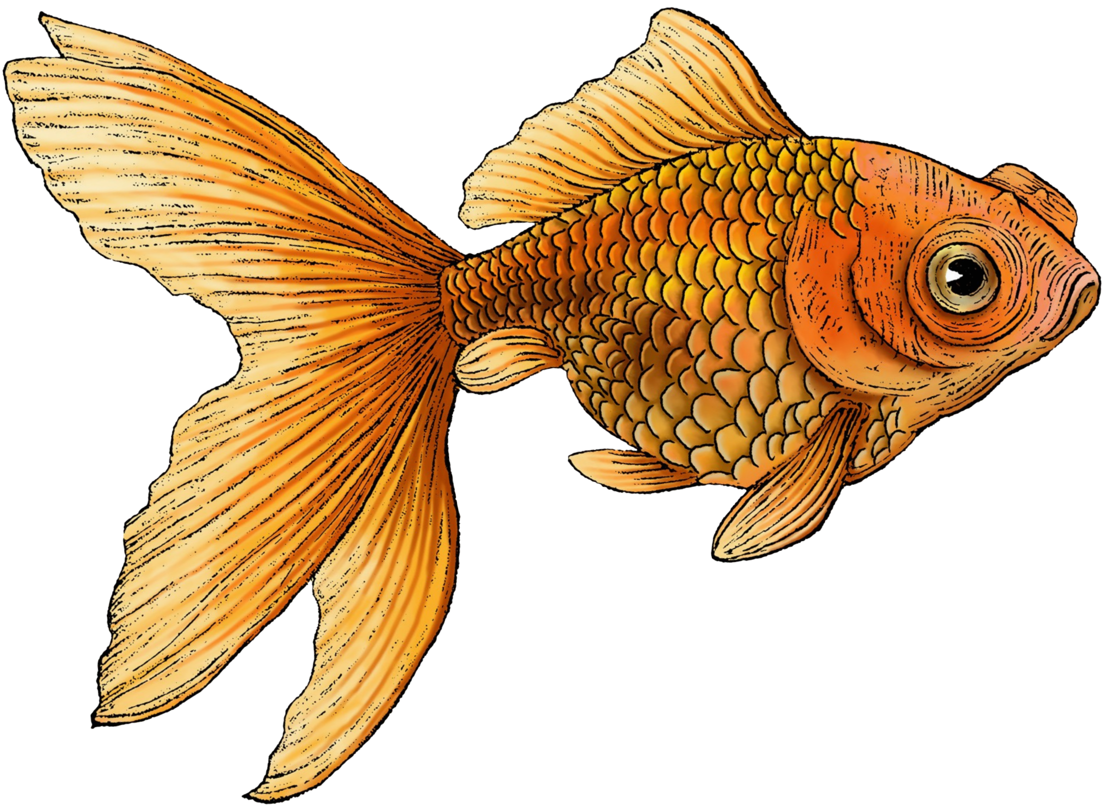
O'REILLY®

Deep Learning for Finance

Creating Machine & Deep Learning Models for Trading in Python



Sofien Kaabar

O'REILLY®

Deep Learning for Finance

Deep learning is rapidly gaining momentum in the world of finance and trading. But for many professional traders, this sophisticated field has a reputation for being complex and difficult. This hands-on guide teaches you how to develop a deep learning trading model from scratch using Python, and it also helps you create and backtest trading algorithms based on machine learning and reinforcement learning.

Sofien Kaabar—financial author, trading consultant, and institutional market strategist—introduces deep learning strategies that combine technical and quantitative analyses. By fusing deep learning concepts with technical analysis, this unique book presents outside-the-box ideas in the world of financial trading. This A-Z guide also includes a full introduction to technical analysis, evaluating machine learning algorithms, and algorithm optimization.

- **Understand and create machine learning and deep learning models**
- **Explore the details behind reinforcement learning and see how it's used in time series**
- **Understand how to interpret performance evaluation metrics**
- **Examine technical analysis and learn how it works in financial markets**
- **Create technical indicators in Python and combine them with ML models for optimization**
- **Evaluate the models' profitability and predictability to understand their limitations and potential**

"This book is a magisterial work that stands as a landmark in the field of quantitative trading, data science, and financial algorithms."

