Index

A
A/B testing, 617, 728, 739, 755, 757, 758
Academic Medical Centers, 765–771, 785, 787–788
ACF, *see* Auto-covariance function (ACF)
Activation Forward Propagation, 538
Ad Exchange, 726, 728
Ad servers, 725
Adaptive-response-rate single exponential smoothing (ARRSES), 392, 409, 411, 415
Adipose tissue problem, 180–181, 189
Adjusted R^2, 200–204, 220–222, 239
Advanced inventory models, 839
Affordable Care Act (ACA), 13, 770
Aggregate planning, 337
Aila, 617
Airbnb, 652
Akaike information criterion (AIC), 221–223, 263–265, 268, 397, 428, 434, 453, 455
Amazon, 78, 105, 106, 509, 556, 565, 602, 652, 728, 744, 803, 804, 824, 826
Amazon Go, 605
Amazon Prime, 550, 652
Amazon S3 (Simple Storage Service), 78, 105
Amazon Web Services (AWS), 105, 106
AmazonDB, 80
AMC, *see* Asset management company (AMC)
Anaconda, 919, 924
Analysis of variance (ANOVA), 4, 137, 166–169, 171, 427, 728, 730–733, 735–738
ANN, *see* Artificial neural network (ANN)
Anscombe's quartet, 114, 115, 243
Apache OpenNLP, 298
Apple, 102, 108, 510, 824, 830
Apple Watch, 766
Application programming interface (API), 22, 34, 36, 37, 80, 96, 97, 99, 106, 285
Arbitrage-free drift, 670
Arbitrage pricing theory (APT), 12, 662
ARCH, *see* Auto-regressive conditional heteroscedasticity (ARCH) model
ARIMA, *see* Auto-regressive integrated moving average (ARIMA)
Arithmetic mean, 27, 958
ARMA, *see* Auto-regressive moving average (ARMA)
ARRSES, *see* Adaptive-response-rate single exponential smoothing (ARRSES)
Artificial intelligence (AI), 71, 509, 527, 533, 848, 867
Artificial neural network (ANN), 10, 11, 383, 537, 540, 570–576, 579, 607, 867
ASSESSOR model, 626
Asset management, 659, 660, 685, 849
Asset management company (AMC), 660, 685, 771–774, 782–784, 787, 788
Asset price estimation, 12, 663–672
Assortment optimization, 603, 613, 614
Assortment planning, 600, 654, 825
Atlas of health and climate, 769
Autocorrelation coefficient, 389
Auto-covariance function (ACF), 389
Auto-encoders, 540, 575
Automated essay grading, 584

© Springer Nature Switzerland AG 2019

B. Pochiraju, S. Seshadri (eds.), *Essentials of Business Analytics*, International Series in Operations Research & Management Science 264,
https://doi.org/10.1007/978-3-319-68837-4

Automatic Region Server failover, 80
Auto-regressive conditional heteroscedasticity (ARCH) model, 396, 397, 667
Auto-regressive integrated moving average (ARIMA), 8, 389, 395–399, 409, 410, 414, 417, 605
Auto-regressive models, 389, 625, 667
Auto-regressive moving average (ARMA), 395–397, 667, 700–702
Axioms of probability, 947, 948
Axis Bank, 744

B
Babylon, 767
Back propagation, 573, 592
Backward feature selection, 519
Bagging, 10, 546, 547, 607
Bag-of-words, 6, 286, 297, 532, 563
Bank for International Settlements, 686
Banner advertisement, 13, 723
Basel Committee on Banking Supervision (BCBS), 672
Baseline hazard model, 625
Basis, 24, 34, 104, 218, 291, 292, 295, 354, 359, 360, 402, 404, 421, 440, 510, 534, 540, 625, 644, 650, 662, 676, 690, 714, 716, 721, 739, 751, 752, 757, 758, 767, 806, 818, 829, 830, 834, 835, 879, 915, 945
Bass model, 626
Batch-wise gradient descent, 573, 583
Bayesian classifier, 10, 488, 493, 529–549
Bayesian estimation, 669
Bayesian information criterion (BIC), 221, 222, 264, 265, 268, 397
Bayes' theorem, 947–948
BCBS, *see* Basel Committee on Banking Supervision (BCBS)
BeautifulSoup, 944
Bernard Madoff, 684
Bernoulli distribution, 491, 952, 954
Best subset regression, 180, 220, 222–223
Best-worst conjoint, 643
Bfonics, 617
Bias-variance-trade off, 520
BIC, *see* Bayesian information criterion (BIC)
Big data, 1, 3, 32, 42, 71–108, 289, 382, 602, 663, 741, 794, 863
Big data techniques, 766–770, 794, 877
Binary hierarchical classifiers, 547–549
Bivariate analysis, 958, 962–963
Bivariate distributions, 945
Black and Scholes, 12
Black-Box, 403
Blended recommendations, 564–565
Blending problem, 337
Blippar, 617
Bloomberg, 685, 687
Bluetooth beacons, 11, 604
Blue Yonder, 618
Boosting, 10, 546–547
Bootstrap, 401, 523, 607, 634
Bootstrap aggregation, 546, 607
Bounding, 366, 367
Box plots, 185, 186, 194, 196, 197, 202, 238, 239, 960–962
Box–Cox transformation, 180, 232–233
Branch-and-bound method, 7, 365–367
Branching, 366–367, 369
Brand specific conjoint, 643
Brickstream, 617
Brownian motion, 660

C
Cancer, 570, 766
Capital asset pricing model (CAPM), 12, 661, 662
Cartesian coordinate system, 344–345
Cassandra, 80
Categorical features, 526, 527
Categorical response variable, 248
Category management, 825
Causation, 963
Celect, 618
Censoring, 8, 439–444, 457, 602
Center for Disease Control (CDC), 284
CentOS Linux, 87
Central limit theorem, 4, 137, 140–141, 148, 316, 957
2^{nd} Central moment, 958
Certainty assumption, 343
CF, *see* Collaborative filtering (CF)
CGE, *see* Computable general equilibrium (CGE)
Chain ratio method, 626
Chi-Square (χ^2) distribution, 4, 149, 150, 174, 254, 255
Choice analytics, 625
Choice based conjoint, 643, 644
CISCO, 827
Citibank, 660
CIV, *see* Customer influence value (CIV)
Class label, 461, 510, 511, 514, 515, 524, 528, 535, 545, 571

