



How to Architect and Build a Machine Learning Solution

Organizations are looking at technologies like machine learning, IoT, and Big Data to be more relevant in the market and attract customers in the competitive age of technologies, process, and the war of talent. However, the only way to achieve their goals is to provide effective products and respond to market needs in a timely manner. In order to effectively and efficiently take advantage of these technologies, an organization must have a strong foundational infrastructure to store data, execute analytical jobs, and defend its data assets from unforeseen modification or compromise.

Cloud-based infrastructures provide organizations with a flexible platform for data storage, management, and processing. They are also easily scalable (horizontally and vertically) based on need. However, to build a good machine learning solution, the requirements, needs, vision, and practical specific use cases must be clear. A well-designed cloud-based infrastructure can easily add new capabilities.

In the recent times, machine learning implementation has become less expensive through the use of cloud infrastructure. Therefore, chances were high that technology would be misapplied. Cloud providers often highlighted that machine learning would provide companies with huge benefits, but this is a subjective statement and depends on multiple factors. Therefore, making a proper, rational decision is the key. The value and benefit of a machine learning solution will not be realized if it is applied to systems where it is not required. For instance, if it is applied to the system where prediction capabilities are not required at all or appropriate data is not available.

Generally, a machine learning solution should not be implemented to replace existing systems or replace an existing data store. However, to build a machine learning solution, organizations need a solid data strategy, infrastructure, architecture, and workflows. These help ensure the high-quality data availability across organizations (LOB, UNITS), which are linked to rapid analysis and do not expose the organization to risk through data compromise. This also helps them meet compliance challenges.

Here are the common steps an organization must take before and after kicking off machine learning projects. These must be addressed for an effective and efficient machine learning project implementation:

- Define the scope.
- Review the existing system and data governance structures.
- Revisit security policies and tune them per current needs.
- Generate a map of existing workflow and relationships between applications and data stores.
- Create infrastructure, process, and technology maps and highlight strong and weak relations among them.
- Define the business vision for machine learning.
- Determine the suitability of technologies for the solution, define business drivers, call out risks, and establish business cases.
- Define migration strategies for the application, infrastructures, and services.
- Secure the budget and resources.
- Define the operational model.
- Create an implementation roadmap.
- Determine the implementation plan.
- Implement it.
- Evaluate the implementation.
- Incorporate feedback iteratively.
- Modify and fine-tune the strategies based on the analysis of feedback.
- Follow the usual cycle.

The primary value of Big Data and machine learning is to bring flexibility and insight through analysis of complex data sets. Many associated technologies will be part of this suite, including predictive analytics tools, IOT, and cloud technologies, as well as data modeling, data quality, and cognitive computing frameworks. However, the most important step of any analytical workflow is an efficient data process.

The following steps should generally be followed to convert data to insight to its consumption. Following these steps involves confirming that high quality data is considered. Also, these would help data scientists analyze the prepared data in the correct manner.

These steps help to convert raw data to insight:

1. **Gather:** Gathering data from heterogeneous sources.
2. **Discover relationship:** Discover the relationship in the existing data set.
3. **Organize:** Organize and then reorganize the data set for efficient and effective utilization.
4. **Analyze:** Identify and analyze the relationship.
5. **Generate insight:** Generate insight after the analysis of data.
6. **Report:** Report the insight in user understandable format.
7. **Consume:** Consume the insight for the business purpose.

Architectural Considerations

The technical architecture for a machine learning solution must be built around need and requirements. Therefore, while designing the machine learning solution, the key design considerations must align with these factors:

- **Know your need:** Identify your need and define the use cases accordingly. This will enable proper prioritization of work, selection of technology, scalability considerations, seamless systems, and data integration.
- **Define operational strategies:** Define the operational strategies beforehand in order to execute the solution effectively.
- **Be optimistic about scalability and performance:** While creating machine learning solutions, an optimistic approach is required. You have to believe as time passes that your organization will grow and evolve. Therefore, while you are designing the solution or architecture, your platform must be capable of incorporating the growing demand of data and must take care of analysis without much change. It must be able to handle data expansion effortlessly.
- **Strong data access and retrieval strategies are key:** Easy to understand interfaces, effective tools for storing data, optimized platforms, and excellent handling capabilities of unstructured data are some of the parameters that must be considered during the design. These parameters facilitate efficient and effective ingestion and processing of data.

- Security controls, logging, and auditing: Security is a key consideration for machine learning solutions. Identity management, auditing, and access controls must be designed to cater to the risk levels of the organization and be efficient enough to handle compliance needs. Access control implementation must be consistent between access methods.
- DevOps and Analytics Ops (refer to Chapter 7 for details on these concepts): Incredible operational value comes from storing and processing heterogeneous sources of data in a machine learning solution, especially in the cloud environment. Therefore, it is difficult to manage them manually. Hence, the preferred solution is to automate deployment and recovery. This would lower the operational problem on the IT team when making changes and responding to incidents.
- Be synchronized with advanced capabilities: Implementing advanced capabilities, like APIs, parallelization, and cognitive capabilities is the key to an effective machine learning solution. Therefore, synchronization with the latest advances is a must.

Cloud Adoption of a Machine Learning Solution

When organizations are planning to put their machine learning solution on the cloud, they need to adopt a few specific steps and strategies. In cloud-based environments, ensuring environment availability, reliability, and scalability becomes important. It is more relevant in a Big Data/machine learning scenario because of the real-time demand for insight. Therefore, to provide analytics and machine learning as a service, this is a high priority.

Blueprinting is a very important activity for designing any architecture. Therefore, a brief description it is provided in the next section. However, these are high-level views. If you need more detail, consider dedicated literature for the appropriate setup.

Blueprinting and Machine Learning Projects

Blueprinting an IT project involves the following:

- Brainstorm with business and technology stakeholders to elaborate, align, and document the scope of the stated business problem
- Translate this problem further to high-level requirements.
- Come up with various possible technical solutions (including high-level costs and timelines).
- Align and secure sign-off with the best solution.

Subsequent to this exercise, the budget is secured for the implementation project and a project team is identified and installed. The team refers to the blueprinting documents as high-level requirements that furnish the technical requirements, followed by subsequent SDLC phases.

It is extremely important to spend time in the brainstorming sessions to understand the business problem in detail from the business stakeholders.

The business stakeholders have a tendency to state the problem at a very high level. Create a questionnaire to probe for the necessary details. That questionnaire should cover aspects such as:

- What is/are the base issue(s)?
- How is it impacting the business or performance of duties?
- Are there any financial impacts of this issue?
- Is the stated functionality actually going to solve the base issue or create new ones in the future?
- Has the business tried to solve these issues before?
- If yes, what were the shortcomings of the previous solution?

For a successful blueprinting exercise, it's imperative to include all the relevant stakeholders. Here is a typical list of stakeholders and the area of contribution.

Individuals/Groups	Contribution
Business teams	Sign off on the requirements and the chosen solution.
Architect/Architectural group	Sign off on proposed solution compliance to organization technology architecture.
Designer/design group	Provide input on current state of the system design and contribute to proposed solutions.
Project manager	Manage budget, resourcing, talent, timeliness, and coordination of the project.
Technology manager	Provide roadmap of the technology to the project. Interact with the architect/architecture groups to make sure technical smoothness of the project.
Business analysts	Help to understand current business logic and process and give the current state and data flow.



A Holistic Machine Learning and Agile-Based Software Methodology

Workforces in the IT/IS industry are comprised of humans and their complex social, moral, emotional, and spiritual behaviors. When employees come to work, they bring their psychological state of minds with them. This affects their interactions with the stakeholders and other activities at work. Having high-level technical skills do not automatically mean high performance, unless the team members are committed, motivated, and enjoy their work. Therefore, organizations and in turn managers and leaders have to understand the importance of team members' psychology and behavior at the workplace, which includes their stress and conflict-handling skills.

Employee communication and collaboration abilities in the workplace (especially in the IT industry) play a critical role in project success. Alignment of employee values, skills, competencies, and goals with the organization and with clients is important for good results. Also, proper understanding of the philosophy of technical methodologies is one of the factors for achieving excellence in the deliverables.

Many academic studies and professional viewpoints have recognized the significance of managing emotions (emotional intelligence, EI) in the workplace. The ability to navigate and facilitate social relationships (social intelligence, SI) and the ability to apply universal principles to one's values and actions (moral intelligence, MI) toward workplace success and organizational efficiency and effectiveness has gained recognition in professional cycles (for details, refer to Chapter 8).

