

---

# Python For Data Engineering Tutorial

---

The Data Warehouse Toolkit  
 Spark: The Definitive Guide  
 Powerful Python  
 Artificial Intelligence with Python  
 Python for Finance  
 Spark in Action  
 Hands-On Data Preprocessing in Python  
 In-Depth Tutorials: Deep Learning Using Scikit-Learn, Keras, and TensorFlow with Python GUI  
 Data Science on AWS  
 Data Science in Production  
 Data Engineering with Google Cloud Platform  
 Python Coding and Programming  
 Python for Data Analysis  
 A Hands-On Introduction to Data Science  
 Database Reliability Engineering  
 Data Science from Scratch  
 Python Programming and Numerical Methods  
 Python Machine Learning  
 The The Data Science Workshop  
 Data Engineering with Python  
 Fundamentals of Data Engineering  
 The Hitchhiker's Guide to Python  
 Python for Everybody  
 Python Data Science Handbook  
 Python Made Easy  
 Learn Python 3 the Hard Way  
 Learning Spark  
 Classic Computer Science Problems in Java  
 Pragmatic AI  
 Machine Learning Guide for Oil and Gas Using Python  
 Scala for Data Science  
 Head First SQL  
 Frank Kane's Taming Big Data with Apache Spark and Python  
 Rise of the Data Cloud  
 Introduction to Python for Engineers and Scientists  
 An Introduction to Statistics with Python  
 Automate the Boring Stuff with Python, 2nd Edition  
 Data Analysis with Python and PySpark  
 The the Data Science Workshop

*Python For Data Engineering Tutorial*

Downloaded from [dev.mabts.edu](http://dev.mabts.edu) by  
 guest

---

## **BARKER XIMENA**

---

**The Data Warehouse Toolkit** "O'Reilly Media, Inc."  
 Build, monitor, and manage real-time data pipelines to create data engineering infrastructure efficiently using open-source Apache projects Key Features Become well-versed in data architectures, data preparation, and data optimization skills with the help of practical examples Design data models and learn how to extract, transform, and load (ETL) data using Python Schedule, automate, and monitor complex data pipelines in production Book Description Data engineering provides the foundation for data science and analytics, and forms an important part of all businesses. This book will help you to explore various tools and methods that are used for understanding the data engineering process using Python. The book will show you how to tackle challenges commonly faced in different aspects of data engineering. You'll start with an introduction to the basics of data engineering, along with the technologies and frameworks required to build data pipelines to work with large datasets. You'll learn how to transform and clean data and perform analytics to

get the most out of your data. As you advance, you'll discover how to work with big data of varying complexity and production databases, and build data pipelines. Using real-world examples, you'll build architectures on which you'll learn how to deploy data pipelines. By the end of this Python book, you'll have gained a clear understanding of data modeling techniques, and will be able to confidently build data engineering pipelines for tracking data, running quality checks, and making necessary changes in production. What you will learn Understand how data engineering supports data science workflows Discover how to extract data from files and databases and then clean, transform, and enrich it Configure processors for handling different file formats as well as both relational and NoSQL databases Find out how to implement a data pipeline and dashboard to visualize results Use staging and validation to check data before landing in the warehouse Build real-time pipelines with staging areas that perform validation and handle failures Get to grips with deploying pipelines in the production environment Who this book is for This book is for data analysts, ETL developers, and anyone looking to get started with or transition to the field of data engineering or refresh their knowledge of data engineering using Python. This book will also be useful for students planning to build a career in

data engineering or IT professionals preparing for a transition. No previous knowledge of data engineering is required.

**Spark: The Definitive Guide** Addison-Wesley Professional Presents an instructional guide to SQL which uses humor and simple images to cover such topics as the structure of relational databases, simple and complex queries, creating multiple tables, and protecting important table data.