—Amaury Goguel
Head of MSc Financial Markets & Investments, SKEMA Business School, Paris

"This is the book I wish I had read when I started developing ML trading algorithms as a quantitative investment strategist."

—Ning Wang
Quantitative Investment Structurer, Barclays

Sofien Kaabar is a financial author, trading consultant, and institutional market strategist specializing in the currencies market with a focus on technical and quantitative topics. Sofien's goal is to make technical analysis objective by incorporating clear conditions that can be analyzed and created with the use of technical indicators.

DATA

US $69.99 CAN $87.99
ISBN: 978-1-098-14839-3

Praise for *Deep Learning for Finance*

As the scientific director of a leading program in market finance for over 10 years, I can testify to the immense quality of this book. *Deep Learning for Finance* is a magisterial work that stands as a landmark in the field of quantitative trading, data science, and financial algorithms. The author's profound knowledge and deep insights are evident throughout the book, which is written with clarity and precision. It will undoubtedly become a reference in this specialized field in which the inherent complexity of the subject is rarely well served in disclosure books. This one is an exception, striking the perfect balance between clarity and precision without becoming oversimplistic or overcomplex. It is an essential read for anyone interested in the cutting edge of quantitative trading/finance, both for master's degree students in finance and for practitioners.

—*Amaury Goguel, Head of MSc Financial Markets & Investments, SKEMA Business School, Paris*

This is the book I wish I had read when I started developing ML trading algorithms as a quantitative investment strategist.

—*Ning Wang, Quantitative Investment Structurer, Barclays*

Sofien is a master, providing the right balance of detail and autonomy, allowing readers to connect the dots themselves.

—*Timothy Kipper, Head of Business Development, Coperniq.io*

Deep Learning for Finance

Creating Machine and Deep Learning Models for Trading in Python

Sofien Kaabar

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Deep Learning for Finance
by Sofien Kaabar

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Michelle Smith
Development Editor: Corbin Collins
Production Editor: Elizabeth Faerm
Copyeditor: Audrey Doyle
Proofreader: Piper Editorial Consulting, LLC
Indexer: WordCo Indexing Services, Inc.
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Kate Dullea

January 2024: First Edition

Revision History for the First Edition
2024-01-08: First Release

See *http://oreilly.com/catalog/errata.csp?isbn=9781098148393* for release details.

978-1-098-14839-3

[LSI]

Table of Contents

Preface

Learning never exhausts the mind.
—Leonardo da Vinci

Machine learning and deep learning have completely changed the finance industry in recent years. The different learning models are well suited to a world where data is abundant and continuous. Data is the new gold, and its value keeps rising as proper analyses lead to key business decisions, which are the driver of economic shifts.

The rise of quantitative funds is living proof that the world of data science has much to offer to the trading world. After fundamental and technical traders, a new breed of leaders of the universe is emerging. These are the quantitative traders who rely on machine-based algorithms with extremely complex operations that seek to forecast and outperform the markets.

This book covers in detail the subject of *deep learning for finance*.

Why This Book?

I have spent my career researching trading strategies, techniques, and all things related to the financial world. Through the years, I have become familiar with a few algorithmic models that have the potential of adding value to the trading framework. In this book, I discuss different learning models and their applications in the trading world, with a focus on deep learning and neural networks. My main aim is to cover them in such a way that everyone understands how they function.

Machines can perform operations and detection better than humans for many reasons, one of which is their objectivity. This means that one of the key skills you will learn is how to use Python to create the algorithms required to do such operations.

As mentioned, my objective is to provide a comprehensive introduction to the use of deep learning in finance. I do this by discussing a wide range of topics, including data science, trading, machine and deep learning models, and reinforcement learning applications for trading.

The book begins with an overview of the field of data science and its role in the finance world. It then delves into the knowledge requirements, such as statistics, math, and Python, before focusing on how to use machine and deep learning in trading strategies.

Who Should Read It?

This book is intended for a wide audience, including professionals and academics in finance, data scientists, quantitative traders, and students of finance of any level. It provides a thorough introduction to the use of machine and deep learning in time series prediction, and it is an essential resource for anyone who wants to understand and apply these powerful techniques.

