

# Index

*Note:* Page numbers followed by “*f*” and “*t*” refer to figures and tables, respectively.

## A

AAA (Authenticate and Authenticate Again), 158  
Adaptive cruise control (ACC), 18  
Advanced Automatic Crash Notification Relay system, 193  
Aggregations, 48  
Amazon  
    Redshift, 125  
    S3, 123  
Analytics workflow, 32  
Anderson–Darling (A–D) test, 243  
    comments on, 244–247  
Anscombe’s Quartet dataset, 34, 34*t*, 167–169, 168*t*  
    scatter plots and linear regression models for, 35*f*  
Antitheft system vulnerabilities, 139  
    attacks, 140–142, 140*t*  
Apache  
    Beam, 120  
    Hadoop, 87–90  
    Tez, 61  
    UIMA project, 60–61  
    Zeppelin, 121–122  
ArcGIS, 6  
Architecture Analysis & Design Language (AADL), 195–202  
    AGREE, 199–200, 206–207  
        annex snippet, 201*f*  
        speed sensor, 206*f*  
    behavior annex, 197–198  
        snippet, 198*f*  
    contract snippet, 197*f*  
        contract implementation snippet, 197*f*  
    error annex, 198–199  
        OutOfRange error, 199  
        snippet, 199*f*, 200*f*  
        StuckValue error, 199  
    functionality, 197  
    language overview, 196–197  
    Resolute annex, 200–202  
        snippet, 201*f*  
        snippet, 201*f*  
Artificial neural networks (ANNs), 295–296, 296*f*  
AssocExplorer, 41  
Asymmetric cryptography, 147  
Attack taxonomy, 137, 137*f*  
Attacks on connected vehicle systems  
    antitheft system attacks, 140–142, 140*t*  
    bypass kits, 141  
    digital signature transponder (DST), 141  
    ECU attack, 142–144, 142*t*

    internal bus attacks, 143, 143*t*  
    Keeloq, 141  
    radio jamming, 141  
    relay attack, 141  
    RollJam, 141–142  
    TPMS attack, 144, 144*t*  
    VANETs attacks, 145, 145*t*, 146*t*  
Autoencoder, 219  
Automated driving system (ADS), 23  
Automated vehicles, 134–135  
AWS Elastic MapReduce, 123  
Azure Blob Storage, 123

## B

Basic Safety Messages (BSMs), 18, 220  
Bayesian Neural Networks (BNN), 225–227  
Big data, defined, 20–21  
Big data analytics, 4–9, 54  
Big data and open data initiatives, 231–233  
Bivariate analysis, 247–253  
Bloom’s taxonomy, 42  
Bluetooth low energy (BLE), 153  
Bluetooth smart, 153  
Bluetooth WPAN, 142–143  
Bokeh, 121–122  
Boxplots, 36, 37*f*  
Building and validation models, 58–60  
Business analytics, 47–53  
    business intelligence (BI), 47–48  
    data mining, 51–53  
    data warehouses, 48–49  
    Extract, Transform, and Load (ETL) tool, 49–51  
    OLAP cubes, 48–49  
    star schema, 48–49, 49*f*  
Business Intelligence (BI), 31, 47–48, 121–122

## C

CAN vulnerabilities, 138  
Capability Maturity Model (CMM), 33  
CarShark, 143  
Cascading, 61  
Cell phone technologies, 4  
Central tendency, 33  
Chi-squared test, 243  
    comments on, 244–247  
Chrysler’s Uconnect dashboard computers, 144  
Cloudera Data Hub (CDH), 62