**Powerful Python** "O'Reilly Media, Inc."

Cut through the noise and get real results with a step-by-step approach to data science Key Features Ideal for the data science beginner who is getting started for the first time A data science tutorial with step-by-step exercises and activities that help build key skills Structured to let you progress at your own pace, on your own terms Use your physical print copy to redeem free access to the online interactive edition Book Description You already know you want to learn data science, and a smarter way to learn data science is to learn by doing. The Data Science Workshop focuses on building up your practical skills so that you can understand how to develop simple machine learning models in Python or even build an advanced model for detecting potential bank frauds with effective modern data science. You'll learn from real examples that lead to real results. Throughout The Data Science Workshop, you'll take an engaging step-by-step approach to understanding data science. You won't have to sit through any unnecessary theory. If you're short on time you can jump into a single exercise each day or spend an entire weekend training a model using sci-kit learn. It's your choice. Learning on your terms, you'll build up and reinforce key skills in a way that feels rewarding. Every physical print copy of The Data Science Workshop unlocks access to the interactive edition. With videos detailing all exercises and activities, you'll always have a guided solution. You can also benchmark yourself against assessments, track progress, and receive content updates. You'll even earn a secure credential that you can share and verify online upon completion. It's a premium learning experience that's included with your printed copy. To redeem, follow the instructions located at the start of your data science book. Fast-paced and direct, The Data Science Workshop is the ideal companion for data science beginners. You'll learn about machine learning algorithms like a data scientist, learning along the way. This process means that you'll find that your new skills stick, embedded as best practice. A solid foundation for the years ahead. What you will learn Find out the key differences between supervised and unsupervised learning Manipulate and analyze data using scikit-learn and pandas libraries Learn about different algorithms such as regression, classification, and clustering Discover advanced techniques to improve model ensembling and accuracy Speed up the process of creating new features with automated feature tool Simplify machine learning using open source Python packages Who this book is for Our goal at Packt is to help you be successful, in whatever it is you choose to do. The Data Science Workshop is an ideal data science tutorial for the data science beginner who is just getting started. Pick up a Workshop today and let Packt help you develop skills that stick with you for life. [Artificial Intelligence with Python](#) John Wiley & Sons Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications delivers a critical training and resource tool to help engineers understand machine learning theory and practice, specifically referencing use cases in oil and gas. The reference moves from explaining how Python works to step-by-step examples of utilization in various oil and gas scenarios, such as well testing, shale reservoirs and production optimization. Petroleum engineers are quickly applying machine learning techniques to their data challenges, but there is a lack of references beyond the

math or heavy theory of machine learning. Machine Learning Guide for Oil and Gas Using Python details the open-source tool Python by explaining how it works at an introductory level then bridging into how to apply the algorithms into different oil and gas scenarios. While similar resources are often too mathematical, this book balances theory with applications, including use cases that help solve different oil and gas data challenges. Helps readers understand how open-source Python can be utilized in practical oil and gas challenges Covers the most commonly used algorithms for both supervised and unsupervised learning Presents a balanced approach of both theory and practicality while progressing from introductory to advanced analytical techniques

*Python for Finance* Notion Press

The Hitchhiker's Guide to Python takes the journeyman Pythonista to true expertise. More than any other language, Python was created with the philosophy of simplicity and parsimony. Now 25 years old, Python has become the primary or secondary language (after SQL) for many business users. With popularity comes diversity—and possibly dilution. This guide, collaboratively written by over a hundred members of the Python community, describes best practices currently used by package and application developers. Unlike other books for this audience, The Hitchhiker's Guide is light on reusable code and heavier on design philosophy, directing the reader to excellent sources that already exist.

*Spark in Action* O'Reilly Media

Python Made Easy: Beginners Guide to Programming and Data Analysis using Python Get comprehensive learning of Python Programming starting from the very basics and going up to utilizing python libraries for data analysis and Visualization. Based on the author's journey to master Python, this book will help you to quickly start with writing programs and solving your problems using Python. It provides an ideal and elegant way to start learning Python, both for a newcomer to the programming world and a professional developer expert in other languages. This book comes loaded with illustrations and real-life examples. It gives you exercises which challenge you to refresh your conceptual clarity and write better codes. It is super easy to follow and will work as a self-paced tutorial to get you started with the latest and best in Python. All the advanced Python features to date are included. • Get to know the history, present, and future of Data Science • Get introduced to the basics of Computer Programming • Explore the exciting world of Python using Anaconda • Learn how to install and use Python on your computer • Create your Variables, Objects and learn Syntax of operations • Explore Python's built-in object types like Lists, dictionaries, Tuples, Strings and sets • Learn to make your codes reusable by using functions • Organize your codes, functions and other objects into larger components with Modules • Explore Classes - the Object-Oriented Programming tool for elegant codes • Write complex codes and learn how to handle Errors and Exceptions • Learn about NumPy arrays and operations on them • Explore data analysis using pandas on a real-life data set • Dive into the exciting world of Visualization with 3 chapters on Visualization and Matplotlib • Experience the Power of What you learnt by 3 projects • Learn to make your own application complete with GUI by using API

**Hands-On Data Preprocessing in Python** Packt Publishing Ltd Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters.

Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

In-Depth Tutorials: Deep Learning Using Scikit-Learn, Keras, and TensorFlow with Python GUI Packt Publishing Ltd

