

INDEX

- Access control, 70
- Axiom, 61, 74
- Ad network, 4
 - clicks, 4
- Advanced Research and Development Activity (ARDA), 64
- Advertiser traffic, discounting
 - of, 13–14
 - anomaly detection, 14
 - click fraud, 13–14
 - impression fraud, 13–14
- ADVISE. *See* Analysis, Dissemination, Visualization, Insight and Semantic Enhancement.
- Alpha testing, 138
- Amazon.com, 21, 44, 189
 - Purchase Circles, 25, 38
- Analysis, Dissemination, Visualization, Insight and Semantic Enhancement (ADVISE), 64, 65
- Analytic tools, 138
- Anchor text, 6
- Anomaly detection, 14
- Anonymized databases, 72
- AOL, 4, 59–61, 62
- ARDA. *See* Advanced Research and Development Activity.
- Artifacts, 15
- Asymptotic behavior, 231–232
- Auction attributes, 39
- Auction behavior
 - opportunistic, 121
 - participatory, 121
 - sniping, 121
- Auction bidder
 - dynamics, 115–122
 - network, 119–121
 - small-world properties, 120
 - subgroup analysis, 121–122
- Auction market structure, 122–125
- Auction online, 105–127
- Audit trail, 70
- Average position, 234
- Bandit policy, 11–12
 - margin, 12
 - regret, 12
- Banner ads, 4
 - cost calculations, 5
 - impressions, 4

- BARISTA. 336–339
 model, 326
- Barnes & Noble, 38
- Bayesian data analysis, 201
 Markov chain Monte Carlo, 201
- Bayesian networks (BN), 143–151,
 164–166
 conditional probability tables, 149
 directed acyclic graph, 148
 equivalent sample size, 151
 graphical models, 148–155
 joint probability distribution, 149
 learning problem, 150
- Beta testing, 138
- Bid behavior, 325–339
 BARISTA model, 326
 characteristics, 326
 bid sniping, 326
 increasing intensity, 326
 self-similarity, 326
 striking similarity, 326
 cumulative distribution
 functions, 326
 general bid process, 327–332
 Poisson process, 332–339
 revising, 325
 sniping, 325, 326
- Bid closing price, prediction
 of, 277–278
- Bidder dynamics, 115–122
 social network analysis, 116–119
- Bid price evolution
 continuous curve, 297
 nonparametrical approach
 monotone splines, 299–300
 smoothing splines, 297–299
 parametrical approach, 300–306
 exponential model, 300–302
 fitting growth models, 304–306
 logistic model, 303
 reflected logistic model, 303–304
- Bid revising, 325
- Bid sniping, 326
- Bidder subgroup analysis
 opportunistic behavior, 121
 participatory behavior, 121
 sniping behavior, 121
- Bid trajectory analysis, 276
- Bidding draught, 372
- Bidding strategies, 225–241
- Bivariate analysis, 405–409
 extensions, 408–409
- BN. *See* Bayesian networks.
- Bonacich's Power, 118–119
- Borders, 38
- Branding, 21–22
- Budget cap, 226
- Business Week*, 13
- CAPPS II. *See* Computer Assisted
 Passenger Profiling System II.
- CART. *See* classification and
 regression trees.
- Categorical variable assessment,
 376–377
- CDF. *See* cumulative distribution
 functions.
- Central Intelligence Agency (CIA), 63
- Choice model, 350–356
 dynamic updating, 353–356
 estimation, 353–356
 parameter estimation, 354, 356
 semiparametric
 mixed vs., 353–354
 spatial, 351–353
- ChoicePoint, 61, 67, 74, 75
- Chow Test, 211–212
 regimes, 211
- CIA. *See* Central Intelligence Agency.
- Cities, electronic commerce research
 and, 23–24
- Classification and regression trees
 (CART), 391
- Classification and Regression
 Trees*, 374
- Classification trees, 411–413
- Click fraud, 5, 13–14
- Click through rate (CTR), 236–239
 estimating of, 8–11
 bid, 8
 cost budgeting, 8
 impression, 9
 problems in estimating, 9
 data sparsity, 9
 data squashing, 11
 massive scale, 9
 ranking, 9
 rarity of clicks, 9