Classification, 5, 101, 264, 265, 269, 273, 292, 293, 373–374, 402, 432, 461, 478, 479, 484, 511–513, 515, 517, 518, 520, 522–549, 554, 570, 571, 574, 576, 577, 579, 588, 859
  plot, 259–260, 265
  table, 259–260, 265
Classifier, 10, 510, 511, 522–523, 567, 571, 579, 592
Clickstream log, 73
Cloud computing, 103–108
Cloudera, 3, 86–90
Cluster analysis, 625
CLV, *see* Customer lifetime value (CLV)
CNNs, *see* Convolutional neural networks (CNNs)
Coefficient of determination, 191–192, 194
COGS, *see* Cost of goods sold (COGS)
Cold start problem, 557, 559, 562
Cole bombing suspects, 722
Cole Haan, 617
Collaborative filtering (CF), 9, 101, 555, 557, 559, 561, 564
Collinearity, 5, 180, 215–224, 228, 239, 241, 243
Combinations, 28, 60, 118, 167, 238, 295, 298, 325, 337, 355, 380, 388, 395, 397, 418, 424, 432, 437, 466, 477, 481, 488, 497, 514, 518, 526, 535, 536, 549, 554, 562, 565, 604, 629, 643, 645, 675, 735, 736, 739, 749, 750, 774, 807, 809, 810, 832–834, 878, 900, 927, 945, 953, 954
Commodity procurement, 14, 829
Common identifier, 31
Comparison of two populations, 161–162
Complement, 33, 234, 630, 720, 744, 761, 946
Complementary slackness conditions of optimality, 361
Complete enumeration, 23, 366
Completely randomized design (CRD), 13, 729–732, 735
Component capital, 679–682
Computable general equilibrium (CGE), 829
Computer simulation, 6, 307, 333
Concordant pairs, 5, 261–262
Conditional mean model, 642, 700–701
Conditional probabilities, 444, 532, 674, 947
Conditional variance, 396, 397, 430, 701
Conditional variance models, 701–702
Confidence intervals, 4, 137, 141–151, 161, 169, 170, 192–195, 203, 204, 233, 256, 271, 315, 316, 329, 388, 451, 634, 961
Conjoint analysis, 12, 624–626, 643–650, 655
Connected cows, 828
Constant return to scale (CRS), 637, 638
Consumer's ratings, 643
Continuous random variables, 948, 949, 951
Contour plots, 344, 348
Control variables, 340
Convex hull, 362, 364, 380
Convexity constraint, 638
Convolution layer (CONV), 10, 579–582
Convolutional neural networks (CNNs), 10, 540, 570, 574, 576–583, 588, 589, 591, 594
Co-occurrence graphs (COG), 289–290, 296, 298, 501
Cook's distance, 212, 213, 430
Copula functions, 12, 672, 673, 676
Correlation, 208–211, 219, 273, 300, 389, 504, 509, 514, 564, 579, 607, 661, 668, 672–677, 679, 683, 696, 715, 716, 739, 826, 962, 963
Cosine similarity, 476, 482, 499, 545, 555, 556, 562
Cost efficiency, 823, 824, 830, 831, 842, 844
Cost of goods sold (COGS), 290, 292, 298–301, 670, 826
Count data regression model, 421–437
Covariance, 187, 188, 212, 242, 492–494, 530, 532–534, 661, 690, 696, 962, 963
Covariates, 9, 183, 188, 427, 428, 441, 446–448, 450, 453, 455–457, 606, 625
Cox and Snell R^2, 257–258, 264, 268
Cox proportional hazard model, 440, 447–449, 456
CRAN repository, 801
CRD, *see* Completely randomized design (CRD)
CreditMetrics™, 670
Crew scheduling, 337
Cross domain recommendations, 565–566
Cross-sectional data, 24, 238, 383, 625, 667
Croston's approach (CR), 8, 400, 401
CRS, *see* Constant return to scale (CRS)
CRV, *see* Customer referral value (CRV)
Cumulative distribution function (CDF), 442, 949
Cumulative hazard, 442, 457
Curvilinear relationships, 12, 630–635
Custom/algorithmic attribution, 747–750
Customer analytics, 12, 75–76, 626, 650–654
Customer identity, 32, 33
Customer influence value (CIV), 12, 653–654
Customer lifetime value (CLV), 12, 440, 624, 650–652, 654
Customer referral value (CRV), 12, 626, 652–654

Customer retention, 600, 651
Customer satisfaction, 14, 623, 637, 823, 828, 830–832, 842–844, 847, 848, 855, 856, 858
Customer service, 284, 306, 308, 824, 825, 830, 844, 855
Cyclical, 388–390

D
Damped exponential smoothing, 394, 413
Data collection, 2, 11, 19–38, 101, 138, 213, 285, 384, 440, 600, 603–605, 607, 616, 617, 644–645, 654, 767, 772, 774, 776–778, 851, 856, 879
Data dimensions, 119, 120, 122, 134
Data driven marketing, 741
Data envelopment analysis (DEA), 12, 624, 625, 636–641, 655
Data mining, 75, 498, 855
Data relationships and graphs, 119
Data warehousing, 600
Database administrator (DBA), 48
Database management system (DBMS), 42–43
Data control language (DCL), 47, 48
Data definition language (DDL), 47–48
Dataframes, 101
Data manipulation language (DML), 47–48
DataNodes, 78, 81–84
DBA, *see* Database administrator (DBA)
DBMS, *see* Database management system (DBMS)
DBN, *see* Deep belief networks (DBN)
DCL, *see* Data control language (DCL)
DCO, *see* Dynamic creative optimization (DCO)
DDL, *see* Data definition language (DDL)
DEA, *see* Data envelopment analysis (DEA)
Decision boundary, 527, 533, 534, 540, 542, 547, 549, 567
Decision making under uncertainty, 7, 305, 317–327
Decision-making units (DMUs), 637, 638
Decision trees (DTs), 10, 11, 182, 273, 519, 524, 526, 527, 535, 545–547
Decision variables, 305, 317–319, 321, 324–327, 338–344, 356–358, 360–369, 375–378, 605, 608, 812, 840
Deep belief networks (DBN), 575
Deep content based recommendations, 562–563
Deep learning, 6, 10, 106, 374, 516, 534, 540, 569–594, 605, 607, 661
Deep neural networks (DNNs), 570, 607
Deep Q-network (DQN), 575
Degenerate, 353–355, 377
Degrees of freedom, 145–147, 149, 150, 160, 164–169, 176, 191, 193–195, 199, 200, 206, 211, 252, 255–257, 263, 264, 268, 270, 426, 427, 477, 733
Deletion diagnostics, 180, 206, 211–215
Delphi method, 626
Demand forecasting, 11, 14, 15, 400, 416, 607, 654, 704, 798, 811, 813, 814, 819, 824–826
Demand side platform (DSP), 726
Demand-side software (DSP), 725
Denoising auto-encoder (DA), 575
Dependent variable, 183, 257, 263, 266, 269, 272, 273, 576, 583, 625–627, 630, 633, 635, 737
Depth-first-search process, 368
Depth-threshold, 525
Derivative security, 660
Descriptive classifiers, 488, 522, 523, 534
Descriptive statistics, 179, 185–186, 769, 945, 952, 958
Designed experiment, 22
Deviance R^2, 25
DFFITS, 212
DHL, 602
Dickey–Fuller test, 305, 666, 693–695
Dictaphones, 768
Digital attribution, 654, 722, 744–750
Digital Diaries, 768
Dimensionality, 29, 122, 286, 344, 345, 467, 482, 531, 559, 572, 606, 616, 661, 663, 683, 685, 690, 691, 696
Dimensionality reduction, 101
Discordant pairs, 261, 262
Discovery, 524, 549, 550, 570
Discrete event simulation (DES), 13, 770, 772, 786
Discrete random variables, 139, 948, 951, 952
Discriminant analysis, 533, 625, 747
Discriminative classifiers, 522, 523, 534, 535
Disjoint, 946, 947
Dispersion, 8, 185, 238, 428, 707, 708, 950, 951, 958, 962
Distributed computing, 77, 78
Distribution centers (DCs), 827, 828, 830, 831, 833, 834, 839, 840, 843
Divide and conquer, 365, 390, 545
Divisibility assumption, 343
DML, *see* Data manipulation language (DML)
DMUs, *see* Decision-making units (DMUs)
DNA, 768
DNNs, *see* Deep neural networks (DNNs)

Document, 6, 283, 286–292, 294–301, 476, 482, 484, 510, 511, 517–519, 532, 545, 670, 852
Document classifier, 510
Document-term matrix (DTM), 288, 292, 296, 300
Dor, 318, 617
DQN, *see* Deep Q-network (DQN)
Dropbox referral program, 652
Dual values, 355, 358
Dummy variables, 5, 180, 224–233, 239, 247, 248, 263, 437, 647, 730, 731, 734
Dynamic creative optimization (DCO), 728