There are many books for those new to Python, new to programming, or both. Powerful Python is different. Written for experienced developers like you, its carefully crafted chapters teach intermediate and advanced strategies, patterns, and tools for modern Python. Focused on Python 3, with full support for 2.7. DRM-free digital upgrade: [powerfulpython.com/book-upgrade](http://powerfulpython.com/book-upgrade) "Feels like Neo learning Jiu jitsu in the Matrix." - John Beauford (@johnbeauford) "I just wanted to let you know what an excellent book this is... I keep going back to your book to learn Python." - Fahad Qazi, London, UK "Thanks. Keep up the good work. Your chapter on decorators is the best I have seen on that topic." - Leon Tietz, Minnesota, USA "Powerful Python is already helping me get huge optimization gains." - Timothy Dobbins (@TmthyDobbins) "What have I found good and valuable about the book so far? Everything honestly. The clear explanations, solid code examples have really helped me advance as a Python coder... Thank you! It has really helped me grasp some advanced concepts that I felt were beyond my abilities." - Nick S., Colorado, USA For data scientists, back-end engineers, web developers, sysadmins, devops, QA testers and more. What's included: An unrelenting selective spotlight on what's most valuable and impactful to working, full-time, professional Python developers Well-researched, detailed, realistic code on almost every page, powerfully illustrating key points. Very little "toy code" How to use decorators to add rich features to functions and classes; untangle distinct, frustratingly intertwined concerns in your code; and build powerful, extensible software frameworks How to use Python in ways that incentivize other developers to use and re-use your code, again and again... amplifying the impact of the code you write, and boosting your reputation among your peers Powerfully and easily weave iterators and generators throughout your applications, making them massively scalable, highly performant, and far more readable and maintainable How to fully leverage Python's exception and error model... giving you a detailed understanding even experienced Pythonistas often lack, and putting some of the most powerfully Pythonic exception-handling patterns in your toolbox How "magic methods" imbue natural, readable, expressive syntax into your classes and objects... and how to "break the rules" to craft stunningly intuitive, compellingly reusable library interfaces Valuable and powerful design patterns, and how Python's special language features give you uniquely powerful implementations not possible in other languages Deep and detailed instruction on how to write practical, realistic unit tests... using test-driven development to easily get into a state of flow... where you find yourself implementing feature after feature, keeping your focus with ease for long periods of time How to rapidly set up effective logging for scripts, sprawling Python applications, and everything in between An enthusiastic and unapologetic focus on Python 3, and what makes it great... with full explanation and support for getting the same results with Python 2.7 More at [PowerfulPython.com](http://PowerfulPython.com).

*Data Science on AWS* Packt Publishing Ltd

Build real-world Artificial Intelligence applications with Python to intelligently interact with the world around you About This Book Step into the amazing world of intelligent apps using this comprehensive guide Enter the world of Artificial Intelligence, explore it, and create your own applications Work through simple yet insightful examples that will get you up and running with Artificial Intelligence in no time Who This Book Is For This book is for Python developers who want to build real-world Artificial Intelligence applications. This book is friendly to Python beginners, but being familiar with Python would be useful to play around with the code. It will also be useful for experienced Python programmers who are looking to use Artificial Intelligence techniques in their existing technology stacks. What You Will Learn Realize different classification and regression techniques Understand the concept of clustering and how to use it to automatically segment data See how to build an intelligent recommender system Understand logic programming and how to use it Build automatic speech recognition systems Understand the basics of heuristic search and genetic programming Develop games using Artificial Intelligence Learn how reinforcement learning works Discover how to build intelligent applications centered on images, text, and time series data See how to use deep learning algorithms and build applications based on it In Detail Artificial Intelligence is becoming increasingly relevant in the modern world where everything is driven by technology and data. It is used extensively across many fields such as search engines, image recognition, robotics, finance, and so on. We will explore various real-world scenarios in this book and you'll learn about various algorithms that can be used to build Artificial Intelligence applications. During the course of this book, you will find out how to make informed decisions about what algorithms to use in a given context. Starting from the basics of Artificial Intelligence, you will learn how to develop various building blocks using different data mining techniques. You will see how to implement different algorithms to get the best possible results, and will understand how to apply them to real-world scenarios. If you want to add an intelligence layer to any application that's based on images, text, stock market, or some other form of data, this exciting book on Artificial Intelligence will definitely be your guide! Style and approach This highly practical book will show you how to implement Artificial Intelligence. The book provides multiple examples enabling you to create smart applications to meet the needs of your organization. In every chapter, we explain an algorithm, implement it, and then build a smart application. *Data Science in Production* "O'Reilly Media, Inc." This old edition was published in 2002. The current and final edition of this book is *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd Edition which was published in 2013 under ISBN: 9781118530801. The authors begin with fundamental design recommendations and gradually progress step-by-step through increasingly complex scenarios. Clear-cut guidelines for designing dimensional models are illustrated using real-world data warehouse case studies drawn from a variety of business application areas and industries, including: Retail sales and e-commerce Inventory management Procurement Order management Customer relationship management (CRM) Human resources management Accounting Financial services Telecommunications and utilities Education Transportation Health care and insurance By the end of the book, you will have mastered the full range of powerful techniques for designing dimensional databases that are easy to understand and provide fast query response. You will also learn how to create an architected framework that integrates the distributed data warehouse using standardized dimensions and facts.