Several emotional intelligence training programs have been established to tune employees to the organizational culture and vision, but no framework has been developed (so far) to address employees' emotional, social, moral, and spiritual competencies holistically.

The rationale behind this appendix is to underline the importance and application of emotional, social, moral, and spiritual intelligence in the software industry and with technology companies. It also highlights how these competencies can contribute positively toward an organization's quality of service and deliverables, which will ultimately lead to organizational success.

It also discusses an integrative software development methodology based on holistic intelligence (= IQ+ SQ+ MQ+ EQ+ Social intelligence+ ingredients of positive psychology). Technologies are changing dynamically. Every day, something big is happening in the technical space. But the software methodologies have not changed over the last few decades. They are forced to fit to the existing ones with the small changes to accommodate new generations of technical project demand.

The Goal

This appendix presents a software methodology that's based on the Agile software development principles and psychological concepts. The proposed methodology is effective in designing, executing, and testing highly technical, complex projects, which often need with motivation, focus, and commitment from all levels, starting from programmers to clients to leaders. Big Data and machine learning projects fall under this umbrella (however, they could be applicable to any project). This framework uses integrative and innovative methods and models to diffuse different leadership techniques with information technology (IT) and software development.

Proposed Software Process and Model

Agile is a fundamental change in how people manage projects if it is compared to waterfall techniques. Delivering workable software on time is what actually measures the success of a project. However, software delivery is not enough. The quality of the delivered project, stakeholder satisfaction, and workforce happiness are also very important. If the workforce is happy, they will be more productive, creative, and innovative and therefore be able to create more robust solutions. Existing software methodologies typically cater to the few administrative and technical success factors of software development projects, like timely delivery of workable software and keeping projects within budgets. Often, behavioral and psychological success factors are not being accommodated in the project in any considerable way.

Thus, the overall existing process, including the software methodologies, need to change the fundamental way of thinking. Organizations have to change or fine-tune the old management style by making it more dynamic and customizable, per project need and demand. One standard management style across the organization will not be of much use. Customized and personalized management styles on a project basis are the need of the hour. The behavior of the programmers—their psychology and skills—needs to be adjusted or modified based on the need of the project. This can be done by using machine learning, Big Data, and cognitive computing techniques.

The machine learning analytics based software development methodology is capable of bringing dynamism and agility to the development methodology itself with continuous learning and adaptability. The expectation is that the model must learn from the continuous changing behavior of the people who are the part of the project and in turn with the organization. The model can analyze all the data available in the organization in all forms and suggest the appropriate customized software process and methodologies based on the availability of skills, processes, technologies, and people.

This includes charting out the guidelines for planning, developing, and executing in an automated way with no or very little human intervention.

The model will also suggest the appropriate customized software methodology based on the requirements. For example, the machine learning and predictive analytics techniques will enable the model to evaluate and analyze the requirements, including inputs from similar projects that the organization delivered in the past. The model can study the scope of work by using SOW documents, accessibility of the self-components, present talent availability, time to deliver the project, character, and behavioral analysis of the stakeholders/team members, and then determine that the project needs a mix of Agile, waterfall, and Kanban methodologies. Then it would generate customized methodologies with the guidelines for implementation, process, and other action items. It then can predict its success percentage on the solid foundation of data, statistical techniques, and analytics.

Problem State

Most software methodologies are based on mathematical and statistical techniques and not on the people factor. Therefore, several problems occur. Although all methodologies are implemented by humans, they do not consider the psychological state of the human being on work during the overall project lifecycle. It is well known that overall productivity and project success depend heavily on the psychological well-being of the individuals and of the group. For any project to be successful, we have to take into consideration the human factors along with the technical factors.

Solution

Holistic intelligence maps the complete personality of humans. People having holistic intelligence can communicate with the outer world effectively and efficiently, and also with their inner world (their own minds) with confidence. The proposed software development methodology is the integration of holistic intelligence with Agile methodology.

Holistic intelligence is applicable everywhere—during the requirements gathering, design, testing, and implementation phases of software development. It deals with humans and their psychological states and human intervention and involvement exists at all phases of the software development lifecycle. Therefore, to make the methodology more robust and efficient, there is a need for holistic thinking, which includes technical, behavioral, and intellectual skills. Proper application of this proposed methodology will increase the knowledge workers' productivity. For example, in the process of implementing software, the programmers' emotional and social competencies become important, because if these competencies are present in the programmers, it will help them better interact with team members, clients, and other stakeholders.

The humans/programmer/knowledge workers are not necessarily equipped with all these competencies. You must be aware of all of their hidden potential at the start of the project. So, once the team is onboarded with the bare minimum qualifications, the desired competencies should be incorporated in the progress of the project. This can be done in an iterative way, whereby the core competencies are cultivated first in the people involved and the rest later... exactly like the Agile methodology.

As projects proceed, training is added to the cycle to improve the holistic intelligence parameters of employees, resulting in more efficient and well-prepared individuals. The training program will be done again in an iterative way: training ► evaluation ► performance ► feedback ► more training. That way, the right feedback can correctly assess individuals' progress and competency without having an adverse effect on the project.

Working

This methodology can be technically implemented in multiple ways. One method of applying it is described next.

The Process

When companies or organizations recruit employees, they ask for all the information to determine if the prospective employee is the right person for the job. The information ranges from grades to professional experience. In time-based projects, it is very important to pick the right set of people; hence, this information matters because it provides insight on the person's technical expertise. These information in turn could be mapped to the professional expertise of the resource to the specific needs of machine learning, cognitive computing, and Big Data analytics-based projects.

Also based on these techniques, a model could be created that provides appropriate suggestions based on the project requirements. This "insight" is generated on the basis of information stored in the database that defines an individual's intellectual capability. It also helps an organization pick people who are suitable for projects in the centralized way.

Many organizations have similar databases; however, they do not contain all the behavioral data of an individual. This ranges from individual details scattered at multiple places, including social media and other unstructured data. This need justifies the importance of integrative database or "organizational professional database". The "integrative database" or "organizational performance database" will contain all the content of the employee/ perceptive employee to the extent of behavioral history, personality type, strength and weakness, and emotional and spiritual parameters.

In Agile methodology, when teams develop software, they use pair programming for coding. Pair programming is a software development technique in which two programmers work closely together at one workstation. The driver transcribes code while the navigator analyzes, accesses, and reviews each line of code as it is typed. In this way, while reviewing, the navigator can deliberate the strategic path to the coding work. The navigator also comes up with ideas for improvements and provides insights on future glitches to address. The driver's main responsibilities are to focus on the "tactical" aspects of implementation of the present task. In this process, the driver generally uses the navigator as a safety net and guide.

The integrative database can be helpful in this type of situation, in order to pick the right resources that would play these roles based on the gathered insights on the data stored. Picking the right resources will result in cooperative and more efficient teams, because team members are selected on the basis of behavioral patterns. This is just one example of applying the database to many development methodologies.

Relevance and Future Direction of the Model

No attempts have been made to date to combine social and spiritual intelligence with Agile software development methodologies. These factors (social and spiritual intelligence) help develop one model/framework and innovative methods/calculations, which will provide a new dimension to software development techniques. To improve this methodology, multiple tools need to be developed. A few are mentioned in Figure B-1.

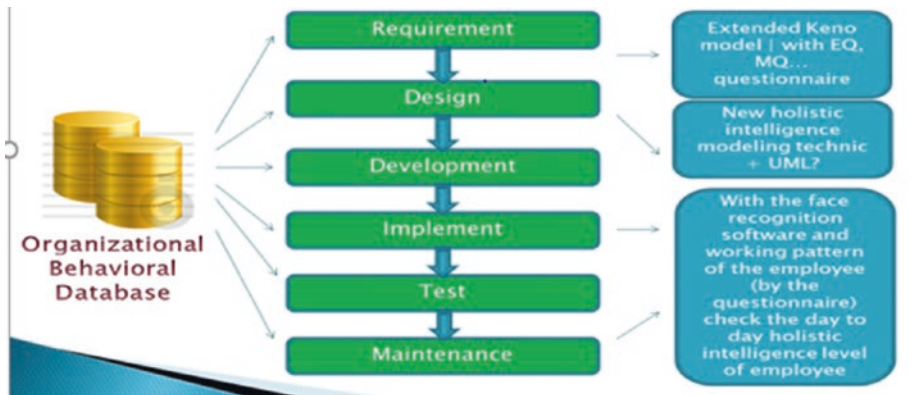


Figure B-1. Combining social and spiritual intelligence with Agile software development methodologies

The best way to change problematic behavior is to understand the issues that drive it and then design a plan to combat it. For example, violent conduct may be motivated by distress and anxiety, by cluelessness, or by a wish to govern and control team members. The workplace should be a place of peace and positivity in order to enable quality work and innovation. Teams/individuals are supposed to contribute toward and maintain that peace. One of the prime reasons for emotional imbalance, including aggression, is fear and insecurity.