- Clicks (pay per click), 4
 - cost calculation of, 4
 - monitoring of, 4–5
 - conversions, 4
 - rarity of, 9
- Closing price predictors, 277–278, 282–284
 - mean square prediction error, 278
- Clustering data
 - error-based, 246–247
 - measurement errors, 245–264
 - problems with, 245–246
- Clustering model parameters, 253–256
- Clustering time series modeling, 259–260
- Clusters, number of, 251–252
- CNN, 4, 21
- Coefficients, parameterizing, 217–222
- Common bond, 42
- Common identity, 42
- Computer Assisted Passenger Profiling System II (CAPPS II), 72
- Conditional probability tables (CPT), 149
- Consideration sets, 22
- Consumer information-seeking, 37–43
- Content Match, 4
- Content, Wikipedia’s maintenance cost, 91–92
- Contextual Advertising, 4
- Continuous change, coefficients and, 217–219
- Contributors to Wipedia, 90–91
- Conversion lag, 230–231
- Conversion rate, 132–133, 239–240
 - changes, 230
- Cost adjustments, 49–50
- Cost budgeting, click through rates and, 8
- Cost calculations, cost per milli, 5
- Cost per click (CPC), 4, 228, 233–234
- Cost per milli (CPM), 5
- Costs, search, 26–27
- Cox proportional hazards model, 199–200
- CPC. *See* cost per click.
- CPM. *See* cost per milli.
- CPT. *See* conditional probability tables.
- Crawler, 6–7
 - dynamic content, 6
 - feature extraction, 7
 - multi-armed bandits, 6
- Creatives, 226
- Cross price elasticity, offline vs online markets, 25
- Cross sectional data modeling, 207–222
 - empirical example, 207–210
- Cross validation (CV), 274
- CTR. *See* click through rates.
- Cumulative distribution functions (CDF), 326
- Cumulative hazard function, 175–179
- Current high bid, 295
- Curve representation, bid price evolution and, 297
- Customer feedback 43–44
 - Amazon.com, 44
 - eBay, 44
- Customer wallet
 - definitions, 385–386
 - REALISTIC, 386
 - SERVED, 385
 - TOTAL, 385
 - estimation, 383–398
 - bottom up, 384
 - quantile modeling, 383–398
 - model evaluation, 387–389
 - high-level indicators, 387
 - quantile loss-based, 387–388
 - survey values, 387
- CUSUM/MOSUM test, 212–214
- CV. *See* cross validation.
- DAG. *See* directed acyclic graph.
- DARPA. *See* Defense Advanced Research Program.
- Data mining, 68–69
 - privacy preserving, 68–69
- Data privacy protection, 67
- Data sources, 151
- Data sparsity, 9
- Data squashing, 11
- Data structures, 204–207
 - panel data, 205–207
 - pooled cross sections, 205–207
 - time series data, 205–207
- Data visualization analysis, 179–189
 - advantages, 190–191
 - IPO entry and exit patterns, 183–189
 - digital and physical products, 186–189
 - product life cycle, 180–181
 - survival and failure theoretical explanations, 181

- Data visualization methods, 200
- Data warehouses, 61, 74–76
 - Axciom, 61, 74
 - ChoicePoint, 61, 74, 75
 - LexusNexus, 61, 74
- Databases
 - anonymized, 72
 - theft of, 75
 - Department of Veteran Affairs, 75
 - National Nuclear Security Administration Center, 75
- Datasets, online auctions and, 108–110
- Dataveillance, 65
- Day parting, 226
- Defense Advanced Research Program (DARPA), 63
 - Total Information Awareness, 63
- Degree, as used in social network analysis, 118
- Del.icio.us, 15
- DEM. *See* differential equation models.
- Department of Veteran Affairs, 75
- Differential equation models (DEM), 364, 366–373
 - phase plane plots, 369–371
 - price dynamics, 371–373
 - bidding draught, 372
- Differential equation trees, 363–379
 - functional, 373–379
 - data analysis, 364
 - modeling, 364
 - price curve modeling, 366
- Digital divide, 52
- Directed acyclic graph (DAG), 148
- Discrete changes, coefficients and, 220–222
- Discrete choice models, 198–199
- Discrimination, 27–28
- DNS. *See* domain names server.
- Domain names server (DNS), 5
- Duration, 197
- Dynamic content (hidden web), 6
- Dynamic spatial choice model, 352–353
- Dynamic spatial models, 341–360
 - case study, online mortgage leads, 345–350
 - choice model, 350–356
 - empirical applications, 357–360
 - geographical, 343–344
 - geo-targeting, 344
 - need for, 343–345
 - predictive performance, 358–360
- Dynamic updating, 356
 - stochastic approximation, 356
- eBay, 23, 44, 270–272, 291–293
 - case study data, 322–324
 - willing to pay values, 271
- Economic transactions
 - cities, 23–24
 - electronic commerce research and, 22–24
 - location, 22
- Electronic commerce privacy issues, 59–76
 - AOL, 59–61, 62
 - data warehousing, 61
 - MySpace.com, 62
 - University of Pittsburgh Medical Center, 62
- Electronic commerce research, 19–31
 - branding, 21–22
 - consideration sets, 22
 - economics
 - cities, 23–24
 - international 22–23
 - transactions, 22–24
 - international economics,
 - eBay, 23
 - MercadoLibre, 23
 - internet communications
 - technologies, 23
 - offline vs online markets, 24–30
 - stockouts, 21
 - word of mouth marketing, 20–21
- Electronic commerce, homeland security, 63–65
- EM. *See* expectation maximization.
- Encryption, 67–69
 - data mining, 68
- Equivalent sample size (ESS), 151
- Error assessment, 232–233
- Error based clustering, 246–247
 - clustering time series modeling, 259–260
 - expectation maximization algorithm, 249
 - hError algorithm, 247
 - kError algorithm, 247
 - clustering, 252–253
 - Markov chain modeling, 260–261, 262