[Data Engineering with Google Cloud Platform](#) Simon and Schuster  
**BOOK 1: LEARN FROM SCRATCH MACHINE LEARNING WITH PYTHON GUI** In this book, you will learn how to use NumPy, Pandas, OpenCV, Scikit-Learn and other libraries to how to plot graph and to process digital image. Then, you will learn how to classify features using Perceptron, Adaline, Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbor (KNN) models. You will also learn how to extract features using Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Kernel Principal Component Analysis (KPCA) algorithms and use them in machine learning. In Chapter 1, you will learn: Tutorial Steps To Create A Simple GUI Application, Tutorial Steps to Use Radio Button, Tutorial Steps to Group Radio Buttons, Tutorial Steps to Use CheckBox Widget, Tutorial Steps to Use Two CheckBox Groups, Tutorial Steps to Understand Signals and Slots, Tutorial Steps to Convert Data Types, Tutorial Steps to Use Spin Box Widget, Tutorial Steps to Use ScrollBar and Slider, Tutorial Steps to Use List Widget, Tutorial Steps to Select Multiple List Items in One List Widget and Display It in Another List Widget, Tutorial Steps to Insert Item into List Widget, Tutorial Steps to Use Operations on Widget List, Tutorial Steps to Use Combo Box, Tutorial Steps to Use Calendar Widget and Date Edit, and Tutorial Steps to Use Table Widget. In Chapter 2, you will learn: Tutorial Steps To Create A Simple Line Graph, Tutorial Steps To Create A Simple Line Graph in Python GUI, Tutorial Steps To Create A Simple Line Graph in Python GUI: Part 2, Tutorial Steps To Create Two or More Graphs in the Same Axis, Tutorial Steps To Create Two Axes in One Canvas, Tutorial Steps To Use Two Widgets, Tutorial Steps To Use Two Widgets, Each of Which Has Two Axes, Tutorial Steps To Use Axes With Certain Opacity Levels, Tutorial Steps To Choose Line Color From Combo Box, Tutorial Steps To Calculate Fast Fourier Transform, Tutorial Steps To Create GUI For FFT, Tutorial Steps To Create GUI For FFT With Some Other Input Signals, Tutorial Steps To Create GUI For Noisy Signal, Tutorial Steps To Create GUI For Noisy Signal Filtering, and Tutorial Steps To Create GUI For Wav Signal Filtering. In Chapter 3, you will learn: Tutorial Steps To Convert RGB Image Into Grayscale, Tutorial Steps To Convert RGB Image Into YUV Image, Tutorial Steps To Convert RGB Image Into HSV Image, Tutorial Steps To Filter Image, Tutorial Steps To Display Image Histogram, Tutorial Steps To Display Filtered Image Histogram, Tutorial Steps To Filter Image With CheckBoxes, Tutorial Steps To Implement Image Thresholding, and Tutorial Steps To Implement Adaptive Image Thresholding. You will also learn: Tutorial Steps To Generate And Display Noisy Image, Tutorial Steps To Implement Edge Detection On Image, Tutorial Steps To Implement Image Segmentation Using Multiple Thresholding and K-Means Algorithm, Tutorial Steps To Implement Image Denoising, Tutorial Steps To Detect Face, Eye, and Mouth Using Haar Cascades, Tutorial Steps To Detect Face Using Haar Cascades with PyQt, Tutorial Steps To Detect Eye, and Mouth Using Haar Cascades with PyQt, Tutorial Steps To Extract Detected Objects, Tutorial Steps To Detect Image Features Using Harris Corner Detection, Tutorial Steps To Detect Image Features Using Shi-Tomasi Corner Detection, Tutorial Steps To Detect Features Using Scale-Invariant Feature Transform (SIFT), and Tutorial Steps To Detect Features Using Features from Accelerated Segment Test (FAST). In Chapter 4, In this tutorial, you will learn how to use Pandas, NumPy and other libraries to perform simple classification using perceptron and Adaline (adaptive linear neuron). The dataset used is Iris dataset directly from the UCI Machine Learning Repository. You will learn: Tutorial Steps To Implement Perceptron, Tutorial Steps To Implement Perceptron with PyQt, Tutorial Steps To Implement Adaline (ADaptive Llinear NEuron), and Tutorial Steps To