These toxic behaviors can be controlled if managed with tolerance and reassurance. Unmanaged emotions worsen all types of problems. Therefore, better understanding of emotions and social behavior is required even in the workplace. The better you comprehend how other individuals see the world and what inspires them, the better you will be able to motivate them to perform in helpful ways. The more you identify, analyze, and understand what inspires individuals of different personality types, the better you will be able to shield yourself and inspire them to collaborate with you and provide the work per the requirements. Toxic behaviors generally create problems in software development environments/teams, because ultimately software development is a team activity. Individual-based projects generally fail. The chances of success of team-based projects are high, no matter how complex they are.

This proposed model can be helpful in understanding difficult people and then correcting their behaviors. The goal is to align them with the organization's vision, mission, and propose. Ultimately, this harmony will percolate down to the project level.

APPENDIX C



Data Processing Technologies

Multiple tools and technologies for data processing are described under the Big Data analytics technology stack in Chapter 4. However, many other technologies exist as well, and it is good to know about them for a complete understanding. Covering all the tools is beyond the scope of this book, but the tables in this appendix provide brief explanations of some of them.

Table C-1. *Data Gathering and Processing Tools and Technologies*

Purpose	Tools
Data integration • Enterprise data warehouse • Enterprise reporting	SAP, BDOS Oracle DI, IBM Data Stage, IBM Information Analyzer, Business Glossary, IBM MDM, DB2
Data modeling • MDM data conversion/migration	Trillium, D3JS, IDL, Riverstand, TIBCO Spotfire, 17 Tableau, SSRS, SSAS, PowerTools
Data profiling • Data quality • Data governance	Erwin, ER/Studio, Visio, DB Designer
Data integration • Enterprise data warehouse • Enterprise reporting	SAP, BDOS Oracle DI, IBM Data Stage, IBM Information Analyzer, Business Glossary, IBM MDM, DB2
Text mining • Data streaming • Complex event processing	SAP MDM, Oracle MDM, 18 Oracle Essbase, 19 IBM, COGNOS, MicroStrategy, SAP BO, Oracle OBIEE
Big Data • Social media	SAP HANA, Oracle Exadata, Oracle Times 10, Terradata, IBM Netezza, HP

Table C-2. *Analytical Tools and Technologies*

Purpose	Tools
Cluster analysis • Statistical testing, such as t-test, chi-square, and ANOVA • Latent class analysis • Discriminant analysis	SAS, SPSS, R, Python, Knex, Knime, Weka, MiniTab, Mahout, ilog, Matlab, Statistica, Evolve
Descriptive analytics • Univariate analytics • Bivariate analytics • Multivariate analytics	Hadoop, HBase, Hive, Spark, Storm, Splunk, Pig, Oozie
Monte Carlo analysis • Conjoint analysis • Retention analysis • Survey analysis • Lifetime value analysis	MondoDB, CouchDB, Neo4J, Infinitie, MarkLogic, Amazon Dynamo, TITAN
Campaign analysis • Pricing analysis • Survival analysis • Pareto analysis • Quality control analysis	UIMA, Rapid Miner, Tresseract Café, Brandwatch, Crimson Hexagon, Radian6, Symosis, Lithium
Chaid analysis • Regression analysis • Decision trees • Neural network	Apache, Cloudera, 16 Hortonworks, MapR, IBM, Cassandra, Hypertable, Amazon

Bibliography

Machine Learning, Big Data, IoT, and Cloud and Cognitive Computing:

1. *Data Science for Dummies*, 2nd Edition.
2. Olsen, Dan. *The Lean Product Playbook: How to Innovate with Minimum Viable Products and Rapid Customer Feedback*. Wiley. 2015.
3. Newport, Cal. *Deep Work: Rules for Focused Success in a Distracted World*. Grand Central. 2016.
4. Christensen, Clayton M., Allworth, James, & Dillon, Karen. *How Will You Measure Your Life*. Harper Business. 2012.
5. Xamarin Continuous Integration and Delivery
6. Schwab, Klaus. "The Fourth Industrial Revolution". 2016.
7. Gertner, Jon. *The Idea Factory: Bell Labs and the Great Age of American Innovation*. Penguin Press HC. 2012.
8. Christensen, Clayton M., Allworth, James, & Dillon, Karen. *Competing Against Luck: The Story of Innovation and Customer Choice*. HarperBusiness. 2016.
9. Stigler, Stephen M. *The History of Statistics: The Measurement of Uncertainty Before 1900*. Belknap Press. 1986.
10. Stigler, Stephen M. *The Seven Pillars of Statistical Wisdom*. Harvard University Press. 2016.
11. Akerlof, George A. & Shiller, Robert J. *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton University Press. 2015.
12. Akerlof, George A., Kranton, Rachel E. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. 2010.
13. Ellenberg, Jordan. *How Not to Be Wrong, The Power of Mathematical Thinking*. Penguin Press HC. 2014.

14. Tapscott, Don & Tapscott, Alex. *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Portfolio. 2016.
15. Tetlock, Philip E. & Gardner, Dan. *Superforecasting: The Art and Science of Prediction*. Crown. 2015.
16. Ford, Martin. *The Lights in the Tunnel Automation, Accelerating Technology and the Economy of the Future*. CreateSpace. 2009.
17. Ford, Martin. *Rise of the Robots Technology and the Threat of a Jobless Future*. Basic Books. 2015.
18. Ross, Alec. *The Industries of the Future*. Simon & Schuster. 2016.
19. Rogers, David L. *The Digital Transformation Playbook: Rethink Your Business for the Digital Age*. Columbia University Press. 2016.
20. Hay, Louise. *Heal Your Body*. Hay House, Inc. 1976.
21. Dispenza, Dr. Joe. *You Are the Placebo_ Making Your Mind Matter*. Hay House, Inc. 2014.
22. Sims, Christopher A. *Tech Anxiety Artificial Intelligence and Ontological Awakening in Four Science Fiction Novels*
23. Starkes, Janet L., Ericsson K. Anders. *Expert Performance in Sports*
24. Ericsson, K. Anders. *Development of Professional Expertise Toward Measurement of Expert Performance and Design of Optimal Learning Environments*
25. Eagleman, David. *The Brain: The Story of You*.
26. Sauro, Jeff. *Customer Analytics For Dummies*.
27. Laursen, Gert H. N. *Business Analytics for Sales and Marketing Managers How to Compete in the Information Age*
28. Artun, Omer; Levin, Dominique. *Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data*.
29. Magy Seif El, Nasr Ph.D., Drachen, Anders Ph.D., Canossa, Alessandro Ph.D. *Game Analytics Maximizing the Value of Pl*.
30. McNeill, Dwight. *A Framework for Applying Analytics in Healthcare What Can Be Learned from the Best Practices in Retail, Banking, Politics, and Sports*.

31. Holtz, Shel, Havens, John C. *Tactical Transparency How Leaders Can Leverage Social Media to Maximize Value and Build their Brand*. J-B International Association of Business Communication.
32. Havens, John. *Hacking Happiness Why Your Personal Data Counts and How Tracking it Can Change the World*.
33. *Enterprise Cloud Strategy*. Microsoft Press ebook.
34. Oram, Andy, Wilson, Greg. *Making Software: What Really Works, and Why We Believe It*.
35. Weil, David. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done To Improve It*.
36. *Scala for Machine Learning*
37. *Spark for Data Science*
38. *DevOps on the Microsoft Stack*
39. *Clojure for Data Science*
40. *Building Arduino Projects for the Internet of Things*. 2016.
41. *Pro SQL Server on Microsoft Azure*
42. *Analyzing the Analyzers*
43. *Softimage Towards a New Theory of the Digital Image*
44. *When Google Met WikiLeaks*
45. *The Intelligent Web*
46. *Machine Learning For Dummies*
47. *Social Machines*
48. Gutierrez, Sebastian. *Data Scientists at Work*.
49. Paul C. Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, David Corrigan, James Giles. *Harness the Power of Big Data: The IBM Big Data Platform*. TMH. 2013.
50. Xin Luna Dong, Divesh Srivastava. *Big Data Integration*. Morgan & Claypool Publishers. 2015.
51. Schmarzo, Bill. *Big Data MBA*. Wiley. 2016.
52. Sinha, Sudhi. *Making Big Data Work for Your Business*. Impact Publishing. 2014.
53. Paul, Jean Isson, Harriott, Jesse S. *People Analytics in the Era of Big Data*. Wiley. 2016.

