



# JUST ENOUGH DATA SCIENCE AND MACHINE LEARNING

ESSENTIAL TOOLS AND TECHNIQUES



MARK LEVENE

|

MARTYN HARRIS

# Just Enough Data Science and Machine Learning

---

# Just Enough Data Science and Machine Learning: Essential Tools and Techniques

## Table of Contents

Cover

Half Title

Title Page

Copyright Page

Contents

List of Figures

Preface

About the Authors

1 What Is Data Science?

2 Basic Statistics

2.1 Introductory Statistical Notions

2.2 Expectation

2.3 Variance

2.4 Correlation

2.5 Regression

2.6 Chapter Summary

3 Types of Data

3.1 Tabular Data

3.2 Textual Data

3.3 Image, Video, and Audio Data

# **Table of Contents**

- 3.4 Time Series Data
- 3.5 Geographical Data
- 3.6 Social Network Data
- 3.7 Transforming Data
- 3.8 Chapter Summary

## **4 Machine Learning Tools**

- 4.1 What Is Machine Learning?
- 4.2 Evaluation
  - 4.2.1 Evaluation for Supervised Models
  - 4.2.2 Evaluation for Unsupervised Models
- 4.3 Supervised Methods
  - 4.3.1 K-Nearest Neighbours
  - 4.3.2 Naive Bayes
  - 4.3.3 Support Vector Machines
  - 4.3.4 Decision Trees and Random Forests
  - 4.3.5 Neural Networks and Deep Learning
- 4.4 Unsupervised Methods
  - 4.4.1 K-Means
  - 4.4.2 Hierarchical Clustering
  - 4.4.3 Principal Components Analysis
  - 4.4.4 Topic Modelling
  - 4.4.5 DBSCAN
- 4.5 Semi-Supervised Methods
- 4.6 Chapter Summary

## **5 Data Science Topics**

- 5.1 Searching, Ranking, and Rating
  - 5.1.1 The Vector Space Model
  - 5.1.2 Ranking with PageRank

# **Table of Contents**

5.1.3 Rating with the Elo System

5.1.4 Recommender Systems and Collaborative Filtering

## **5.2 Social Networks**

5.2.1 The Basics of Social Networks

5.2.2 Centrality Measures

5.2.3 Power Laws and the 8020 Rule

5.2.4 SIS and SIR Models for the Spread of Disease

## **5.3 Three Natural Language Processing Topics**

5.3.1 Sentiment Analysis

5.3.2 Named Entity Recognition

5.3.3 Word Embeddings

## **5.4 Chapter Summary**

## **6 Selected Additional Topics**

6.1 Neuro-Symbolic AI

6.2 Conversational AI

6.3 Generative Neural Networks

6.4 Trustworthy AI

6.5 Large Language Models

6.6 Epilogue

## **7 Further Reading**

7.1 Basic Statistics

7.2 Data Science

7.3 Machine Learning

7.4 Deep Learning

7.5 Research Papers

7.6 Python

## **Bibliography**

# **Table of Contents**

Index