Implement Adaline with PyQt. In Chapter 5, you will learn how to use the scikit-learn machine learning library, which provides a wide variety of machine learning algorithms via a user-friendly Python API and to perform classification using perceptron, Adaline (adaptive linear neuron), and other models. The dataset used is Iris dataset directly from the UCI Machine Learning Repository. You will learn: Tutorial Steps To Implement Perceptron Using Scikit-Learn, Tutorial Steps To Implement Perceptron Using Scikit-Learn with PyQt, Tutorial Steps To Implement Logistic Regression Model, Tutorial Steps To Implement Logistic Regression Model with PyQt, Tutorial Steps To Implement Logistic Regression Model Using Scikit-Learn with PyQt, Tutorial Steps To Implement Support Vector Machine (SVM) Using Scikit-Learn, Tutorial Steps To Implement Decision Tree (DT) Using Scikit-Learn, Tutorial Steps To Implement Random Forest (RF) Using Scikit-Learn, and Tutorial Steps To Implement K-Nearest Neighbor (KNN) Using Scikit-Learn. In Chapter 6, you will learn how to use Pandas, NumPy, Scikit-Learn, and other libraries to implement different approaches for reducing the dimensionality of a dataset using different feature selection techniques. You will learn about three fundamental techniques that will help us to summarize the information content of a dataset by transforming it onto a new feature subspace of lower dimensionality than the original one. Data compression is an important topic in machine learning, and it helps us to store and analyze the increasing amounts of data that are produced and collected in the modern age of technology. You will learn the following topics: Principal Component Analysis (PCA) for unsupervised data compression, Linear Discriminant Analysis (LDA) as a supervised dimensionality reduction technique for maximizing class separability, Nonlinear dimensionality reduction via Kernel Principal Component Analysis (KPCA). You will learn: Tutorial Steps To Implement Principal Component Analysis (PCA), Tutorial Steps To Implement Principal Component Analysis (PCA) Using Scikit-Learn, Tutorial Steps To Implement Principal Component Analysis (PCA) Using Scikit-Learn with PyQt, Tutorial Steps To Implement Linear Discriminant Analysis (LDA), Tutorial Steps To Implement Linear Discriminant Analysis (LDA) with Scikit-Learn, Tutorial Steps To Implement Linear Discriminant Analysis (LDA) Using Scikit-Learn with PyQt, Tutorial Steps To Implement Kernel Principal Component Analysis (KPCA) Using Scikit-Learn, and Tutorial Steps To Implement Kernel Principal Component Analysis (KPCA) Using Scikit-Learn with PyQt. In Chapter 7, you will learn how to use Keras, Scikit-Learn, Pandas, NumPy and other libraries to perform prediction on handwritten digits using MNIST dataset. You will learn: Tutorial Steps To Load MNIST Dataset, Tutorial Steps To Load MNIST Dataset with PyQt, Tutorial Steps To Implement Perceptron With PCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Perceptron With LDA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Perceptron With KPCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Logistic Regression (LR) Model With PCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Logistic Regression (LR) Model With LDA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Logistic Regression (LR) Model With KPCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement , Tutorial Steps To Implement Support Vector Machine (SVM) Model With LDA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Support Vector Machine (SVM) Model With KPCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Decision Tree (DT) Model With PCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Decision Tree (DT) Model With LDA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement

