

Index

Notes:

- *PA* stands for *predictive analytics*.
- Page numbers in *italics* followed by “*i*” refer to pages within the Central Tables insert.
- The “*n*” after a page number refers to an entry that appears in a footnote on that page.

A

- Abbott, Dean, 71, 91, 199, 200, 204, 303
- AB testing, 261
- Accident Fund Insurance, 7*i*
- accommodation bookings, 2*i*, 8*i*
- actuarial approach, 152
- advertisement targeting, predictive, 31, 296
- advertising. *See* marketing and advertising
- Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (Harcourt), 82
- AI. *See* artificial intelligence (AI)
- Airbnb, 2*i*, 8*i*
- Air Force, 18*i*
- airlines and aviation, predicting in, 8, 14*i*, 128
- Albee, Edward, 113
- Albrecht, Katherine, 53
- algorithmic trading. *See* blackbox trading
- Allen, Woody, 10
- Allstate, 7*i*, 9, 191
- Amazon.com
- employee security access needs, 12*i*
 - machine learning and predictive models, 8
 - Mechanical Turk, 76
 - personalized recommendations, 5*i*
 - sarcasm in reviews, 8
- American Civil Liberties Union (ACLU), 67, 99
- American Public University System, 9, 16*i*
- analytics, 16
- Analytics Revolution, The* (Franks), 305
- Ansari X Prize, 187
- Apollo 11, 223, 244
- Apple, Inc., 115, 223
- Apple Mac, 118
- Apple Siri, 215, 292
- Applied Predictive Analysis* (Abbott), 303
- “Are Orange Cars Really not Lemons?” (Elder, Bullard), 143n
- Argonne National Laboratory, 13*i*
- Arizona Petrified Forest National Park, 259–260
- Arizona State University, 9, 16*i*
- artificial intelligence (AI)
- Amazon.com Mechanical Turk, 76
 - mind-reading technology, 8, 19*i*
 - possibility of, the, 209, 217–219, 248
 - Watson computer and, 213–219
- Asimov, Isaac, 139
- astronomy, 113, 189

- AT&T Research BellKor Netflix Prize teams, 188–189, 192–197
- Australia, *14i*, *16i*, 106, 120, 178, 191
- Austria, 122, 193
- automatic suspect discovery (ASD), 87–101
 approach to, 88–91
 arguments for and against, 96–99
 assumptions about NSA’s use of, 94–95
 challenges of, 99–100
 defined, 88
 example patterns, 91–94
 privacy issues and, 87–88, 100
- automobile insurance
 crashes, predicting, 9, 191
 credit scores and accidents, 120–121
 driver inattentiveness, 9, *18i*, 191
 fraud predictions for, *11i*, 70
- Averitt, *18i*
- aviation incidents, 9, *14i*
- Aviva Insurance (U.K.), *11i*
- AWK computer language, 111
- Ayres, Ian, 306
- B**
- backtesting, 39–41. *See also* test data
- Baesens, Bart, 182
- bagging (bootstrap aggregating), 200–203
- Baker, Stephen, 306
- Baltimore, *11i*
- Bangladesh, 140, 167, 174
- Barbie dolls, 118
- Bari, Anasse, 304
- Bayes, Thomas (Bayes Network), 31
- Beane, Billy, 43
- Beano, 57
- Beaux, Alex, 61–62, 63
- behavioral predictors, 115
- Bella Pictures, *5i*
- BellKor, 188–189, 193, 195–197
- BellKor Netflix Prize teams, 188–189, 192–197
- Ben Gurion University (Israel), 79, 125
- Bernstein, Peter, 151
- Berra, Yogi, 268
- Berry, J. A., 303
- Big Bang theory, 4, 113
- Big Bang Theory*, *The* (TV show), 18
- Big Brother, 85
- BigChaos team, 192–197
- big data
 about, xviii, 16, 105, 114
 The Data Effect and, 115, 135–145, 295
 “wider” data and, 140–141
- billing errors, predicting, *10i*
- black-box trading, *8i*, 24, 38–44
- Black Swan*, *The* (Taleb), xviii, 168
- Blue Cross Blue Shield of Tennessee, *7i*, *10i*
- BMW, 23, 79
- BNSF Railway, *13i*
- board games, predictive play of, 76, 297
- Bohr, Niels, 13
- Boire, Richard, 305
- Bollen, Johan, 107n
- books, about Predictive Analytics, 303–304, 305, 306
- book titles, testing, 262
- Bowie, David, 32
- brain activity, predicting, *19i*
- Brandeis, Louis, 55
- Brazil, 98
- breast cancer, predicting, *9i*, 10
- Breiman, Leo, 178, 200
- Brigham Young University, 10, *10i*
- British Broadcasting Corporation, *16i*, *18i*
- Brobst, Stephen, 151
- Brooks, Mel, 38
- Brynjolfsson, Eric, 111–112
- buildings, predicting fault in, *13i*
- Bullard, Ben, 143n
- burglaries, predicting, *11i*, 67
- business rules, decision trees and, 161, 164
- buying behavior, predicting, *3i*, 48–49, 121
- C**
- Cage, Nicolas, 28, 262
- Canadian Automobile Association, *14i*
- Canadian Tire, *7i*, 121

- car crashes and harm, predicting, 9, *10i*, 191
- CareerBuilder, *20i*, 191
- Carlin, George, 212
- Carlson, Gretchen, 52
- Carnegie Mellon University, 57, 228, 285
- cars, “orange lemons,” 104–109, 142–143
- car services passenger destination, predicting, *14i*
- CART decision trees, 178–181, 200–203
- causality, 35, 131–135, 158, 257
- cell phone industry
- consumer behavior and, 120
 - dropped calls, predicting, 8–9, *14i*, 204
 - GPS data and location predicting, *2i*, 57, 84
 - Telenor (Norway), *5i*, 252–254, 281
- CellTel (African telecom), 125
- Centers for Medicare and Medicaid Services, *11i*
- Cerebellum Capital, *8i*
- Chabbert, Martin, 185–189, 192, 301
- Chaffee, Alexander Day, 171, 244
- CHAID, 178
- Chaouchi, Mohamed, 304
- Charlotte Rescue Mission, *15i*
- Chase Bank
- actuarial approach used by, 152
 - churn modeling, 166, 252n
 - data, learning from, 153–156
 - economic recession and risk, 181–182
 - learning from data, 162–163
 - mergers and growth of, 149, 181
 - microrisks faced by, 149–153
 - mortgage risk decision trees, 163–165, 179–181
 - mortgage risks, predicting, *5i*, *7i*, 149–150
 - mortgage value estimation, 180–181
 - predictive models and success, 181
 - risks faced by, 149–153, 181
 - Steinberg as leader, 3, 147–148, 153–156, 175–176, 178–181
- cheating, predicting, 11–12, *18i*, 76
- check fraud, predicting, 11, *11i*
- chess-playing computers, 76, 222–223
- Chicago Police Department, 12, *12i*
- Chopra, Sameer, 304
- Chu–Carroll, Jennifer, 244, 246
- churn modeling, 166, 252–254, 277, 281, 298, 299
- Cialdini, Robert, 259–260
- Citibank, 8, *18i*
- Citigroup, *7i*
- Citizens Bank, 11, *11i*, 69
- City of Boston, *15i*
- City of Chicago, *14i*, *15i*
- City of New York, *12i*, *14i*, *16i*
- city power, predicting fault in, *13i*
- city regulations, crime predictions and, *11i*
- civil liberties, risks to, 47–48, 80
- Clarke, Arthur C., 209
- clicks, predicting, 7, *20i*, 25, 29, 262
- clinical trial recruitment, predicting, *10i*
- Clinton, Hillary, 8, 289
- cloning customers, 156n
- Coase, Ronald, 167
- Colbert, Stephen, 53, 283
- Colburn, Rafe, 74
- collection agencies, PA for, 10, *13i*
- collective intelligence, 197–199
- collective learning, 17–18
- Columbia University, 12, 79, 125, 248
- Commander Data, 77
- company networks, predicting fault in, *13i*
- Competing on Analytics: The New Science of Winning* (Davenport and Harris), 37–38, 305
- competitions. *See* PA competitions
- computational linguistics, 210. *See also* natural language processing (NLP)
- computer “intelligence.” *See* artificial intelligence (AI)
- computer programs, decision trees and, 161–162

