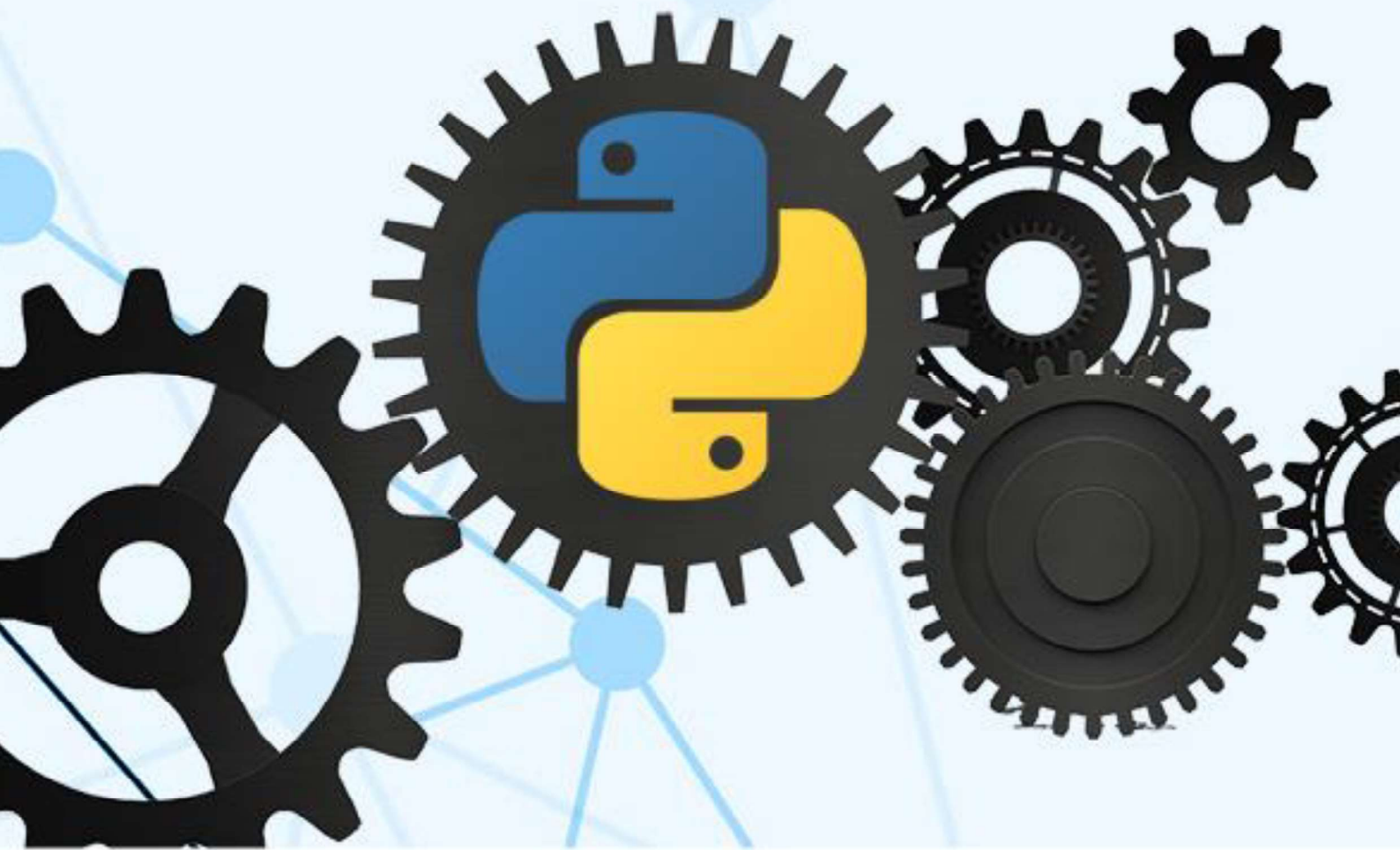




Machine Learning with Python



tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com



<https://www.facebook.com/tutorialspointindia>



<https://twitter.com/tutorialspoint>

About the Tutorial

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do. In simple words, ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method. The key focus of ML is to allow computer systems to learn from experience without being explicitly programmed or human intervention.

Audience

This tutorial will be useful for graduates, postgraduates, and research students who either have an interest in this subject or have this subject as a part of their curriculum. The reader can be a beginner or an advanced learner.