- Cloudera Manager, 123
  - Clustering (cluster analysis), 53, 298–299
  - CODASYL, 45
  - Coefficient of determination, 252
  - Cognition, 45, 55
  - Cognitive analytics, 31, 54–55
  - Cognitive architecture, 55
  - Cognitive computing system, 45, 55
  - Cognitive model, 55
  - Cognitive processes, 45, 55
  - Cognitive Science, 45
  - Collaborative Adaptive Sensing of the Atmosphere (CASA) radars, 105
  - Columnar data format, 114
  - Comma separated values (CSV) file, 76
  - Commercial/Media Wholesale Web Portal (CWWP), 2
  - Computer Emergency Response Team (CERT), 137
  - Confusion matrix, 58, 58*t*, 59*t*
  - Connected transport system (CTS), 113
    - data infrastructure, 115–117, 116*f*, 274*f*
    - data ingest and stream processing, 119–120
      - message broker, 119–120, 124*f*
      - processing engines for streaming, 120
      - storage formats, 120
    - use cases, 118*t*
  - Connected vehicle networks and vehicular applications, 3–4, 132–135, 134*f*, 191
    - attack taxonomy, 137, 137*f*
    - development scenario, 202–209
    - external networks, 133
    - gateway ECU, 151
    - handover security, 158
    - innovative vehicular applications, 133–135
    - intrusion detection system (IDS), 151
    - in-vehicle networks, 132–133
    - privacy measurement of sensor data, 157–158
      - Personal identifiable information (PII), 157
      - Privacy Preserving Data Mining (PPDM), 157
      - sensitivity (anomaly) detection and measurement of privacy/sensitivity, 158
      - unpredictable anomalous events, 157
    - rootkit traps, 150–151
    - secure OTA ECU firmware update, 155–157
    - secure VANETs, 153–154
    - security analysis, 137–145
      - attacks, 140–145
      - future directions, 158–159
      - network and protocol vulnerability analysis, 138–140
    - security and privacy solutions, 146–158
      - accountability, 147
      - choice, 147
      - data minimization, de-identification, & retention, 147
      - data security, 147
      - integrity and access, 147
      - respect for context, 147
      - transparency, 147
    - security solutions for bus communications, 148–151
      - authentication, 148–149
      - code obfuscation, 148
      - confidentiality, 149
      - integrity, 149–150
    - stakeholders and assets, 135–136, 136*f*
    - WPAN security and privacy, 152–153
      - Bluetooth security checklist, 152
      - enabling data privacy in WPAN, 153
      - secure WPAN, 152–153
  - Connected Vehicle Reference Implementation Architecture (CVRIA), 6, 18, 114–115, 202
    - collision detection behavior, 205*f*
    - Do Not Pass Warning (DNPW), 203–209, 204*f*
    - EMV2 error annex, 208, 208*f*
    - goals and requirements, 203–204
    - information flow characteristics, 20*t*
    - resolute power draw validation, 207*f*, 208*f*
  - Continuous Air-interface for Long and Medium Range (CALM), 135
  - Continuous Engineering, 194–195
    - pipeline, 195*f*
  - Cooperative Adaptive Cruise Control (CACC), 18, 19*f*
  - COPA, 42
  - Correlation coefficient, 43–44
  - Crash data, 228
  - Crash injury severity modeling, 217–218
  - Cryptography, 147–148
    - asymmetric, 147
    - hash function, 148
    - symmetric, 148
  - Curbee, 106–107, 106*f*
  - Cybersecurity, 235
- ## D
- Data analytics
    - architecture, 202–203
    - descriptive, 242–249
    - functional facets, 32–45
    - future directions, 62–63
    - predictive, 249–259
    - R programming environment for, 70–71
    - tools and resources, 60–62
  - Data cleaning, 47
  - Data driven highway patrol plan, 218–219
  - Data infrastructure, 115–117, 116*f*
    - data collection and ingest, 116
    - data processing engines, 116
    - higher-level infrastructure, 117–122
    - low-level infrastructure, 116, 122–125
    - machine learning and analytics, 116