E
EBay, 728, 804
E-commerce, 73, 331, 507, 549, 553, 554, 563, 599, 601, 616, 743, 803, 804
Economies of scale, 343, 757, 833, 834
EDA, *see* Exploratory data analysis (EDA)
Efficiency model, 830
Efficient market hypothesis (EMH), 662, 665
EGARCH, 685, 701–703, 705, 707–709, 713, 716
Eigenvalues, 470, 471, 590, 688–690
Eigenvectors, 470, 688–690
Elliptical distribution, 662
EMH, *see* Efficient market hypothesis (EMH)
Engagement ads, 724
Engagement bias, 564
Engagement matrix, 551, 556–561
Ensemble, 10, 527, 545–547, 549, 607, 886
Enthought Canopy, 919, 920, 923
Entity integrity, 43
Entropy, 525, 587, 589
Epidemiology, 766
ERP systems, 825, 858
Error back propagation, 539
Estimates, 5, 72, 74, 137–151, 163, 164, 167, 169, 170, 179, 180, 183, 190–195, 199–208, 211–215, 218–221, 223, 225, 228, 229, 232, 233, 239, 241, 242, 247, 249–256, 258, 261–268, 273–276, 278, 284, 292, 296, 309, 310, 313, 315–320, 328, 331, 332, 369–371, 373, 381, 389, 390, 394, 396–398, 408, 409, 425, 427, 428, 431, 432, 434, 436, 440, 442–445, 448, 449, 451, 453, 455, 488, 489, 524, 529, 531, 532, 556, 571, 584, 599, 601, 605, 606, 608, 613, 614, 616, 628, 629, 633, 634, 638, 640–642, 645, 647–649, 652, 662, 663, 665, 668–670, 683–686, 690, 692, 702, 708, 709, 715, 728, 732, 735, 736, 739, 747, 748, 776, 797, 801, 805, 806, 811, 816, 817, 819, 826, 829, 833, 834, 840, 841, 850, 854, 857, 876–878, 886
Estimation, 7, 9, 12, 33, 139, 180, 186, 189, 193, 195, 199, 211, 223, 233, 250–251, 263, 264, 275, 276, 278, 279, 328, 338, 369, 371, 383, 388, 397, 398, 441, 443, 449, 459, 462, 488–491, 494, 504, 508, 512, 577, 602, 603, 610, 611, 616, 617, 628, 641, 642, 645, 652, 663–665, 667–669, 683, 685, 696, 702, 805, 806, 829, 879, 886, 944
Estimation of parameters, 189, 199, 223, 250–251, 275
Estimators, 139–141, 148, 149, 190, 191, 199, 223, 396, 444, 490, 606, 616
Euclid analytics, 617
Euclidian distances, 476, 528
Eugene Fama, 662
Excessive zeros, 430
Exogeneity, 188
Expectation-maximization (EM) algorithm, 480, 613
Expected value, 140, 147, 148, 151, 186, 307, 328, 424, 425, 429, 660, 950, 958
Experiment, 13, 22, 87, 422, 434, 608, 729–733, 736, 737, 739, 752–754, 773, 775, 778–781, 783, 787–789, 877, 946–948, 950–953
Experimental design, 644, 647, 649, 723, 728, 729, 735, 739, 750, 785
Explanatory variables, 182, 247–249, 251–253, 255, 257, 262–266, 268–270, 421, 422, 427, 435, 447–449, 451, 642, 886
Exploratory data analysis (EDA), 427
Exponential smoothing, 391–395, 401, 410, 413, 605
Exponomial choice (EC) model, 610
Exposure variable, 430
ExtenSim version 9.0, 786
External data, 23, 908
Extreme point, 351, 352

F
Face detection, 570
Facebook, 13, 26, 71, 72, 75, 76, 284, 285, 478, 499, 512, 545, 550, 553, 565, 570, 601, 654, 719, 721, 724, 726, 728, 739, 748–750, 762
Factor analysis, 295, 625, 663, 690–691
Factorial designs, 644, 736, 739
Fama–French three-factor model, 667

Fashion retailer, 6, 7, 306, 309, 317, 318, 321, 323, 336, 826
F distribution, 4, 165, 167, 176, 195, 212, 335, 685, 698, 806, 959
Feature centric data scientists, 516
Feature distributions, 514
Feature engineering, 10, 516–519, 532, 537, 574, 577
Feature normalization, 517
Feature selection, 519, 606
Feature space, 486, 488, 522, 524–526, 531, 533, 537, 545, 562
Feature transformation, 517
Federal Reserve Bank (Fed), 679, 686
FedEx, 825, 843
Filter methods, 519
Financial instrument, 660, 661
First interaction/first click attribution model, 746
First normal form (1NF), 44
Fisher discriminant analysis, 468, 515
FitBit, 766
Fit-for-purpose visualization, 124
Flash, 723, 826
Flipkart, 556, 744
Flume, 101
Footmarks, 617
Ford, 132, 828
Forecast/prediction intervals, 7, 193, 195, 204, 385, 406, 408
Forecasting, 7, 8, 11, 14, 15, 75, 180, 338, 381–418, 512, 518, 521, 601, 605–607, 625, 626, 647, 654, 796, 805–806, 814, 824–826
Forecasting intermittent demand, 399–400
Foreign key, 43, 64, 66–67
Forward feature selection, 519
Forward method, 263
Fourier transformations, 517
Fraudulent claims detection, 76
Full profile conjoint, 643
Fully connected layers (FC), 579, 582, 592
Function, 5, 7–9, 61, 102, 149, 185, 191, 210, 234, 249–263, 266, 269, 270, 275, 276, 279, 300, 308–310, 312, 317, 319, 320, 323, 325, 336, 337, 339, 341, 342, 344, 346–359, 370–378, 382, 389, 390, 396, 397, 408, 411–414, 426, 431, 434, 440, 442, 443, 446–448, 451, 453, 455, 456, 463, 464, 470, 472, 476, 480, 482, 485–490, 492–494, 524, 528–537, 540–547, 552, 563, 571, 574, 576, 581, 582, 586, 587, 591, 592, 601, 607, 609, 610, 638, 640–642, 662, 667, 669, 672–724, 761, 767, 771, 796, 797, 805, 829, 833, 834, 839, 875–877, 880, 883–887, 896, 902, 903, 905, 907–909, 913–915, 925, 936–942, 944, 948–951, 961
Functional magnetic resonance imaging, 34
Fundamental theorem of asset pricing, 660
Fundamental theorem of linear programming, 352

G

Gaana, 550
Gasoline consumption, 181–182, 201, 206–208, 210
Gaussian copulas, 672–676, 715
GE, 307
Gender discrimination, 182, 224, 229
Generalized autoregressive conditional heteroscedasticity (GARCH) model, 397, 667, 668, 684, 685, 696–699, 701–706, 708–713, 715
Generalized linear models (GLM), 425, 536, 544
Generating random numbers, 333–334
Generative adversarial networks, 540
Generic conjoint, 643
Geo-fencing, 655, 722
Geometric Brownian motion (GBM), 660
Ginni index, 524
GJR, 685, 700–704, 706–709, 712, 716
GLM, *see* Generalized linear models (GLM)
Goldman Sachs, 660
Google, 6, 37, 38, 77, 105, 106, 284, 301, 383, 498, 499, 509, 520, 570, 601, 720, 722–725, 728, 741, 742, 744, 750, 755, 760, 762, 769, 877
Google Adwords, 720, 741, 762
Google Analytics, 720
Google Big Query, 106
Google Brain, 769
Google Cloud Platform (GCP), 106
Google Compute Engine, 106
Google Display Networks, 724
Google DoubleClick, 725
Google Maps, 37, 38, 520, 723
Google Prediction API, 106
Google Traffic, 722
Google Trends, 769
Gradient descent, 573, 581, 583, 594, 607
Graph processing, 98, 101
Graphics processing units (GPUs), 482, 516, 575

