Index

access control, 12, 16, 24, 28, 49
accommodation, 99
accuracy, 29, 40–41, 50
adjacency matrix, 130, 135
advanced persistent threat, 5, 69, 84–85, 139
anomalous window, 118
anomaly, 18, 27–28, 34, 40, 42, 49, 81, 91–93, 95–101, 104, 107, 110–112, 116–118, 120–123, 129, 147, 158
anomaly detection, 27, 34, 36, 91, 93, 100, 104, 116–117, 140
application security, 6
Apriori algorithm, 57–58
APT, 5, 79, 84–89, 139
assets, 1–3, 14, 16, 24, 79
association rule mining, 18–19, 36, 41, 59
asymmetric coefficient, 39
attack maps, 3, 113–114
attack vector, 85
autocorrelation, 67, 113–115, 117–118, 120, 122
autonomous systems, 95

Bayesian, 24
behavioral, 11
behavioral data, 23
betweenness, 131
BGP anomalies, 95
big data, 9, 20, 60–61, 63, 65–66, 69–70, 72, 74–76
blackhole attack, 79
Border Gateway Protocol, 95

categorical data, 32
central nodes, 70, 72, 132–133
centrality, 21, 131, 133
centroid, 34, 43, 110–111
CLARANS, 44–45
classification, 29–30, 40–41, 49–51, 55, 75, 90, 101, 152, 157
client server, 14
closeness, 131
cloud, 2–3, 6, 15–16, 76
clustering, 11, 25, 34, 37, 40, 42–45, 48–49, 53, 75, 92, 101, 105, 108–110, 117–118, 155
common vulnerabilities and exposures, 25
common vulnerability scoring system, 25, *See* CVSS
communication, 21
computational data model, 81–82, 84
computer networks, 3, 7–8, 11, 13, 16, 18, 27, 67, 124, 148, 154
conditional probability, 42, 55, 57
confidence, 30, 41, 57
consistent, 22, 28, 71–72, 89, 134
consistent nodes, 71–72
context, 93
contextual anomalies, 91, 94
contingency matrix, 38
CPS, 8, 67–68, 147, 149
critical infrastructure, 3, 80
curse of dimensionality, 96–97, 109
cyberattacks, 5, 24–26, 62, 68–70, 138–139, 144, 156
cyber espionage, 5
cyberphysical, 7, 16, 147
cyberphysical systems, 8, 16, 67, 147
cybersecurity, 1–2, 6, 14–15, 27, 60–61, 112, 123, 128–129, 133–134, 147
cybersecurity data, 28
cyber threats, 1

© in this web service Cambridge University Press & Assessment www.cambridge.org

data analytics, 7–9, 12, 60, 70, 137, 147
data cleaning, 33
data Mining, 8, 29, 36
data preprocessing, 31–32
data reduction, 35
data storage, 3, 6, 76
data transformation, 34
decision tree, 50–52, 54
deep learning, 157–158, 161
degrees, 22, 71, 132
denial of service, 4, 79, 125
densification, 133
density based, 42, 46–48
detection, 1, 13, 16–18, 20, 23, 25, 27, 34, 42, 46, 59, 65, 67, 70–71, 73–74, 81, 84, 88, 91, 93–97, 99–101, 103–105, 108–109, 111–112, 116–118, 120–123, 129, 139, 141, 147, 154, 158
diameter, 21, 69, 71, 127–128, 134
discord, 91, 121
discordancy, 34, 102–104, 108–109
dissimilarity, 37, 119
distance, 34, 37–38, 42–43, 46–49, 93, 101, 105, 107–109, 111, 113, 116, 122, 124, 131, 134
distance based, 105, 108
distance based outlier detection, 34
distributed control systems, 147
distributed denial of service (DDoS), 4, 125
distributed intrusion detection system, 73
domain name system, 14

e-commerce, 16
edges, 21, 45, 72, 122, 124, 128, 130–131, 133, 148, 151
eigenvector, 131
electronic assets, 1–2
ensemble, 50, 53–54, 75, 162
entropy, 52–53
equal frequency binning, 34
Euclidean distance, 37, 109, 118
evaluation, 27, 29–30, 37, 40, 72, 111, 139, 142, 149
exception, 91
explainability, 159
exploits, 86
eye tracker, 24

false negative, 40
false negative rate, 41
false positive, 40

firewall, 14, 19
first law of geography, 113
F-measure, 40–41
FP tree, 58
frequent itemset, 58
frequent pattern, 55–56, 89

Geary's, 113
generative adversarial network, 160–161
geocontextual features, 152
geographic, 11
geolocations, 19, 27, 94, 115, 126, 141
geological, 3
georeferenced, 68, 113
geospatial data, 152–153
graph data, 21
graphs, 21, 69, 129, 133–134
ground truth, 40

hacktivism, 5
Hausdorff distance, 38
header, 20
heartbleed, 5
heterogeneous, 9, 30–31, 37, 60, 67, 73, 120, 151
hierarchical, 42
high dimensional, 36, 47, 67, 96, 109
honeypots, 100
human behavior, 18

IDS, 17
IDS logs, 11
imbalanced learning, 53
inconsistent nodes, 71
industrial control systems, 3, 147–148
insider threat, 2, 24, 138, 140–142
Intellectual property, 4
intercluster, 42
interface design, 143
Internet of Things. *See* IoT
interquartile range, 34, 104
intracluster, 42
intrusion detection systems. *See* IDS
IoT, 8, 149–150
IP reputation scoring, 152
IQR, 34, 104