- Data integration, 47
- Data lake, 115
- Data lifecycle, 56–57, 92, 95–102
  - comparison of models, 102*r*
  - DataONE data lifecycle model, 98–99, 98*f*
  - Digital Curation Center (DCC) curation model, 96–98, 97*f*
    - community watch and participation, 97
    - conceptualization and data collection, 97–98
    - curation and preservation, 97
    - description and representation, 97
    - disposition of data, 97–98
    - full lifecycle actions, 97
    - preservation planning, 97
    - transformation of data, 97–98
  - future directions, 107–108
  - SEAD Research Object lifecycle model, 99–102, 100*f*, 105–106
    - interrelated components, 99
    - Live Object (LO), Curation Object (CO), or Publishable Object (PO), 101
    - relationships, 100–101
    - Research Object (RO) approach, 99
    - states, 100–101
  - U.S. Geological Survey (USGS) model, 95–96, 95*f*
    - acquiring, 96
    - analysis stage, 96
    - cross-cutting activities in, 96
    - planning, 95–96
    - preservation and publication of data, 96
    - processing, 96
- Data marts, 49–51
- Data mining, 51–53
  - classification, 40–41
  - clustering (cluster analysis), 53
  - evolution analysis, 52
  - frequent patterns, 51–52
    - itemset, 52
    - subsequence, 52
    - substructure, 52
  - outlier detection, 56
- Data pipelines, 102–107
  - data lifecycle and, 103–104
  - examples, 104*f*, 105
- Data quality, 57
- Data routing and processing, 103
- Data sampling, 56–57
- Data science, 45, 55–60, 115, 117, 121–122
- Data use cases, 92–95
  - importance of data to advancements in ITS, 94
  - in-vehicle sensor data, 94
  - sensor data, 94
  - social media data, 94
  - weather and climate data, 94
- Data variability, 92–95
- Data visualizations, 165
  - case study, 181–185
  - classifying systems, 171–172
    - data types, 171–172
    - interaction and distortion techniques, 172
    - visualization techniques, 172
  - computer graphics, 166
  - fundamental challenge in designing effective, 172–175
    - data quantity reduction, 173–174
    - miniaturizing visual glyphs, 174–175
  - interactive, 165–166
  - navigation strategies, 175–177
    - focus + context strategy, 177
    - overview + detail strategy, 176
    - zoom and pan operations, 176
  - pipeline, 169–171, 170*f*
  - principles for designing effective, 179–181
    - clear understanding in captions, 181
    - data-pixel ratio, 180
    - focus + context scheme, 180–181
    - graphical excellence, 179
    - graphical integrity, 179
    - multifunctioning graphical elements, 180
    - optimal quantitative scales, 180
    - reference lines, 180
    - support multiple concurrent views, 180
    - techniques for alleviating over-plotting issues, 181
  - using graphical symbols, or glyphs, 166
  - visual interaction strategies, 177–179
    - dynamic query technique, 178
    - filtering operations, 178–179
    - linking techniques, 178
    - rearranging and remapping, 179
    - selecting items, 177–178
  - visual mappings, 170
- Database Management Systems (DBMS), 45
  - IBM Information Management System (IMS), 45
  - Relational DBMS (RDBMS), 45–47, 61
  - System R, 45–46
- Databricks cloud, 121–122
- Dataframe abstractions, 120
- Dataframes, 120–121
- Decision trees, 52
- Deep learning, 122
- Delimited file, 75–78
- Density curve, 39–40
- Department of Motor Vehicles (DMV), 230
- Descriptive analytics, 6, 32–41
- Diagnostic analytics, 32, 41–43
  - case studies, 41–43
    - COPA, 42
    - Student Success System (S3), 42
    - teaching and learning, 42–43
- Digest, 148

Digital signature transponder (DST), 141  
 Dispersion, 33  
 Distribution of a variable, 33–34  
 Distributive Collaborative Adaptive Sensing (DCAS), 105

