

Table of Contents

Preface	1
Chapter 1: Machine Learning for Cybersecurity	7
Technical requirements	8
Train-test-splitting your data	8
Getting ready	8
How to do it...	9
How it works...	10
Standardizing your data	10
Getting ready	11
How to do it...	11
How it works...	12
Summarizing large data using principal component analysis	12
Getting ready	13
How to do it...	13
How it works...	15
Generating text using Markov chains	15
Getting ready	16
How to do it...	16
How it works...	17
Performing clustering using scikit-learn	18
Getting ready	19
How to do it...	19
How it works...	22
Training an XGBoost classifier	22
Getting ready	22
How to do it...	23
How it works...	24
Analyzing time series using statsmodels	24
Getting ready	24
How to do it...	24
How it works...	27
Anomaly detection with Isolation Forest	28
Getting ready	28
How to do it...	28
How it works...	33
Natural language processing using a hashing vectorizer and tf-idf with scikit-learn	34
Getting ready	34

How to do it...	35
How it works...	37
Hyperparameter tuning with scikit-optimize	37
Getting ready	38
How to do it...	38
How it works...	41
Chapter 2: Machine Learning-Based Malware Detection	43
Technical requirements	44
Malware static analysis	44
Computing the hash of a sample	44
Getting ready	45
How to do it...	45
How it works...	46
YARA	47
Getting ready	47
How to do it...	48
How it works...	48
Examining the PE header	49
Getting ready	49
How to do it...	49
How it works...	51
Featurizing the PE header	52
Getting ready	52
How to do it...	52
How it works...	54
Malware dynamic analysis	54
Getting ready	54
How to do it...	55
How it works...	59
Using machine learning to detect the file type	59
Scraping GitHub for files of a specific type	59
Getting ready	60
How to do it...	60
How it works...	62
Classifying files by type	62
Getting ready	62
How to do it...	63
How it works...	64
Measuring the similarity between two strings	65
Getting ready	65
How to do it...	65
How it works...	66
Measuring the similarity between two files	67
Getting ready	67
How to do it...	68
How it works...	69

Extracting N-grams	69
Getting ready	69
How to do it...	70
How it works...	71
Selecting the best N-grams	72
Getting ready	72
How to do it...	72
How it works...	74
Building a static malware detector	75
Getting ready	75
How to do it...	75
How it works...	80
Tackling class imbalance	81
Getting ready	81
How to do it...	81
How it works...	85
Handling type I and type II errors	85
Getting ready	86
How to do it...	86
How it works...	88
Chapter 3: Advanced Malware Detection	89
Technical requirements	90
Detecting obfuscated JavaScript	90
Getting ready	90
How to do it...	91
How it works...	92
Featurizing PDF files	93
Getting ready	93
How to do it...	94
How it works...	95
Extracting N-grams quickly using the hash-gram algorithm	95
Getting ready	95
How to do it...	96
How it works...	98
See also	99
Building a dynamic malware classifier	99
Getting ready	99
How to do it...	100
How it works...	104
MalConv – end-to-end deep learning for malicious PE detection	105
Getting ready	105
How to do it...	106
How it works...	108
Tackling packed malware	109

Using packers	109
Getting ready	109
How to do it...	110
How it works...	111
Assembling a packed sample dataset	111
Getting ready	111
How to do it...	111
How it works...	112
Building a classifier for packers	112
Getting ready	112
How to do it...	113
How it works...	116
MalGAN – creating evasive malware	116
Getting ready	116
How to do it...	117
How it works...	118
Tracking malware drift	119
Getting ready	119
How to do it...	119
How it works...	123
Chapter 4: Machine Learning for Social Engineering	125
Technical requirements	126
Twitter spear phishing bot	126
Getting ready	126
How to do it...	127
How it works...	130
Voice impersonation	130
Getting ready	131
How to do it...	131
How it works...	133
Speech recognition for OSINT	134
Getting ready	134
How to do it...	134
How it works...	136
Facial recognition	136
Getting ready	136
How to do it...	137
How it works...	140
Deepfake	140
Getting ready	140
How to do it...	141
How it works...	144
Deepfake recognition	144
Getting ready	144
How to do it...	144