- models, 249
 - hError clustering algorithm, 250–252
 - Mahalanobis mean, 250
 - parameters, 253–256
 - multiple linear regression, 258–259
 - probability modeling, 248–249
 - real-world datasets, 261–263
 - sample averaging, 257–258
- ESS. *See* equivalent sample size.
- Estimation method, 391
- Evolution function, 217
- Evolving bid trajectory analysis, 276
- Excite, 188
- Expectation maximization (EM), algorithm, 249, 354–355
- Exponential modeling, 300–302

- Facebook, 73–74
- Factual data analysis, 63
- Failure process, 197
- FDA. *See* functional data analysis.
- Feature extraction, 7
- Fine art auctions, 105–127
- Firm generated online content, 35–53
- Fitting exponential model, 305–306
- Fitting growth models, 304–306
 - exponential, 305–306
 - logarithmic, 306
 - logistic, 306
 - reflected-logistic, 306
- Fitting logarithmic growth, 306
- Fitting logistic growth, 306
- Fitting price curves, smoothing, 367–369
- Fitting reflected-logistic growth, 306
- Fixed, mixed modeling and, 354
- Flickr, 15
- FPCA. *See* functional principal component analysis.
- Fraud
 - anomaly detection, 14
 - click, 13–14
 - impression, 13–14
- Friction costs, 46
- Functional data analysis (FDA), 110, 200–201, 364
 - penalized smoothing splines, 111
- Functional differential equation trees, 373–379
 - functional trees, 374
 - model-based functional differential equation trees, 377
 - model-based recursive partitioning, 374–377
- Functional equation trees, model-based, 377–378
- Functional principal component analysis (FPCA), 270
 - recovering longitudinal trajectories, 272–276
- Functional trees, 374

- GAM. *See* generalized additive model.
- GAO. *See* General Accounting Office.
- GBD. *See* general bid process.
- GCV. *See* generalized cross validation.
- General Accounting Office (GAO), 63
- General bid behavior
 - multibidder auction, 331–332
 - opportunists, 329
 - participators, 329
 - single-bidder auction, 329–331
- General bid process (GBD), 327–332
- Generalized additive model (GAM), 284
- Generalized cross validation (GCV), 274
- Generalized first price, 227
- Generalized second price, 226–227
- Geo targeting, 226, 344
- Gibbs sampler algorithm, 201
- GNU Free Documentation License, 83
 - invariant sections, 83–84
- Goal/question/metric. *See* GQM.
- Google, 4, 39, 61
- GQM (goal/question/metric), 139–141
 - definition of, 140–141
- Graphical ads, 4
- Graphical models, 148–155
 - data sources, 151
 - mental activities, time analysis of, 152
 - misleading links, 155
 - page readability, 154–155
 - task related mental activity time analysis, 153–154
 - usability
 - diagnostics, 153
 - problem indicators, 153
 - visitor's response time analysis, 152–153
 - website usability attributes, 151–152
- Greedy heuristic, 251