Jaccard coefficient, 39, 135
joint count, 113

K nearest neighbor, *See* K-NN
key logging, 17

CAMBRIDGE

Cambridge University Press & Assessment
978-1-108-41527-9 — Data Analytics for Cybersecurity
Vandana P. Janeja
Index
More Information

K-means, 43–44, 66
K-NN, 47, 109

lift, 30, 41, 84
linear semantic-based scan, 118
links, 16, 47, 69, 124, 128, 137, 139, 142, 148–149
logical errors, 5

macroneighborhoods, 120
malfunctions, 8, 150
malicious intent, 2, 81, 95, 142
malware, 19–20, 24, 28, 78, 81, 86, 124, 152
malware detection, 24
Manhattan distance, 37
Markov model, 122
Markov stationary distribution, 123
masking, 104
measures of similarity, 37
medoids, 43–45
microneighborhood, 119
MILES, 53–54
Minkowski distance, 37
missing data, 33
missing data values, 33
misuse, 6, 11, 27, 142
mobile ad hoc, 79
mobile devices, 3, 61
Moran’s, 113
multilevel mining, 73
multipronged attack, 9
multivariate, 34

naive Bayes, 55
National Vulnerability Database, 25
neighborhood, 46, 119–120, 124
network security, 7
network topology, 23, 134
network traffic, 21
nodes, 22–23, 28, 69–71, 76, 122–123, 125, 127–128, 131–133, 135, 148, 159
NodeXL, 21
noisy data, 34
normal distribution, 101–102
normalization, 35
numeric data attributes, 32
NVD, 25

OPTICS, 48–49, 105, 118
ordinal, 32
outlier detection, 34, 67, 101, 105, 108–109, 117
outliers, 34, 37, 46, 67, 91, 96–97, 99–105, 107–109, 111, 117, 121

page rank, 21
parametric, 101–102
partitioning, 42–44
password attack, 79
password theft, 81–83
payload, 19
peculiarity, 91
personal assets, 2
phishing, 5, 10, 25, 78, 80, 84–85, 124, 126, 138–139, 144, 154
physical security, 2, 11, 27, 68, 142
power grid, 2, 124
precision, 40–41, 151
predict, 14
prevention, 1, 13, 17, 129
probabilistic label, 55
probe, 81, 145
psychological profiling, 140
psychology, 15
public assets, 2

rare patterns, 56, 58–59
recall, 40
recover, 1
rejection, 100
replay attack, 79
respond, 1, 6, 137, 141
router, 14

scan statistic, 118
security indicators, 25
security perimeter, 9
security policies, 12, 19, 25, 138, 142, 152
semantic relationships, 120
semisupervised, 91
sensor networks, 3, 8, 16, 67, 124, 154
sequential, 103
silhouette coefficient, 42, 45
similarity, 32, 34, 38–40, 43, 104, 119–120, 122, 129, 135, 156, 161
simple matching coefficient, 39
social engineering, 5, 78, 80
social media, 16
software vulnerability, 25
spam, 126, 152
spatial, 11, 27, 32, 46, 67–68, 94, 108–109, 112–114, 116–118, 120–122, 124–125, 151

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-108-41527-9 — Data Analytics for Cybersecurity
Vandana P. Janeja
Index
More Information

spatial autocorrelation, 113–114
spatial data, 67, 112
spatial data mining, 67
spatial heterogeneity, 113
spatial neighborhoods, 32, 119, 124
spatial relationships, 117, 119, 125
spatiotemporal scan statistic, 118
spoofed, 121
spoofing, 19
SQL injection, 79, 86
standard deviation, 34, 83, 105
Stuxnet, 124
sum of squared error, 30, 42
supervised, 91
supervisory control and data acquisition, 147
support, 12, 41, 55, 57–59, 94, 132, 140
support vector machines, 55
swamping, 104

temporal, 11, 25, 27–28, 67, 69, 88, 94, 112, 115, 118, 121, 123, 125–126, 133, 145, 149, 151, 155
threat hunting, 7
threat taxonomy, 78
threats, 1, 6–7, 9, 12, 14–15, 17–18, 20, 23, 25, 28, 63, 70, 79, 85, 88–90, 112, 123, 129, 136–137, 139, 141, 146–147, 150–151, 154, 162
time series discords, 121
Titan Rain attack, 5
topology analysis, 140
Trojan, 87, 100
true negative, 40
true negative rate, 41
true positive, 40
true positive rate, 41

unauthorized access, 1–2, 4–6, 19, 70, 81, 154
univariate data, 34
unsupervised, 34, 91
unusual window, 118
user interaction, 143
user's perception, 139

variety, 9
velocity, 9
veracity, 9
vertices, 21, 124, 128, 130–131
volume, 9
vulnerability, 9, 25, 38–39, 69, 80, 82, 86, 125, 148, 150
vulnerable, 2, 84–85, 136–137, 143, 162

walk, 130
watering hole, 85
weighted degree, 72
white listing, 129
wireless networks, 8
Wireshark, 20

zero day, 86, 129, 154
z-score, 35

© in this web service Cambridge University Press & Assessment www.cambridge.org