

Index

A

- Amazon Mechanical Turk, 40
- Artificial neural networks, 111–113
 - hidden nodes, 112
- Authentic websites, 32
- Automated theorem proving (ATP), 123
- Autonomous systems, 161
- Avalanche effect, 162

B

- Backtracking, 132
- Backward chaining, 131
- Bayes theorem, 103
- Big Data
 - algorithms, 2, 4
 - analytics, 1
 - areas, 3
 - businesses, 1
 - characteristics, 9
 - computing, 1
 - developments, 3
 - falsity and inaccuracy, 65
 - fission, 3
 - fusion, 3
 - human, 2–3

- identify patterns, 4
- Naïve Bayes technique, 4
- OSN, 1
- overfitting, 4
- RDBMS, 5
- relationship, V's, 8
- sources, 2
- support vector machine model, 4
- training data, 4
- uncertainty, 1
- valuation, 7
- value, 6
- variability, 6
- variety, 6, 8
- velocity, 6
- veracity (*see* Veracity data)
- viability, 7
- visualization, 7
- volume, 5
- Web 2.0, 2
- Blockchain
 - asset, 162
 - avalanche effect, 162
 - Big Data veracity, 167–168
 - concept, 156
 - consensus-based truth, 157
 - cryptocurrency, 61, 155

INDEX

Blockchain (*cont.*)

- cryptographic hash
 - functions, 161
- cryptographic hash value, 166
- cryptography, 156
- database, 156
- data structure, 161
- fake news, 61
- genesis block, 162
- hash value, 166
- hit-and-trial process, 167
- implementation, 159, 169
- Internet of Things (IoT), 157
- ledger forms, 157–158
- math puzzle, 166
- Merkle Path, 164
- Merkle tree, 163–165
- mining, 167
- NIST, 160
- nonce value, 166
- online monopoly, 161
- P2P network, 156, 161, 164
- pre-image resistance, 163
- problems, 158
- puzzle, 166
- quality of data, 169
- requirements, design, 158
- schematic, 162
- SHA values, 160
- smart contract feature, 161
- spam emails, 166
- technology, 61
- tracing Big Data, 155
- tracking, 162

- traditional financial systems, 157
- transaction data, 156
- transactions, 161

Boolean Logic, 139

- sound, 120

Bots, 21

Brown corpus, 69

C

Central Intelligence Agency (CIA), 30

Change detection

- CUSUM algorithm, 43, 46
- estimation/statistical
 - techniques, 43
- feedback loop, 44
- fine-tuning parameters, 45
- Kalman Filter, 43–45
- loss of integrity, 43
- microblogging websites, 42
- real-time data, 46
- sensor measurements, 42
- veracity, 47

Clickstream data, 24

Closed World Assumption (CWA), 121, 122, 133, 156

Cognitive hacking, 19, 22

Collaborative filtering (CF), 59–60

- ColdStart problem, 150
- formal settings, 145
- k-NN algorithm, 149
- measures, 148
- microblogging users, 146

- Pearson's Correlation
 - Coefficient, 148
 - process, 146–147
 - QA system, 146
 - truthfulness, 148
 - truth matrix, 147
 - user based/claim based, 148
 - users, 149
 - veracity domain, 146
 - Conditional probability, 103
 - Confusion matrix, 48
 - Connected cars, hacking, 167
 - Context Free Grammar (CFG), 53
 - Contingents, 122
 - Contradiction, 122
 - Corollaries, 122
 - Corpus of Contemporary American
 - English (COCA), 69
 - Crash-avoidance system, 14
 - Crowdsourcing, 40
 - Cryptocurrencies, 155, 157
 - Cryptographic hash functions,
 - 160–161
 - Cryptographic hash value, 164, 166
 - Cryptography, 164
 - Cumulative LLR (CLLR), 72–73
 - CUSUM technique
 - algorithm, 74, 76, 78
 - data values, 77
 - development, 75
 - distribution, data, 75
 - finance and process control, 74
 - Gaussian PDF, 76
 - process, 77
 - representative state, 74
 - simple and complex, 74
 - and SPRT, 74
 - stopping time, 76
 - threshold, 76
 - values, 78, 79
 - Cybersquatting, 25
- ## D
- Dark Web, 25
 - Database, Blockchain, 156
 - DeepDream project, 87
 - Deep learning, 114
 - Detection techniques
 - advantages, 69
 - analogy, 68
 - change, 66
 - CUSUM (*see* CUSUM technique)
 - Gaussian Assumption, 68
 - hacked attacks, 67
 - Internet of Things (IoT), 65
 - interpretation, falsity, 67
 - Kalman Filter, 80
 - manifestation, 65
 - microblogs, 66
 - monitoring, 68, 70
 - quality of data, 65
 - statistics, 68, 69
 - untruthfulness, 67
 - Digital universe, 21
 - Discriminative approach, 103–104
 - Disjunction Syllogism, 128