## E

Elastic MapReduce (EMR), 62, 123  
 Electronic Control Units (ECUs), 131, 149–150  
   rootkit vulnerabilities, 150–151  
 Elliptic Curve Integrated Encryption Scheme algorithm, 153  
 EMC Isilon, 123  
 Enterprise data warehouse, 49–51  
 ERTICO Automated Driving Roadmap, 23  
 Evolution analysis, 52  
 Exploratory data analysis (EDA), 34–36  
   case studies, 40–41  
   discovery process, 36  
   illustration, 36–40  
   presentation process, 35  
   visual exploration process, 35–36  
 Exploratory Data analysis Environment (EDEN) visual  
   analytics, 171, 175, 175*f*, 181  
   dynamic variable summarization via embedded  
     visualizations, 183  
   dynamic visual queries through direct manipulation, 182  
   multiple coordinated views, 183–185, 184*f*, 185*f*  
   multivariate visualization using interactive parallel  
     coordinates, 182  
 Extract, Transform, and Load (ETL) tool, 49–51  
 Extreme outliers, 36

## F

Feature selection, 57  
 Feature vector, 52  
 Federal Motor Carrier Safety Administration (FMCSA), 218  
 First-Order Reliability Method (FORM), 219–220  
 5Vs of Big Data, 1–2  
 Freight Advanced Traveler Information Systems (FRATIS),  
   241–242  
 Fuzzy regression, 256–259  
   fuzzy parameters, 257

## G

Georgia Navigator, 14–15, 15*f*  
 Geostationary Operational Environmental Satellite system  
   (GOES), 94  
 Ggplot2, 121–122  
 Global positioning systems (GPS), 3–4  
 Goodness-of-fit (GOF) test, 243  
 Google, 134–135  
   BigQuery, 125

Cloud Storage, 123  
 Dataflow, 120  
 DistBelief, 122  
 Dremel, 120  
 Graph mining, 52  
   application, 56  
 GraphFrames, 120–121

## H

Hadoop, 113  
   the cloud, 123–125  
   deployment options, 124*t*  
   Distributed File System (HDFS), 61, 120, 123  
   ecosystem, 61  
   MapReduce, 61, 87–89, 117–119  
   Platform-as-a-Service (PaaS) distribution, 62  
   Yet Another Resource Negotiator (YARN), 120, 123  
 Handovers, security, 158  
 Hash function, 148  
 HAWQ, 120  
 HAZMAT security, 14  
 High-level abstractions for data analytics, 113  
 Histograms, 33–34  
 Hive, 61, 120  
 Hortonworks, 62  
 Hortonworks Data Platform (HDP), 62  
 Hours-of-Service (HOS) rules, 218  
 Hypercube, 48

## I

IBM Watson's Jeopardy! game championship, 31  
 Identity Resolution Key (IRK), 153  
 If-then rules, 52  
 Image processing tools, 6  
 Impala, 120  
 Incident management, 4  
 Information Mural, 175  
 Information visualization, 165–166  
 Integrated Database Management System (IDMS), 45  
 Integrity constraints (ICs), 57  
 Intelligent transportation system (ITS), 91, 131, 191  
   architecture, 9–14  
     logical, 11  
     physical, 11–12  
     security, 14  
     service packages, 12–13  
     standards, 13–14  
     system components, 9  
   Transit Management Subsystem, 12–13  
   transit vehicle tracking service package, 12, 13*f*  
   US national ITS architecture, 10, 10*f*  
   user services and user service requirements, 10–11

- data flow diagram, 12*f*
- data sources and data collection technologies, 3–4, 5*t*
- data system, 2–3
- data-intensive applications of, 1–4
- 5Vs of Big Data, 1–4
  - foundation layer, 2–3
- infrastructure to support, 4–9
- Joint Program Office (JPO), 13–14
- 1960s and 1970s, 21
- 1980s and 1990s, 21–22
- overview of its applications, 14–20, 16*t*
  - data analytics, 18–20
  - mobility, safety and environmental, 15–18
  - variable speed limits system, 16–18, 17*f*
- Strategic Plan, 167
- three Is, 3
- 2000s, 22–23
- 2010s, 23–24
- Intermodal freight transportation
  - ITS-enabled
    - data analytics, 242
- Internet of Things (IoT), 31–32, 107–108, 191
- Internet protocol (IP), 2–3
- Interquartile (IQ) range, 36
- Intrusion detection system (IDS), 151
- In-vehicle infrastructure, 114
- Inverse document frequency (idf), 53