- Hazard rate, 197
- hError algorithm, 247
- hError clustering algorithm, 250–252
 - hierarchical greedy heuristic, 251
 - number of clusters, 251–252
- Hidden web, 6
- Hierarchical greedy heuristic, 251
- High-level indicators, 387
- Homeland security, 63–65
 - Analysis, Dissemination, Visualization, Insight and Semantic Enhancement, 64, 65
 - Central Intelligence Agency, 63
 - Computer Assisted Passenger Profiling System II, 72
 - Defense Advanced Research Program, 63
 - factual data analysis, 63
 - General Accounting Office, 63
 - Information Awareness Prototype System, 64, 65
 - predictive analytics, 63
- Horizontal partition, 68
- HTML. *See* Hypertext Markup Language page.
- Hyperlinks, 5–6
 - anchor text, 6
 - PageRank algorithm, 6
- Hypertext Markup Language (HTML) page, 5
 - hyperlinks, 5–6
- IAPS. *See* Information Awareness Prototype System.
- IBM, 392–397
 - market alignment program, 393–394
- ICT. *See* internet communications technologies.
- Impressions, 4, 9, 234, 236
 - fraud, 13–14
- Increasing intensity, 326
- Indexing, inverted, 7
- Inference control, 70–72
 - k-anonymity, 71
- Information Awareness Prototype System (IAPS), 64, 65
- Information retrieval, 7–8
 - SIGIR, 7
- Information searches
 - cost of, 46–48
 - adjustments, 49–50
 - friction costs, 46
 - rational inattention, 46
- Long Tail phenomenon, 50–51
- processing textual content, costs of, 48–49
- Infoseek, 188
- Initial public offering (IPO), 175, 179
 - Amazon.com, 189
 - entry and exit patterns, 183–189
 - digital and physical products, 186–189
 - irrational exuberance, 185–186
 - Excite, 188
 - Infoseek, 188
 - Lycos, 188
 - N2K Inc., 189
 - Yahoo!, 188
- Integration testing, 138
- International economics, electronic commerce research and, 22–23
- Internet communications technologies (ICT), 23
- Internet firm survival and failure, 173–194
 - cumulative hazard function, 175–179
 - data on, 174–175
 - data visualization analysis, 179–189
 - exits, 175
 - hybrid analytical methods, 189–192
 - initial public offerings, 175
 - Kaplan-Meier curve, 175–179
- Invariant sections, 83–84
- Inverted index, 7
- IPO. *See* initial public offerings.
- Irrational exuberance, 185–186
- Joint probability distribution (JPD), 149
- JPD. *See* joint probability distribution.
- K anonymity, 71
- Kaplan Meier curves, 175–179
 - subgroup comparisons, 176–179
 - business sectors, 177–178
 - digital and physical products, 178
 - IPO timing, 179
 - market entry, 179
 - market sectors, 178
- Kaplan Meier estimator, 198

- kError algorithm, 247
- kError clustering algorithm, 252–253
- Keyword selection, 226
- k*-nearest neighbor, 390–391

- Least squares error (LSE), 391
- LexisNexus, 61, 74
- Linear quantile regression, 389
- Link diagnostics, 143
- Linkages, Wikipedia's content, 96–99
- Links, misleading, 155
- Location, economic transactions and, 22
- Log bid analysis, 279–282
 - pooled adjacent violators algorithm, 282
- Log file analysis, 141
- Log price increments, 284–286
- Logarithmic model, 302
- Logistic model, 303
- Logistic regression, 198–199
 - odds ratio, 199
- Long Tail phenomenon, 50–51
 - price search costs, 50
 - product search costs, 50
- Longitudinal trajectories, recovering, 272–276
- LSE. *See* least squares error.
- Lycos, 188

- Mahalanobis mean, 250
- Margin, 12
- Market alignment program (MAP), 393–394
- Marketing, word of mouth, 20–21
- Markov chain modeling, 260–261, 262
- Markov chain Monte Carlo (MCMC), 201
 - Gibbs sampler algorithm, 201
- Markov Chain, 158–164
- Markov Processes, 143, 144–145
- Massive scale, 9
- MATRIX. *See* Multistate Anti-Terrorism Information Exchange.
- MCMC. *See* Markov chain Monte Carlo.
- Mean square prediction error (MSPE), 278
- Measurement error, 273
 - clustering data, 245–264
- Mental activities, 143, 145–148
 - barriers to page usability, 146–147
 - page evaluation, 145–146
 - page usability attributes, 147–148
 - task related analysis, 153–154
 - time analysis of, 152
- MercadoLibre, 23
- Misleading links, 155
- Mixed modeling
 - fixed, 354
 - random effects, 354
 - semiparametric vs., 353–354
- Model based clustering, 248–249
- Model based functional differential equation trees, 377–378
 - applications of, 378–379
- Model based recursive partitioning, 374–377
 - parameter instability, 375–377
 - splitting, 377
- Model price dynamics, differential equation trees, 363–379
 - functional data analysis, 364
 - models, 364
 - price curves, 366
- Monotone splines, 299–300
- Mortgage leads case study, 345–350
- Moving regression. *See* rolling regression.
- Moving window regression. *See* rolling regression.
- MSN, 4, 39
- MSPE. *See* mean square prediction error.
- Multi armed bandit, 6
 - problem, 11–12
- Multibidder auction, 331–332
- Multidimensional scaling, 122–125
- Multiple linear regression, 258–259
- Multistate Anti-Terrorism Information Exchange system (MATRIX), 63–65
 - dataveillance, 65
- MySpace.com, 62, 73–74

