:

- What are Association Rules?
- Association Rule Parameters
- Calculating Association Rule Parameters
- Recommendation Engines
- How Recommendation Engines work?
- Collaborative Filtering
- Content Based Filtering

Hands On:

- Apriori Algorithm
- Market Basket Analysis

Reinforcement Learning

Goal: In this module, you will learn about developing a smart learning algorithm such that the learning becomes more and more accurate as time passes by. You will be able to define an optimal solution for an agent based on agent environment interaction.

Objective: At the end of this module, you should be able to

- Explain the concept of Reinforcement Learning
- Generalize a problem using Reinforcement Learning
- Explain Markov's Decision Process
- Demonstrate Q Learning

Topics:

- What is Reinforcement Learning
- Why Reinforcement Learning
- Elements of Reinforcement Learning
- Exploration vs Exploitation dilemma
- Epsilon Greedy Algorithm
- Markov Decision Process (MDP)
- Q values and V values
- Q – Learning
- α values

Hands On:

- Calculating Reward
- Discounted Reward
- Calculating Optimal quantities
- Implementing Q Learning
- Setting up an Optimal Action

Time Series Analysis

Goal: In this module, you will learn about Time Series Analysis to forecast dependent variables based on time. You will be taught different models for time series modelling such that you analyse a real time dependent data for forecasting.

Objective: At the end of this module, you should be able to:

- Explain Time Series Analysis (TSA)

- Discuss the need of TSA
- Describe ARIMA modelling
- Forecast the time series model

Topics:

- What is Time Series Analysis?
- Importance of TSA
- Components of TSA
- White Noise
- AR model
- MA model
- ARMA model
- ARIMA model
- Stationarity
- ACF & PACF

Hands on:

- Checking Stationarity
- Converting a non-stationary data to stationary
- Implementing Dickey Fuller Test
- Plot ACF and PACF
- Generating the ARIMA plot
- TSA Forecasting

Model Selection and Boosting

Goal: In this module, you will learn about selecting one model over another. Also, you will learn about Boosting and its importance in Machine Learning. You will learn on how to convert weaker algorithms to stronger ones.

Objective: At the end of this module, you should be able to:

- Discuss Model Selection

- Define Boosting
- Express the need of Boosting
- Explain the working of Boosting algorithm

Topics:

- What is Model Selection?
- Need of Model Selection
- Cross – Validation
- What is Boosting?
- How Boosting Algorithms work?
- Types of Boosting Algorithms
- Adaptive Boosting

Hands on:

- Cross Validation
- AdaBoost