- computers
 creation of, 18–19
 Deep Blue computer, 76, 222–223
 human brain *versus*, 19–20
 viruses on, predicting, 12*i*, 73
See also artificial intelligence; machine learning; Watson computer *Jeopardy!*
 challenge
- computer science, 74
- Con Edison, 6, 13*i*
- conferences, 305
- Congo massacres, 125
- consumer behavior, insights on, 118–119
- contests. *See* PA competitions
- Continental Airlines, 14*i*
- control sets, 267, 268
- corporate roll-ups, 195n
- coupons, predictive, 6–7
- Coursera, 153n
- Cox Communications, 4*i*
- Craig, Roger, 17*i*, 225, 302
- credit cards
 fraud detection for, 69
 payment systems, predicting fault in, 10, 13*i*
- credit risk
 credit scores and, 149–151
 typing and, 121–122
- Crichton, Michael, 172
- crime fighting and fraud detection
 cheating in board games, 76
 credit card fraud, 69
 in financial institutions, 69–71
 fraud, defined, 68
 fraudulent transactions, 69
 IRS tax fraud, 11*i*, 12, 71, 204
 machine injustice, 79–80
 network intrusion, 12*i*, 72
 PA application, 70, 297
 PA for, 8*i*–9*i*, 11, 68–77
 prejudice in, risk of, 81–82
 spam filtering, 74
- crime prediction for law enforcement
 automatic suspect discovery (ASD) and NSA, 87–101
 cybercrime, 73
 ethical risks in, 79–82
 fighting, PA for, 8*i*–9*i*, 11, 67
 judicial decisions and, 77–80, 160–161
 murder, predicting, 11, 11*i*, 12*i*, 18*i*, 78
 PA application, 66–67, 297
 PA examples and insights, 125
 prediction models for, pros and cons of, 83–86
 recidivism, 12*i*, 77–80
 Uber and behavior, 118
- crowdsourcing
 collective intelligence and, 197–199
 Kaggle PA crowdsourcing contests, 189–191, 204
 noncompetitive crowdsourcing, 190n
 PA and, 185, 187–191, 197–200, 224
- Cruise, Tom, 79, 152
- customer need, predicting fault in, 14*i*
- customer retention
 cancellations and predicting, 4*i*, 8
 with churn modeling, 166, 252–254, 277, 298
 with churn uplift modeling, 277, 299
 contacting customers and, 119–120
 feedback and dissatisfaction, 18*i*
- customization, perils of, 30–31
- cybercrime, 73
- cybernetics, xvii
- D**
- D’Arcy, Aoife, 304
- data, sharing and using
 automatic suspect discovery (ASD) and NSA, 87–101
 battle over mining, 56–57
 choosing what to collect, 6–12, 56–57, 249, 254–256
 growth and rates of expansion, 109–114
 policies for, 55

- privacy concerns and, 52–54, 58–60, 83–86
- data, types and sources of
 - big data, xviii, 16, 105, 114
 - buzzwords for, 114
 - consumption and, 6–7
 - employee data, 61
 - fake data, 68–69
 - free public data, 110
 - learning data, 162–163
 - location data, 2*i*, 57, 84
 - medical data, 85
 - personal data, 54–55, 57
 - social media posts, 9*i*
 - textual data, 210, 214
- data, value of
 - about, 4–5
 - learning from data, 153–156
 - machine learning and, 4–5
 - personal data, 54–55, 57
 - as “the new oil,” 115
- Dataclysm* (Rudder), 306
- Data Effect, The, 115, 135–145, 295
- data gluts, 112
- Data.gov, 110
- data hustlers and prospectors, 57, 85
- data mining, 17, 36, 58, 114
- Data Mining for Managers* (Boire), 305
- Data Mining Techniques* (Linoff, Berry), 303
- data preparation phase, 162–163
- data science, 16–17, 114*n*
- Data-sim* (Lohr), 306
- Data Smart* (Foreman), 303
- data storage, 56–57, 112
- dating websites
 - attractiveness ratings and success, 127
 - consumer behavior on, 118–119
 - predicting on, 3*i*, 7, 110, 127, 259, 291–292
- Davenport, Thomas H., xix, 37–38, 305
- death predictions, 9*i*, 10*i*, 10–11, 84–85, 306
- “Deathwatch” (Siegel), 306
- deception, predicting, 11–12, 18*i*
- Decision Management Systems* (Taylor), 305
- decision trees
 - CART decision trees, 178–181, 200–203
 - circle single model, 201–202
 - decision boundaries in, 202
 - getting data from, 175–179
 - in machine learning, 156–162
 - methods competing with, 171–172
 - mortgage risk decision trees, 163–165, 180–181
 - overlearning and assuming, 167–169
 - random forests, 201
 - uplift trees, 275
- deduction *versus* induction, 169–170
- Deep Blue computer, 76, 222–223
- DeepQA, 247
- deliveries, predicting, 14*i*
- Delta Financial, 38–41
- Deming, William Edwards, 4
- Democratic National Committee (DNC), 283–288
- Deshpande, Bala, 304
- Dey, Anindya, 59, 61, 301
- Dhar, Vasant, 17
- diapers and beer, behavior and, 117
- Dick, Philip K., 28
- differential response modeling. *See* uplift modeling
- discrimination, risks of, 81–82
- disease, predicting, 9*i*
- Disraeli, Benjamin, 167
- divorce, 7, 53–54
- dolls and candy bars, behavior and, 118
- Domingoes, Pedro, 305
- donations and giving, predicting, 15*i*, 204
- do-overs, 257–259
- Dormehl, Luke, 306
- downlift, 273
- driver inattentiveness, predicting, 9, 18*i*, 191
- driverless cars, 23, 79
- Drucker, Peter, 149
- drugs effects and use, predicting, 10*i*