- N2K Inc., 189
- National Nuclear Security Administration Center, 75
- Natural language processing (NLP), 48
- Neighbor, 391
- Network data analysis
 - Facebook, 73–74
 - MySpace, 73–74
 - transaction based, 72–74
- New York Times*, 13, 59
- NLP. *See* natural language processing.

- Noncompliance estimates, 409–410
- Nonparametric smoothing model,
 parametric growth vs., 312–315
- Numerical variable assessment, 376
- Observable characteristics, 352
- Odds ratio, 199
- Offline markets, online vs.,
 cross price elasticity, 25
 discrimination, 27–28
 electronic commerce research
 and, 24–30
 sales tax distortion measurement, 28–30
 search costs, 26–27
 store openings, 25–26
 substitution between, 24–26
 vertical organization, 28
- Online auction bid history, 106–107
- Online auction bidder dynamics, 115–122
 social network analysis, 116–119
- Online auction bidder network, 119–121
- Online auction bidder subgroup analysis,
 121–122
 opportunistic behavior, 121
 participatory behavior, 121
 sniping behavior, 121
- Online auction multidimensional scaling,
 122–125
- Online auction price dynamics, 110–115
 functional data analysis, 110
- Online auction price level, 114–115
- Online auction price process
 auction attributes, 293
 case study, 295–296
 current high bid, 295
 data points, 293–208
 eBay, 291–293
 growth curves, 315–318
 growth models, 291–321
 model selection metrics, 307
 weighted sum-of-squares standardized
 by the range, 307
 weighted sum-of-squares standardized
 by the variance, 307
 model selection procedures, 308–312
 parametric growth functions, 293
 price evolution, 297
 nonparametrical approach, 297–300
 second-price auctions, 295
 smoothing methods comparison,
 parametric growth vs.
 nonparametric, 312–315
 sniping, 296
- Online auction price velocity, 114–115
- Online auction proxy-bid system, 108
- Online auction value affiliation, 126
- Online content
 consumer information-seeking, 37–43
 economic value of, 36–37
 purchase behavior, geographical location,
 38–39
 user generated, 35–53
 social information, 41–43
- Online distribution channel, 28
- Online learning, 11–13
 multi-armed bandit problem, 11
- Online markets, offline vs.
 cross price elasticity, 25
 discrimination, 27–28
 electronic commerce research
 and, 24–30
 sales tax distortion measurement, 28–30
 search costs, 26–27
 store openings, 25–26
 substitution between, 24–26
 vertical organization, 28
- Online mortgage leads, case study,
 345–350
- Online purchase behavior, 37–39
 geographical location
 Amazon.com, 38
 Barnes & Noble, 38
 Borders, 38
 Target, 38
 Walmart, 38
- Online search advertising, 39–41
- Opportunistic behavior, 121
- Opportunists, 329
- Outliers, 229–230
- PACE. *See* principal analysis through
 conditional expectation.
- Page diagnostics, 142–143
- Page evaluation, 145–146
 visitor's reaction to, 146
- Page hit attributes, 157
- Page readability time analysis, 154–155
- Page tagging, 141