How it works...	146
Lie detection using machine learning	147
Getting ready	147
How to do it...	147
How it works...	152
Personality analysis	152
Getting ready	152
How to do it...	152
How it works...	154
Social Mapper	154
Getting ready	155
How to do it...	155
How it works...	157
Fake review generator	157
Training a fake review generator	157
Getting ready	157
How to do it...	158
How it works...	160
Generating fake reviews	160
Getting ready	160
How to do it...	161
How it works...	163
Fake news	163
Getting ready	163
How to do it...	164
How it works...	167
Chapter 5: Penetration Testing Using Machine Learning	169
Technical requirements	170
CAPTCHA breaker	170
Processing a CAPTCHA dataset	171
Getting ready	172
How to do it...	172
How it works...	175
Training a CAPTCHA solver neural network	177
Getting ready	178
How to do it...	178
How it works...	182
Neural network-assisted fuzzing	182
Getting ready	182
How to do it...	183
How it works...	185
DeepExploit	185
Getting ready	186
How to do it...	186
How it works...	190

Web server vulnerability scanner using machine learning (GyoiThon)	191
Getting ready	192
How to do it...	193
How it works...	194
Deanonymizing Tor using machine learning	195
Getting ready	195
How to do it...	195
Collecting data	196
Training	197
Predicting	198
How it works...	199
IoT device type identification using machine learning	199
Getting ready	199
How to do it...	200
How it works...	201
Keystroke dynamics	201
Getting ready	202
How to do it...	202
How it works...	204
Malicious URL detector	204
Getting ready	204
How to do it...	205
How it works...	207
Deep-pwning	207
Getting ready	208
How to do it...	208
How it works...	210
Deep learning-based system for the automatic detection of software vulnerabilities	211
Getting ready	211
How to do it...	212
How it works...	214
Chapter 6: Automatic Intrusion Detection	217
Technical requirements	218
Spam filtering using machine learning	218
Getting ready	218
How to do it...	218
How it works...	220
Phishing URL detection	220
Getting ready	221
How to do it...	222
How it works...	222
Capturing network traffic	224

Getting ready	224
How to do it...	224
How it works...	225
Network behavior anomaly detection	226
Getting ready	226
How to do it...	227
How it works...	231
Botnet traffic detection	231
Getting ready	232
How to do it...	232
How it works...	233
Insider threat detection	233
Feature engineering for insider threat detection	234
Getting ready	234
How to do it...	235
How it works...	239
Employing anomaly detection for insider threats	240
Getting ready	240
How to do it...	240
How it works...	244
Detecting DDoS	245
Getting ready	245
How to do it...	245
How it works...	247
Credit card fraud detection	247
Getting ready	248
How to do it...	248
How it works...	250
Counterfeit bank note detection	250
Getting ready	251
How to do it...	251
How it works...	252
Ad blocking using machine learning	252
Getting ready	252
How to do it...	253
How it works...	254
Wireless indoor localization	255
Getting ready	255
How to do it...	256
How it works...	257
Chapter 7: Securing and Attacking Data with Machine Learning	259
Technical requirements	259
Assessing password security using ML	260
Getting ready	260
How to do it...	260

How it works...	262
Deep learning for password cracking	263
Getting ready	263
How to do it...	264
How it works...	265
There's more	265
Deep steganography	265
Getting ready	266
How to do it...	266
How it works...	269
ML-based steganalysis	270
Getting ready	273
How to do it...	273
How it works...	275
ML attacks on PUFs	276
Getting ready	277
How to do it...	277
How it works...	278
There's more	278
Encryption using deep learning	279
Getting ready	279
How to do it...	279
How it works...	280
HIPAA data breaches – data exploration and visualization	281
Getting ready	281
How to do it...	281
How it works...	287
Chapter 8: Secure and Private AI	289
Technical requirements	289
Federated learning	290
Getting ready	290
How to do it...	291
How it works...	294
Encrypted computation	294
Getting ready	294
How to do it...	295
How it works...	296
Private deep learning prediction	296
Getting ready	297
How to do it...	297
How it works...	299
Testing the adversarial robustness of neural networks	300
Getting ready	301
How to do it...	301

How it works...	307
Differential privacy using TensorFlow Privacy	307
Getting ready	308
How to do it...	308
How it works...	310
Appendix	311
Other Books You May Enjoy	321
Index	325