GraphX, 101
Gumbel distribution, 610

H
H1N1, 6, 284
Hadoop, 3, 72, 77–85, 87, 89, 90, 92–99, 101, 105–108
Hadoop architecture, 80–82
Hadoop distributed file system (HDFS), 3, 78, 80–85, 94, 101
Hadoop ecosystem, 78–s81
Hadoop streaming, 89, 90, 92, 93
Harry Markowitz, 661
Harvard, 684, 685
HARVEY, 307
Hat matrix, 205
Hazard ratio, 447, 449, 451
HBase, 80, 94
HBO, 724
HDFS, *see* Hadoop distributed file system (HDFS)
Healthcare operations, 766, 768
Heat maps, 125, 126, 617
Hermitian matrix, 677
Heterogeneous, 292, 483, 495, 498
Heteroscedasticity, 5, 180, 215, 233–237, 239, 249, 667–669
Hive, 80, 94, 101
Holdout data, 387
Holt exponential smoothing (HES), 392–393, 410
Holt's method, 394, 413
Holt–Winters' trend and seasonality method, 393–394
Homogenous, 291, 666, 817
Hortonworks, 87
Hosmer and Lemeshow test, 252, 257, 263, 264, 271
Human visual perception, 115
Hurdle models, 422, 434–435
Hypothesis testing, 4, 137, 151–156, 169–171

I
IAAS, *see* Infrastructure as a service (IAAS)
IBM, 43, 47, 78, 87, 570
IBM Watson, 570
ICICI bank, 636, 637
IDC Manufacturing Insights, 827
IDE, *see* Integrated development environment (IDE)
Ideal sample size, 144, 148, 149
Image classification, 570, 577, 579, 588
ImageNet challenge, 570, 578
Imputation techniques, 519
Incremental capital, 679–681
Indentation, 897, 923–925
Independent events, 946
Independent samples, 163–166, 169, 170
Independent variables, 6, 183, 251, 252, 254, 255, 266, 272, 275, 347, 373, 424, 434, 625–628, 630, 655, 962
Infeasible region implies, 346
Inferential statistics, 958
Influential observation, 211–213, 232, 239
Information extraction, 510
Infrastructure as a service (IAAS), 104, 105
In-memory computing, 96, 97
Innovation analytics, 626
Input features, 467, 511, 512, 517, 518, 536, 537, 547, 569
Instagram, 654
Instantaneous hazard, 442
Insurance, 8, 15, 71, 76, 213, 284, 338, 421–423, 430–434, 437, 440, 441, 659, 766, 769, 770, 843, 847–861
Integer optimization, 614
Integer programming, 7, 343, 362, 365, 374
Integrated development environment (IDE), 891, 892, 919
Integrated distribution, 824, 831
Integrates R with Python, 944
Interaction effect, 11, 227, 626–632, 655, 736–739, 755, 759
Interactive mode, 897, 923
Interactive Python (IPython), 944
Internet of things (IoT), 73, 105, 460, 507, 604
Interpolation, 240, 273, 557, 584, 688
Interpreted language, 890, 919
Intersection, 32, 125, 345, 346, 348, 350, 353, 503, 504, 946
Interval, 2, 4, 7, 8, 24, 27, 83, 98, 131, 137, 139, 141–151, 157, 161, 169, 170, 192, 193, 195, 203, 204, 233, 256, 271, 313, 315, 316, 329, 373, 385, 388, 399, 400, 406, 408, 414, 429, 441, 442, 444, 451, 574, 619, 634, 661, 663, 674, 961
Invariant features, 517
Inventory management, 824, 826, 827
Inventory optimization, 14, 337, 609, 826–827, 829
Inventory turnover, 826, 830
IoT, *see* Internet of things (IOT)
IPad, 776
IRIS dataset, 467, 514
Irregular/error components, 388
Item-item collaborative filtering, 555, 556

J
JavaScript, 723
Java virtual machine (JVM), 99
JioCinema, 556
Joint probability density function, 371, 531
JPMorgan, 660, 681, 682

K
Kafka, 101
Kaplan–Meier estimates, 444–445
Kate Spade, 617
Kernel functions, 489, 490, 528, 544, 545
Key performance indicators (KPIs), 600
Klarna, 601
Kline, 605
K-nearest neighbor classifier (k-NN), 10, 274, 527–529, 532, 535, 540, 545
Kohl's, 603
Kurtosis, 698–700, 714, 959, 960

L
Lagrange multiplier, 541–543
Language processing, 6, 283, 286, 294, 297–299, 569, 570
Language translation, 570
Laplacian smoothing, 524, 532
Large-sample confidence intervals, 147–148
Last AdWords click attribution model, 746
Last interaction/last click attribution model, 13, 745, 746
Last non-direct Click attribution model, 745
Least absolute shrinkage and selection operator (LASSO) regression, 11, 180, 220, 223, 224, 338, 606
Leverage values, 205–206
LexisNexis, 769, 854
Likelihood ratio test, 5, 252, 255, 448
Linear attribution model, 13, 746
Linear discriminant analysis (LDA), 10, 533–535
Linear inequalities, 345
Linear optimization, 7, 338, 339, 614, 828
Linear programming (LP), 7, 339–347, 349, 351–369, 374, 375, 377, 378, 380, 610, 637, 638, 809, 812–814, 817
  formulation, 340–342, 353
Linear regression, 4, 5, 7, 114, 179, 180, 184–186, 189, 193, 198, 201, 204, 205, 208, 211, 215, 230, 231, 234–244, 247, 248, 250, 258, 264, 273, 275, 369, 371, 373, 421–424, 429, 437, 535, 606, 626, 645
Lines of business (LoB), 681, 716
Linkedln, 652
Linux, 87, 892
List-wise deletion, 272
Location-based data capture, 32, 34, 603
Location bias, 565
Log odds, 251, 535
Logistic regression, 7, 249–252, 254, 256–258, 260–264, 266, 269–271, 274, 441, 517, 518, 535–537, 540, 544, 545, 547, 594, 607, 747
Logistics planning, 831, 844
Log-logistic distribution, 454
Long short-term memory (LSTM), 574, 590
Lord & Taylor, 617
Loyalty cards, 600
LP formulation, *see* Linear programming (LP)
LSTM, *see* Long short-term memory (LSTM)

M
Machine learning (ML), 9–11, 15, 71, 95, 98, 101, 106, 374, 459–505, 507–570, 600, 601, 605–607, 618, 769, 848, 944, 947
Machine translation, 10, 583, 584, 588–590
Macy's, 602, 605
MANOVA, 625
Map Advertisements, 723
MapR, 87
MapReduce, 3, 77, 78, 80, 81, 84–98, 108, 501–504
Market share forecasting, 626, 647–648
Markov chain, 613
Master–Slave architecture, 78, 81
Matplotlib, 944
Matrix factorization approaches, 557–559, 561
Mattel, 297
Maximum likelihood estimate (MLE), 5, 12, 247, 250, 266, 275–279, 371–372, 397, 398, 425, 427, 432, 445, 448, 449, 616, 634, 668, 669, 685
MBC, *see* Multimodal Bayesian classifier (MBC)
McFadden R2, 257–258, 268
Measure of central tendency, 185, 238, 950, 958
Measure of dispersion, 185, 238, 708, 950, 951, 958, 962
Measures of purity, 524, 525
Measure the efficiency, 636, 637, 639
Median, 27, 32, 108, 185, 193, 194, 201, 208, 209, 214–216, 223, 225, 227, 229, 230, 235, 237, 273, 405, 425, 427, 433, 435, 454, 456, 488, 808, 952, 955, 958, 961