This tutorial has been prepared for the students as well as professionals to ramp up quickly. This tutorial is a stepping stone to your Machine Learning journey.

Prerequisites

The reader must have basic knowledge of artificial intelligence. He/she should also be aware of Python, NumPy, Scikit-learn, Scipy, Matplotlib.

If you are new to any of these concepts, we recommend you to take up tutorials concerning these topics, before you dig further into this tutorial.

Copyright & Disclaimer

© Copyright 2019 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

| | |
|---|-----------|
| About the Tutorial | ii |
| Audience..... | ii |
| Prerequisites..... | ii |
| Copyright & Disclaimer | ii |
| Table of Contents | iii |
| 1. Machine Learning with Python – Basics | 1 |
| What is Machine Learning? | 1 |
| Need for Machine Learning | 1 |
| Why & When to Make Machines Learn? | 1 |
| Machine Learning Model..... | 2 |
| Challenges in Machines Learning | 4 |
| Applications of Machines Learning..... | 4 |
| 2. Machine Learning with Python – Python Ecosystem | 6 |
| An Introduction to Python..... | 6 |
| Strengths and Weaknesses of Python | 6 |
| Installing Python..... | 7 |
| Why Python for Data Science? | 9 |
| Components of Python ML Ecosystem..... | 10 |
| Jupyter Notebook | 10 |
| Types of Cells in Jupyter Notebook | 12 |
| 3. Python Machine Learning – Methods for Machine Learning | 17 |
| Different Types of Methods | 17 |
| Tasks Suited for Machine Learning..... | 20 |
| 4. Machine Learning with Python – Data Loading for ML Projects | 22 |
| Consideration While Loading CSV data | 22 |
| Methods to Load CSV Data File | 23 |

| | |
|--|-----------|
| Load CSV with NumPy | 24 |
| Load CSV with Pandas | 25 |
| 5. Machine Learning with Python – Understanding Data with Statistics | 27 |
| Introduction..... | 27 |
| Looking at Raw Data | 27 |
| Checking Dimensions of Data | 29 |
| Getting Each Attribute’s Data Type | 29 |
| Statistical Summary of Data | 30 |
| Reviewing Class Distribution | 31 |
| Reviewing Correlation between Attributes | 32 |
| Reviewing Skew of Attribute Distribution | 33 |
| 6. Machine Learning with Python – Understanding Data with Visualization | 35 |
| Introduction..... | 35 |
| Univariate Plots: Understanding Attributes Independently | 35 |
| Density Plots | 37 |
| Box and Whisker Plots | 38 |
| Multivariate Plots: Interaction Among Multiple Variables | 39 |
| Correlation Matrix Plot | 39 |
| Scatter Matrix Plot..... | 41 |
| 7. Machine Learning with Python – Preparing Data | 43 |
| Introduction..... | 43 |
| Why Data Pre-processing? | 43 |
| Data Pre-processing Techniques | 43 |
| Normalization | 44 |
| Types of Normalization..... | 45 |
| Binarization..... | 46 |
| Standardization | 48 |
| Data Labeling | 49 |

| | |
|---|-----------|
| What is Label Encoding? | 49 |
| 8. Machine Learning with Python – Data Feature Selection | 51 |
| Importance of Data Feature Selection | 51 |
| Feature Selection Techniques | 51 |
| Recursive Feature Elimination | 53 |
| Principal Component Analysis (PCA) | 54 |
| Feature Importance | 55 |
| MACHINE LEARNING ALGORITHMS – CLASSIFICATION | 56 |
| 9. Classification – Introduction | 57 |
| Introduction to Classification | 57 |
| Types of Learners in Classification | 57 |
| Building a Classifier in Python | 57 |
| Classification Evaluation Metrics | 61 |
| Confusion Matrix | 61 |
| Various ML Classification Algorithms | 63 |
| Applications | 63 |
| 10. Classification Algorithms – Logistic Regression | 64 |
| Introduction to Logistic Regression | 64 |
| Types of Logistic Regression | 64 |
| Logistic Regression Assumptions | 64 |
| Binary Logistic Regression model | 65 |
| Implementation in Python | 66 |
| Multinomial Logistic Regression Model | 69 |
| Implementation in Python | 69 |
| 11. Classification Algorithms – Support Vector Machine (SVM) | 71 |
| Introduction to SVM | 71 |
| Working of SVM | 71 |