54. Dacosta, Francis. *Rethinking the Internet of Things*. Apress Open. 2016.
55. Kelly, Kevin. *The Inevitable Understanding the 12 Technological Forces That Will Shape Our Future*. Viking. 2016.
56. Ross, Alec. *The Industries of the Future*. Simon & Schuster eBook. 2016.
57. Lyons, Dan. *Disrupted: My Misadventure in the Start-Up Bubble*. Hachette Books. 2016.
58. Duhigg, Charles. *Smarter Faster Better: The Secrets of Being Productive in Life and Business*. Penguin Random House LLC. 2016.
59. Franks, Bill. *Taming the Big Data Tidal Wave Finding Opportunities in Huge Data Streams with Advanced Analytics*. Wiley & SAS Business Series. 2012.
60. Bullard, Brittany. *Style & Statistics: The Art of Retail Analytics*. Wiley. 2016.
61. Rackley, Jerry. *Marketing Analytics Roadmap Methods, Metrics, and Tools*. Apress. 2015.
62. Dykes, Brent. *Web Analytics Action Hero Using Analysis to Gain Insight and Optimize Your Business*. Adobe Press. 2012.
63. Sironi, Paolo. *FinTech Innovation: From Robo-Advisors to Goal Based Investing and Gamification*. Wiley. 2016.
64. McGuire, Kelly A., Wood Jr., Dexter E. *The Analytic Hospitality Executive: Implementing Data Analytics in Hotels and Casinos*. Wiley. 2017.
65. Boobier, Tony. *Big Data and Analytics for Insurers*. Wiley. 2016.
66. Baesens, Bart, Van Vlasselaer, Veronique, Verbeke, Wouter. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. Wiley. 2015.
67. Sathi, Arvind. *Cognitive (Internet of) Things*. Palgrave Macmillan. 2016.
68. Sathi, Dr. Arvind. *Big Data Analytics: Disruptive Technologies for Changing the Game*. IBM Corporation. 2012.
69. Pries, Kim H., Dunnigan, Robert. *Big Data Analytics: A Practical Guide for Managers*. CRC Press. 2015.
70. Rice, Daniel M. *Calculus of Thought: Neuromorphic Logistic Regression in Cognitive Machines*. Elsevier. 2014.

71. Kelleher, John D., Mac Namee, Brian, Darcy, Aoife. *Fundamentals of Machine Learning for Predictive Data: Analytics Algorithms, Worked Examples, and Case Studies*. The MIT Press. 2015.
72. Winston, Wayne L. *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football*. Princeton University Press. 2009.
73. Camm, Jeffrey D., James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann, & David R. Anderson. *Essentials of Business Analytics*. Cengage Learning. 2014.
74. *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing, and Presenting Data*. EMC Education Services. Wiley. 2015.
75. Subhashini Sharma Tripathi. *Learn Business Analytics in Six Steps Using SAS and R*. Apress. 2016.
76. Manoochehri, Michael. *Data Just Right: Introduction to Large Scale Data & Analytics*. Addison-Wesley Professional. 2013.
77. Owen, Sean, Anil, Robin, Dunning, Ted, Friedman, Ellen. *Mahout in Action*. Manning. 2012.
78. *Dunning, Ted, Friedman, Ellen, Real World Hadoop*. O'Really. 2015.
79. MacCormick, John, Bishop, Chris. *Nine Algorithms That Changed the Future The Ingenious Ideas That Drive Today's Computers*.
80. Alex Smola, S.V.N. Vishwanathan. *Introduction to Machine Learning*. Cambridge University Press. 2008.
81. Gilchrist, Alasdair. *Industry 4.0_ The Industrial Internet of Things*. Apress. 2016.
82. Brandewinder, Mathias. *Machine Learning Projects for .NET Developers*. Apress. 2016.
83. Shui Yu, editor. *Networking for Big Data*. CRC Press. 2016.
84. Morton, John. *Big Data Opportunities and Challenges*. BCS. 2016.
85. Van Rijmenam, Mark. *Think Bigger: Developing a Successful Big Data Strategy For Your Business*. American Management Association. 2014.
86. Thomas, Rob, McSharry, Patrick. *Big Data Revolution: What Farmers, Doctors, and Insurance Agents Teach Us About Discovering Big Data Patterns*.

87. Arthur, Lisa. *Big Data Marketing: Engage Your Customers More Effectively and Drive Value*. Wiley. 2013.
88. Shroff, Gautam. *The Intelligent Web Search, Smart Algorithms and Big Data*. Oxford University Press. 2013.
89. Glass, Russell, Callahan, Sean. *The Big Data Driven Business*. Wiley. 2015.
90. Mosco, Vincent. *To the Cloud: Big Data in a Turbulent World*. Paradigm Publishers. 2014.
91. Zikopoulos, Paul, Eaton, Chris. *Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data*. TMH. 2012.
92. Camm, Jeffrey D., James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann, & David R. Anderson. *Essentials of Business Analytics. Cengage Learning*. 2014.
93. Domingos, Pedro, Markov, Daniel Lowd. *Logic: An Interface Layer for Artificial Intelligence*. University of Washington. 2009.
94. Nazemoff, Valeh. *The Four Intelligences of the Business Mind: How to Rewire Your Brain and Your Business for Success*. Apress. 2014.
95. Raj, Pethuru. *Cloud Enterprise Architecture*. Auerbach Publications. CRC Press. 2012.
96. McGrath, Michael P. *Understanding PaaS*. O'Reilly Media. 2012.
97. McGrath, Michael J., Scanail, Cliodhna N_. *Sensor Technologies: Healthcare, Wellness and Environmental Applications*. Apress. 2013.
98. Larose, Daniel T., Larose, Chantal D. *Data Mining and Predictive Analytics*. Wiley. 2015.
99. Lohr, Steve. *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. HarperBusiness. 2015.
100. Brown, Meta S. *Data Mining For Dummies*. Wiley. 2014.
101. Barlow, Mike. *Learning to Love Data Science*. O'Reilly Media. 2015.
102. Hoberman, Steve. *Data Model Scorecard: Applying the Industry Standard on Data Model Quality*. Technics Publications. 2015.

103. Iafrate, Fernando. *From Big Data to Smart Data*. Wiley, ISTE. 2015.
 104. Gendron, Michael S. *Business Intelligence and the Cloud*. Wiley. 2014.
 105. Cobb, Charles G. *The Project Manager's Guide to Mastering Agile*. Wiley. 2015.
 106. S.P.T. Krisnan, Jose L Ugia Gonzalez. *Building Your Next Big Thing with Google Cloud Platform*. Apress. 2016.
 107. Berkun, Scott. *The Myths of Innovation*. O'Reilly Media. 2007.
 108. Spiliopoulou, Myra, Schmidt, Lars, Thieme, Ruth Janning. *Data Analysis, Machine Learning and Knowledge Discovery*. Springer. 2014.
 109. Mueller, John Paul, Massaron, Luca. *Machine Learning For Dummies*. Wiley. 2016.
 110. Peters, Hannah, Vogel, Mia, Editors. *Machine Learning Research Progress*. Nova Science Publishers, Inc. 2010.
- Team Building, EQ, SQ, Psychology, and So On:
111. *Scientific American*. March, 2015. USA Edition.
 112. Keller, Scott, Price, Colin. *Beyond Performance How Great Organizations Build Ultimate Competitive Advantage*.
 113. Chandler, Steve. *100 Ways to Motivate Yourself: Change Your Life Forever*.
 114. Fubini, David, Price, Colin, Maurizio Zollo. *Mergers Leadership, Performance and Corporate Health*. INSEAD Business Press.
 115. Chandler, Steve, Richardson, Scott. *100 Ways to Motivate Others, Third Edition: How Great Leaders Can Produce Insane Results Without Driving People Crazy*.
 116. Chandler, Steve. *100 Ways to Motivate Yourself*. 1996.
 117. Sapolsky, Robert M.. *Behave: The Biology of Humans at Our Best and Worst*. Penguin Press. 2017.
 118. Ericsson, Anders, Pool, Robert. *Peak Secrets from the New Science of Expertise*.
 119. Van Edwards, Vanessa. *The Science of Succeeding with People*.
 120. Davidowitz, Seth Stephens. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. Dey Street Books. 2017.

