

CANOPUS DATAWEB

MACHINE LEARNING SYLLABUS

BASICS OF DATA SCIENCE

Module 1: Introduction to Data Science (Duration-1hr)

- What is Data Science?
- What is Machine Learning?
- What is Deep Learning?
- What is AI?
- Data Analytics & it's types

Module 2: Introduction to Python (Duration-1hr)

- What is Python?
- Why Python?
- Installing Python
- Python IDEs
- Jupyter Notebook Overview

Module 3: Python Basics (Duration-5hrs)

- Python Basic Data types
- Lists
- Slicing
- IF statements
- Loops
- Dictionaries
- Tuples
- Functions
- Array
- Selection by position & Labels

Module 4: Python Packages (Duration-2hrs)

- Pandas
- Numpy
- Sci-kit Learn
- Mat-plot library

Module 5: Importing data (Duration-1hr)

- Reading CSV files
- Saving in Python data
- Loading Python data objects
- Writing data to CSV file

Module 6: Manipulating Data (Duration-1hr)

- Selecting rows/observations
- Rounding Number
- Selecting columns/fields
- Merging data
- Data aggregation
- Data munging techniques

Module 7: Statistics Basics (Duration-11hrs)

- Central Tendency
 - Mean
 - Median
 - Mode
 - Skewness
 - Normal Distribution
- Probability Basics
 - What does it mean by probability?
 - Types of Probability
 - ODDS Ratio?
- Standard Deviation
 - Data deviation & distribution
 - Variance
- Bias variance Tradeoff

- Underfitting
 - Overfitting
- Distance metrics
 - Euclidean Distance
 - Manhattan Distance
- Outlier analysis
 - What is an Outlier?
 - Inter Quartile Range
 - Box & whisker plot
 - Upper Whisker
 - Lower Whisker
 - catter plot
 - Cook's Distance
- Missing Value treatments
 - What is an NA?
 - Central Imputation
 - KNN imputation
 - Dummification
- Correlation
 - Pearson correlation
 - Positive & Negative correlation
- Error Metrics Duration-3hr
 - Classification
 - Confusion Matrix
 - Precision
 - Recall
 - Specificity
 - F1 Score
- Regression
 - MSE
 - RMSE
 - MAPE

CORE MACHINE LEARNING TOPICS

Module 8: Introductory to Machine Learning

Module 9: Supervised Learning (Duration-6hrs)

- Linear Regression
 - Linear Equation
 - Slope
 - Intercept
 - R square value
- Logistic regression
 - ODDS ratio
 - Probability of success
 - Probability of failure
 - ROC curve
 - Bias Variance Tradeoff

Module 10: Unsupervised Learning (Duration-4hrs)

- K-Means
- K-Means ++
- Hierarchical Clustering

Module 11: Other Machine Learning algorithms (Duration-10hrs)

- K - Nearest Neighbor
- Naïve Bayes Classifier
- Decision Tree - CART
- Decision Tree - C50
- Random Forest