Mediation analysis, 12, 625, 633, 635, 656
Medicinal value in a leaf, 182–183
Memory-based recommendation engine, 552, 555, 557
Mesos, 99
Meta-rules of data visualization, 4, 116–133
Method of least squares, 4, 180, 189, 191, 199, 223, 232, 233, 236, 275
Microsoft, 43, 46, 47, 106, 297, 344, 375, 405, 570, 589
Microsoft Azure, 106
Microsoft Cognitive Toolkit, 106
Millennials demographic, 602
MIN price, 62
Missing data, 5, 31, 247, 272–274, 372, 384, 526, 687
Missing features, 519, 526
Mitigation strategy, 828
Mixed integer optimization, 614
Mixture-of-Gaussians (MoG), 483, 489, 492–494, 530, 531, 535, 545
ML, *see* Machine learning (ML)
MLE, *see* Maximum likelihood estimate (MLE)
MLPs, *see* Multi-layered perceptrons (MLPs)
MNIST data, 10, 471, 515, 575–576, 582–583
MNL, *see* Multinomial logit (MNL)
Mobile advertising, 725
Mobile auto-correct, 584
Mode, 27, 30, 99, 219, 273, 549, 550, 611, 719, 723, 858, 897, 923, 937, 938, 955, 958
Model centric data scientists, 516–517
Model validation, 237–239
Moderation, 625, 628, 635
MoG, *see* Mixture-of-Gaussians (MoG)
Monte Carlo simulation, 6, 12, 306, 669–672
Mordor Intelligence, 766
Morgan Stanley, 660
Movie recommendation, 563
Moving average (MA) methods, 8, 388, 389, 391, 396, 397
MR, *see* Multiple regression (MR)
Multi-layered perceptrons (MLPs), 10, 570–576, 579, 581–585, 592–594
Multimodal Bayesian classifier (MBC), 530–531
Multinomial logistic regression (MNL), 5, 15, 266–269
Multinomial logit (MNL), 610–613, 806, 817, 867, 870
Multiple regression (MR), 225, 252, 396, 626, 627, 634
Multiple variance ratio test, 666
Multivariate analysis, 625, 958
Multivariate Cauchy distribution, 662
Multivariate exponential, 662
Multivariate normal, 662
Multivariate statistical analysis, 625
Multivariate student t-distribution, 662
MySQL, 1, 3, 42, 43, 47–51, 53, 54, 68, 80, 102, 106

N

Nagelkerke R2, 257–258, 268
Naïve Bayes classifier, 10, 531–532
Naïve method (NF), 8, 390–391
Named entity recognition (NER), 286, 298, 300
NameNode, 81–84
Namespace, 83
Natural language processing (NLP), 6, 283, 286, 287, 294, 297–300, 569
Natural language toolkit (NLTK), 298
Near field communication (NFC), 604, 751
Negative binomial distribution, 8, 431
Negative binomial regression model, 431
Neocognitron, 579
NER, *see* Named entity recognition (NER)
Nested logit model, 11, 613
Netflix, 19–20, 509, 512, 550, 556, 826
Net present value (NPV), 12, 19, 661
Network
  analytics, 722
  and influence diagrams, 722
  planning, 337
Neural networks (NN), 10, 11, 338, 374, 383, 537–540, 545, 582, 583, 586, 591, 607, 855
New product
  design, 626
  development, 11, 75, 402, 624, 643
Newspaper Problem, 181
News recommendation, 563
Nextel, 830
Next word prediction, 10, 584, 585, 587–588, 594
NFC, *see* Near field communication (NFC)
NLP, *see* Natural language processing (NLP)
9/11, 722
Nobel prize, 660, 662
Nominal, 2, 24, 25, 27, 132, 247, 266, 670, 730
Non-linear analysis, 369–374
Non-linear optimization, 369–374
Non-negative matrix factorization, 558
Nonparametric classifier, 527
Nonparametric envelopment frontier, 637
Nonparametric resampling procedures, 634

Nonparametric tools, 11, 625
Non-random walk hypothesis, 666
Non-sampling errors, 23
Normal distribution, 141, 149, 156, 157, 185, 187, 193, 231, 249, 275, 276, 278, 310, 311, 313, 319, 371, 430, 466, 491, 529, 669, 670, 673, 685, 696–698, 702, 708, 816, 839, 955–957, 960
Normality, 194, 196, 202, 208, 215, 231–233, 238, 239, 466, 634, 685, 698, 699
Normalization, 43–46, 517, 528
NoSql database, 80
NPV, *see* Net present value (NPV)
Null deviance, 425, 427, 428, 433, 435
Null hypothesis, 151–154, 156–161, 163–167, 170, 171, 193, 195, 206, 253, 255–257, 426, 434, 694
Numerical Python (NumPy), 944

O
Oak Labs, 617
Objective function, 7, 9, 337, 339, 341, 342, 344, 346–352, 354, 355, 357, 375, 376, 378, 463–465, 470, 472, 480, 486, 487, 490, 493, 535, 536, 540–543, 640, 887
Objective function coefficients, 341, 357, 378
Object-oriented programming, 918
Object recognition, 510
Observational data, 636
Observational equations, 187
Occam's razor, 520, 574
Odds ratio, 249, 251, 535
Office365, 105
Offset regressor, 430
Offset variable, 422, 430, 432
OLS, *see* Ordinary least square (OLS)
Omni-channel retail, 11, 616
One-tailed hypothesis test, 159–161
Oozie, 80
Open source, 1, 3, 77, 78, 80, 286, 298, 606, 890, 892, 915, 918
  language, 890, 918
Operational efficiency, 640, 830
Operations research (OR), 339, 773, 774, 778, 785
Optical illusion, 126
Optimality, 7, 338, 344, 356, 359–362, 364, 365, 527
Optimization, 7, 9, 14, 21, 34, 49, 97, 325–327, 337–380, 386, 391, 393, 397, 398, 414, 463–465, 470–475, 479, 480, 489, 490, 493, 497, 521, 535, 541, 542, 549–551, 568, 587, 603, 608–610, 613, 614, 672, 747, 748, 750, 793–798, 806, 808–811, 815, 817, 824, 826–829, 887, 944
Ordinal, 2, 24, 25, 27, 180, 230, 266, 269, 512
Ordinal logistic regression models, 266
Ordinary least square (OLS), 4, 8, 236, 249, 369–371, 373, 634
ORION, 307
Orthogonal GARCH, 683, 684, 696
Orthogonality, 468, 469, 739–740
Outliers, 5, 8, 205–206, 213, 243, 384, 388, 403, 404, 462, 488, 490, 493, 517, 545, 963
Output feature ratios, 518
Over-defined, 354
Overdispersion, 8, 430–433, 436
Ozon.ru, 602

P
PACF, *see* Partial autocorrelation function (PACF)
Paired-observation comparisons, 162–163
Pairwise classifier, 547–549
Pair-wise Deletion, 272
Palisade DecisionTools Suite, 310, 311
Pandas, 944
Parallel computing, 3, 77, 81, 84
Parameters, 4, 5, 30, 33, 138–139, 141, 142, 145, 151, 156, 164, 179, 183, 184, 187–191, 199–201, 204, 212, 220, 223, 232, 241, 247, 249–251, 264, 266, 275, 276, 278, 369–371, 374, 377, 392, 397, 398, 400, 410–414, 422, 423, 425–428, 430, 431, 433–435, 442, 443, 447–449, 453, 455, 463–465, 473, 476, 480–482, 486, 489, 490, 492, 493, 517, 524, 530, 531, 535, 536, 542, 544, 545, 558, 572, 574–576, 579–583, 585–587, 592, 606, 607, 610, 613, 615, 629, 642, 650, 667–669, 685, 686, 701, 702, 708, 715, 726, 728, 742, 773, 778, 779, 787, 797, 801, 805, 806, 833–835, 858, 877, 896, 915, 940, 941, 955
Parametric density function, 489, 490
Parametric methods, 456, 634
Parametric model(ing), 443, 447, 449, 453–455, 457, 535
Partial autocorrelation function (PACF), 8, 389, 395–397
Partial profile conjoint, 644
Partial regression coefficients, 202, 203, 241
Parzen Window classifier (PWC), 10, 492, 528–529, 531, 540
PasS, *see* Platform as a Service (PasS)