Decision Tree (DT) Model With KPCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Random Forest (RF) Model With PCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Random Forest (RF) Model With LDA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement Random Forest (RF) Model With KPCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement K-Nearest Neighbor (KNN) Model With PCA Feature Extractor on MNIST Dataset Using PyQt, Tutorial Steps To Implement K-Nearest Neighbor (KNN) Model With LDA Feature Extractor on MNIST Dataset Using PyQt, and Tutorial Steps To Implement K-Nearest Neighbor (KNN) Model With KPCA Feature Extractor on MNIST Dataset Using PyQt. BOOK 2: THE PRACTICAL GUIDES ON DEEP LEARNING USING SCIKIT-LEARN, KERAS, AND TENSORFLOW WITH PYTHON GUI In this book, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to implement deep learning on recognizing traffic signs using GTSRB dataset, detecting brain tumor using Brain Image MRI dataset, classifying gender, and recognizing facial expression using FER2013 dataset In Chapter 1, you will learn to create GUI applications to display line graph using PyQt. You will also learn how to display image and its histogram. In Chapter 2, you will learn how to use TensorFlow, Keras, Scikit-Learn, Pandas, NumPy and other libraries to perform prediction on handwritten digits using MNIST dataset with PyQt. You will build a GUI application for this purpose. In Chapter 3, you will learn how to perform recognizing traffic signs using GTSRB dataset from Kaggle. There are several different types of traffic signs like speed limits, no entry, traffic signals, turn left or right, children crossing, no passing of heavy vehicles, etc. Traffic signs classification is the process of identifying which class a traffic sign belongs to. In this Python project, you will build a deep neural network model that can classify traffic signs in image into different categories. With this model, you will be able to read and understand traffic signs which are a very important task for all autonomous vehicles. You will build a GUI application for this purpose. In Chapter 4, you will learn how to perform detecting brain tumor using Brain Image MRI dataset provided by Kaggle (<https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection>) using CNN model. You will build a GUI application for this purpose. In Chapter 5, you will learn how to perform classifying gender using dataset provided by Kaggle (<https://www.kaggle.com/cashutosh/gender-classification-dataset>) using MobileNetV2 and CNN models. You will build a GUI application for this purpose. In Chapter 6, you will learn how to perform recognizing facial expression using FER2013 dataset provided by Kaggle (<https://www.kaggle.com/nicolejyt/facialexpressionrecognition>) using CNN model. You will also build a GUI application for this purpose. BOOK 3: STEP BY STEP TUTORIALS ON DEEP LEARNING USING SCIKIT-LEARN, KERAS, AND TENSORFLOW WITH PYTHON GUI In this book, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to implement deep learning on classifying fruits, classifying cats/dogs, detecting furnitures, and classifying fashion. In Chapter 1, you will learn to create GUI applications to display line graph using PyQt. You will also learn how to display image and its histogram. Then, you will learn how to use OpenCV, NumPy, and other libraries to perform feature extraction with Python GUI (PyQt). The feature detection techniques used in this chapter are Harris Corner Detection, Shi-Tomasi Corner Detector, and Scale-Invariant Feature Transform (SIFT). In Chapter 2, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform classifying fruits using Fruits 360 dataset provided by Kaggle

(<https://www.kaggle.com/moltean/fruits/code>) using Transfer Learning and CNN models. You will build a GUI application for this purpose. In Chapter 3, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform classifying cats/dogs using dataset provided by Kaggle (<https://www.kaggle.com/chetankv/dogs-cats-images>) using CNN with Data Generator. You will build a GUI application for this purpose. In Chapter 4, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform detecting furnitures using Furniture Detector dataset provided by Kaggle (<https://www.kaggle.com/akkithetechie/furniture-detector>) using VGG16 model. You will build a GUI application for this purpose. In Chapter 5, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform classifying fashion using Fashion MNIST dataset provided by Kaggle (<https://www.kaggle.com/zalando-research/fashionmnist/code>) using CNN model. You will build a GUI application for this purpose. BOOK 4: Project-Based Approach On DEEP LEARNING Using Scikit-Learn, Keras, And TensorFlow with Python GUI In this book, implement deep learning on detecting vehicle license plates, recognizing sign language, and detecting surface crack using TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries. In Chapter 1, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform detecting vehicle license plates using Car License Plate Detection dataset provided by Kaggle (<https://www.kaggle.com/andrewmvd/car-plate-detection/download>). In Chapter 2, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform sign language recognition using Sign Language Digits Dataset provided by Kaggle (<https://www.kaggle.com/ardamavi/sign-language-digits-dataset/download>). In Chapter 3, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform detecting surface crack using Surface Crack Detection provided by Kaggle (<https://www.kaggle.com/arunrk7/surface-crack-detection/download>). BOOK 5: Hands-On Guide To IMAGE CLASSIFICATION Using Scikit-Learn, Keras, And TensorFlow with PYTHON GUI In this book, implement deep learning-based image classification on detecting face mask, classifying weather, and recognizing flower using TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries. In Chapter 1, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform detecting face mask using Face Mask Detection Dataset provided by Kaggle (<https://www.kaggle.com/omkargurav/face-mask-dataset/download>). In Chapter 2, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform how to classify weather using Multi-class Weather Dataset provided by Kaggle (<https://www.kaggle.com/pratik2901/multiclass-weather-dataset/download>). In Chapter 3, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform how to recognize flower using Flowers Recognition dataset provided by Kaggle (<https://www.kaggle.com/alxmamaev/flowers-recognition/download>). BOOK 6: Step by Step Tutorial IMAGE CLASSIFICATION Using Scikit-Learn, Keras, And TensorFlow with PYTHON GUI In this book, implement deep learning-based image classification on classifying monkey species, recognizing rock, paper, and scissor, and classify airplane, car, and ship using TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries. In

Chapter 1, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform how to classify monkey species using 10 Monkey Species dataset provided by Kaggle

(<https://www.kaggle.com/slothkong/10-monkey-species/download>).