- “Dry Bones” (song), 114
 DTE Energy, 8i
 Dubner, Stephen J., 27, 72, 81, 91n, 92, 95
 Duhigg, Charles, 51–52
 dunnhumby, 191
 Dyson, George, 214
- E**
- Earthlink.com, 119
 Echo Nest, 126
 economic recession, 181–182
 education
 grades, predicting, 16i, 191
 guided studying for targeted learning, 225, 299
 PA for, 8, 15i–17i
 student dropout risk, predicting, 9, 16i
 student knowledge and performance, predicting, 17i, 191
 eHarmony, 7
 Eindhoven University, 16i
 Eindhoven University, Netherlands, 9
 Einstein, Albert, 4, 175
 Elder, John, 8i, 143n, 300, 304
 about, 23, 39, 43–44
 “Are Orange Cars Really not Lemons?,” 143n
 black-box trading systems, 24, 38–44
 on employee death predictions, 84
 on generalization paradox, 204–205
 in Netflix Prize competition, 192
 on passion for science, 44–45
 on power of data, 54
 risks taken by, 38–41
 on “vast search,” 140
 Elder Research, Inc., 45, 71, 132, 143n, 192
 elections, crime rates and, 125
 electoral politics
 Hillary for America 2016 Campaign, 15i
 musical taste political affiliation, 125
 Obama for America 2012 Campaign, 15i
 Obama for America Campaign, 8, 282–288, 286–288
 uplift modeling applications for, 277, 286–288
 voter persuasion, predicting, 15i, 160, 282–288
 electronic equipment, predicting fault in, 13i
Elements of Statistical Learning, The (Hastie, Tibshirani, Friedman), 304
 Elie Tahari, 4i
 e-mail
 consumer behavior and addresses for, 119–120
 government storage of, 113
 Hotmail.com, 34, 109, 119
 phishing e-mails, 74
 spam filtering for, 74
 emotions
 in blog posts, 107
 human behavior and, 106–108
 mood predictions and, 106–108
 See also human behavior
 employee longevity, predicting, 20i
 employees and staff
 job applications and positions, 20i, 191
 job performance, predicting, 20i, 204
 job promotions and retention, 128
 job skills, predicting, 20i
 LinkedIn for career predictions, 7, 20i
 privacy concerns and data on, 84
 quitting, predicting, 9, 20i, 58–64, 128
 Energex (Australia), 6, 16i
 energy consumption, predicting, 16i
 Ensemble Effect, The, 205, 275n, 295
 Ensemble Experts, 192
 ensemble models
 about, 204
 automatic suspect discovery (ASD) and, 93n
 CART decision trees and bagging, 178–181, 200–203
 collective intelligence in, 199–200
 complexity in, 204

- crowdsourcing and, 185, 187–191, 197–200, 224
- generalization paradox and, 204
- IBM Watson question answering
computer and, 18*i*, 204
- IRS (tax fraud), 11*i*, 12, 71, 204
- meta-learning and, 193–195
- Nature Conservancy (donations), 15*i*, 204
- Netflix (movie recommendations), 5*i*, 6, 183, 186, 204
- Nokia-Siemens Networks (dropped calls), 14*i*, 204
- University of California, Berkeley (brain activity), 19*i*, 204
- for uplift modeling, 275*n*
- U.S. Department of Defense (fraudulent invoices), 11*i*
- U.S. Department of Defense Service (fraudulent invoices), 204
- U.S. Special Forces (job performance), 20*i*, 204
- Ensemble team, 197
- Epagogix, 8*i*
- erectile dysfunction, 10*i*
- Experian, 150
- Exxon Mobil Corp., 115
- F**
- “Fab Four” inventors, 178, 200
- Facebook
data glut on, 112
data on, 4, 29, 55
fake data on, 55
friendships, predicting, 2*i*, 191
happiness as contagious on, 124
job performance and profiles on, 20*i*
social effect of, 124
student performance PA contest, 17*i*
- facial recognition, 2*i*
- Failure of Risk Management, The (Hubbard), 151–152
- false conclusions, avoiding, 104–109, 142–143
- false positives (false alarms), 79, 140
- family and personal life, PA for, 2*i*–3*i*, 7, 54, 124
- Farrell, Colin, 79
- fault detection for safety and efficiency, PA for, 13*i*–14*i*
- Federal Trade Commission, 69
- FedEx, 4*i*, 9
- Femto-photography, 112
- Ferguson, Andrew, 101
- Ferrucci, David, 217, 243, 246
- FICO, 10*i*, 150
- Fidelity Investments, 275
- finance and accounting, fraud detection in, 11, 69–71
- finance websites, behavior on, 118–119
- financial risk and insurance, PA for, 7*i*–8*i*, 11
- Fingerhut, 4*i*
- Finland, 124
- fire, predicting, 16*i*
- First Tennessee Bank, 3*i*, 4*i*
- Fisher, Ronald, 135, 136*n*
- Fleming, Alexander, 139
- flight delays, predicting fault in, 14*i*
- Flight Risks, predicting, 8, 47, 58–64
- Flirtback computer program, 111
- Florida Department of Juvenile Justice, 12*i*
- fMRI brain scans, 19*i*
- Foldit, 190*n*
- Food and Drug Administration (FDA), 279
- Fooled by Randomness* (Taleb), 136–137, 138, 168
- Ford Motor Co., 9, 18*i*, 191
- forecasting, 17, 284
- Foreman, John W., 303
- Formula, The* (Dormehl), 306
- Fox & Friends* (TV show), 52
- Franklin, Benjamin, 29, 75
- Franks, Bill, 305
- fraud, defined, 68
- fraud detection, 48, 61, 68–77, 297. *See also* crime fighting and fraud detection

Freakonomics Radio, 27, 72, 81, 140
Freakonomics Radio, 140
 frequency, 116
 Friedman, Jerome, 178, 304
 friendships, predicting, 3*i*, 191
 Fukuman, Audrey, 159
 Fulcher, Christopher, 66
Fundamentals of Machine Learning for Predictive Data Analytics (Kelleher, MacNamee, D'Arcy), 304
 fund-raising, predicting in, 16*i*
 Furnas, Alexander, 56, 83
 future, views on
 human nature and knowing about, xxxi
 predictions for 2022, 291–293
 uncertainty of, 12–13