## J

- JavaScript Object Notation (JSON)-type data, 273

## K

- Keeloq, 141
- Keim's classification scheme, 171
- Kernel, 39–40
- Kernel density plot, 33–34
- K-fold cross validation, 300
- Kolmogorov–Smirnov (KS) test, 158, 243
  - comments on, 244–247
- KullbackLeibler (KL) distance, 158
- Kurtosis, 33–34

## L

- LAK dataset, 41
- Lambda architecture for scalable real-time data systems, 274–276
- Latent Dirichlet Allocation (LDA) model, 40
- LIN vulnerabilities, 138
- Linear regression, 291–292
- Linear regression line, 43, 44*f*
- LinkedUp project, 41

- Loadable kernel module (LKM) rootkit, 150–151
- Long-term pavement performance (LTPP) data, 288–289
- Lustre, 123

## M

- Machine learning, 114, 122, 283
  - algorithms
    - clustering, 298–299
    - decision trees, 293–295, 294*f*
    - development approach, 287*f*
    - evaluating performance, 299–300
    - neural networks (NNs), 295–297
    - regression methods, 290–293, 291*f*
    - support vector machines (SVMs), 297–298
  - example, 300–302, 302*f*, 303*f*
  - methods and algorithms, 283–286
    - supervised learning, 284–285
    - unsupervised learning, 285–286
  - understanding data, 286–290
    - data collection, 287–288
    - data fusion, 288–289
    - data preprocessing, 289–290
    - problem definition, 286–287
- Magellan, 119
- Matlab, 6
- Matplotlib, 121–122
- Message Authentication Codes (MACs), 149–150, 150*f*
- Microsoft
  - HDInsight, 123
  - Power BI, 121–122
- Mild outliers, 36
- Miniaturizing visual glyphs, 174–175
- MongoDB, 105
- MOST vulnerabilities, 138–139
- MPP (Massively Parallel Processing) databases, 125
- Mtcars, 36
- Multicube, 48
- Multidimensional scaling (MDS), 173–174
- Multivariate analysis, 253–256
  - correlation values, 254–255
- Multivariate regression, 292
- MySQL database, 105

## N

- Naturalistic driving study, 230–231
- Natural language processing tools, 6
- Natural Language Understanding (NLU) systems, 60–61
- Near-crash events/extreme driving behaviors, 234
- Network and protocol vulnerability analysis, 138–140
  - antitheft system vulnerabilities, 139
  - CAN vulnerabilities, 138
  - LIN vulnerabilities, 138

Network and protocol vulnerability analysis (*Continued*)  
 MOST vulnerabilities, 138–139  
 TPMS vulnerabilities, 139  
 VANETs vulnerabilities, 139  
 WPAN vulnerabilities, 139  
 Neural network models for crash data modeling, 225  
   fully connected multilayer feed-forward, 226*f*  
 Neural networks (NNs), 52, 122, 295–297  
 Non-linear feedback shift register (NLFSR), 141  
 Normal distribution, 36–37  
 NoSQL database, 105

## O

On Board Equipment (OBE), 202  
 On-Board Diagnostic (OBD) system, 233  
 Online analytical processing (OLAP), 31, 46–47  
   cubes, 48–49  
   functions, 46–47  
   hybrid OLAP (HOLAP), 51  
   multidimensional OLAP (MOLAP), 51  
   relational OLAP (ROLAP), 51  
   servers, 51  
   specialized SQL DBMS, 51  
 Online transaction processing (OLTP) applications, 45, 47  
 Open Archives Initiative Object Reuse and Exchange (OAI-ORE), 107  
 Open Systems Interconnection network model, 2–3  
 Oracle, 45–46, 119  
 OTA ECU firmware update, 155–157, 155*f*  
   availability, 155  
   code confidentiality, 155  
   code integrity, 155  
   code origin authenticity, 155  
   command freshness, 155  
   Hardware Security Module (HSM), 156  
   message source authenticity, 155  
   Remote Diagnosis phase, 157  
   secure protocol for, 156, 156*f*  
   update metadata confidentiality, 155  
 Outlier, 36  
   detection, 55