- Page transition analysis, 157
- Page usability
 - attributes, 147–148
 - barriers to, 146–147
- Page usage statistics, 157–158
- PageRank algorithm, 6
- Panel data, 205–207
- Parameter estimation, 354, 356
 - dynamic updating, 356
 - EM algorithm, 354–355
- Parameter functions, 27
- Parameter instability, 375–377
 - categorical variable assessment, 376–377
 - numerical variable assessment, 376
- Parameterizing coefficients
 - continuous change, 217–219
 - evolution function, 217
 - parameter functions, 217
 - process function, 217
 - transition equations, 217
- Parametric growth curves, 315–318
 - integration of, 318
 - rug plots, 315–317
- Parametric growth functions, 293
- Parametric growth model, nonparametric
 - vs., 312–315
- Participators, 329
- Participatory behavior, 121
- PAVA. *See* pooled adjacent violators algorithm, 282
- Pay per click (PPC), 4, 225–227
 - bidding, 226
 - budget cap, 226
 - creatives, 226
 - day parting, 226
 - generalized first price, 227
 - generalized second price, 226–227
 - geo-targeting, 226
 - keyword selection, 226
- Penalized smoothing splines, 111
- Phase plane plots (PPP), 369–371
- Poisson bid process, 332–339
 - BARISTA, 336–339
 - self-similar bid process, 335
 - time-invariant departure probability, 334–335
- Pooled adjacent violators algorithm (PAVA), 282
- Pooled cross sections, 205–207
- Postrandomization. *See* PRAM.
- PPC. *See* pay per click.
- PPDM. *See* privacy preserving data mining.
- PPP. *See* phase plane plots.
- PRAM (postrandomization), 402
- Predicted page usability, 136
 - limitations of, 136–137
- Predictive analytics, 63
- Predictive performance, dynamic spatial
 - models and, 358–360
- Price curve modeling, 366
 - fitting price curves, smoothing, 367–369
- Price dynamics, online auctions and, 110–115
 - functional data analysis, 110
- Price evolution nonparametrical approach
 - monotone splines, 299–300
 - smoothing splines, 297–299
- Price evolution parametrical approach, 300–306
 - exponential model, 300–302
 - fitting growth models, 304–306
 - logarithmic model, 302
 - logistic model, 303
 - reflected-logistic model, 303–304
- Price level, effects of, 114–115
- Price process, 270
- Price search costs, 50
- Price velocity, 114–115
- Pricing, differential equation models
 - and, 371–373
- Principal analysis through conditional expectation (PACE), 270
- Privacy appliances, 70
 - access control, 70
 - audit trail, 70
 - inference control, 70–72
- Privacy issues, 59–76
 - anonymized databases, 72
 - AOL, 59–61, 62
 - data warehousing, 61
 - encryption, 67–69
 - MySpace.com, 62
 - protection, 67
 - record-linkage methods, 65–67
 - record-matching systems, 65–67
 - safe releases, 72
 - selected revelation, 69–72

- Privacy issues (*Continued*)
 - Technology and Privacy Advisory Committee, 69
 - transaction based network data analysis, 72–74
 - Transportation Security Administration, 75
 - University of Pittsburgh Medical Center, 62
- Privacy preserving data mining (PPDM), 68–69
 - horizontally partitioned, 68
 - privacy-preserving statistical databases, 68
 - secure multiparty computation, 68, 69
 - vertically partitioned, 68
- Privacy preserving statistical databases, 68
- Probability modeling, 248–249
- Process function, 217
- Processing textual content,
 - costs of, 48–49
- Product descriptions, 45–46
- Product life cycle, 180–181
- Product reviews, 44–45
- Product search costs, 50
- Protection, Wikipedia’s content, 92–94
- Proxy-bid auction systems, 108
- Pruning method, 391–392
- Purchase behavior, 37–39
 - geographical location, 38–39
 - Google, 39
 - MSN, 39
 - online search advertising, 39–41
 - social networks, 40–41
 - web search 39–41
 - Yahoo, 39
- Purchase Circles, 25, 38
- Quantile loss based evaluations, 387–388
- Quantile modeling, 383–398
 - applications, 397
 - types, 389–392
 - k -nearest neighbor, 390–391
 - linear quantile regression, 389
 - quanting, 389–390
 - regression tree, 391–392
- Quantile regression tree, 391–392
 - classification and regression trees, 391
 - estimation method, 391
 - pruning method, 391–392
 - splitting criterion, 391
- Quanting, 389–390
- Radial smoothing splines, 353
- Random effects, 354
- Randomized response methodology,
 - 401–414
 - bivariate analysis, 405–409
 - classification trees, 411–413
 - noncompliance estimates, 409–410
 - PRAM, 402
 - statistical disclosure control, 402–403
 - univariate analysis, 404–405
- Range of inattention, 46
- Ranking, 9
- Rarity of clicks, 9
- Rational inattention, 46
 - range of inattention, 46
- Reach, Wikipedia’s, 86
- REALISTIC, 386
- Record linkage methods, 65–67
- Record matching systems, 65–67
 - ChoicePoint, 67
 - SearchSystems.net, 67
- Recovering longitudinal trajectories,
 - 272–276
 - closing price prediction, 277–278
 - cross-validation, 274
 - evolving bid trajectory analysis, 276
 - generalized cross-validation, 274
 - measurement error, 273
- Recursive partitioning, model based,
 - 374–377
- Reflected logistic model, 303–304
- Regimes, 211
- Regression tree, quantile, 391–392
- Regret 12
- Response time analysis, 152–153
- Revising, 325
- Revision, Wikipedia’s content, 94–96
- Right censored, 197
- Rolling regression testing methods,
 - 214–217
 - step size, 214
 - window size, 214
- Routers, 5
- Rug plots, 315–317
- Safe releases, 72
- Sales oriented analytics, 142

