

ADDISON
WESLEY
DATA &
ANALYTICS
SERIES



An Introduction to
Cloud-Based
Machine Learning

Pragmatic AI



NOAH GIFT

Praise for *Pragmatic AI*

“[This is] a sweeping guide that bridges the gap between the promise of AI and solutions to the gritty problems that must be solved to deploy real-world projects. Clear and usable, *Pragmatic AI* covers much more than just Python and AI algorithms.”

—**Christopher Brousseau**, founder and CEO of Surface Owl, the Enterprise AI platform

“A fantastic addition for any technology enthusiast! There is so much you can say about this book! Noah Gift really made this a practical guide for anyone involved with machine learning in the industry. Not only does it explain how one can apply machine learning on large data sets, it provides a valuable perspective on technology feedback loops. This book will benefit many data science and development teams so they can create and maintain their applications efficiently right from the beginning.”

—**Nivas Durairaj**, technical account manager, AWS (Certified Professional Architect AWS)

“A great read if you want insights into building production-quality ML pipelines and tools that truly help your data engineering, data science, or data DevOps team. Even experienced developers often find themselves spinning their wheels on low-productivity tasks. Oftentimes, software books and university classes don’t explain the steps needed to go into production. Noah has a gift for finding pragmatic approaches to software deployments that can really accelerate the development and delivery process. He has a focus and passion for enabling rapid software solutions that is very unique.

“The key to building production-quality ML pipelines is automation. Tasks and steps, which engineers may do manually during the research or prototype phase, must be automated and scaled in order to create a production system. This book is full of practical and fun exercises that will help any Python developer automate and extend their pipelines into the cloud.

“I’m currently working with big data, ML pipelines, Python, AWS, Google cloud, and Azure at Roofstock.com, an online real estate company. Our analytics database is approaching 500 million rows! Within this book I found many practical tips and exercises that will immediately improve my own productivity. Recommended!”

—**Michael Vierling**, lead engineer, Roofstock

Pragmatic AI: An Introduction to Cloud-Based Machine Learning

Table of Contents

Cover

Half Title

Title Page

Copyright Page

Dedication

Contents

Preface

Acknowledgments

About the Author

I: Introduction to Pragmatic AI

1 Introduction to Pragmatic AI

Functional Introduction to Python

Procedural Statements

Printing

Create Variable and Use Variable

Multiple Procedural Statements

Adding Numbers

Adding Phrases

Complex Statements

Strings and String Formatting

Adding and Subtracting Numbers

Multiplication with Decimals

Using Exponents

Table of Contents

Converting Between Different Numerical Types

Rounding Numbers

Data Structures

Dictionaries

Lists

Functions

Using Control Structures in Python

for Loops

While Loops

If/Else

Intermediate Topics

Final Thoughts

2 AI and ML Toolchain

Python Data Science Ecosystem: IPython, Pandas, NumPy, Jupyter Notebook,
Sklearn

R, RStudio, Shiny, and ggplot

Spreadsheets: Excel and Google Sheets

Cloud AI Development with Amazon Web Services

DevOps on AWS

Continuous Delivery

Creating a Software Development Environment for AWS

Integrating Jupyter Notebook

Integrating Command-Line Tools

Integrating AWS CodePipeline

Basic Docker Setup for Data Science

Other Build Servers: Jenkins, CircleCI, and Travis

Summary

3 Spartan AI Lifecycle

Pragmatic Production Feedback Loop

AWS SageMaker

AWS Glue Feedback Loop

Table of Contents

AWS Batch

Docker-based Feedback Loops

Summary

II: AI in the Cloud

4 Cloud AI Development with Google Cloud Platform

GCP Overview

Colaboratory

Datalab

Extending Datalab with Docker and Google Container Registry

Launching Powerful Machines with Datalab

BigQuery

Moving Data into BigQuery from the Command Line

Google Cloud AI Services

Classifying my Crossbreed Dog with the Google Vision API

Cloud TPU and TensorFlow

Running MNIST on Cloud TPUs

Summary

5 Cloud AI Development with Amazon Web Services

Building Augmented Reality and Virtual Reality Solutions on AWS

Computer Vision: AR/VR Pipelines with EFS and Flask

Data Engineering Pipeline with EFS, Flask, and Pandas

Summary

III: Creating Practical AI Applications from Scratch

6 Predicting Social-Media Influence in the NBA

Phrasing the Problem

Gathering the Data

Collecting Challenging Data Sources

Collecting Wikipedia Pageviews for Athletes

Collecting Twitter Engagement for Athletes

Exploring NBA Athlete Data

Table of Contents

Unsupervised Machine Learning on NBA Players

Faceting Cluster Plotting in R on NBA Players

Putting it All Together: Teams, Players, Power, and Endorsements

Further Pragmatic Steps and Learnings

Summary

7 Creating an Intelligent Slackbot on AWS

Creating a Bot

Converting the Library into a Command-Line Tool

Taking the Bot to the Next Level with AWS Step Functions

Getting IAM Credentials Set Up

Working with Chalice

Building Out the Step Function

Summary

8 Finding Project Management Insights from a GitHub Organization

Overview of the Problems in Software Project Management

Exploratory Questions to Consider

Creating an Initial Data Science Project Skeleton

Collecting and Transforming the Data

Talking to an Entire GitHub Organization

Creating Domain-specific Stats

Wiring a Data Science Project into a CLI

Using Jupyter Notebook to Explore a GitHub Organization

Pallets GitHub Project

Looking at File Metadata in the CPython Project

Looking at Deleted Files in the CPython Project

Deploying a Project to the Python Package Index

Summary

9 Dynamically Optimizing EC2 Instances on AWS

Running Jobs on AWS

Table of Contents

Spot Instances

Summary

10 Real Estate

Exploring Real Estate Values in the United States

Interactive Data Visualization in Python

Clustering on Size Rank and Price

Summary

11 Production AI for User-Generated Content

The Netflix Prize Wasnt Implemented in Production

Key Concepts in Recommendation Systems

Using the Surprise Framework in Python

Cloud Solutions to Recommendation Systems

Real-World Production Issues with Recommendations

Real-World Recommendation Problems: Integration with Production APIs

Cloud NLP and Sentiment Analysis

NLP on Azure

NLP on GCP

Exploring the Entity API

Production Serverless AI Pipeline for NLP on AWS

Summary

A: AI Accelerators

B: Deciding on Cluster Size

Index