

HYATT SALEH

THE MACHINE LEARNING WORKSHOP

SECOND EDITION

GET READY TO DEVELOP YOUR OWN
HIGH-PERFORMANCE MACHINE LEARNING
ALGORITHMS WITH SCIKIT-LEARN

Packt>

DATA SCIENCE & ARTIFICIAL INTELLIGENCE

Table of Contents

Preface	i
Chapter 1: Introduction to Scikit-Learn	1
Introduction	2
Introduction to Machine Learning	3
Applications of ML	3
Choosing the Right ML Algorithm	4
Scikit-Learn	6
Advantages of Scikit-Learn	8
Disadvantages of Scikit-Learn	9
Other Frameworks	10
Data Representation	10
Tables of Data	10
Features and Target Matrices	12
Exercise 1.01: Loading a Sample Dataset and Creating the Features and Target Matrices	13
Activity 1.01: Selecting a Target Feature and Creating a Target Matrix	16
Data Preprocessing	18
Messy Data	18
Missing Values	19
Outliers	21
Exercise 1.02: Dealing with Messy Data	23
Dealing with Categorical Features	30
Feature Engineering	30

Exercise 1.03: Applying Feature Engineering to Text Data	31
Rescaling Data	33
Exercise 1.04: Normalizing and Standardizing Data	34
Activity 1.02: Pre-processing an Entire Dataset	37
Scikit-Learn API	38
How Does It Work?	38
Estimator	38
Predictor	40
Transformer	40
Supervised and Unsupervised Learning	42
Supervised Learning	42
Unsupervised Learning	44
Summary	46
Chapter 2: Unsupervised Learning – Real-Life Applications	49
Introduction	50
Clustering	50
Clustering Types	51
Applications of Clustering	52
Exploring a Dataset – Wholesale Customers Dataset	53
Understanding the Dataset	54
Data Visualization	55
Loading the Dataset Using pandas	56
Visualization Tools	57
Exercise 2.01: Plotting a Histogram of One Feature from the Circles Dataset	59

THE MACHINE LEARNING WORKSHOP

Machine learning algorithms are an integral part of almost all modern applications. To make the learning process faster and more accurate, you need a tool flexible and powerful enough to help you build machine learning algorithms quickly and easily. With *The Machine Learning Workshop, Second Edition*, you'll master the scikit-learn library and become proficient in developing clever machine learning algorithms.

The Machine Learning Workshop, Second Edition, begins by demonstrating how unsupervised and supervised learning algorithms work by analyzing a real-world dataset of wholesale customers. Once you've got to grips with the basics, you'll develop an artificial neural network using scikit-learn and then improve its performance by fine-tuning hyperparameters. Towards the end of the workshop, you'll study the dataset of a bank's marketing activities and build machine learning models that can list clients who are likely to subscribe to a term deposit. You'll also learn how to compare these models and select the optimal one.

By the end of *The Machine Learning Workshop, Second Edition*, you'll not only have learned the difference between supervised and unsupervised models and their applications in the real world, but you'll also have developed the skills required to get started with programming your very own machine learning algorithms.

Packt

ISBN 978-1-83921-906-1



9 781839 219061