■ BIBLIOGRAPHY

121. Govindarajan, Vijay . *The Three-Box Solution: A Strategy for Leading Innovation*. Harvard Business Review Press. 2016.
122. Govindarajan, Vijay, Trimble, Chris. *The Other Side of Innovation Solving the Execution Challenge*.
123. Newberg, Andrew, Waldman, Mark Robert. *Words Can Change Your Brain: 12 Conversation Strategies to Build Trust, Resolve Conflict, and Increase Intimacy*.
124. Newberg, Andrew, M. D., Waldman, Mark Robert. *How God Changes Your Brain*.
125. Kotter, John P. *Leading Change*.
126. Morin, Amy. *13 Things Mentally Strong People Dont Do Take Back Your Power, Embrace Change, Face Your Fears, and Train Your Brain for Happiness and Success*.
127. Davies, Jacqueline, Kourdi, Jeremy. *The Truth About Talent: A Guide to Building a Dynamic Workforce, Realising Potential, and Helping People To Succeed*.
128. Kourdi, Jeremy. *Business Strategy: A Guide to Taking Your Business Forward*. The Economist.
129. Kourdi, Jeremy. *100 Great Business Ideas From Leading Companies Around the World*. 100 Great Ideas.
130. Jeff Dyer, Hal Gregersen, Clayton M. Christensen. *The Innovators DNA Mastering the Five Skills of Disruptive Innovators*.
131. William G. Dyer, W. Gibb, Jr. Dyer, Jeffrey H. Dyer Team. *Building Proven Strategies for Improving Team Performance*.
132. Friedman, Thomas L. *Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations*.

Companies whose whitepapers, reports, articles, websites, case studies, and other resources are used for the reference purposes:

cortexlogic.com	<i>GE</i>	<i>Apteco</i>	<i>Cignus Web Services</i>
<i>Genpact</i>	<i>4i</i>	<i>Artelnic</i>	<i>Cogent Data Analytics</i>
<i>PWC</i>	<i>Intel</i>	<i>AsiaAnalytics</i>	<i>Cognitro Analytics</i>
<i>Symphony</i>	<i>7Puentes</i>	<i>ASID Analytics</i>	<i>CogNova Technologies</i>
<i>Deloitte</i>	<i>AbsolutData</i>	<i>Austin</i>	<i>Computer Associates</i>
<i>Accenture</i>	<i>Acheron Analytics</i>	<i>BADR</i>	<i>Connolly Consultants LLC</i>
<i>TCS</i>	<i>Aiaioo Labs</i>	<i>Bayesia</i>	<i>Core Services</i>
<i>Wise.io</i>	<i>Rose Technology</i>	<i>Bentley University</i>	<i>comSysto GmbH</i>
<i>JPMorgan Chase & Co.</i>	<i>Algotlytics</i>	<i>Center for</i>	<i>Creascience</i>
<i>UnitedHealth Group</i>	<i>Algoritmica</i>	<i>Quantitative</i>	<i>CRM Software Solutions</i>
<i>Mphasis</i>	<i>Alpha Analysis</i>	<i>Analysis</i>	<i>CVM Data Sciences Inc</i>
<i>Infosys</i>	<i>Consulting</i>	<i>Beyond the Arc, Inc</i>	<i>CVM Marketing</i>
<i>Wipro</i>	<i>Alta Plana</i>	<i>BGStats Consulting</i>	<i>Custom Analytics</i>
<i>HCL</i>	<i>AltexSoft</i>	<i>Biclast solutions</i>	<i>Walmart</i>
<i>Cognizant</i>	<i>Altius</i>	<i>Big Data Landscape</i>	<i>Netflix</i>
<i>Mindtree</i>	<i>Analytica</i>	<i>Big Learning</i>	<i>Rolls-Royce</i>
<i>HappiestMind</i>	<i>Analytics Advisory</i>	<i>BitRefine group</i>	<i>Shell</i>
<i>Honeywell</i>	<i>Group</i>	<i>bitsnbrains</i>	<i>Apixio</i>
<i>Csc</i>	<i>Analytics Center</i>	<i>Bizooki</i>	<i>Facebook</i>
<i>ADP</i>	<i>Analytics Iowa LLC</i>	<i>BI Solutions</i>	<i>John Deere</i>
<i>Google</i>	<i>Analyze Partner</i>	<i>Blue Guava</i>	<i>RoyalBankofScotland</i>
<i>EY</i>	<i>Analytical People</i>	<i>Boire Filler Group</i>	<i>Axiom</i>
<i>Microsoft</i>	<i>Analytics Trainings</i>	<i>Brains and Business</i>	<i>Data and Analytic</i>
<i>IBM</i>	<i>Anderson Analytics</i>	<i>BRIDGEi2i</i>	<i>Solutions Inc.</i>
<i>HP</i>	<i>Anthem Marketing</i>	<i>BVA Data sciences</i>	<i>Data Cowboys</i>
<i>Citi</i>	<i>Solutions</i>	<i>Business Data</i>	<i>DecisionCraft Analytics</i>
<i>Dell</i>	<i>APower Solutions</i>	<i>Miners</i>	<i>Decision Systems</i>
<i>Data Learning</i>	<i>Data Shaping</i>	<i>BusinessMinds</i>	<i>Definitive Logic</i>
<i>Consultant</i>	<i>Solutions</i>	<i>C2G Partners</i>	<i>Consultants</i>
<i>Data Mind</i>	<i>DataInsight</i>	<i>CAC group</i>	<i>Dextro Analytics</i>
<i>Data Miners</i>	<i>DataLab</i>	<i>Cambio</i>	<i>Differentiated Analytics</i>
<i>Datalya Consulting</i>	<i>DataMine Lab</i>	<i>Technologies</i>	<i>Directing</i>
<i>Data Mining Experts</i>	<i>DataMineIt(SM)</i>	<i>Capgemini</i>	<i>Discovery Corps Inc.</i>
<i>Data Mining</i>	<i>Datamotus</i>	<i>Caserta Concepts</i>	<i>divis intelligent solutions</i>
<i>Technologies</i>	<i>Datascope</i>	<i>CBIG Consulting</i>	<i>DLabs</i>
<i>Data Science</i>	<i>Analytics</i>	<i>CDO Advisors</i>	<i>Nest</i>
<i>Partnership</i>	<i>DAYCHE Data</i>	<i>Centre of Excellence</i>	<i>Fitzgerald Analytics</i>
<i>Etsy</i>	<i>Mining Group</i>	<i>CIBIT</i>	<i>Forefront Economics</i>
<i>NarrativeScience</i>	<i>Airbnb</i>	<i>Chicago Data</i>	<i>Foris</i>
<i>BBC</i>	<i>Sprint</i>	<i>Strategy Group</i>	<i>Foviance</i>
<i>MiltonKeynes</i>	<i>Fitbit</i>	<i>Experian</i>	<i>Fractal Analytics</i>
	<i>Zynga</i>		
	<i>Autodesk</i>		

Equity Decision Systems	<i>eKPI solutions</i>	<i>Drug Analytics</i>	<i>FrontAnalytics</i>
Evolutionary Systems	<i>Elder Research</i>	<i>Dwinnell Consulting</i>	<i>Gmid Associates</i>
Exclusive Ore	<i>Elite Analytics</i>	<i>East Agile</i>	<i>Gray Matter Analytics</i>
Experfy	<i>Eleks provides</i>	<i>EDS Business</i>	<i>GroupVisual</i>
Exsilon Data & Statistical Solutions	<i>offshore</i>	<i>Intelligence Services</i>	<i>GSTAT</i>
Faunus Analytics	<i>ENQIO</i>	<i>EffectiveSoft</i>	<i>Haave</i>
InEdge	<i>Entopix</i>	<i>Engineering Consulting Services</i>	
InfoCentricity	<i>Epsilon</i>	<i>Haensel AMS</i>	
Analytic Services	<i>Isoft</i>	<i>Hamilton Numbers</i>	
Intended Motion	<i>CogniMem</i>	<i>Houston Analytics</i>	
Intelegencia	<i>Technologies</i>	<i>Hybridstat</i>	
Intelligent Business Systems	<i>Xtra IQ Inc.</i>	<i>Predictive Analytics</i>	
Intelligent Datalytics	<i>Friendly</i>	<i>Hummer Winblad</i>	
iThinkTank			