The book assumes you have basic background knowledge in both Python programming (professional Python users will find the code very straightforward) and financial trading. I take a clear and simple approach that focuses on the key concepts so that you understand the purpose of every idea.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic
: Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`
: Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

`Constant width bold`
: Shows commands or other text that should be typed literally by the user.

`Constant width italic`
: Shows text that should be replaced with user-supplied values or by values determined by context.

This element signifies a tip or suggestion.

This element signifies a general note.

This element indicates a warning or caution.

Using Code Examples

Supplemental material (code examples, exercises, etc.) is available for download at *https://github.com/sofienkaabar/deep-learning-for-finance*.

If you have a technical question or a problem using the code examples, please send email to *bookquestions@oreilly.com*.

This book is here to help you get your job done. In general, if example code is offered with this book, you may use it in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but generally do not require, attribution, which usually includes the title, author, publisher, and ISBN. For example: "*Deep Learning for Finance* by Sofien Kaabar (O'Reilly). Copyright 2024 Sofien Kaabar, 978-1-098-14839-3."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at *permissions@oreilly.com*.

O'Reilly Online Learning

For more than 40 years, *O'Reilly Media* has provided technology and business training, knowledge, and insight to help companies succeed.

Our unique network of experts and innovators share their knowledge and expertise through books, articles, and our online learning platform. O'Reilly's online learning platform gives you on-demand access to live training courses, in-depth learning paths, interactive coding environments, and a vast collection of text and video from O'Reilly and 200+ other publishers. For more information, visit *https://oreilly.com*.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-889-8969 (in the United States or Canada)
707-829-7019 (international or local)
707-829-0104 (fax)
support@oreilly.com
https://www.oreilly.com/about/contact.html

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at *https://oreil.ly/deep-learning-for-finance*.

For news and information about our books and courses, visit *https://oreilly.com*.

Find us on LinkedIn: *https://linkedin.com/company/oreilly-media*.

Follow us on Twitter: *https://twitter.com/oreillymedia*.

Watch us on YouTube: *https://youtube.com/oreillymedia*.

Acknowledgments

Nothing would be the same without the support of my parents, which is why I can't help but acknowledge their direct and indirect impact on the book.

I would also like to acknowledge the debt I owe to the editors, Michelle Smith and Corbin Collins, as well as to the production editor, Elizabeth Faerm, for their continued support, the amazing job they do, and their patience. Similarly, I would like to thank every person at O'Reilly who was involved in the production of this book.

Additionally, my special thanks go to the great technical reviewers for their immense contributions. They had a sizable impact on making this book readable, useful, and straightforward. I could not ask for better people to review my book.

Finally, I am deeply grateful to you, the reader, for investing your time into reading my work and for placing your trust in my research. I hope you find it useful.

CHAPTER 1
Introducing Data Science and Trading

The best way to begin learning about a complex topic is to break it down into smaller parts and understand those pieces first. Understanding deep learning for finance requires knowledge of data science and financial markets.

This chapter lays the building blocks needed to thoroughly understand data science and its uses, as well as to understand financial markets and how trading and forecasting can benefit from data science.

By the end of the chapter, you should know what data science is, what its applications are, and how you can use it in finance to extract value.

Understanding Data

It is impossible to understand the field of data science without first understanding the types and structures of data. After all, the first word for the name of this immense field is *data*. So, what is data? And more importantly, what can you do with it?

Data in its simplest and purest form is a collection of raw information that can be of any type (numerical, text, boolean, etc.).

The final aim of collecting data is decision-making. This is done through a complex process that ranges from the act of gathering and processing data to interpreting it and using the results to make a decision.

Let's take an example of using data to make a decision. Suppose you have a portfolio composed of five different equally weighted dividend-paying stocks, as detailed in Table 1-1.

Table 1-1. Stocks and their dividend yields

Stock	Dividend yield
A	5.20%
B	3.99%
C	4.12%
D	6.94%
E	5.55%

> A *dividend* is the payment made to shareholders from a company's profits. The *dividend yield* is the amount distributed in monetary units over the current share price of the company.