In Chapter 2, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform how to recognize rock, paper, and scissor using 10 Monkey Species dataset provided by Kaggle

(<https://www.kaggle.com/sanikamal/rock-paper-scissors-dataset/download>). In Chapter 3, you will learn how to use TensorFlow, Keras, Scikit-Learn, OpenCV, Pandas, NumPy and other libraries to perform how to classify airplane, car, and ship using Multiclass-image-dataset-airplane-car-ship dataset provided by Kaggle

(<https://www.kaggle.com/abtabm/multiclassimagedatasetairplane-car>).

**Python Coding and Programming** Gulf Professional Publishing Learn how to code while you write programs that effortlessly perform useful feats of automation! The second edition of this international fan favorite includes a brand-new chapter on input validation, Gmail and Google Sheets automations, tips for updating CSV files, and more. If you've ever spent hours renaming files or updating spreadsheet cells, you know how tedious tasks like these can be. But what if you could have your computer do them for you? Automate the Boring Stuff with Python, 2nd Edition teaches even the technically uninclined how to write programs that do in minutes what would take hours to do by hand—no prior coding experience required! This new, fully revised edition of Al Sweigart's bestselling Pythonic classic, Automate the Boring Stuff with Python, covers all the basics of Python 3 while exploring its rich library of modules for performing specific tasks, like scraping data off the Web, filling out forms, renaming files, organizing folders, sending email responses, and merging, splitting, or encrypting PDFs. There's also a brand-new chapter on input validation, tutorials on automating Gmail and Google Sheets, tips on automatically updating CSV files, and other recent feats of automations that improve your efficiency. Detailed, step-by-step instructions walk you through each program, allowing you to create useful tools as you build out your programming skills, and updated practice projects at the end of each chapter challenge you to improve those programs and use your newfound skills to automate similar tasks. Boring tasks no longer have to take to get through—and neither does learning Python!

[Python for Data Analysis](#) "O'Reilly Media, Inc."

You Will Learn Python 3! Zed Shaw has perfected the world's best system for learning Python 3. Follow it and you will succeed—just like the millions of beginners Zed has taught to date! You bring the discipline, commitment, and persistence; the author supplies everything else. In Learn Python 3 the Hard Way, you'll learn Python by working through 52 brilliantly crafted exercises. Read them. Type their code precisely. (No copying and pasting!) Fix your mistakes. Watch the programs run. As you do, you'll learn how a computer works; what good programs look like; and how to read, write, and think about code. Zed then teaches you even more in 5+ hours of video where he shows you how to break, fix, and debug your code—live, as he's doing the exercises. Install a complete Python environment Organize and write code Fix and break code Basic mathematics Variables Strings and text Interact with users Work with files Looping and logic Data structures using lists and dictionaries Program design Object-oriented programming Inheritance and composition Modules, classes, and objects Python packaging Automated testing Basic game development Basic web development It'll be hard at first. But soon, you'll just get it—and that will feel great! This course will