G

Galileo, 109
 Gates, Bill, 137
 generalization paradox, 204
 Ghani, Rayid, 284–285, 302
 Gilbert, Allen, 99
 Gladwell, Malcolm, 30–31
 GlaxoSmithKline (U.K.), 10*i*
 Goethe, Johann Wolfgang von, 36
 Goldbloom, Anthony, 190, 204
 Gondek, David, 207, 224, 227–231, 234–237, 240n, 245–248, 302
 Google
 mouse clicks, measuring for predictions, 7, 29, 262
 privacy policies, 84
 Schmidt, 83, 84
 searches for playing *Jeopardy!*, 214
 self-driving cars, 23, 79
 spam filtering, 74
 Google Adwords, 30, 262
 Google Flu Trends, 9*i*, 123
 Google Page Rank, 199
 government
 data storage by, 110
 fraud detection for invoices, 11*i*, 71, 204

PA for, 15*i*–17*i*
 public access to data, 110
 See also individual names of U.S. government agencies
 GPS data, 2*i*, 57, 84
 grades, predicting, 16*i*, 191
 grant awards, predicting, 16*i*, 191
 Greenwald, Glenn, 98
 Grockit, 17*i*, 191
Groundhog Day (film), 258
 Grundhoefer, Michael, 267, 268–269

H

hackers, predicting, 12*i*, 73
 HAL (intelligent computer), 209–210
 Halder, Gitali, 59, 61, 301
Handbook of Statistical Analysis and Data Mining Applications (Nisbet, Elder, Miner), 304
 Hansell, Saul, 182
 happiness, social effect and, 107, 124
 Harbor Sweets, 4*i*, 6
 Harcourt, Bernard, 82
 Harrah's Las Vegas, 4*i*
 Harris, Jeanne, 37–38, 305
 Harvard Medical School, 9
 Harvard University, 123
 Hastings, Reed, 197
 healthcare
 death predictions in, 9*i*, 10–11, 84–85
 health risks, predicting, 10*i*
 hospital admissions, predicting, 10*i*
 influenza, predicting, 9*i*, 124
 insurance companies, predicting, 7*i*, 9*i*
 medical research, predicting in, 124
 medical treatments, risks for wrong predictions in, 267–269, 277, 279
 medical treatments, testing persuasion in, 262
 PA for, 9*i*–10*i*, 10–11, 124
 personalized medicine, uplift modeling applications for, 277, 279
 health insurance companies, PA for, 9*i*–10*i*, 10–11, 84–85

- Hebrew University, *18i*
- Heisenberg, Werner Karl, 262
- Helle, Eva, 280, 281, 282, 302
- Helsinki Brain Research Centre, 124
- Hennessey, Kathleen, 282, 286
- Heraclitus, 257
- Heritage Health Prize, 191
- Heritage Provider Network, 10, *10i*
- Hewlett Foundation, *16i*, 189, 191
- Hewlett-Packard (HP)
- employee data used by, 61
 - financial savings and benefits of PA, 61, 64
 - Global Business Services (GBS), 62
 - quitting and Flight Risks, predicting, 8, *20i*, 48, 58–64, 128
 - sales leads, predicting, *5i*
 - turnover rates at, 60
 - warranty claims and fraud detection, 11, *11i*
- Hillary for America 2016 Campaign, 8, *15i*
- HIV progression, predicting, *9i*, 191
- HIV treatments, uplift modeling for, 279
- Hollifield, Stephen, 67
- Holmes, Sherlock, 33, 34, 83, 86
- Hopper, *8i*
- hormone replacement, coronary disease and, 133
- hospital admissions, predicting, *10i*, 10–11
- Hotmail.com, 34, 109, 119
- House* (TV show), 83
- “How Companies Learn Your Secrets” (Duhigg), 51–52
- Howe, Jeff, 187, 189
- HP. *See* Hewlett-Packard (HP)
- Hubbard, Douglas, 109, 151–152
- human behavior
- collective intelligence, 197–199
 - consumer behavior insights, 119–120
 - emotions and mood prediction, 199–200
 - mistakes, predicting, 11–12
 - social effect and, 11–12, 120, 124
- human genome, 113
- human language
- natural language processing (NLP), 210, 219–221, 227–231
 - PA for, *18i–19i*
 - persuasion and influence in, 259–260
- human resources. *See* employees and staff
- I**
- IBM
- corporate roll-ups, 195n
 - Deep Blue computer, 76, 224
 - DeepQA project, 247
 - Iambic IBM AI, 248
 - mind-reading technology, 8
 - natural language processing research, 227
 - sales leads, predicting, *5i*
 - student performance PA contest, *17i*
 - T. J. Watson Research Center, 224
 - value of, 222–223
 - See also* Watson computer *Jeopardy!* challenge
- ID3, 178
- impact modeling. *See* uplift modeling
- Imperium, *18i*, 191
- inappropriate comments, predicting, *18i*
- incremental impact modeling. *See* uplift modeling
- incremental response modeling. *See* uplift modeling
- India, 125
- Indiana University, 107
- Induction Effect, The, 179, 295
- induction *versus* deduction, 169–170
- inductive bias, 170
- ineffective advertising, predicting, *20i*
- infidelity, predicting, 122
- Infinity Insurance, *7i*, *16i*
- influence. *See* persuasion and influence
- influenza, predicting, *9i*, 124
- information technology (IT) systems, predicting fault in, *13i*
- InnoCentive, 190n
- insults, predicting, *18i*, 191

- insurance claims
 automobile insurance fraud, predicting, 7*i*, 10–11, 70, 121, 191
 death predictions and, 7*i*, 10–11, 84–85
 financial risk predicting in, 8*i*
 health insurance, 9*i*, 10–11, 84–85
 life insurance companies, 11
 life nsurance companies, 7*i*
 Integral Solutions Limited, 195n
 Internal Revenue Service (IRS), 11*i*, 12, 71, 106, 204
 International Conference on Very Large Databases, 114
 Iowa State University, 9, 16*i*
 iPhone. *See* Apple Siri
 Iran, 13*i*
 Israel Institute of Technology, 12*i*
- J**
 Japan, 124
 Jennings, Ken, 226, 239, 240, 246
Jeopardy! (TV show). *See* Watson computer
Jeopardy! challenge
 Jevons, William Stanley, 169
 Jewell, Robert, 248
 jobs and employment. *See* employees and staff
 Jones, Chris, 238
Journal of Computational Science, 107n
 JPMorgan Chase. *See* Chase Bank
 judicial decisions, crime prediction and, 79, 160–161
 Jung, Tommy, 304
Jurassic Park (Crichton), 172
 Just Giving, 15*i*
- K**
 Kaggle, 189–191, 204
 Kane, Katherine, 276
 Kasparov, Garry, 76, 223
 KDnuggets, 84, 110
 Keane, Bil, xxix
 “keep it simple, stupid” (KISS) principle, 178, 200
 Kelleher, John D., 304
 Khabaza, Tom, 115
 killing, predicting, 11, 11*i*, 18*i*
 King, Eric, 248, 269n
 Kiva, 8*i*
 Kmart, 7
 knee surgery choices, 124
 knowledge (for education), predicting, 13*i*
 Kotu, Vijay, 304
 Kretsinger, Stein, 132, 133, 189
 Kroger, 7
 Kuneva, Meglena, 4, 115
 Kurtz, Ellen, 78, 82
 Kurzweil, Ray, 113
- L**
 language. *See* human language
 Lashkar-e-Taiba, 92
 law enforcement. *See* crime prediction for law enforcement
 lead poisoning from paint, predicting, 16*i*
 learning
 about, 17–21
 collective learning, 17–18
 education—guided studying for targeted learning, 225, 299
 learning from data, 162–163
 memorization *versus*, 169
 overlearning, avoiding, 167–169, 175, 200, 204
 Leinweber, David, 167
 Leno, Jay, 12
 Levant, Oscar, 217
 Levitt, Stephen, 72, 81, 91n, 92, 95
 Lewis, Michael, 15
 lies, predicting. *See* deception, predicting
Lie to Me (TV show), 12
 life insurance companies, PA for, 7*i*, 11
 Life Line Screening, 4*i*
 lift, 166, 266n
 Lindbergh, Charles, 223
 LinkedIn
 friendships, predicting, 3*i*

