



**UNIVERSITÉ
DE GENÈVE**

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Livre

2022

Summary

Open Access

This file is a(n) Summary of:

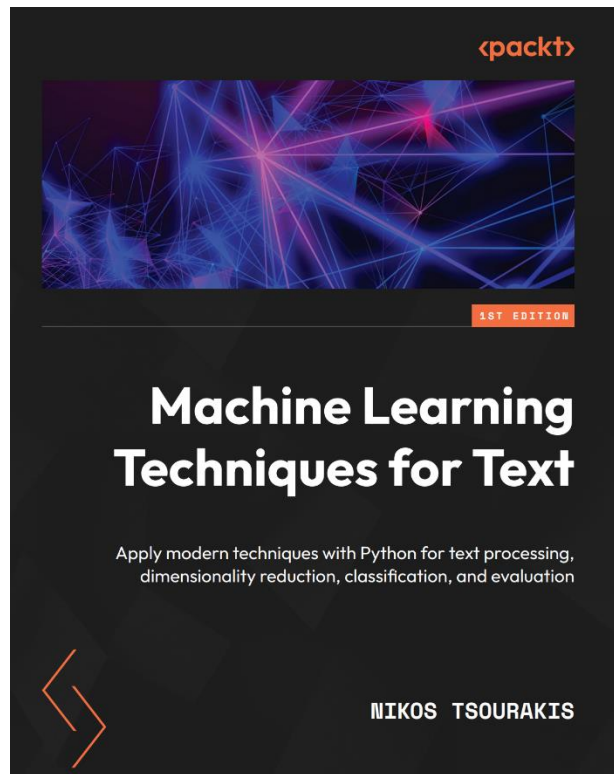
Machine Learning Techniques for Text: Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation

Tsourakis, Nikolaos

This publication URL:

<https://archive-ouverte.unige.ch/unige:166016>

© This document is protected by copyright. Please refer to copyright holders for terms of use.



Take your Python text processing skills to another level by learning about the latest natural language processing and machine learning techniques with this full color guide

Key Features

- Learn how to acquire and process textual data and visualize the key findings
- Obtain deeper insight into the most commonly used algorithms and techniques and understand their tradeoffs
- Implement models for solving real-world problems and evaluate their performance

Book Description

With the ever-increasing demand for machine learning and programming professionals, it's prime time to invest in the field. This book will help you in this endeavor, focusing specifically on text data and human language by steering a middle path among the various textbooks that present complicated theoretical concepts or focus disproportionately on Python code.

A good metaphor this work builds upon is the relationship between an experienced craftsman and their trainee. Based on the current problem, the former picks a tool from the toolbox, explains its utility, and puts it into action. This approach will help you to identify at least one practical use for each method or technique presented. The content unfolds in ten chapters, each discussing one specific case study. For this reason, the book is solution-oriented. It's accompanied by Python code in the form of Jupyter notebooks to help you obtain hands-on experience. A recurring pattern in the chapters of this book is helping you get some intuition on the data and then implement and contrast various solutions.

By the end of this book, you'll be able to understand and apply various techniques with Python for text preprocessing, text representation, dimensionality reduction, machine learning, language modeling, visualization, and evaluation.

What you will learn

- Understand fundamental concepts of machine learning for text
- Discover how text data can be represented and build language models
- Perform exploratory data analysis on text corpora
- Use text preprocessing techniques and understand their trade-offs

- Apply dimensionality reduction for visualization and classification
- Incorporate and fine-tune algorithms and models for machine learning
- Evaluate the performance of the implemented systems
- Know the tools for retrieving text data and visualizing the machine learning workflow

Who this book is for

This book is for professionals in the area of computer science, programming, data science, informatics, business analytics, statistics, language technology, and more who aim for a gentle career shift in machine learning for text. Students in relevant disciplines that seek a textbook in the field will benefit from the practical aspects of the content and how the theory is presented. Finally, professors teaching a similar course will be able to pick pertinent topics in terms of content and difficulty. Beginner-level knowledge of Python programming is needed to get started with this book.

Table of Contents

1. Introducing Machine Learning for Text
2. Detecting Spam Emails
3. Classifying Topics of Newsgroup Posts
4. Extracting Sentiments from Product Reviews
5. Recommending Music Titles
6. Teaching Machines to Translate
7. Summarizing Wikipedia Articles
8. Detecting Hateful and Offensive Language
9. Generating Text in Chatbots
10. Clustering Speech-to-Text Transcriptions