

Python Machine Learning Blueprints

Intuitive data projects you can relate to

An approachable guide to applying advanced machine learning methods to everyday problems

Alexander T. Combs



BIRMINGHAM - MUMBAI

Table of Contents

Preface	1
Chapter 1: The Python Machine Learning Ecosystem	7
The data science/machine learning workflow	8
Acquisition	9
Inspection and exploration	9
Cleaning and preparation	9
Modeling	10
Evaluation	10
Deployment	10
Python libraries and functions	10
Acquisition	11
Inspection	11
The Jupyter notebook	12
Pandas	13
Visualization	20
The matplotlib library	21
The seaborn library	27
Preparation	29
Map	29
Apply	30
Applymap	32
Groupby	33
Modeling and evaluation	36
Statsmodels	36
Scikit-learn	39
Deployment	45
Setting up your machine learning environment	45
Summary	46
Chapter 2: Build an App to Find Underpriced Apartments	47
Sourcing the apartment listing data	48
Pulling listing data using import.io	48
Inspecting and preparing the data	52
Analyzing the data	62
Visualizing the data	69
Modeling the data	71

Forecasting	74
Extending the model	78
Summary	78
Chapter 3: Build an App to Find Cheap Airfares	79
<hr/>	
Sourcing airfare pricing data	80
Retrieving the fare data with advanced web scraping techniques	82
Parsing the DOM to extract pricing data	84
Identifying outlier fares with clustering techniques	87
Sending real-time alerts using IFTTT	99
Putting it all together	103
Summary	107
Chapter 4: Forecast the IPO Market using Logistic Regression	109
<hr/>	
The IPO market	110
What is an IPO?	110
Recent IPO market performance	111
Baseline IPO strategy	120
Feature engineering	121
Binary classification	131
Feature importance	137
Summary	141
Chapter 5: Create a Custom Newsfeed	143
<hr/>	
Creating a supervised training set with the Pocket app	144
Installing the Pocket Chrome extension	144
Using the Pocket API to retrieve stories	145
Using the embed.ly API to download story bodies	151
Natural language processing basics	153
Support vector machines	156
IFTTT integration with feeds, Google Sheets, and e-mail	159
Setting up news feeds and Google Sheets through IFTTT	159
Setting up your daily personal newsletter	169
Summary	174
Chapter 6: Predict whether Your Content Will Go Viral	175
<hr/>	
What does research tell us about virality?	176
Sourcing shared counts and content	177
Exploring the features of shareability	187
Exploring image data	187
Exploring the headlines	192

Exploring the story content	197
Building a predictive content scoring model	199
Summary	205
Chapter 7: Forecast the Stock Market with Machine Learning	207
<hr/>	
Types of market analysis	208
What does research tell us about the stock market?	209
How to develop a trading strategy	210
Extending our analysis period	218
Building our model with a support vector regression	221
Evaluating our model's performance	224
Modeling with Dynamic Time Warping	231
Summary	235
Chapter 8: Build an Image Similarity Engine	237
<hr/>	
Machine learning on images	238
Working with images	239
Finding similar images	242
Understanding deep learning	249
Building an image similarity engine	252
Summary	262
Chapter 9: Build a Chatbot	263
<hr/>	
The Turing test	263
The history of chatbots	264
The design of chatbots	269
Building a chatbot	274
Summary	286
Chapter 10: Build a Recommendation Engine	287
<hr/>	
Collaborative filtering	288
User-to-user filtering	289
Item-to-item filtering	293
Content-based filtering	296
Hybrid systems	297
Building a recommendation engine	298
Summary	315
Index	317
<hr/>	