- job skills, predicting, 7, 20i
- Linoff, Gordon S., 303
- Linux operating systems, 190n
- Lloyds TSB, 5i
- loan default risks, predicting, 7i, 8i
- location data, 2i, 57, 84
- logistic regression, 236
- Lohr, Steve, 306
- London Stock Exchange, 8i
- Los Angeles Police Department, 12i, 67
- Lotti, Michael, 83
- Loukides, Mike, 17
- love, predicting, 7, 110–111, 126–127, 259, 291–292
- Lynyrd Skynyrd (band), 253

- M**
- MacDowell, Andie, 258
- machine learning
 - about, 4–5, 19–21, 147–148, 156–159, 171–173, 204
 - courses on, 153n
 - in crime prediction, 79–80
 - data preparation phase for, 162–163
 - decision trees in, 156–162
 - induction and, 169
 - induction *versus* deduction, 169–170
 - learning data, 162–163
 - learning from mistakes in, 156
 - learning machines, building, 153–156
 - overlearning, 167–169, 175, 200, 204
 - predictive models, building with, 36, 41
 - silence, concept of, 20n
 - testing and validating data, 173–175
 - univariate *versus* multivariate models, 154–156, 156–157
 - See also* Watson computer *Jeopardy!* challenge
- machine risk, 79–80
- MacNamee, Brian, 304
- macroscopic risks, 182
- Mac *versus* Windows users, 118
- Madrigal, Alexis, 54
- Magic 8 Ball toy, 81
- Mao, Huina, 107n
- maritime incidents, predicting, 14i
- marketing and advertising
 - banner ads and consumer behavior, 119
 - mouse clicks and consumer behavior, 7, 29, 262
 - targeting direct marketing, 26, 296
- marketing models
 - do-overs in, 257–258
 - messages, creative design for, 259–262
 - Persuasion Effect, The, 276
 - quantum humans, influencing, 262–266
 - response uplift modeling, 270–271, 275
- marketing segmentation, decision trees and, 161–162
- marriage and divorce, predicting, 7, 53–54, 122
- Mars Climate Orbiter, 39–40, 245
- Martin, Andres D., 160
- Maryland, crime predictions in, 11, 11i, 78
- Massachusetts Institute of Technology (MIT), 227
- Master Algorithm, The* (Domingos), 305
- Match.com, 3i, 7, 291–292
- Matrix, The* (film), 213
- McCord, Michael, 231n
- McKinsey reports, 190–191
- McNamara, Robert, 269
- Mechanical Turk, 76
- medical claims, fraudulent, 11i
- medical treatments. *See* healthcare
- memorization *versus* learning, 169
- Memphis (TN) Police Department, 12, 12i, 67
- metadata, 91–92, 106, 106n
- meta-learning, 193–195
- Mexican Tax Administration, 71
- Miami Police Department, 12i
- Michelangelo, 175
- microloan defaults, predicting, 16i
- Microsoft, 2i, 223
- Milne, A. A., xxix

- Mimoni (Mexico), 7i
 mind-reading technology, 8, 19i
 Miner, Gary, 304
Minority Report (film), 79
 Missouri, 79
 Mitchell, Tom, 57, 170, 177, 178, 234, 304
 mobile operators. *See* cell phone industry
 moneyballing, concept of, 15, 224–226, 282
 mood labels, 106–108
 mood prediction, blogs and, 107
 mortgage prepaids and risk, predicting, 7i, 149–151
 mortgage risk decision trees, 163–165, 180–181
 mortgage value estimation, 180–181, 298
 mouse clicks, predicting, 7, 29, 262
 movie hits, predicting, 8i, 20i
 movie recommendations, 6, 183, 186, 204, 298
 movies, 20i
 MTV, 21i
 MultiCare Health System (Washington State), 10i
 “multiple comparisons problem”/multiple comparisons trap”, 140
 multivariate models, 154–156, 156–157
 murder, predicting, 11, 11i, 12i, 18i, 78
 Murray, Bill, 258
 music, stroke recovery and, 124
 musical taste, political affiliation and, 126
 Muslims, 81–82
- N**
- Naïve Bayes, 31
Naked Future, The (Tucker), 306
 NASA
 Apollo 11, 223, 244
 Mars Climate Orbiter, 39–40, 245
 PA contests sponsored by, 187, 189
 on space exploration, 76
 National Insurance Crime Bureau, 69
 National Security Agency (NSA), 12i, 87–101
 National Transportation Safety Board, 9
 natural language processing (NLP), 210, 219–221, 227–231
 Nature Conservancy, 15i, 204
 Nazarko, Edward, 222
Nerds on Wall Street (Leinweber), 167
 Netflix movie recommendations, 5i, 6, 183, 186, 204
 Netflix Prize
 about, 185–187
 competition and winning, 192
 crowdsourcing and PA for, 185–187, 187–191, 197–200, 223
 meta-learning and ensemble models in, 193–195, 204
 PragmaticTheory team, 188–189, 196–197
 Netherlands, 16i
 net lift modeling. *See* uplift modeling
 net response modeling. *See* uplift modeling
 net weight of evidence (NWOE), 272
 network intrusion detection, 12i, 72
 New South Wales, Australia, 191
 New York City Medicaid, 11i
 New York State, 11i
New York Times, The, 279
Next (Dick), 28, 262
 Ng, Andrew, 153n
 Nightcrawler (superhero), 54
Nineteen Eighty-Four (Orwell), 85
 99designs, 190n
 Nisbet, Robert, 304
 “no free lunch” theorem, 170n
 Nokia, 2i
 Nokia-Siemens Networks, 14i, 204
 nonprofit organizations, PA for, 15i–17i
 Noonan, Peggy, 286
No Place to Hide (Greenwald), 98
 Northwestern University Kellogg School of Management, 117
 nuclear reactors, predicting fault in, 13i
 null hypothesis, 136n
Numerati, The (Baker), 306