Name of the websites used extensively for the references (reports, whitepapers, and other resources):

- www.mckinsey.com
- <https://www.forbes.com>
- <https://www.linkedin.com>
- <https://home.kpmg.com>
- <https://artificialintelligencenow.com>

Additional links:

- <http://www.softwebsolutions.com/resources/posts.html?blogtype=insights>
- <http://www.bridgei2i.com/Campaigns/Insurance/BRIDGEi2i-Insurance-Case-study.pdf>
- http://www.iftf.org/fileadmin/user_upload/downloads/th/IFTF_TH_2015_AutomatedWorldMap.pdf
- <https://www.mu-sigma.com/sites/default/files/blog/Innovation%20in%20Decision%20Sciences.pdf>
- https://www.cipd.co.uk/Images/the-rise-of-hr-wisdom-from-73-thought-leaders_2015_tcm18-15549.pdf
- <https://www.greenbook.org/pdfs/2015GRITweb.pdf>
- <https://upxacademy.com/wp-content/uploads/2017/01/The-Ultimate-Guide-on-Big-Data-Analytics-Data-Science-2.pdf>

- <http://www.genpact.com/exclusive-content/The-rise-of-intelligent-operations-BFS.pdf>
- http://www.assima.net/downloads/presentations/IBM_SAP_Digital_Transformation_with_ASSIMA.pdf
- <http://info.deloitte.dk/rs/539-HCP-749/images/Robotics-%26-Cognitive-Automation-GNG-Apr-17.pdf>
- [https://static.rainfocus.com/oracle/oow16/session/1464872663994001BfKj/ppt/OOW%2716%20-%20Oracle%20Big%20Data%20Predictive%20Analytics%20and%20Machine%20Learning%20Strategy%20and%20Roadmap%20\[CON6500\]%20V2%20minus%20Futures.pdf](https://static.rainfocus.com/oracle/oow16/session/1464872663994001BfKj/ppt/OOW%2716%20-%20Oracle%20Big%20Data%20Predictive%20Analytics%20and%20Machine%20Learning%20Strategy%20and%20Roadmap%20[CON6500]%20V2%20minus%20Futures.pdf)
- <https://dattatreysindol.com/>
- <https://www.linkedin.com/pulse/sway-skype-yammer-delve-video-communication-tools-365-richardson>
- <https://www.mssqltips.com/sqlservertip/4604/overview-of-intelligence-pillar-in-the-cortana-intelligence-suite/>
- <https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/operations/lu-intelligent-automation-business-world.pdf>
- <https://static1.squarespace.com/static/562a32b0e4b0e6f4ec3104ae/t/58cef6c9be6594524c72c291/1489958608504/1650+Amir+Tabakovic.pptx.pdf>
- http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf
- <http://ece.eng.umanitoba.ca/undergraduate/ECE4850T02/Lecture%20Slides/MachineLearningIntro.pdf>
- <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>
- https://www.jefferies.com/CMSFiles/Jefferies.com/files/Insights/ChinaTech_PrepforNewTide.pdf
- <https://www.businessesgrow.com/free-slide-presentations/>
- <https://www.businessesgrow>
- http://cdn.chiefmarketer.com/wp-content/uploads/2013/11/CM_090113_Infographic1.pdf?_ga=1.107489986.1562322025.1491555186
- https://hbr.org/2016/11/how-to-tell-if-machine-learning-can-solve-your-business-problem&ab=Article-Links-End_of_Page_Recirculation

■ BIBLIOGRAPHY

- <http://data-informed.com/how-machine-learning-will-improve-retail-and-customer-service/>
- <https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-data-science-plan-sample-scenarios>
- <https://blogs.msdn.microsoft.com/shishirs/2015/01/26/big-data-machine-learning-scenarios-for-retail/>
- <https://blogs.dxc.technology/tag/big-data/> <https://machinelearningmastery.com/4-steps-to-get-started-in-machine-learning/>
- <http://www.embedded-computing.com/embedded-computing-design/analytics-driven-embedded-systems-an-introduction>
- <https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-choice>
- <http://wireless.sys-con.com/node/3932415> <https://aws.amazon.com/articles/> www.codeproject.com
- <https://www.dezyre.com/article/top-10-machine-learning-algorithms/202>
- <https://www.forbes.com/sites/quora/2016/10/19/these-are-the-most-elegant-useful-algorithms-in-machine-learning/#7ea0fe65183e>
- <http://epaperbeta.timesofindia.com/Archives.aspx> (last three years of Economics Times, Times of India)
- <http://www.informationweek.com> (for articles and papers) <https://www.pinterest.com> (for graphics and ideas)
- www.Quora.com (for questions and answers , ideas and insights)
- <https://phrasee.co/3-extremely-cool-uses-for-machine-learning/> <https://phrasee.co/what-you-need-to-know-about-machine-learning-part-one/> <https://practicalanalytics.co/predictive-analytics-101/>
- <https://quantdare.com/machine-learning-a-brief-breakdown/>
- <http://go.softwebsolutions.com/resources/live-webinar-on-machine-learning.html>

- http://www.informationweek.com/strategic-cio/executive-insights-and-innovation/11-cool-ways-to-use-machine-learning/d/d-id/1323375?image_number=3
- <https://www.linkedin.com/pulse/real-world-applications-cognitive-computing-stephen-deangelis> <https://tdwi.org/art>

Index

■ A

- Absentia, 246
- AbsolutData company, 122
- Advertising, 86, 226
- Agile
 - customized and personalized management, 326
 - pair programming, 328
 - social and spiritual intelligence, 329
- Agile approach
 - defined, 126
 - disadvantages, 128
 - reason for, 126–127
 - SDLC, 127
 - usage, 128
- Agriculture, 227
- AiCure, 205
- Alexa, 129
- Algorithms, 36, 72, 91–92
- AlphaGo, 58
- Amazon Alexa, 247
- Amazon's Deep Scalable Sparse Tensor Network Engine (DSSSTNE), 177
- Ambari technology, 153
- Analytical tools and technologies, 332
- Analytics, 15–16
- Analytics layer
 - Apache Solr, 166
 - Apache Spark, 167
 - Apache Storm, 166
 - Hadoop MapReduce, 158–159
 - HBase, 162–165
 - HDInsight, 167
 - Hive, 160–161
 - MangoDB, 165–166
 - Pig, 160
- Android Things, 144
- Anomaly detection
 - defined, 98
 - pros and cons, 99
 - semi-supervised, 99
 - supervised, 99
 - unsupervised, 99
 - use cases, 100
- Anticipative cognitive computing, 184–185
- Apache
 - Flume, 174
 - Hive, 160–162
 - Kafka, 174–175
 - Mahout, 176–177
 - Pig, 160
 - Solr, 166
 - Spark, 167
 - Sqoop, 175
 - Storm, 166
- Application layer, 168
- Applications, products, and services (APS)
 - advertising, 60
 - autonomous systems, 56
 - data-driven approach, 56
 - deep learning, 58
 - digital assistance, 59
 - emotions and sentiment analysis, 58
 - financial and insurance services, 60–61
 - manufacturing, 64
 - oil and gas, 64
 - photo-tagging, 60
 - professional services, 62
 - public sector, 62
 - retail and wholesale, 63
 - self-learning machines products, 58
 - telecom network, 61

Applications, products, and services (APS) (*cont.*)
 transport, 63–64
 virtual digital assistance, 60
 Application Programming Interface (API), 149
 Apriori
 advantages, 120
 applications, 120
 disadvantages, 120
 overview, 120
 SAP HANA, 121
 Architecture, 41–42
 Artificial intelligence (AI), 5, 37
 Artificial neural network (ANN), 97
 Bayesian network and
 advantages, 124
 applications, 124
 disadvantages, 124
 overview, 123
 training examples, 123
 Augmented reality (AR), 246–247
 Automated identification
 approach, 88
 Autonomous systems, 56
 Aviation, 225
 Avro technology, 153