Analyzing this data can help you understand the average dividend yield you are receiving from your portfolio. The average is basically the sum divided by the quantity, and it gives a quick snapshot of the overall dividend yield of the portfolio:

$$\text{Average dividend yield} = \frac{5.20\% + 3.99\% + 4.12\% + 6.94\% + 5.55\%}{5} = 5.16\%$$

Therefore, the average dividend yield of your portfolio is 5.16%. This information can help you compare your average dividend yield to other portfolios so that you know whether you have to make any adjustments.

Another metric you can calculate is the number of stocks held in the portfolio. This may provide the first informational brick in constructing a wall of diversification. Even though these two pieces of information (average dividend yield and the number of stocks in the portfolio) are very simple, complex data analysis begins with simple metrics and may sometimes not require sophisticated models to properly interpret the situation.

The two metrics you calculated in the previous example are called the *average* (or mean) and the *count* (or number of elements). They are part of a field called *descriptive statistics* discussed in Chapter 3, which is also itself part of data science.

Let's take another example of data analysis for inferential purposes. Suppose you have calculated a yearly correlation measure between two commodities, and you want to predict whether the next yearly correlation will be positive or negative. Table 1-2 has the details of the calculations.

Table 1-2. Yearly correlation measures

Year	Correlation
2015	Positive
2016	Positive
2017	Positive
2018	Negative
2019	Positive
2020	Positive
2021	Positive
2022	Positive
2023	Positive

Correlation is a measure of the linear reliance between two time series. A *positive correlation* generally means that the two time series move on average in the same direction, while a *negative correlation* generally means that the two time series move on average in opposite directions. Correlation is discussed in Chapter 3.

From Table 1-2, the historical correlation between the two commodities was mostly (i.e., 88%) positive. Taking into account historical observations, you can say that there is an 88% probability that the next correlation measure will be positive. This also means that there is a 12% probability that the next correlation measure will be negative:

$$E(\text{Positive correlation}) = \frac{8}{9} = 88.88\%$$

This is another basic example of how to use data draw inferences from observations and make decisions. Of course, the assumption here is that historical results will exactly reflect future results, which is unlikely in real life, but occasionally, to predict the future all you have is the past.

Now, before discussing data science, let's review what types of data can be used and segment them into different groups:

Numerical data
: This type of data is composed of numbers that reflect a certain type of information that is collected at regular or irregular intervals. Examples can include market data (OHLC,[1] volume, spreads, etc.) and financial statements data (assets, revenue, costs, etc.).

1 OHLC refers to the four essential pieces of market data: open price, high price, low price, and close price.

Categorical data

Categorical data is data that can be organized into groups or categories using names or labels. It is qualitative rather than quantitative. For example, the blood type of patients is a type of categorical data. Another example is the eye color of different samples from a population.

Text data

Text data has been on the rise in recent years with the development of *natural language processing* (NLP). Machine learning models use text data to translate, interpret, and analyze the sentiment of the text.

Visual data

Images and videos are also considered data, and you can process and transform them into valuable information. For example, a *convolutional neural network* (CNN) is a type of algorithm (discussed in Chapter 8) that can recognize and categorize photos by labels (e.g., labeling cat photos as cats).

Audio data

Audio data is very valuable and can help save time on transcriptions. For example, you can use algorithms on audio to create captions and automatically create subtitles. You can also create models that interpret the sentiment of the speaker using the tone and volume of the audio.

Data science is a transdisciplinary field that tries to extract intelligence and conclusions from data using different techniques and models, be they simple or complex. The data science process is composed of many steps besides just analyzing data. The following summarizes these steps:

1. *Data gathering*: This process involves the acquisition of data from reliable and accurate sources. A widely known phrase in computer science generally credited to George Fuechsel goes "Garbage in, garbage out," and it refers to the need to have quality data that you can rely on for proper analysis. Basically, if you have inaccurate or faulty data, then all your processes will be invalid.
2. *Data preprocessing*: Occasionally, the data you acquire can be in a raw form, and it needs to be preprocessed and cleaned for the data science models to be able to use it. For example, dropping unnecessary data, adding missing values, or eliminating invalid and duplicate data can be part of the preprocessing step. Other, more complex examples can include *normalization* and *denoising* of data. The aim of this step is to get the data ready for analysis.
3. *Data exploration*: During this step, basic statistical research is conducted to find trends and other characteristics in data. An example of data exploration is to calculate the mean of the data.