- Sales tax distortion measurement, 28–30
- Sample averaging, 257–258
- SDC. *See* statistical disclosure control.
- Search costs, 26–27
- Search engine marketing
 - advantages of, 227
 - bidding strategies, 225–241
 - case studies, 233–240
 - average position, 234
 - click-through rate (CTR), 236, 238–239
 - conversion rate, 239–240
 - cost per click, 233–234
 - impressions, 234, 236
 - modeling, 227–233
 - asymptotic behavior, 231–232
 - conversion lag, 230–231
 - conversion rate changes, 230
 - cost per click, 228
 - error assessment, 232–233
 - outliers, 229–230
 - simultaneous regression bias, 231–232
 - sparse data, 232
 - pay per click, 225–227
- Search engines, 5–8
 - crawler, 6–7
 - domain name servers, 5
 - Hypertext Markup Language, 5
 - indexing, 7
 - information retrieval, 7–8
 - routers, 5
- SearchSystems.net, 67
- Second price auctions, 295
- Secure Flight program, 75
- Secure multiparty computation (SMC), 68, 69
- Security, homeland, 63–65
- Selected revelation, 69–72
- Selected revelation, privacy appliances, 70
- Self similar bid process, 335
- Self similarity, 326
- Semantic orientation, 44
- Semiparametric modeling, mixed vs., 353–354
- Semiparametric spatial choice model, 351–353
 - observable characteristics, 352
 - radial smoothing splines, 353
 - spatial smoothing, 353
 - unobservable characteristics, 352
- Sentiment analysis techniques, 42–43
- Sequential design, 11–13
- SERVED, 385
- Server log files, 138, 156
 - user activity, 156–157
- Server log, 141
- SIGIR, 7
- Simultaneous regression bias, 231–232
- Single bidder auction, 329–331
- Small world properties, 120
- SMC. *See* secure multiparty computation.
- Smooth functional object, 367
- Smoothing methods comparison,
 - parametric growth vs. nonparametric, 312–315
- Smoothing splines, 297–299
 - radial 353
- SNA. *See* social network analysis.
- Sniping, 269–296
 - behavior, 121
- Social network analysis (SNA), 116–119
 - Bonacich's Power, 118–119
 - degree, 118
- Social networks, 40–41
- Social search, 14–15
- Social search, tagging, 15
- Sparse auction data
 - case study, 278–286
 - closing price predictors, 282–284
 - generalized additive model, 284
 - log price increments, 284–286
 - log-bid analysis, 279–282
 - time-varying approach, 282–284
 - eBay, 270–272
 - functional data analysis, 269–287
 - functional principal component analysis, 270
 - price process, 270
 - principal analysis through conditional expectation, 270
 - sniping, 269
- Sparse data, 232
- Spatial models, dynamic, 351–360
- Spatial smoothing, 353
- Splitting criterion, 391
 - least squares error, 391
- Splitting, 377
- Sponsored Search advertising, 4
 - ad network, 4
- Statistical analysis, 143