## P

Pearson's product-moment correlation coefficient, 252  
   properties, 252–253  
 Photogrammetry, 4  
 Pig, 61  
 Pig Latin, 61  
 Polynomial regression, 292  
 Postgres, 119  
 Predictive analytics, 32–33, 43–44  
   correlation coefficient, 43–44

  use cases, 44  
 Prescriptive analytics, 32–33, 45  
 Principal component analysis (PCA), 173–174  
 Privacy, 9  
 Probability density function (PDF), 39–40  
 Public clouds, 123  
 Python, 6, 62, 69, 105, 121–122  
   data ecosystem, 122

## Q

Quadratic model, 292  
 Quantile, 36  
 Quantile–quantile (Q–Q) plot, 36–37, 37*f*  
 Quartiles, 36  
 Queensland Hospital Admitted Patient Data Collection (QHAPDC), 40  
 QuickStarts, 62

## R

R programming, 70–71  
   big data processing, 87–89  
   Chi-squared test, 246  
   data frames and list, 72–75  
   fit of the data against a log-normal distribution,  
     determining, 246  
   Graphical User Interface (GUI), 71*f*  
   importing data from external files, 75–84  
     delimited file, 75–78  
     SQL (Structured Query Language) files, 83–84  
     XML files, 78–82  
   multiple linear regression model, 255–256  
   platforms for large-scale social media data, 84–87  
     dynamic live streaming, 86–87  
     package for static search, 85–86  
   RStudio, 70, 71*f*  
 Random access queries, 121  
 Real-time, speech-to-speech translation  
   systems, 63  
 Regression coefficients or weights, 291–292  
 Regression methods, 290–293  
   choosing suitable, 293  
 Representational State Transfer (REST)  
   architecture, 105  
 Research data exchange (DRE), 72  
 Residual sum of squares (RSS), 291–292  
 Resilient Distributed Datasets (RDD), 117–119  
 Resource drains, 33  
 Roadside units (RSUs), 3–4  
 Roadway data, 229–230  
 Roadway data collection technologies, 3  
 Ross, Thomas, 283  
 Round Table Automated Driving (RTAD), 23