PayPal, 652
PayTM, 565
PCA, *see* Principal components analysis (PCA)
PDF, *see* Probability density function (PDF)
Pearson's correlation, 555–557
Percentiles, 193, 212, 517, 958
Perceptron, 10, 534–536, 540, 570–572, 575, 591, 592
  algorithm, 534–536, 570, 571, 592
  based classifiers, 535
Perceptual Maps, 625
PERCH, 617
Performance analysis, 659
Performance index (PI), 828
Permutations, 749, 750, 895, 945, 953
Personalization, 33, 549, 550, 566, 600, 618, 654, 741
Personalized education, 549, 566
Pfizer, 768
Physical simulation, 6, 307
Pig, 80
Pig Latin, 80
PIMCO, 660
Platform as a Service (PasS), 105, 106
Platform independent, 890, 919, 943
PMF, *see* Probability mass function (PMF)
Point-forecast, 7, 384
Point-of-Sale (POS), 11, 30, 75, 138, 476, 477, 495, 511, 603, 617, 744, 824, 827
Poisson distribution, 8, 423, 424, 430, 431, 491, 611, 614, 954–955
Poisson histogram, 423
Poisson regression model, 8, 421–425, 431, 434, 435, 437
Poisson variance, 423
Polytomous models, 266
Ponzi, 684
Pooling layers (POOL), 10, 579–583, 594
Popularity bias, 565
Population, 4, 13, 23, 31, 132, 137–151, 156, 160–167, 169, 183, 195, 203, 213, 237, 240, 276, 278, 316, 320, 423, 466, 730, 769–771, 805, 829, 830, 850, 876, 878, 886, 948, 957, 958
Population mean, 138–144, 146, 147, 149, 151, 156, 163, 164, 166, 167, 276, 316
Population proportion, 138, 147, 148, 161
Portfolio management, 659, 663, 664, 683
Portfolio selection, 337
POS, *see* Point-of-Sale (POS)
Position based attribution model, 746–747
Posterior probability, 492, 532, 549
Power Curve, 155, 156
P-quants, 12, 659–663, 685, 709, 714
Predictive analytics, 75, 338, 381, 382, 407, 601, 824, 855, 859
Preference bias, 564
Price and revenue optimization, 337, 796
Primal perceptive processes, 113
Primary data, 21, 31, 284, 683
Primary key, 33, 43, 53, 64–65
Prime numbers, 946
Principal components analysis (PCA), 9, 462, 466, 468–472, 475, 480, 515, 516, 535, 663, 683, 688, 690–692, 696, 709
Probability, 2, 5, 138–140, 142, 143, 147, 152–157, 159, 161, 163, 185, 186, 203, 242, 248–252, 254, 255, 257–262, 265, 269, 272, 275–277, 295, 296, 299, 309, 310, 313, 331, 334, 336, 371–373, 399, 431, 442, 444, 445, 448, 449, 462, 464, 482, 488, 490, 492, 498–500, 512, 524, 525, 529, 531–533, 549, 552, 583–585, 610, 612, 619, 652, 662, 665, 667–669, 671–676, 694, 697, 701, 709, 715, 736, 801, 802, 808, 809, 816, 817, 819, 828, 839, 848, 945–963
Probability density function (PDF), 276, 371, 372, 431, 442, 531–533, 839, 949–951
Probability distributions, 138, 140, 242, 299, 334, 336, 373, 482, 585, 662, 665, 672, 676, 701, 945, 954, 958
Probability mass function (PMF), 948–950, 954
Probit, 625
Process map of clinic visit, 13
Production function, 640, 641
Production planning, 337, 379
Product life cycle analysis, 626, 804
Profile based recommendation engine, 559–567
Programmatic advertising, 13, 725–728
Programmatic display, 725–728
Proportional hazard model, 9, 440, 447–449, 456, 625
Proportionality assumption, 343
Pruning, 366, 367, 496, 497
Pure regions, 522, 524, 533, 540
Purity-threshold, 525
p-value, 157–159, 161, 163, 165, 166, 168, 193–195, 203, 206–209, 214–216, 221, 223, 225, 227, 229, 230, 235, 237, 255–257, 263, 268, 270, 426, 431–434, 451, 733, 738
PWC, *see* Parzen Window classifier (PWC)
P-world, 659–663, 665, 714

Python, 1–3, 47, 49, 68, 84, 87–90, 92, 93, 96, 99, 101, 108, 298, 344, 889, 891, 897, 899, 904, 917–944

Q
QDA, *see* Quadratic discriminant analysis (QDA)
Q-quants, 12, 659–661, 663
QR codes, *see* Quick response (QR) codes
Quadratic discriminant analysis (QDA), 10, 533–535
Quantitative finance, 12, 659, 663
Quantitative supply chain analysis, 844
Quartiles, 185, 196, 231, 958, 960
Query length bias, 518
Quick response (QR) codes, 604, 750
Q-world, 659–661, 663, 669, 672

R
R^2, 191, 192, 194, 200–204, 212, 215, 220–222, 237, 238, 240, 257–258, 263, 268
Radio frequency identification (RFID), 11, 31, 34, 604
Rakuten, 825
Random forest, 10, 527, 546, 607, 867
Random sample, 137, 138, 181, 308–310, 312, 315, 317, 319, 327, 336, 401, 443, 886, 957
Random variables, 140–142, 147, 149, 162, 308, 309, 312, 313, 315, 317, 318, 321, 323, 324, 327, 334, 343, 370, 372, 672–674, 772, 787, 801, 948–952
Random walk hypothesis, 665
Rank based conjoint, 643
Rating based conjoint, 643
Ratio, 2, 24, 27, 122, 131, 164, 191, 194, 212, 249, 251–253, 255, 267, 330, 426, 433, 440, 448, 449, 451, 518, 520, 535, 536, 573, 626, 637, 638, 640, 666, 682, 750, 802, 848, 850, 856, 960
Ratio features, 518
RBI, *see* Reserve Bank of India (RBI)
RDBMS, *see* Relational data base management systems (RDBMS)
RDD, *see* Resilient Data Distribution (RDD)
Read–Eval–Print Loop (REPL), 97
Real time bidding (RTB), 13, 725–727, 874
Real-time decision making, 827
Real-time location systems (RTLS), 777
Real time translation, 570
Receiver operating characteristics (ROC) curve, 5, 261
Recency, frequency, monetary value (RFM) analysis, 626, 650
Recommendation for Cross-Sell, 564
Recommendation for lifetime value, 564
Recommendation for loyalty, 564
Recommendation for Preventing Churn, 564
Recommendation for Upsell, 564
Recommendation paradigm, 461, 512–513
Recommendation score, 521, 552, 556, 557, 562–565, 567
Recommendation systems, 33, 550, 553, 555, 570, 859
Rectified linear units layer (RELU), 10, 579–581
Recurrent neural networks (RNNs), 10, 540, 570, 571, 574, 583–590, 593, 594
RedHat Linux, 87
Redshift, 105
Referential integrity, 43, 64, 66
Regressand, 183
Regression, 5, 183–184, 222
  analysis, 114, 179–245, 247–281, 381, 384, 441, 606, 730, 731, 734, 735, 738
  models, 5, 8, 11, 13, 179, 180, 184, 185, 187, 195, 252–257, 373, 423–425, 441, 446–447, 512, 518, 527, 625, 730, 734
  paradigm, 512, 513
Regressors, 4, 5, 179–180, 182–185, 187–206, 208, 210–212, 216–223, 228, 232–234, 238–241, 426, 430, 436, 449, 454, 456
Regularization, 558, 606, 610
Reinforcement learning, 10, 460, 508, 575
Relational data base management systems (RDBMS), 3, 41–68, 76
Reliability, 31, 81, 441, 830
RELU, *see* Rectified linear units layer (RELU)
Reserve Bank of India (RBI), 679, 686, 714
Residual(s), 4, 179, 190, 191, 193–194, 196, 199–202, 204–206, 209, 212–216, 220, 221, 223, 225, 227, 229–232, 234–239, 241, 243, 258, 370, 397, 425–427, 429, 430, 432, 433, 435, 449, 470, 558, 665, 667, 682
  deviance, 425–428, 432, 433, 435
  plots, 179, 197, 205–209, 230, 239, 241, 243
Resilient Data Distribution (RDD), 99–102
Response variable, 5, 183–185, 188, 189, 191–195, 197, 200, 202, 205, 208, 211, 219, 220, 231–233, 237–241, 244, 247–249, 251, 252, 369, 371–373, 424, 435, 439, 606, 739