O

Obama for America 2012 Campaign, 8, *15i*, 282–288

observation, power of, 33–36

Occam's razor, 178, 200

O'Connor, Sandra Day, 160

office equipment, predicting fault in, *13i*

Oi (Brasil Telecom), *8i*

oil flow rates, predicting, *13i*

oil refinery safety incidents, predicting, *13i*

OkCupid, *3i*, 7, 126–127, 259

Oklahoma State University, 9, *16i*

O'Leary, Martin, 189

Olshen, Richard, 178

1-800-FLOWERS, 71

1-sided equality of proportions hypothesis test, 137n

Online Privacy Foundation, *18i*, 191

Oogway, xxix

open data movement, 110

open question answering, 213–222, 226, 238

open source software, 190n

Optus (Australia), *4i*, 106, 120

“orange lemons” (cars), 104–109, 142–143

Orbitz, 118

Oregon, crime prediction in, 11, *12i*, 78

organizational learning, 17–18

organizational risk management, 28

Orwell, George, 85

Osco Drug, 117

overfitting. *See* overlearning

overlearning, 167–169, 175, 200, 204

Oz, Mehmet, 27

P

PA (predictive analytics)

about, xxx–xxxii, 15–17, 116

assumption about NSA's use of, 94–96

choosing what to predict, 6–12, 55–56, 249, 254–256

in crime fighting and fraud detection, *8i–9i*

crowdsourcing and, 185, 187–191, 197–200, 224

defined, 15, 152

in family and personal life, *2i–3i*

in fault detection for safety and efficiency, *13i–14i*

in finance and accounting fraud detection, 11, 69–71, 121

in financial risk and insurance, *7i–8i*

forecasting *versus*, 17, 284

frequently asked questions about, xxii–xxvii

in government, politics, nonprofit, and education, *15i–17i*

in healthcare, *9i–10i*, 10–11, 124

in human language understanding, thought, and psychology, *18i–19i*

launching and taking action with PA, 23–26

in law enforcement and fraud detection, *11i–12i*

in marketing, advertising, and the Web, *4i–6i*

“orange lemons” and, 104–109, 142–143

overview, xxi–xxii

risk-oriented definition of, 152

text analytics, 209, 214

in workforce (staff and employees), *20i*

PA (predictive analytics) applications

black-box trading, 24, 43

blog entry anxiety detection, 107

board games, playing, 76, 292

credit risk, 149–151, 277

crime prediction, 66–67, 297

customer retention with churn modeling, 166, 252–254, 277, 298, 299

customer retention with churn uplift modeling, 277, 299

defined by, 26

education—guided studying for targeted learning, 226, 299

employee retention, 58–64, 297

fraud detection, 70–71, 297

- PA (predictive analytics) applications
(continued)
 mortgage value estimation, 180, 298
 movie recommendations, 186, 298
 network intrusion detection, 72, 297
 open question answering, 220, 238
 political campaigning with voter
 persuasion modeling, 285, 299
 predictive advertisement targeting, 31,
 296
 pregnancy prediction, 48, 296–297
 recidivism prediction for law
 enforcement, 78, 298
 spam filtering, 74, 297
 targeting direct marketing, 26–27, 296
 uplift modeling applications, 277
See also Central Tables insert
- PA (predictive analytics) competitions and
 contests
 in astronomy and science, 189
 for design and games, 190n
 for educational applications, 189
 Kaggle crowdsourcing contests, 189–191,
 204
 Netflix Prize, 185–187, 204
- PA (predictive analytics) insights
 consumer behavior, 119–120
 crime and law enforcement, 125
 finance and insurance, 121
 miscellaneous, 126–130
- Palmisano, Sam, 247
- Panchoo, Gary, 198
- Pandora, 21i
- parole and sentencing, predicting, 11, 12i,
 77–78
- Parsons, Christi, 282, 286
- PAW (Predictive Analytics World)
 conferences, xxii, xxvii, 48–49, 61,
 70–71, 115, 197–199, 227, 240n, 305
- payment processors, predicting fault in, 13i
- PayPal, 8, 18i, 70
- penicillin, 139
- Pennsylvania, 12i
- personalization, perils of, 30–31
- persuasion and influence
 observation and, 255–257
 persuadable individuals, identifying,
 262–266, 274, 286–288
 predicting, 255–256, 255–276
 scientifically proving persuasion,
 259–260
 testing in business, 262
 uplift modeling for, 266–267
 voter persuasion modeling, 285, 299
- Persuasion Effect, The, 276, 295
- persuasion modeling. *See* uplift modeling
- Petrified Forest National Park, Arizona,
 259–260
- Pfizer, 10i
- Philadelphia (PA) Police Department, 77
- photographs
 caption quality and likability, 129
 growth of in data glut, 112
- Piotte, Martin, 185–189, 301
- Pitney Bowes, 195n, 273
- Pittsburgh Science of Learning, 17i
- Pole, Andrew, 48–49, 301
- police departments. *See* crime prediction for
 law enforcement
- politics, PA for, 15i–17i. *See also* electoral
 politics
- Porter, Daniel, 287, 289, 302
- Portrait Software, 195n
- Post hoc, ergo propter hoc*, 130
- Power of Habit: Why We Do What We Do in
 Life and Business* (Duhigg), 52
- Pragmatic Theory team, 188–189, 196–197
- prediction
 benefits of, 3, 28
 choosing what to predict, 6–12, 55–56,
 249, 254–256
 collective obsession with, xxi
 future predictions, 291–293
 good *versus* bad, 83–86
 limits of, 12–14
 organizational learning and, 17–18