- Statistical disclosure control (SDC), 402–403
- Step size, 214
- Stochastic approximation, 356
- Stockouts, 21
 - Amazon.com, 21
 - CNN.com, 21
 - Yahoo!.com, 21
- Striking similarity, 326
- Structural change testing methods, 211–214
 - Chow Test, 211–212
 - CUSUM/MOSUM, 212–214
- Subjective page usability, 139
- Survey values, 387
- Survival analysis
 - Bayesian statistics, 201
 - concepts, 197–198
 - Cox proportional hazards model, 199–200
 - data visualization methods, 200
 - discrete choice models, 198–199
 - duration, 197
 - failure process, 197
 - functional data analysis, 200–201
 - hazard rate, 197
 - Kaplan Meier estimator, 198
 - logistic regression, 198–199
 - right-censored, 197
- Tagging, 15
 - artifacts, 15
 - Del.icio.us, 15
 - Flickr, 15
 - Technocrati, 15
- TAPAC. *See* Technology and Privacy Advisory Committee.
- Target Stores, 38
- Tax distortion measurement, 28–30
- Technocrati, 15
- Technology and Privacy Advisory Committee (TAPAC), 69
- Terrorist Information Program (TIA)
 - dataveillance, 65
 - Multistate Anti-Terrorism Information Exchange system, 63–65
- Textual information, user generated, 43–46
- TIA. *See* Total Information Awareness *or* Terrorist Information Program.
- Terrorist Information Program (TIA), 63
- Time series data, 205–207
- Time series modeling, 259–260
- Time variant departure probability, 334–335
- Time varying approach, 282–284
- Time varying coefficients, 203–223
 - cross-sectional data modeling, 207–222
 - empirical example, 207–210
 - data structures, 204–207
 - panel data, 205–207
 - pooled cross sections, 205–207
 - time series data, 205–207
 - testing methods, 211–222
 - discrete change, 220–222
 - parameterizing coefficient, 217–222
 - rolling regression, 214–217
 - structural change, 211–214
- Top-down, 384
- Total Information Awareness (TIA), 63
- TOTAL, 385
- Transition equations, 217
- Transportation Security Administration (TSA), Secure Flight program, 75
- TSA. *See* Transportation Security Administration.
- Univariate analysis, 404–405
 - estimation, 405
- University of Pittsburgh Medical Center (UPMC), 62
- Unobservable characteristics, 352
- UPI identification, 157
- UPMC. *See* University of Pittsburgh Medical Center.
- Usability assurance, 133–135
- Usability diagnostics, methodology of, 153
- Usability problem indicators, 153
- Usability research, 135–136
- Usability validation, 137
- Usability
 - definition of, 132
 - predicted page, 136
- User activity, 156–157
- User generated online content, 35–53
- User generated social information, 41–43
 - common bond, 42
 - common identity, 42
 - sentiment analysis techniques, 42–43

- User generated textual information
 - customer feedback, 43–44
 - economic value of, 43–46
 - product descriptions, 45–46
 - product reviews, 44–45
 - semantic orientation, 44
- Value affiliation, 126
- Vertical organization, 28
 - online distribution channel, 28
- Vertical partition, 68
- Wallet estimation, 383–398
 - case study, IBM, 392–397
 - market alignment program, 393–394
 - top-down, 384
- Walmart, 38
- Washington Post*, 61
- Web analytics, 139–142
 - GQM, 139
 - log file technology, 141
 - page tagging, 141
 - sales oriented, 142
 - server logs, 141
 - statistical software, 142
 - web log analysis software, 141
- Web log analysis software, 141–142
- Web searching, 39–41
- Web site usability attributes, quantification of, 151–152
- Web statistical software, 142
- Web usability diagnostics, 131–169
 - alpha testing, 138
 - analytic tools, 138
 - barriers to, 133
 - beta testing, 138
 - conversion rates, 132–133
 - designer's impact on, 136
 - integration testing, 138
 - modeling of, 142–155
 - prediction limitations, 136–137
 - research based design, 135–136
 - server log files, 138
 - subjective page usability, 139
 - usability assurance, 132, 133–135
 - validation, 137
 - web analytics, 139–142
- Web usability modeling, 142–155
 - case studies, 158–166
 - Bayesian networks, 164–166
 - Markov Chain, 158–164
 - implementation framework, 155–158
 - WebTest analysis, 156–158
 - limitations, 166
 - link diagnostics, 143
 - modeling types, 143–155
 - Bayesian networks, 143–151
 - Markov Processes, 143, 144–145
 - mental activities, 143, 145–148
 - statistical analysis, 143
 - page diagnostics, 142–143
- WebTest analysis, 156–158
 - page hit attributes, 157
 - page transition analysis, 157
 - server log files, 156
 - user activity, 156–157
 - statistical comparison, 158
 - UPI identification, 157
 - usage statistics, 157–158
- Weighted sum-of-squares standardized by the range (WSSER), 307
- Weighted sum-of-squares standardized by the variance (WSSEV), 307
- Wikipedia
 - background of, 83–84
 - content, 89–101
 - contributors, 90–91
 - functionality evolution, 99–101
 - linkages, 96–99
 - maintenance cost, 91–92
 - protection of, 92–94
 - revision management, 94–96
 - English version, 84–89
 - GNU Free Documentation License, 83
 - micro-growth of, 89
 - network analysis of, 81–101
 - reach of, 86
- Willing to pay (WTP), 271
- Window size, 214
- Word of mouth marketing, 20–21
- WSSER. *See* weighted sum-of-squares standardized by the range.
- WSSEV. *See* weighted sum-of-squares standardized by the variance.
- WTP. *See* willing to pay.
- Yahoo!, 4, 21, 39, 188

