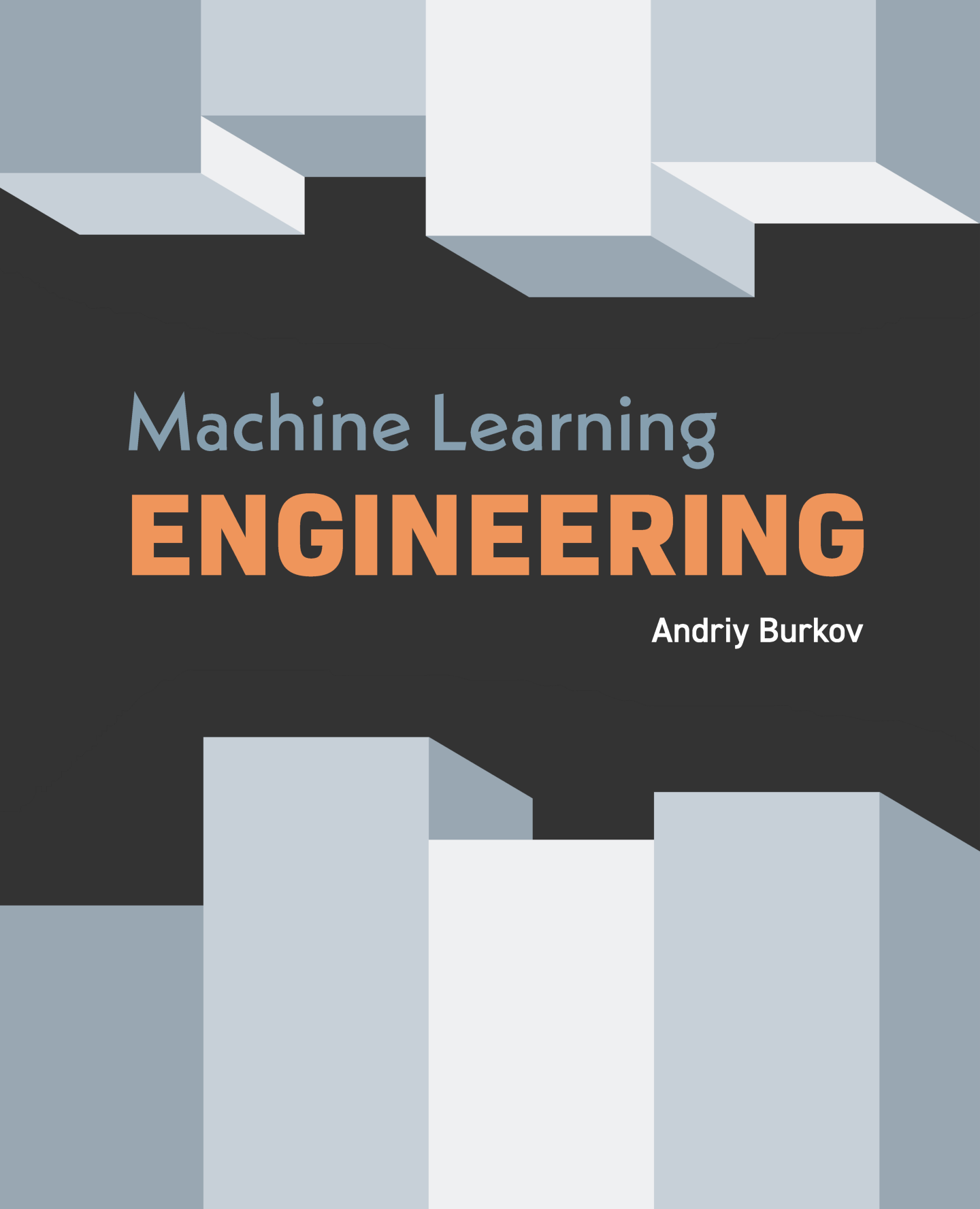
Machine Learning
ENGINEERING
Andriy Burkov

“In theory, there is no difference between theory and practice. But in practice, there is.”

— Benjamin Brewster

“The perfect project plan is possible if one first documents a list of all the unknowns.”

— Bill Langley

“When you’re fundraising, it’s AI. When you’re hiring, it’s ML. When you’re implementing, it’s linear regression. When you’re debugging, it’s printf().”

— Baron Schwartz

The book is distributed on the “read first, buy later” principle.

Preface

During the past several years, machine learning (ML), for many, has become a synonym for artificial intelligence. Even though machine learning, as a field of science, has existed for several decades, only a handful of organizations in the world have fully harnessed its potential. Despite the availability of modern open-source machine learning libraries, packages and frameworks supported by the leading organizations and broad communities of scientists and software engineers, most organizations are still struggling to apply machine learning for solving practical business problems.

One difficulty lies in the scarcity of talent. However, even when they have access to talented machine learning engineers and data analysts, in 2020, most organizations[1] still spend between 31 and 90 days deploying one model, while 18 percent of companies are taking longer than 90 days — some spending more than a year productionizing. The main challenges organizations face when developing ML capabilities, such as model version control, reproducibility, and scaling, are rather engineering than scientific.

There are plenty of good books on machine learning, both theoretical and hands-on. From a typical machine learning book, you can learn the types of machine learning, major families of algorithms, how they work, and how to build models from data using those algorithms.

A typical machine learning book is less concerned with the engineering aspects of implementing machine learning projects. Such questions as data collection, storage, preprocessing, feature engineering, as well as testing and debugging of models, their deployment to and retirement from production, runtime and post-production maintenance, are often left outside the scope of machine learning books.

This book intends to fill that gap.

Who This Book is For

I assume that the reader of this book understands machine learning basics and is capable of building a model, given a properly formatted dataset using a favorite programming language or a machine learning library. If you don't feel comfortable applying machine learning algorithms to data and don't clearly see the difference between logistic regression, support vector machine, and random forest, I recommend starting your journey with The Hundred-Page Machine Learning Book, and then move to this book.

The target audience of this book is data analysts who lean towards a machine learning engineering role, machine learning engineers who want to bring more structure to their work, machine learning engineering students, as well as software architects who happen to deal with models provided by data analysts and machine learning engineers.

[1] "2020 state of enterprise machine learning", Algorithmia, 2019.

How to Use This Book

This book is a comprehensive review of machine learning engineering best practices and design patterns. I recommend reading it from beginning to end. However, you can read chapters in any order as they cover distinct aspects of the machine learning project lifecycle and do not have direct dependencies.

Should You Buy This Book?

Like its companion and precursor The Hundred-Page Machine Learning Book, this book is distributed on the "read-first, buy-later" principle. I firmly believe that readers must be able to read a book before paying for it; otherwise, they buy a pig in a poke.

The "read-first, buy-later" principle implies that you can freely download the book, read it, and share it with your friends and colleagues. If you read and liked the book, or found it helpful or useful in your work, business, or studies, then buy it.

Now you are all set. Enjoy your reading!

Andriy Burkov