- prediction, effects of and on
 about, 12–14, 295
 Data Effect, The, 115, 135–145, 295
 Ensemble Effect, The, 205, 275n, 295
 Induction Effect, The, 179, 295
 Persuasion Effect, The, 276, 295
 Prediction Effect, The, xxx–xxxii, 1–21,
 26–27, 295
- prediction markets, 199
- predictive analytics. *See* PA (predictive analytics)
- Predictive Analytics* (Siegel), website of, 303
- Predictive Analytics and Data Mining* (Kotu, Deshpande), 304
- Predictive Analytics Applied (training program), 304
- Predictive Analytics for Dummies* (Bari, Chaouchi, Jung), 304
- Predictive Analytics Guide, xxvii, 303
- Predictive Analytics Times*, xxvii, 303, 304, 305, 306
- Predictive Analytics World (PAW), xxvii, 305
- Predictive Analytics World (training programs), 304
- Predictive Analytics World (PAW) conferences, xxii, xxvii, 48–49, 61, 70–71, 115, 197–199, 227, 240n, 305
- Predictive Marketing and Analytics* (Strickland), 304
- predictive models
 about, 23–24
 action and decision making, 36–38
 causality and, 35, 131–135, 158, 257
 defined, 34, 154–155
 deployment phase, 32
 Elder’s success in, 41–45
 going live, 24–26, 30–31
 machine learning and building, 36, 41
 marketing models, 257–267, 270–271, 273–276
 observation and, 33–36
 overlearning and assuming, 167–169
 personalization and, 30–31
 response modeling, 255–256, 268–270, 270–271
 response uplift modeling, 270–271, 275, 277, 299
 risks in, 38–41
 univariate *versus* multivariate, 154–156, 156–157
 uplift modeling, 266–267
See also ensemble models
- PredictiveNotes.com, xvi, xxiii, xxvi, 95, 103, 147, 191, 249, 306
- predictive technology, 3–4. *See also* machine learning
- predictor variables, 116
- pregnancy and birth, predicting
 customer pregnancy and buying behavior, 7, 48–52, 296–297
 premature births, 10, 10i
- prejudice, risk of, 81–82
- PREMIER Bankcard, 4i, 7i
- prescriptive analytics, xviii, 267n
- privacy, 47–48, 56–57, 83–86, 186
 Google policies on, 84
 insight *versus* intrusion regarding, 60–61
 predicted consumer data and, 47–52
- probability, The Data Effect and, 137–139
- profiling customers, 156n
- Progressive Insurance, 70
- Psych* (TV show), xxx
- psychology
 predictive analysis in, 18i
 schizophrenia, predicting, 10, 18i
 psychopathy, predicting, 18i, 191
 purchases, predicting, 4i, 48–49, 121
 p-value, 136n
- Q**
- Quadstone, 195n, 273, 280

R

Radcliffe, Nicholas, 267
 Radica Games, 19i
 Ralph's, 7
 random forests, 201
 Rebellion Research, 8i
 recency, 116
 recidivism prediction for law enforcement, 11–12, 12i, 77–78, 298
 recommendation systems, 7, 57, 185–187, 223
 Reed Elsevier, 5i, 17i
 reliability modeling, 13i
 REO Speedwagon (band), 58
 response modeling
 drawbacks of, 268–270
 examples of, 4i
 targeted marketing with, 255–256, 270–271, 277, 299
 response rates, 268, 276–277
 response uplift modeling, 255–256, 270–271, 275, 277
 restaurant health code violations, predicting, 16i
 retail websites, behavior on, 118–119
 retirement, health and, 122
 Richmond (VA) Police Department, 12, 12i, 67, 78
 RightShip, 14i
 Rio Salado Community College, 16i
 risk management, 28, 137–139, 149–151, 152
 Riskprediction.org.uk, 9i
 risk scores, 149–151
Risky Business (film), 152
 Romney, Mitt, 285
 Rousseff, Dilma, 98
 Royal Astronomy Society, 189
 R software, 190n
 Rudder, Christian, 306
 Russell, Bertrand, 135
 Rutter, Brad, 241, 247

S

safety and efficiency, PA for, 13i–14i
 Safeway, 7
 sales leads, predicting, 7i
 Salford Systems, 178
 “Sameer Chopra’s Hotlist of Training Resources for Predictive Analytics” (*Predictive Analytics Times*), 304
 Sanders, Bernie, 289
 Santa Cruz (CA) Police Department, 12i, 67
 sarcasm, in review, 18i
 Sartre, Jean-Paul, 116
 SAS, 195n
 satellites, predicting fault in, 13i
 satisficing, 269n
 Schamberg, Lisa, 108
 Scheer, Robert, 96
 schizophrenia, predicting, 10, 18i
 Schlitz, Don, 238
 Schmidt, Eric, 83, 84
 Schwartz, Ari, 86
Science magazine, 57
 scientific discovery, automating, 139–141
Seattle Times, 104
 security levels, predicting, 12i
 self-driving cars, 23, 79
 Selfridge, Oliver, 161
 Semisonic (band), 41
 sepsis, predicting, 9i
Sesame Street, 109
 Sesenbrenner, James, 96
 Sessions, Roger, 175
 Shakespeare, William, 169, 209
 Shaw, George Bernard, 156
 Shearer, Colin, 110
 Shell, 13i, 20i
 shopping habits, predicting, 4i, 48–49, 121
 sickness, predicting, 9i, 10–11
 Siegel, Eric, 300, 303, 306, 311
 silence, concept of, 20n
 Silver, Nate, 27, 140, 238, 283
Simpsons, The (TV show), 247, 254
 Siri, 202, 215

- Sisters of Mercy Health Systems, *9i*
 smoking and smokers
 health problems and causation for, 135
 motion disorders and, 123, 131–132
 social effect and quitting, 9, 123
 Snowden, Edward, 87, 98
 Sobel, David, 55
 social effect, 10–11, 120, 123
 social media networks
 data glut on, 112
 happiness as contagious on healthcare,
 124
 LinkedIn, *3i*, 7, *20i*
 PA for, *3i*
 spam filtering on, 74
 Twitter, 69, 112
 YouTube, 112
 See also Facebook
 sociology, uplift modeling applications for,
 277
 SpaceShipOne, 187
 spam, predicting, *20i*
 spam filtering, 74, 297
Spider-Man (film), xviii, 83
 sporting events, crime rates and, 125
 sports cars, 119
 Spotify, *21i*
 Sprint, *4i*
 SPSS, 195n
 staff behavior. *See* employees and staff
 Standard & Poor's (S&P) 500, 39
 Stanford University, *9i*, 10, 153n, 178
 staplers, hiring behavior and, 118
Star Trek (TV shows and films), 41, 77, 196,
 209, 220
 statistics, 5, 16, 78, 112, 125, 135–145,
 136–139, 167, 186
 StatSoft, 178
 stealing, predicting, 11–12
 Steinberg, Dan, 3, 147–148, 149, 153,
 175–176, 178–181, 301
 stock market predictions
 black-box trading systems, *8i*, 24, 38–44
 Standard & Poor's (S&P) 500, 39
 Stone, Charles, 178
 Stop & Shop, 6
 street crime, predicting, *11i*
 Strickland, Jeffrey, 304
 student dropout risks, predicting, 9, *16i*
 student performance, predicting, *16i*, 191
 suicide bombers, life insurance and, 125
 Sun Microsystems, *5i*
Super Crunchers (Ayres), 306
SuperFreakonomics (Levitt and Dubner), 72,
 81, 91n, 92, 95
 supermarket visits, predicting, *4i*, 191
 surgical site infections, predicting, *9i*
 Surowiecki, James, 197, 199
 Sweden, 124
 system failures, *13i*
 Szarkowski, John, 112
- T**
 Taleb, Nassim Nicholas, xviii, 136–137,
 138, 168
 Talking Heads (band), 52
 Target
 baby registry at, 49–50
 couponing predictively, 6
 customer pregnancy predictions, *3i*, 7,
 48–52, 296–297
 privacy concerns in PA, 52, 85
 product choices and personalized
 recommendations, *6i*
 purchases and target marketing
 predictions, *4i*
 targeted marketing with response uplift
 modeling, 255–256, 270–271, 277,
 299
 targeting direct marketing, 26–27, 296
 target shuffling, 143n
 taxonomies, 161
 tax refunds, *11i*, 12, 71, 204
 Taylor, James, 305
 TCP/IP, 73
 Telenor (Norway), *5i*, 252–254, 281
 Teragram, 195n
 terrorism, predicting, *11i*, 11–12, 81–82