INDEX

E

- Effects, Web information, 24, 26
- Electronic media, 17
- Entropy, 142
- Error matrix, 48
- European Union, 27
- Extract, Transform, Load (ETL) functions, 9
- Existential generalization, 136

F

- Face2Face, 23
- False positives, 46
- Feedback Control System, 83
- First Order Logic (FOL), 133, 134
- Formal methods, 55–57
 - Blockchain technology, 120
 - closed world assumption, 121
 - expert systems, 121
 - handling semantics, 119
 - lemma, 122
 - mathematical logic, 120
 - mathematical techniques, 120
 - process of entailment, 121
 - quantification, 56
 - RDF, 119
 - representation of information, 119
 - tautology, 122
 - terminology, 122–123
 - weak method, 121
 - Zeroth-Order-Logic, 55

Forward chaining, 130

Fuzzy logic, 57–58

- approximate reasoning, 58
- Boolean Logic, 139, 142
- elastic constraints, 139
- entropy, 142
- Gaussian membership function, 140
- granulation and graduation, 142
- logical operators, 142
- Machine Learning algorithms, 139
- membership function, 58, 139
- neural networks, 139
- probability density function, 141
- Propositional Logic and Predicate Calculus, 142
- soft computing, 139
- trapezoidal membership function, 140–141

G

- Generative approach, 103–104
- Genome research, 3
- Genomic data science, 3
- Google Flu Trends (GFT), 10

H

- Hacked attacks, 67
- Hash chains, 62
- Hidden layer, 112

Higher order logic (HOL), 133
Horn clauses, 132

I, J

India, 11

Information retrieval
 techniques, 59–61
 Text Mining, 59

TF. IDF, 60

Vector Space Model, 59

Internet of Things (IoT), 157, 167

Intractable computation, 127

Iterative algorithm, 116

K

Kalman Filter
 accuracy, 83
 components, 81
 data values, 81–82
 Feedback Control System, 82–83
 Gaussian assumption, 80
 interpretations, 83
 Kalman Gain, 81
 linear state-space approach, 80
 mean squared error, 82
 measurement update, 83
 negative values, 85
 noise, 80
 observation/measurement, 81
 stock market, 81
 time update, 83
 training data, 82
 values, 84

Kernel functions, 110
K-means clustering, 114–117

 Canberra Metric, 115
 centroid, 116

k-NN algorithm, 150

Knowledge Base (KB), 132, 156

L

L&T Infotech, 52

Lizard Squad, 26

Logistic regression, 39, 42

 logistic function, 98

 log-likelihood, 102

 log odds, 101

 multidimensional space, 98

 optimization techniques, 102

 parameter estimation, 102

 plot of, 99

 sigmoid functions, 99, 101

 step function, 99–100

Log-likelihood ratio (LLR), 71

M

Machine Learning algorithms, 12, 82
 Amazon Mechanical Turk, 40
 cluster, 89
 decision tree, 88
 deterministic methods, 88
 features, 36, 40, 87
 frequentist approach, 88
 ground truth, 95–97
 hallucinating machines, 87
 k-means, 90

INDEX

Machine Learning

- algorithms (*cont.*)
 - labeling, 96
 - learning process, 88
 - logistic function, 39
 - logistic regression
 - model, 39, 42
 - Maximum Likelihood, 40
 - Misinformation Containment (MC), 36
 - Naive Bayes, 88
 - n-dimensional space, 93
 - neural networks, 89–90
 - neurons, 89
 - polynomial, 36
 - probability distribution, 37
 - random variable, 37
 - sentiment analysis, 41
 - statistical techniques, 42
 - supervise and unsupervised process, 89–90
 - test set, 38, 88
 - training set, 38, 87
 - weight, 39
- Maximum Likelihood, 40
- Maximum Likelihood Estimate (MLE), 102
- Maximum margin classifier, 109
- Mean Squared Error (MSE), 82
- Merkle tree, 163–165
- Microblogging users, 146
- Microblogging websites
 - classifier, 90, 95
 - features, 91–92

hypersurface, 92–93

mathematical

abstraction, 91

multidimensional space, 93

vector, 92

veracity of, 94

Microwave nutrients, 97

Misinformation Containment

(MC), 12

Mobile Virtual Reality, 26

N

Naïve Bayes classifier

conditional probability, 103

discriminative approach,

103–104

generative approach, 103–104

independence

assumption, 105

logistic regression, 103

probabilities, 106

veracity problem, 105

Naïve Bayes technique, 4

Natural language processing

CFG, 53

characteristics of

authenticity, 53

Language Model, 52

linguistic formality, 53

Parse Thicket, 54

Probabilistic Context Free

Grammar, 53

statistical distribution, 52

systematic web, 52
 veracity problem, 52
 Zipf's law, 52
 Neural networks, 89–90

O

Online Social Networks (OSN), 1
 Optimization techniques
 confusion matrix, 48
 cost function, 49
 costs, 47
 False Negative condition, 49
 fine-tuning parameters, 47
 logistic regression model, 50
 Machine Learning model, 51
 type II error, 49