| | |
|--|------------|
| Implementing SVM in Python..... | 72 |
| SVM Kernels..... | 76 |
| Pros and Cons of SVM Classifiers..... | 79 |
| 12. Classification Algorithms – Decision Tree | 80 |
| Introduction to Decision Tree..... | 80 |
| Implementing Decision Tree Algorithm..... | 81 |
| Building a Tree | 81 |
| Implementation in Python..... | 82 |
| 13. Classification Algorithms - Naïve Bayes..... | 86 |
| Introduction to Naïve Bayes Algorithm | 86 |
| Building model using Naïve Bayes in Python..... | 86 |
| Pros & Cons | 88 |
| Applications of Naïve Bayes classification | 89 |
| 14. Classification Algorithms – Random Forest | 90 |
| Introduction..... | 90 |
| Working of Random Forest Algorithm..... | 90 |
| Implementation in Python..... | 91 |
| Pros and Cons of Random Forest | 93 |
| MACHINE LEARNING ALGORITHMS - REGRESSION..... | 95 |
| 15. Regression Algorithms – Overview..... | 96 |
| Introduction to Regression | 96 |
| Types of Regression Models | 97 |
| Building a Regressor in Python | 97 |
| Types of ML Regression Algorithms | 100 |
| Applications..... | 100 |
| 16. Regression Algorithms – Linear Regression | 101 |
| Introduction to Linear Regression | 101 |

| | |
|--|------------|
| Types of Linear Regression | 102 |
| Multiple Linear Regression (MLR) | 106 |
| Python Implementation | 107 |
| Assumptions | 108 |
| MACHINE LEARNING ALGORITHMS – CLUSTERING | 110 |
| 17. Clustering Algorithms - Overview..... | 111 |
| Introduction to Clustering | 111 |
| Cluster Formation Methods | 111 |
| Measuring Clustering Performance | 112 |
| Silhouette Analysis | 112 |
| Analysis of Silhouette Score | 112 |
| Types of ML Clustering Algorithms..... | 113 |
| Applications of Clustering..... | 114 |
| 18. Clustering Algorithms – K-means Algorithm..... | 115 |
| Introduction to K-Means Algorithm | 115 |
| Working of K-Means Algorithm | 115 |
| Implementation in Python..... | 116 |
| Advantages and Disadvantages | 119 |
| Applications of K-Means Clustering Algorithm..... | 120 |
| 19. Clustering Algorithms – Mean Shift Algorithm | 121 |
| Introduction to Mean-Shift Algorithm..... | 121 |
| Working of Mean-Shift Algorithm | 121 |
| Implementation in Python..... | 121 |
| Advantages and Disadvantages | 123 |
| 20. Clustering Algorithms – Hierarchical Clustering | 124 |
| Introduction to Hierarchical Clustering | 124 |
| Steps to Perform Agglomerative Hierarchical Clustering | 124 |

| | |
|--|------------|
| Role of Dendrograms in Agglomerative Hierarchical Clustering | 124 |
| MACHINE LEARNING ALGORITHMS - KNN ALGORITHM | 130 |
| 21. KNN Algorithm – Finding Nearest Neighbors | 131 |
| Introduction..... | 131 |
| Working of KNN Algorithm | 131 |
| Implementation in Python..... | 132 |
| KNN as Classifier | 133 |
| KNN as Regressor | 135 |
| Pros and Cons of KNN | 136 |
| Applications of KNN..... | 136 |
| 22. Machine Learning Algorithms – Performance Metrics..... | 137 |
| Performance Metrics for Classification Problems | 137 |
| Performance Metrics for Regression Problems..... | 141 |
| 23. Machine Learning with Pipelines – Automatic Workflows | 143 |
| Introduction..... | 143 |
| Challenges Accompanying ML Pipelines..... | 144 |
| Modelling ML Pipeline and Data Preparation | 144 |
| Modelling ML Pipeline and Feature Extraction | 145 |
| 24. Machine Learning – Improving Performance of ML Models | 148 |
| Performance Improvement with Ensembles..... | 148 |
| Ensemble Learning Methods | 148 |
| Bagging Ensemble Algorithms | 149 |
| Boosting Ensemble Algorithms..... | 152 |
| Voting Ensemble Algorithms | 154 |
| 25. Machine Learning – Improving Performance of ML Model (Contd...) | 157 |
| Performance Improvement with Algorithm Tuning | 157 |
| Performance Improvement with Algorithm Tuning | 157 |