- Tesco (U.K.), *5i*, 6
- test data, 173–175
- test preparation, predicting, *17i*
- text analytics, 189, 195n, 209
- Text Analytics World, 305
- textbooks, 304
- text data, 209, 214
- text mining. *See* text analytics
- They Know Everything about You* (Scheer), 96
- thought and understanding, PA for, 8, *18i–19i*
- thoughtcrimes, 85
- Tibshirani, Robert, 304
- Titanic* (ship), 130
- T. J. Watson Research Center, 224
- tobacco. *See* smoking and smokers
- Tolstoy, Leo, 122
- traffic, predicting, *14i*, 191
- training data, 49, 61, 148–149, 162–163, 170, 276. *See also* learning
- training programs, 304
- train tracks, predicting fault in, *13i*
- Trammps (band), 130
- travel websites, behavior on, 118–119
- Trebek, Alex, 207, 209, 224, 230, 245
- TREC QA (Text Retrieval Conference—Question Answering), 222
- truck driver fatigue, predicting, *18i*
- true lift modeling. *See* uplift modeling
- true response modeling. *See* uplift modeling
- TTX, *13i*
- Tucker, Patrick, 306
- Tumblr, 112
- Turing, Alan and the Turing test, 74, 77, 222
- Twenty Questions game, decision trees and, 162
- Twilight Zone* (TV show), 262
- Twitter
- 2001: A Space Odyssey (film), 209
 - data glut on, 112
 - fake accounts on, 69
 - mood prediction research via, 107n
 - person-to-person interactions saved by, 106
- 2degrees (New Zealand), *5i*
- typing, credit risk and, 121–122
- ## U
- Uber, *2i*, 118
- uncertainty principle, 262
- understanding and thought, predictive analysis in, *18i–19i*
- univariate models, 154–156, 156–159
- University of Alabama, 9, *16i*
- University of Buffalo, 12, *18i*
- University of California, Berkeley, *19i*, 204
- University of Colorado, 125
- University of Helsinki, 124
- University of Iowa Hospitals & Clinics, *9i*
- University of Massachusetts, 227
- University of Melbourne, *16i*, 191
- University of New Mexico, 122
- University of Phoenix, *16i*
- University of Pittsburgh Medical Center, 10, *10i*
- University of Southern California, 227
- University of Texas, 227
- University of the District of Columbia, 101
- University of Utah, 10, *10i*, *15i*
- University of Zurich, 122
- uplift modeling
- customer retention with churn uplift modeling, 252–254, 276–279, 281, 299
 - downlift in, 273
 - influence across industries, 276–279
 - mechanics and workings of, 266–267, 270–275
 - Obama for America Campaign and, 277, 286–288
- The Persuasion Effect, and, 276
- response uplift modeling, 270–271, 275
 - targeted marketing with response uplift modeling, 255–256, 277
- Telenor using, 252, 254, 281

- U.S. Bank using, 6, 7*i*, 119–120, 266–276
- uplift trees, 275
- UPS, 14*i*
- U.S. Armed Forces, 12*i*
- U.S. Bank, 5*i*, 6, 119–120, 266–276, 273*n*, 274*n*, 275
- U.S. Department of Defense, 92
- U.S. Department of Defense Finance and Accounting Service, 11*i*, 71, 204
- U.S. Food and Drug Administration (FDA), 279
- U.S. government. *See* government
- U.S. National Institute of Justice, 67
- U.S. National Security Agency, 113
- U.S. Naval Special Warfare Command, 20*i*
- U.S. Postal Service, 11*i*, 16*i*, 71
- U.S. Social Security Administration, 16*i*
- U.S. Special Forces, 20*i*, 204
- U.S. Supreme Court, 160–161
- Utah Data Center, 113
- V**
- variables. *See* predictor variables
- “vast search,” 140
- Vermont Country Store, 4*i*, 6
- Vineland (NJ) Police Department, 12*i*, 67
- viral tweets/posts, predicting, 20*i*
- Virginia, crime prediction in, 12, 12*i*, 67, 78
- Volinsky, Chris, 188
- voter persuasion, predicting, 15*i*, 160, 282–288
- W**
- Wagner, Daniel, 287
- Walgreens, 57
- Wall Street Journal*, *The*, 134
- Walmart, 118
- Wanamaker, John, 26
- warranty claim fraud, predicting, 11, 11*i*
- washing machines, fault detection in, 13*i*
- Watson, Thomas J., 224, 227*n*
- Watson computer *Jeopardy!* challenge about, 18*i*, 24, 207–209
- artificial intelligence (AI) and, 217–219
- candidate answer evidence routines, 232
- confidence, estimation of, 238–240
- Craig’s question predictions for, 17*i*, 225
- creation and programming of, 207–215, 225, 226–227
- ensemble models and evidence, 234–235
- Jeopardy!* questions as data for, 207–209, 222
- language processing and machine learning, 236–237
- language processing for answering questions, 18*i*, 204, 210, 219–234
- moneyballing *Jeopardy!*, 224–226
- natural language processing (NLP) and, 210, 219–221, 227–231
- open question answering, 213–222, 226, 238, 244
- playing and winning, 8, 18*i*, 24, 204, 241–247
- praise and success of, 247–248
- predictive models and predicting answers, 226–227, 231–234
- predictive models for predicting answers, 17*i*, 18*i*
- predictive models for question answering, 17*i*, 18*i*, 219–221, 226–227, 231–234
- Siri *versus* Watson, 215–216
- speed in answering for, 241
- Web browsing, behavior and, 120
- Webster, Eric, 152
- Wells Fargo, 20*i*
- Whiting, Rick, 291
- Who Wants to Be a Millionaire?* (TV show), 199
- “wider” data, 140–141
- Wiener, Norbert, xvii
- Wikipedia, 20*i*
- editor attrition predicting, 9, 20*i*, 191
- entries as data, 214, 227
- noncompetitive crowdsourcing in, 190*n*

Wilde, Oscar, 156
Wilson, Earl, 12
Windows *versus* Mac users,
 118

Winn-Dixie, 6
Wired magazine, 191
Wisdom of Crowds, The (Surowiecki), 197,
 199

WolframAlpha, 216
WordPress, 112
workplace injuries, predicting, 7*i*
Wright, Andy, 159
Wright brothers, 32
Wriston, Walter, 54

X
X Prize, 187

Y
Yahoo!, 119
Yahoo! Labs, 19*i*
*Yes! 50 Scientifically Proven Ways to Be
 Persuasive* (Cialdini et al.), 260
yoga, mood and, 124
YouTube, 112

Z
Zeng, Xiao-Jun, 107n
Zhou, Jay, 69