P

PageRank, 30
 Parameter estimation, 102
 Peer-to-Peer (P2P) network,
 161, 164
 Power law distributions, 69
 Predicate Calculus
 Big Data, 133
 constant symbols, 134
 domain of discourse, 137
 existential elimination, 136
 existential generalization, 136
 introduction and elimination
 rules, 135
 quantification, 134

quantifiers, 134
 teacher's claim, 137–138
 universal quantifier, 135
 universe of discourse, 137
 Probability density function
 (PDF), 71
 Prolog program, 132
 Propositional logic
 backtracking, 132
 backward chaining, 131
 conjunctive and
 operator, 125
 contradiction, 127
 converse, inverse and
 contrapositive, 126
 Disjunction Syllogism, 128
 disjunctive or operator, 125
 equivalence operator, 126
 forward chaining, 130
 generalization, 128
 HOL and FOL, 133
 horn clauses, 132
 identities, 129
 implication operator, 125
 inference rules, 127, 130
 Knowledge Base (KB), 132
 literals, 132
 modus ponens, 56, 127
 notation, 124
 operators, 123–124
 Prolog program, 132
 second order logic, 133
 sequent calculus, 131
 specialization, 128

INDEX

Q

Quantifiers, 134

R

RDBMS, *see* Relational Database Management Systems (RDBMS)

Recognizing Textual Entailment (RTE), 54

Relational Database Management Systems (RDBMS), 5

Resource description framework (RDF), 119

Rules of inference, 122

S

Secure Hash Algorithm (SHA), 160

Securities and Exchange Commission (SEC), 19

Self-driving, 167

Semantic web, 119

Sensor data, 10, 14

Sensors, 65

Sequent calculus, 131

Sequential Probability Ratio Test (SPRT), 43

cumulative LLR, 72–73

decision boundaries, 73

errors, 72

hypotheses, 71

independent and identically distributed, 71

LLR, 71–72

monitoring, 73

null hypothesis, 71

PDF, 71

SAT, 70

statistical hypothesis testing, 71

threshold value, 74

type I and type II errors, 72

Smart contract, 161

Social media, 15, 18, 28

Spam emails, 166

Speech synthesis software, 23

Speed versus Accuracy tradeoff (SAT), 70

Spoofing websites, 20

Stock market, 81

Support vector machine (SVM), 4

class boundaries, 111

hypersurface, 109–110

kernel functions, 110

linear classifiers, 108

maximum margin classifier, 109

Support Vectors, 110

training data, 107

tweet falls, 107

vectors, 109

T

Tautology, 122

Tax evasion, 52

Theory of probability, 83

Threat Intelligence Platform (TIP), 30

Training data, 4, 82–84
 Triple stores, 119
 Trusted websites, 31–32
 Trustless network, 156
 Twitter, 12

U

Uncertainty, 9

V

Valuation, Big Data, 7
 Variability, data, 6
 Vector Algebra, 92
 Vector space model (VSM), 59, 60

- measures, 153
- metrics, 154
- multidimensional space, 151
- TF.IDF value, 152
- three-dimensional
 - space, 151–152
- true/fraudulent
 - documents, 151

 Veracity data

- Big Data, 6, 167–168
- factors, 11
- machine data, 13
- machine learning
 - algorithms, 12
- microblogs, 12
- models and methods, 11
- OSN, 10
- poor data quality, 10, 14

- quality, 13
- sensor data, 14
- social media, 13
- sources, 14
- value, 9
- Web information (*see* Web
 - information)

Viability, Big Data, 7

Volume, data, 5

W, X, Y

Web information

- Bayes' theorem, 29
- biometric identification, 30
- causes, 21–24
- characteristics, 31–32
- cost, 19
- crimes, 28
- earthquake, 18
- effective, 27
- effects, 24, 26
- electronic media, 17
- false tweets, 19
- fraud, 27
- machine learning
 - algorithms, 30
- object-oriented paradigm, 30
- privacy protections, 29
- problem, 20
- “Pump-and-Dump”
 - pattern, 25
- ranking, 30
- remedies, 27, 29–31

INDEX

Web information (*cont.*)

schemes, 19

social media, 18, 28

source, 17

user behavior, 19

vicious cycle, 27

Wikipedia, 31

Wikipedia, 31

Z

Zipfian probability distribution, 69