Restricted Boltzmann machines (RBM), 575
Retail analytics, 11, 12, 337, 599–619, 654, 826
RetailNext, 617
Retrieval paradigm, 461, 513
RFID, *see* Radio frequency identification (RFID)
RFM analysis, *see* Recency, frequency, monetary value (RFM) analysis
Ridge regression, 180, 220, 223–224, 606
Right defaults, 519
@Risk, 310–313, 316, 319, 325, 334, 335
Risk aggregation, 672
Risk-averse, 12, 659, 661, 677
Risk management, 12, 659, 660, 663, 664, 672–683, 686
RiskMetrics™, 681
Risk-neutral world, 659, 661
Rite Aid, 605
RNNs, *see* Recurrent neural networks (RNNs)
ROC curve, *see* Receiver operating characteristics (ROC) curve
Ronald Fisher, 728, 736
RPy, 944
RStudio, 291, 292, 296, 891–893, 897, 899
RTB, *see* Real time bidding (RTB)
Rudyard Kipling, 660
Rue La La, 826
Rule-based classifiers, 523–524

S
Saavn, 550
Sales analytics, 626
Salesforce, 27, 28, 626
Samarthan, 826
Sample, 138–140
  distributions, 140–142, 147, 149, 156–158, 163
  error, 23, 309, 316
  estimates, 138, 170, 433
  mean, 4, 39, 138, 140–143, 145, 146, 148, 152, 156, 167, 310, 316
  proportion, 138, 140, 147, 148, 161
  size determination, 144, 148–149
  space, 946, 948
  statistic, 139, 140, 151, 156, 157, 160, 161
  survey, 23
Samsung, 744, 824
SAP HANA, 827
Satellite imagery, 570
Saturated model, 426, 432, 434, 437
SBA, *see* Syntetos and Boylan approximation (SBA)
Scale effects, 638
Scatter Plots, 4, 31, 188, 189, 197, 198, 205, 210, 211, 216–218, 466–467, 477, 514, 515, 952, 960, 962
Scikits, 944
SciPy, 922, 944
Scoring throughput, 520, 526
Scree plot, 292
Scripts mode, 897, 923
Search engine optimization (SEO), 721
Search methods, 338, 365, 393
Seasonal, 8, 37, 331, 388–391, 393, 395, 517
Seasonal and cyclical adjustment, 389
SEBI, *see* Securities and Exchange Board of India (SEBI)
Secondary data, 2, 21–23, 31
Second normal form (2NF), 44–45
Securities and Exchange Board of India (SEBI), 679, 686
Securities and Exchange Commission (SEC), 679, 686
Seed, 312, 314–316, 319, 322, 333, 336, 493, 566, 783
Segmentation, 11, 30, 283, 291–295, 577, 624, 625, 655, 793, 794
Selfie walls, 602
Semantic feature, 518
Semi-structured data, 73, 74
Sensitivity, 7, 260, 261, 355–360, 376, 377, 379, 480–481, 783, 787, 795, 813
Sensitivity analysis, 7, 355–360, 787, 813
Sentiment analysis, 6, 13, 283, 291–295, 298, 570, 601, 721
Sequential processing, 95, 98
Serendipity, 549, 550
Service level optimization, 337
SES, *see* Simple exponential smoothing (SES)
Set theory, 496, 905, 945–947
SFA, *see* Stochastic frontier analysis (SFA)
Shadow prices, 7, 355–361, 364, 375–378, 813, 814, 816
Shakespeare, 286
Shapiro Wilk test, 698
Shared-nothing architecture, 84
Shelfbucks, 617
Shiny, 290–293, 296, 299
Shipping cost rate, 833, 834
Signal-to-noise-ratio, 520
Simple exponential smoothing (SES), 391–392, 395, 399, 400, 409–411, 415
Simple linear regression model, 273, 369
Simplex algorithm, 339–340, 351, 354, 355, 361

Simulation, 6, 7, 12–15, 305–337, 614, 615, 619, 669–672, 678, 686, 687, 765–770, 772–774, 780–785, 876–878, 883–886, 944
Single regressor, 4, 188–197, 199, 200, 202–205, 210, 234
Singular value decomposition (SVD), 557
Size-threshold, 525
Skewness, 234, 238, 426–427, 958, 959
Sklar's theorem, 672
Sku IQ, 617
Slack variables, 353, 357, 542–544
Sliding motion, 349
Smoothing, 8, 390–395, 400, 401, 410, 413, 414, 524, 532, 557, 584, 605
Social bias, 565
Social graph, 32, 33, 565
Software as a Service (SaaS), 105
Solution space, 345
Song recommendation, 563
Spark, 3, 71, 72, 94–102, 105, 108
Spark ecosystem, 99–100
SPARK MLib, 101
SPARK SQL, 101, 102
SPARK Streaming, 101
Specificity, 260–261
Speech recognition, 517, 569, 570, 584
Speech tagger, 510
Spell checkers, 584
Spotify, 601
Sprint, 830
SQL, *see* Structured query language (SQL)
Sqoop, 80
Standalone capital, 680
Standard conjoint, 643
Standard error, 4, 114, 140, 141, 144, 146, 148, 152, 160, 161, 163, 166, 180, 191–194, 199, 201, 209, 211, 212, 214–217, 219, 223, 225, 227, 229, 230, 235, 237, 256, 268, 270, 426, 428, 429, 431, 445, 451, 634, 708, 782
Standard normal distribution, 143, 145–147, 163, 171, 231, 670, 674, 955, 957
Stationarity, 389, 395, 397, 692–695
Stationary, 389, 396, 397, 665
Statistical sampling, 309, 316, 327
Statsmodels, 944
Stem-word, 287
Stephen Ross, 662
Stepwise method, 221, 263
Stepwise regression, 180, 220, 222, 223
Stitch Fix, 618
Stochastic frontier analysis (SFA), 625, 626, 640–642
Stochastic gradient descent, 573
Stress test(ing), 159, 170, 676
Structural equation modeling (SEM), 634
Structured data, 73, 76
Structured query language (SQL), 43, 47–68, 80, 99, 101
Subjective-bias, 523, 553
Subset selection problem, 519
Supervised learning, 9–10, 374, 460–461, 464, 468, 478, 504–505, 507–513, 521, 527, 570, 571, 574, 575, 607, 608, 618
Supply chain disruption, 14, 827–828
Supply chain management, 14, 823, 824, 829
Supply side platform (SSP), 726
Support vector machines (SVM), 10, 518, 540–542, 544, 545, 747, 896
Survival analysis, 8, 9, 439–457, 625, 748
Survival curve, 439, 443–446, 449, 451, 454, 456, 457
Survival distribution, 439
SVD, *see* Singular value decomposition (SVD)
SVM, *see* Support vector machines (SVM)
Swachha Bharat (clean India) program, 826
Swasthya Slate, 767
Syntetos and Boylan approximation (SBA), 8, 400, 410, 414, 418
System thinking, 844