■ **B**

Balanced scorecard (BSC), 266
 Bayesian network, 122
 Behavior prediction, 45
 Big Data, 3–4, 133
 analytics layer (*see* Analytics layer)
 anti-money laundering and, 170
 decision makers, 171
 definitions, 13
 layers, 140
 machine learning, 172
 sources, 13–14
 unstructured data, 12
 variety, 15
 velocity, 14
 volume, 15
 Big data analytics
 data acquisition and storage layer
 (*see* Data acquisition and storage layer)
 list of technologies, 153
 three Vs, 151

Binary classification
 model, 72
 Biology, 68–69
 Bitcoin, 278
 Blockchain, 277
 Bluetooth network, 147
 Bolts, 166
 Bots, 230–231
 Business models, 316
 Business opportunity, 47–50

■ **C**

C4.5, 109
 CART, 109
 Cassandra, 164
 Cellular mobile network, 147
 Central Nervous System for the Earth (CeNSE), 145
 Chatbots, 228, 229, 232
 Chat management system, 179
 Classification algorithms
 ANN, 97
 decision tree, 97
 ensemble learning, 97
 pros and cons, 98
vs. regression algorithm, 96
 use cases, 98
 Climate models, 104
 Cloud adoption, 256–259
 Cloud-based infrastructures, 319
 Cloud computing, 2–3, 141, 180, 185
 challenges, 21
 characteristics, 17–18
 community cloud, 19
 definition, 17
 deployment models
 benefits and risks, 19
 hybrid, 18
 private, 18
 public, 18
 service models
 benefits, 20
 IaaS, 19
 PaaS, 20
 SaaS, 20
 Cloud specific offering, 165
 Clustering algorithms
 exclusive, 95
 hierarchical, 94
 non-hierarchical, 94

- overlapping, 95
 - use cases, 95
 - Cognitive analytics, 4
 - Cognitive computing, 141
 - characteristics, 30
 - definition, 27–28
 - machine learning, 29
 - process, 29
 - smartphones, 30
 - systems
 - analytics, machine learning and
 - cognitive layer, 183
 - anticipative, 184–185
 - data gathering layer, 183
 - data preparation, extraction, and conversion layer, 183
 - data processing layer, 182–183
 - deduction and machine learning layer, 182
 - vs.* machine learning, 181–182
 - monetization, 184
 - presentation and application layer, 183
 - sensory perception layer, 183
 - smart digital assistants, 184
 - use cases, 28
 - Collaborative filtering (CF), 176
 - Communication and transportation layers
 - exchanging information, 146
 - gateway, 146
 - protocols, 147
 - wireless networking options, 147–148
 - Community cloud model, 19
 - Computing Machinery and Intelligence, 36
 - Connected cars, 241–242
 - Connectivity-based economy, 131
 - Connector layer, Apache
 - Flume, 174
 - Kafka, 174–175
 - Sqoop, 175
 - logic apps, 174
 - MQTT, 174
 - Cortana, 128
 - Cross-validation method, 102
 - Customer machine learning analytics, 220–223
 - Customer support systems
 - applications and products, 55
 - APS, 56
 - automated machine learning, 50–51
 - customer acquisition, 53–54
 - customer retention systems, 52
 - ML benefits, 54
 - retain customers, 55
 - Cybersecurity, 66
- **D**
- Data acquisition and storage layer
 - Hadoop architecture, 155
 - HDFS, 154–156
 - S3, 156
 - Data analysis, 43
 - Data analytics, 43
 - Data-based culture, 300–302
 - Data gathering layer, 183
 - Data island, 38
 - Data lake, 38
 - Data mining, 44
 - DataNodes server, 155
 - Data processing layer, 182
 - analytics, 148
 - cloud APIs, 149
 - collection and management, 148
 - Data processing tools and technologies, 331
 - Data science, 43
 - Data scientist, 305–306
 - Decision tree algorithms
 - advantages, 108
 - applications, 109
 - C4.5, 109
 - CART, 109
 - disadvantages, 109
 - ecosystems, 108
 - flow chart, 108
 - stock market, 110
 - Decision trees, 97, 106
 - Deep Blue, 129
 - Deep learning, 44, 58
 - Deep/machine learning
 - engineer, 303–304
 - Deep Speech 2, 107
 - Detecting money laundering, 272–273
 - Device and sensor layer
 - CeNSE, 145
 - characteristics, 145
 - defined, 143
 - Fitbit, 145
 - microelectromechanical systems, 145
 - DevOps, 312–313
 - Digital adoption, 179

Directed acyclic graph (DAG), 166
Disaster and hazards
 management, 224–225
Distributed file system (DFS), 154
Distributed machine learning, 77
Donna, 129
Driverless cars, 243–244

■ E

Economy, 282
Economy-based business models, 130
Email spam filtering, 117
Emotional intelligence (EI)
 and quotient, 283
 training programs, 325
Entertainment, 227
Evidence-based decision making, 10
Excel/Excel BI/Power BI, 168

■ F

Facebook, 246, 248
Face recognition, 84, 85
Fashion, 226
Finance and banking analytics
 business strategies, 200
 buying habits analysis, 200
 complications, 199
 customer service, 202
 fraud detection, 202
 insurance, 203
 loan identification and
 prediction, 201
 risk management, 201
 sentiment analysis, 203
 trading algorithm, 202
 use case, 203–204
Fitbit device, 145
Fitness, 226
Flume, Apache, 174
Folds, 103

■ G

Gamification, 294
Gateways, 146
Google Home, 130, 247
Google Maps, 169
Google Now, 248
Google TensorFlow, 177

■ H

Hadoop, 153
Hadoop cluster, 155
Hadoop distributed file system
 (HDFS), 154
 availability, 156
 cluster, 155
 DataNodes, 155
 distributed storage, 156
 fast, 156
 flexible, 156
 MapReduce, 159
 NameNode, 155
 open source project, 156
 reliability and fault tolerance, 156
 S3, 156
 scalable, 156
 technology, 153
Hadoop MapReduce, 158
HBase
 NoSQL database (*see* NoSQL
 database)
 scales, 162
 technology, 153
HCatalog technology, 154
HDInsight, 167
Healthcare, 85
Healthcare analytics
 applications
 compliance, 208
 drug development, 208
 imaging analytics, 208
 intelligent analytics, 210
 monitoring and care of
 patients, 209
 public-health intimidations, 210
 treatment and drug, 210
 unstructured documents, 209
 machine learning analytics, 206
 treatment, 207
 use case, 211–212
Hidden Markov model (HMM), 121
Hinge, 187
Hive, 153
 features, 161
 vs. MapReduce, and pig, 162
HiveMall, 161
Hive QL, 153
Holistic intelligence, 327–328
HortonWorks Data Platform (HDP), 163

Hospitality, 225
Human intelligence, 40

■ I

IBM Cognos, 169
IBM Watson, 238–239
Image-based recognition, 84
Indian software services, 252–253
Industry 4.0 model
 agriculture, 134
 energy, 135
 healthcare, 134
 IoT, 134
 manufacturing, 134
 transportation, 135
 Wikipedia, 133
Infographics, 169
Infrastructure-as-a-Service (IaaS), 19, 185
Insurance, 227
Intellectual assets
 adaptability to change, 285
 bottom-up innovation, 284
 customer focus, 285
 spirituality, 285
 teamwork and knowledge sharing, 285
Intelligence, benefits, 293–294
Internet of Things (IoT), 3, 130, 132, 134
 advantages, 138
 Android Things, 144
 challenges, 25–26
 characteristics, 24
 communication and transportation
 (see Communication and transportation layers)
 data processing (see Data processing layer)
 definitions, 23
 device and sensor layer (see Device and sensor layer)
 edge devices and sensors, 26
 gateways, 26
 heterogeneous data access, 23
 layers, 140, 142
 machine learning, 151
 middle management, 142
 in organizations, 22, 143
 presentation and application layer
 (see Presentation and application layer)

solution, 150
tiers, 142
virtualization, 27
Windows 10, 144

■ J

Jargon Buster, 37–38, 86
Jarvis, 129

■ K

Kafka, Apache, 174–175
Key performance indicators (KPIs)
 benefits, 269
 BSC, 266
 guidelines, 265
 measurement
 categories, 267–268
 vs. metrics, 271
 metrics measurement
 framework, 265
 organization/enterprise specific, 270
 parameters, 264
 stock and customer analytics, 270–271
K-means clustering algorithms
 advantages, 118
 applications, 119
 disadvantages, 119
 overview, 117–118
 salesforce, 119