reward you for every minute you put into it. Soon, you'll know one of the world's most powerful, popular programming languages. You'll be a Python programmer. This Book Is Perfect For Total beginners with zero programming experience Junior developers who know one or two languages Returning professionals who haven't written code in years Seasoned professionals looking for a fast, simple, crash course in Python 3 [A Hands-On Introduction to Data Science](#) "O'Reilly Media, Inc." In this book, you will create two MariaDB and PostgreSQL driven projects using PyQt. The step-by-step guide in this book is expected to help the reader's confidence to become a programmer who can solve database programming problems. A progressive project is provided to demonstrate how to apply the concepts of MariaDB and PostgreSQL using Python. In second chapter, you will learn PyQt that consists of a number of Python bindings for cross-platform applications that combine all the strengths of Qt and Python. By using PyQt, you can include all Qt libraries in Python code, so you can write GUI applications in Python. In other words, you can use PyQt to access all the features provided by Qt through Python code. Because PyQt depends on the Qt libraries at run time, you need to install PyQt. In third chapter, you will learn: How to create the initial three tables project in the School database: Teacher, Class, and Subject tables; How to create database configuration files; How to create a Python GUI for inserting and editing tables; How to create a Python GUI to join and query the three tables. In fourth chapter, you will learn how to: Create a main form to connect all forms; Create a project will add three more tables to the school database: Student, Parent, and Tuition tables; Create a Python GUI for inserting and editing tables; Create a Python GUI to join and query over the three tables. In this chapter, you will join the six classes, Teacher, TClass, Subject, Student, Parent, and Tuition and make queries over those tables. In chapter five, you will create dan configure PotgreSQL database. In this chapter, you will create Suspect table in crime database. This table has eleven columns: suspect\_id (primary key), suspect\_name, birth\_date, case\_date, report\_date, suspect\_status, arrest\_date, mother\_name, address, telephone, and photo. You will also create GUI to display, edit, insert, and delete for this table. In chapter six, you will create a table with the name Feature\_Extraction, which has eight columns: feature\_id (primary key), suspect\_id (foreign key), feature1, feature2, feature3, feature4, feature5, and feature6. The six fields (except keys) will have a VARCHAR data type (200). You will also create GUI to display, edit, insert, and delete for this table. In chapter seven, you will create two tables, Police and Investigator. The Police table has six columns: police\_id (primary key), province, city, address, telephone, and photo. The Investigator table has eight columns: investigator\_id (primary key), investigator\_name, rank, birth\_date, gender, address, telephone, and photo. You will also create GUI to display, edit, insert, and delete for both tables. In chapter eight, you will create two tables, Victim and Case\_File. The Victim table has nine columns: victim\_id (primary key), victim\_name, crime\_type, birth\_date, crime\_date, gender, address, telephone, and photo. The Case\_File table has seven columns: case\_file\_id (primary key), suspect\_id (foreign key), police\_id (foreign key), investigator\_id (foreign key), victim\_id (foreign key), status, and description. You will create GUI to display, edit, insert, and delete for both tables as well.

**Database Reliability Engineering** Packt Publishing Ltd Think big about your data! PySpark brings the powerful Spark big data processing engine to the Python ecosystem, letting you seamlessly scale up your data tasks and create lightning-fast pipelines. In Data Analysis with Python and PySpark you will learn how to: Manage your data as it scales across multiple machines,

Scale up your data programs with full confidence, Read and write data to and from a variety of sources and formats, Deal with messy data with PySpark's data manipulation functionality, Discover new data sets and perform exploratory data analysis, Build automated data pipelines that transform, summarize, and get insights from data, Troubleshoot common PySpark errors, Creating reliable long-running jobs. Data Analysis with Python and PySpark is your guide to delivering successful Python-driven data projects. Packed with relevant examples and essential techniques, this practical book teaches you to build pipelines for reporting, machine learning, and other data-centric tasks. Quick exercises in every chapter help you practice what you've learned, and rapidly start implementing PySpark into your data systems. No previous knowledge of Spark is required. Data Analysis with Python and PySpark helps you solve the daily challenges of data science with PySpark. You'll learn how to scale your processing capabilities across multiple machines while ingesting data from any source--whether that's Hadoop clusters, cloud data storage, or local data files. Once you've covered the fundamentals, you'll explore the full versatility of PySpark by building machine learning pipelines, and blending Python, pandas, and PySpark code.

[Data Science from Scratch](#) SPARTA PUBLISHING

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into the discipline without actually understanding data science. In this book, you'll learn how many of the most fundamental data science tools and algorithms work by implementing them from scratch. If you have an aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today's messy glut of data holds answers to questions no one's even thought to ask. This book provides you with the know-how to dig those answers out. Get a crash course in Python Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science Collect, explore, clean, munge, and manipulate data Dive into the fundamentals of machine learning Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering Explore recommender systems, natural language processing, network analysis, MapReduce, and databases

**Python Programming and Numerical Methods** Data Engineering with Python

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with

SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

*Python Machine Learning* "O'Reilly Media, Inc."

An introductory textbook offering a low barrier entry to data science; the hands-on approach will appeal to students from a range of disciplines.

[The Data Science Workshop](#) Michael Learn

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets Spark's core APIs through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation

*Data Engineering with Python* Packt Publishing Ltd

Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn how to solve real-world data analysis problems with thorough, detailed examples

Related with Python For Data Engineering Tutorial:

[© Python For Data Engineering Tutorial Is Coursera Data Science Specialization Worth It](#)

[© Python For Data Engineering Tutorial Is Diana Gabaldon Writing Another Outlander Book](#)

[© Python For Data Engineering Tutorial Is Home Economics Cancelled](#)