**S**

- Safety analysis methods
  - issues and future directions, 233–235
  - statistical methods, 221–225
    - categorical data modeling, 222–225
    - count data modeling, 221–222
    - Generalized Additive Models (GAMs), 222
    - Generalized Linear Models (GLMs), 221–222
    - latent class logit (LCL) model, 223
    - standard MNL model and mixed MNL model, 224–225
- Safety applications of data and data analytics
  - big and heterogeneous data for safety, 219
  - commercial vehicle safety, 218
  - connected vehicles and traffic safety, 220
  - crash count/frequency modeling, 216–217
  - crash injury severity modeling, 217–218
  - data driven highway patrol plan, 218–219
  - effectiveness of safety countermeasures, 217
  - human factors, 215–216
  - real-time traffic operation and safety monitoring, 219–220
- Safety data, 227–233
  - big data and open data initiatives, 231–233
  - crash data, 228
  - naturalistic driving study, 230–231
  - NDS data for addressing safety issues, 232
  - roadway data, 229–230
  - traffic data, 228–229
  - vehicle and driver data, 230
  - weather data, 230
- Safety Pilot Model Deployment (SPMD), 220, 231
- Samuel, Arthur, 283
- Santayana, George, 93
- Satisfiability Modulo Theorem (SMT), 200
- Scalable data processing, 117–119
- Scatter plot matrix, 37–39, 38*f*, 43
- Scientific visualization, 165–166
- Scikit-learn machine learning package, 58
- Seaborn, 121–122
- Search-based analytics, 121
- Security challenges in automobile industry, 131–132
- SeeSoft system, 175
- Shneiderman's task taxonomy system, 171
- Short-running data management, 121
- Simple linear regression, 43–44, 249
- Skewness, 33–34
- Smart phone Apps, 134–135
- Social media data in transportation, 263–264
  - application, 270–272
    - traffic information dissemination, 272
    - traffic management during planned events, 271
    - traffic management during unplanned events, 271–272
    - traffic prediction, 270–271
    - transportation planning projects, 270
  - characteristics, 264–267
    - value, 266–267
    - variety, 266
    - veracity, 266
    - volume and velocity, 265–266
  - data analysis, 267–270
  - future research issues/challenges, 272–277
  - supplemental traffic data source, 273–277
  - supplemental transportation data source, 272–273
- Society of Automotive Engineers (SAE) standard, 195
- Software Defined Networking (SDN), 158
- Software metrics and measurements, 33
- Space-Filling Multidimensional Data Visualization (SFMDVis), 41
- Spark, 61, 117–119, 119*f*, 121–123
  - dataframes, 120–121
  - MLlib, 121
  - SQL, 120
  - Streaming, 120
- SQL (Structured Query Language) analytics, 46–47, 83–84, 120–121
- Static database schema analysis, 57
- Statistical and spatiotemporal analysis tools, 6
- Stop words, 53
- Storm, 61
- Streaming engines, 120
- Streaming infrastructure for deployment, 114–115
- Stream-processing engines (SPE)
  - data provenance, 277
  - dynamic provision of computing resources from distributed locations, 277
  - evaluation and selection, 276
  - integration of infrastructure elasticity, 276
  - integration of multiple, 276
- Student Success System (S3), 42
- Success index, 42
- Supervised learning, 284–285
  - classification, 285
  - regression, 285
- Support vector machines (SVMs), 297–298
- Symmetric cryptography, 148
- Systems Development V Model, 192–194
  - construction, 194
  - operations and maintenance, 194
  - plans, specifications, and estimates, 193–194
  - preliminary engineering, 193
  - project closeout, 194
  - project initiation, 193
- Systems engineers, 191–192, 192*f*

**T**

- Teaching analytics, 42–43
- Telematics Control Units (TCUs), 133–134

Temperature sensors, 104  
 TensorFlow, 60–61  
 Term frequency (tf), 53  
 Text Insight via Automated, Responsive Analysis (TIARA), 40  
 Traffic data, 228–229  
 Traffic management center (TMC), 6  
 Traffic monitoring, 4  
 Transmission control protocol (TCP), 2–3  
 Transport Layer Security (TLS) for communication sessions, 158  
 Traveler-based data collection technology, 3  
 Tripwire, 150–151

## U

UC Irvine Machine Learning Repository, 60  
 United States DOT (USDOT), 3–4, 6, 63–64, 72, 167, 220  
 Univariate analysis, 242–247  
 Unsupervised learning, 285–286
 

- association, 286
- clustering, 285–286

 US National Environmental Satellite, Data, and Information Service (NESDIS), 94  
 US national highway system, long-haul trucking traffic data, 7f, 8f

## V

VANETs
 

- applications, 135

security requirements, 153–154
 

- DSRC, 154
- Efficient Conditional Privacy Preservation Protocol (ECPP), 154
- k*-anonymity, 154
- vulnerabilities, 139
  - attacks, 145, 145f, 146f

 Vehicle and driver data, 230  
 Vehicle platooning, 134–135  
 Vehicle-based data collection technologies, 3–4  
 Vehicle–driver interface design, 235  
 Vehicle-to-vehicle (V2V) communication, 18, 139  
 Vehicle-to-vehicle (V2V) infrastructure, 114  
 Visual analytics, 35–36, 53–54
 

- software products, 54
- techniques, 166

 Visualization cube, 40

## W

Whiskers, 36  
 Wide area data collection technology, 3  
 Win-loss chart, 42  
 Wireless internal networks, 137–138  
 Workflow, 103  
 World Atlas of Language Structures (WALS), 60

## X

XML files, 78–82