■ L

Large-scale machine learning, 77
Learning algorithms, 40
Learning analytics, 4
Legal activities, 65
Linear regression, 110, 112
Logic Apps, 174
Logistic regression algorithms
 advantages, 111
 applications, 112
 defined, 110
 disadvantages, 112
 overview, 110
 PayPal, 113
 regression-based forecasting, 111
Lyft company, 130

■ **M**

Machine and human brain interfaces, 245

Machine intelligence, 38

Machine learning algorithms

 automobile industry, 132

 business value, 103

 classification, 93–94

 comparison of algorithms and
 models, 102

 considerations, 105

 cross-validation, 102

 data investigation, 102

 economy of wearables, 130

 macro-level, 132

 online banking, 131

 overview, 92

 performance, 103

 pros, 104

 relationships between variables, 102

 requirement phases, 126

 selection, 100–101

 stages of application, 125–126

 tool selection, 103

Machine learning analytics, 31

 benefits, 193

 complications, 194

 conceptual view, 192

 cost and revenue, 194

 value-based service/products, 193

Machine learning (ML)

 and algorithms, 315

 application, 6–7, 35

 and Big Data, 320

 blueprinting exercise, 323

 business drivers, 12

 business opportunity, 47–50

 characteristics, 10

 cloud adoption, 322

 cloud infrastructure, 319

vs. cognitive computing
 systems, 181–182

 cognitive system, 31

 complex and heterogeneous data, 10

 computation phase, 7

 definition, 5, 39

 digitation, 48

 dynamic business scenarios, 11

 features, 255

 financial benefits, 46–47

 framework, 76

IOT, 151

layers, 141

 Big Data, 32

 cloud infrastructure, 32

 cognitive computing, 33

 presentation and reporting, 33

 project implementation steps, 320

 raw data conversion, 321

 service layer, 186

 technical architecture, 321–322

 text analytics, 87

vs. traditional programming, 6

 unpredictable system behavior, 11

 wisdom pyramid, 48–49

Mahout, Apache, 176–177

MangoDB, 165

Manufacturing analytics

 Big Data and IoT, 196

 complications, 195

 integration of legacy systems, 195

 machine learning application, 197

 security challenges, 196

 unified data model, 196

 use case, 198

MapReduce, Hadoop

 architecture, 158

 defined, 158

 functions, 159

 Google, 159

vs. hive, and pig, 162

MapReduce technology, 153

Marketing analytics

 applications, 214–215

 business drivers, 214

 complications, 213

 social technologies, 212

 use cases, 216

Markov models

 advantages, 122

 disadvantages, 122

 overview, 121

Mathematical variables, 46

Medical science, 67

“Me first” approach, 40

MEMS. *See* Microelectromechanical
 systems (MEMS)

Metropolitan area network (MAN), 147

Microelectromechanical systems
 (MEMS), 145

Microservices, 278

Microsoft Azure ML, 173

Microsoft Cognitive Toolkit, 177
 Microsoft Cortana, 239–241
 Microsoft SQL server 2005, 117
 Middle management, 142
 MLLib, 178
 Model and runtime layer
 cognitive toolkit, 177
 DSSTNE, 177
 Mahout, Apache, 176–177
 TensorFlow, 177
 Money laundering, 66
 Moral intelligence (MI), 284
 MQTT, 174
 Multiclass classification model, 73

■ **N**

Naïve Bayes
 advantages, 116
 disadvantages, 116
 document categorization, 117
 email spam filtering, 117
 Microsoft SQL server 2005, 117
 overview, 115
 sentiment analysis, 116
 NameNode server, 155
 Nanotechnology sensors, 150
 Natural language processing (NLP), 88
 Netflix, 107
 SVM, 115
 Neuralink, 245
 Neural processing, 39–40
 NoSQL database
 Cassandra, 164
 cloud-based applications, 162
 column based, 164
 document based, 164
 graph based, 164
 key/value based, 164
 MangoDB, 165
 RDBMS *vs.*, 163
 uses, 165

■ **O**

Oil and gas, 226
 Oozle technology, 154
 Oracle Business Intelligence Enterprise
 Edition (OBIEE), 168
 Organizations
 candidates characteristics, 310–311

clients, 263
 co-creation, 263
 effectiveness, 302
 implementation, 3
 innovation and automation, 262
 machine learning, 264
 vendor, 263

■ **P**

PageRank mechanism, 117
 Pair programming, 328
 Pattern-recognition techniques, 272
 PayPal, 113
 Performance drivers
 competency metrics, 287
 EI, 289
 hierarchy, 288
 holistic and comprehensive
 view, 292, 293
 MI, 289
 SI, 290
 social intelligence, 291
 Personal area network (PAN), 147
 Physics, 68
 Pig technology
 description, 153
 vs. MapReduce, and hive, 162
 overview, 160
 Platform as a service (PaaS), 20, 186
 Predictive analytics, 43
 Presentation and application layer, 149,
 168
 Programming languages, 78
 Project management
 principles, 309–310
 value, 308
 Python, 82–83

■ **Q**

Quantum machine learning (QML),
 254–255
 Qubit, 254

■ **R**

R, 79–80
 Random forest algorithm
 advantages, 107
 decision trees, 106

Random forest algorithm (*cont.*)
 Deep Speech 2, 107
 disadvantages, 107
 Netflix, 107

RDBMS
vs. NoSQL, 163
 uses, 165

Regression algorithms
 linear, 95
 logistic, 95
 polynomial, 95
 pros and cons, 96
 ridge regression, 95
 use cases, 96

Regression-based forecasting, 111

Regression model, 73

Reinforcement learning, 69–70

Retail industry
 complications, 217–218
 machine learning application, 219
 use cases, 220

Ridge regression, 95

Risk-management
 assessment, 275
 monitor and control, 275
 response plan, 275
 risk identification, 274
 types, 274

Robotic process automation (RPA), 38

■ S

Salesforce Einstein, 250–251

SAP HANA, 121

SAP Leonardo, 248–250

Scala, 80–82

Security
 incident management, 253
 Indian software services, 252–253
 use cases, 253–254

Semi-supervised learning, 71

Sensory perception layer, 183

Serverless, 277

Shallow learning, 44

Simultaneous localizations and mapping (SLAM), 245

Siri, 128, 237

Social intelligence (Social I), 284

Software as a service (SaaS), 20

Solr, Apache, 166

Space science, 67

Spark, Apache, 167

Spectacles, 187

Spiritual intelligence (SI), 283, 284

Spouts (Storm), 166

SQL Server Reporting Services (SSRS), 168

Sqoop, Apache, 175

Statistical modeling technique, 95

Stock market, 110

Stock-picking models, 104

Storm, Apache, 166

Superposition, 254

Supervised learning, 8, 71

Support vector machine (SVM) algorithms
 advantages, 114
 applications, 114
 disadvantages, 114
 Netflix, 115
 overview, 113–114

SVM model. *See* Support vector machine (SVM) algorithms

■ T

Team building
 CEO, 307–308
 criteria, 295–296
 organizational leader, 299
 team leader, 297
 team member, 298
 technical roles, 306
 technology manager, 298, 300

Technology stack
 big data analytics (*see* Big data analytics)
 cognitive computing, 181–183
 connected view, 140
 IOT, 142–149, 151
 layer specific, 139–140
 machine learning
 cloud computing, 180
 connector layer, 173–175
 facial recognition techniques, 180
 model and runtime layer, 176–178
 presentation and application layer, 178–180
 processing layer, 175
 storage layer, 175
 software application, 138

Telecommunications, 227

TensorFlow, Google, 177

Tez technology, 67

Thinking machine, 36
Tools, 73–75
Toxic behaviors, 329
Trading algorithm, 202
Traditional and data based economics,
190–191
Training ML models, 72
Transportation companies, 225
Travel and communication, 85

■ **U**

Unsupervised learning, 9, 71

■ **V**

Value, 308
Velocity, variety, and volume (Vs), 151
Video analytics, 178
Video games, 224
Visabots, 228

■ **W, X**

Waterfall software development lifecycle
(SDLC), 127
Watson, 129
Wide area network (WAN), 147
WiFi, 147
Windows 10, 144
Wireless networking
options, 147–148
Workplace, 282

■ **Y**

Yarn technology, 153

■ **Z**

ZigBee network, 147
ZooKeeper technology, 153
Z-wave network, 147