T

Tableau, 101
Target, 11–13, 20, 27, 33, 292, 316, 331, 375, 405, 512, 535, 557, 589, 604, 617, 624, 625, 636, 643, 655, 720–722, 724, 725, 727, 728, 741, 747, 757, 794, 818, 826, 839, 873–880
Tata Motors, 652
TCL, *see* Traction control language (TCL)
t distribution, 145–147, 160, 163, 173, 193, 206, 662, 702, 703, 707–709
Teaching content recommendations, 563
Tech bubble burst, 827
Telemedicine, 13, 766, 767
Tensor product representation (TPR), 574
Teradata, 87
Term-document matrix (TDM), 6, 286, 288, 291, 294, 296, 298, 299
Term Frequency—-Inverse Document Frequency (TFIDF), 288, 289, 300, 476, 482, 545
Testing of hypotheses, 192–193
Tests of significance, 200
Text analytics, 5–6, 283–301, 928
Text corpus, 287, 291, 295, 584, 595

Theta model, 395, 410, 413
Third industrial revolution, 71
Third normal form (3NF), 45–46
Thomson Reuters, 687
Threshold persistence (TP), 664, 676, 685–687, 703, 707, 709
Timberland, 617
Time Decay Attribution Model, 746–747
Time-series, 7, 8, 24, 132, 383–384, 386–388, 403, 404, 605, 625, 667, 692, 694, 867
Times Inc, 724
Time-to-event, 8, 439–441, 443
Time to recovery (TTR), 828
Tobin's separation theorem, 662
Tobit, 625
Tokenization, 286, 287, 298, 300
Topic mining, 6, 295–297, 299
Traction control language (TCL), 47, 48
Trade-off analysis, 626, 647–648
Trend, 13, 76, 113, 132, 208, 210, 220, 273, 388–395, 397, 403, 404, 408, 410, 412, 565, 601, 602, 604, 607, 630, 666, 667, 692, 694, 720, 721, 724, 741, 768, 794, 795, 815, 824, 826, 847
Trilateration, 722
Trimmed means, 388
TripAdvisor, 721
Twitter, 13, 36, 37, 71, 550, 553, 565, 601, 719, 721, 724, 739, 748, 756, 762
Two-fund theorem, 662
Two-level factorial designs, 736, 739
Type I error, 4, 152, 153, 156, 157, 159, 171
Type II error, 4, 153, 156

U
Uber, 20, 21, 35, 75, 479, 652, 794, 803
Unbalanced, 24
Unbounded feasible region, 345–347
Uncertainty modeling, 308
Unconstrained optimization, 338
Unidimensional objects, 906
Unimodal Bayesian classifier (UBC), 530
Union, 495, 946
Unit root, 389, 395, 694
Univariate analysis, 625, 958
Universal set, 946
University of California, Berkeley's AMP Lab, 95
Unpunctuality, 772, 780, 781, 787, 788
Unstructured data, 23, 31, 73, 74, 284, 604, 605, 607, 752, 891, 922
Unsupervised learning, 9, 10, 292, 295, 459–505, 508, 510, 530, 532, 575, 688
UPS, 307, 825
Urban Outfitters, 617
User–user collaborative filtering, 555

V
Validation, 5, 31–32, 180, 237–239, 244, 270, 521, 574, 867
Valuation, 12, 19–21, 440, 624, 659, 663, 796, 802, 804–805, 815
Value at Risk (VaR), 12, 664, 676–680, 685, 686, 696, 702, 703, 707–709, 714–716
VaR, *see* Value at Risk (VaR)
Variable return to scale (VRS), 637, 638
Variance, 4, 5, 12, 27, 32, 147–151, 161, 164–167, 170, 180, 186, 188, 189, 192, 193, 195, 198, 200, 201, 204, 205, 211–213, 218, 233, 239, 252, 273, 310, 337, 371, 389, 396, 397, 400, 423, 427, 428, 430, 431, 433, 434, 462, 468–471, 474, 475, 493, 515, 517, 518, 546, 606, 607, 611, 642, 653, 662, 665, 666, 668, 694, 697, 699–702, 709, 776, 777, 784, 799, 817, 867, 879, 950–952, 954, 955, 958, 959, 962
Variance decomposition proportions (VP), 218, 219, 241
Variance inflation factors(VIF), 218, 219, 239
Variation between samples, 168
Variation within samples, 168
Variety, 30, 31, 73–75, 117, 291, 330, 387, 459, 460, 465, 467, 477, 484, 494, 498, 504, 507, 508, 534, 536, 540, 545, 557, 567, 577, 582, 588, 600, 611, 660, 766, 769, 786, 809, 853, 890, 908, 918
VASTA, 14, 830–844
Vector auto regression (VAR) models, 625, 667
Velocity, 31, 72–74
Velocity feature, 518
Veracity, 73, 74
Verizon, 830
Video
  analytics, 617
  captioning, 570
  search, 570
  surveillance, 570
VIF, *see* Variance inflation factors(VIF)
Virtual machine (VM), 3, 86–90, 99
Virtual mirror, 306, 308
Virtual Physiological Human Institute, 768
Vistaprint, 601
Visual cortex system, 579
VM, *see* Virtual machine (VM)

Volume, 31, 73, 74, 76–79, 181, 202, 203, 307, 309, 318, 323, 384, 399, 507, 531, 580, 581, 593, 601, 663, 682, 720, 782, 784, 793, 824, 832–837, 839, 840, 849, 877
VRS, *see* Variable return to scale (VRS)

W
Wage balance problem, 182
Wald's test, 5, 252, 255–256, 270
Wallpaper advertisements, 724
WalMart, 300, 604, 794, 823, 843
WarbyParker.com, 616
Wavelet transformations, 517
Wearable technologies, 13, 766, 767
Web scraping, 35, 300, 890, 918
Website development, 890, 918
Weibull distribution, 455, 457
Whole numbers, 946
Wikipedia, 300, 383, 570
Winsorized means, 388
Wordcloud, 6, 288, 289, 292, 294, 296, 297, 299–301
WordNet, 576, 577
Word2vec, 584, 587
Workflow based recommendations, 566–567
Workforce optimization, 337
Work units, 772
World Health Organization, 769
World War II, 307
Wrapper methods, 519
Write-once Read-many (WORM), 81

X
XGBoost, 527, 547

Y
Yale, 684, 685
Yet Another Resource Negotiator (YARN), 3, 78, 80, 81, 93–95, 98, 99
Yield to maturity, 665, 666
Youden's Index, 5, 259–261, 271
YouTube, 71, 74, 291, 478, 495, 509, 512, 513, 550, 553, 564, 744

Z
Zero-coupon bond, 665
Zero-inflated models, 422
Zero inflated Poisson (ZIP) and zero inflated negative binomial (ZINB) Models, 434–435
Zomato, 509