

Business Analytics

Subjects: **Computer Science, Interdisciplinary Applications**

Contributor: Shiyu Liu , Ou Liu , Junyang Chen

Business analytics has been widely used in various business sectors and has been effective in increasing enterprise value. With the advancement of science and technology in the Big Data era, business analytics techniques have been changing and evolving rapidly.

business analytics

descriptive analytics

predictive analytics

prescriptive analytics

1. Introduction

In recent decades, data have been rapidly changing the world. Especially in the era of Big Data, data are cheap and ubiquitous, but what makes datum a valuable asset is how it is used to obtain useful information. Since there are many different types of business objectives, different analytics techniques are needed to achieve them. These techniques have many applications in the business area and “business analytics” enables the business application of Big Data. Since the emergence of the term business analytics, it is growing by leaps and bounds, reflecting the increasing importance of data in terms of volume, variety and velocity ^[1]. Although there is no uniform definition of business analytics, the existing definitions can be summarized into several dimensions, such as a movement, a transformation process, a capacity set and so on ^[2].

Interest in analytics and data science is growing as business organizations are using business analytics extensively to improve their business value. Business analytics has evolved into an important part of the business decision-making process, using data to drive decisions and support decision-makers in making strategic, operational and tactical decisions ^[3]. Specifically, business analytics can help companies to leverage the value of historical data by harnessing the power of statistical and mathematical models and advanced techniques such as artificial intelligence algorithms. Through these models and algorithms, enterprises can integrate disparate data sources for trend prediction, decision optimization and more. As business analytics continues to evolve, its applications continue to broaden. It is adapted in some functional departments within the enterprise and some non-business areas.

2. Definitions of Business Analytics

At present, there is still no uniform definition of business analytics. Scholars in different fields have defined the term business analytics from several perspectives. Holsapple concluded 18 definitions of analytics in 6 dimensions ^[2].

First, from the perspective of techniques, business analytics is considered an application of any data analytics [4] or data science [5] in business fields, which uses tools and techniques statistically and quantitatively to analyze a huge collection of data sources to support decisions for business [6]. More specifically, business analytics can be viewed as 'a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions' [7]. With the continuous emergence of new technologies, business analytics can also be viewed as a combination of operation research, artificial intelligence (machine learning) and information systems [1].

Second, from the process perspective, business analytics is an encapsulation of tools to convert data into actionable insights through a scientific/mathematical/intelligent process [8]. The Institute for Operations Research and the Management Sciences (INFORMS) defined it as 'a scientific process of transforming data into insight for making better decisions' [9].

Third, from the practice perspective, business analytics is defined as 'an ability of firms and organizations to collect, manage, and analyze data from a variety of sources to enhance the understanding of business processes, operations, and systems' [10]. Business analytics refers to 'the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions' [11].

Finally, from the perspective of the management, business analytics is a qualitative methodology to derive valuable meanings based on data [12] and is 'a paradigm shifter of models, technologies, opportunities, and capabilities used to scrutinize a corporation's data and performance to transpire data-driven decision-making analytics for the corporation's future direction and investment plans' [3].

Overall, regardless of the perspective from which it is defined, it can conclude that the implementers of business analytics are enterprises; the approaches to achieve business analytics are various techniques; and the ultima goal of business analytics is to improve enterprise values.

3. Techniques of Business Analytics

Business analytics is generally identified into three types: descriptive analytics, predictive analytics and prescriptive analytics [8]. Descriptive analytics is used to provide a summary of descriptive statistics as a straightforward presentation of facts. Predictive analytics is used to discover what is likely to happen in the future based on current data. Prescriptive analytics focuses on identifying optimal actions in the decision-making process.

3.1. Descriptive Analytics Techniques

Descriptive statistics is a process of characterizing historical data. There are two core techniques of descriptive statistics: data visualization and data analysis. Data visualization produces graphical images of data or concepts, which helps decision making [13]. Data analysis consists of common statistical techniques, including mean, median,

standard deviation, range, stem, histogram and advanced data mining techniques used to describe hidden patterns in the data.

3.1.1. Data Visualization

Over the years, many data visualization techniques have been developed to represent large amounts of information and examine them. These methods include bar charts, box, and whiskers, bubble charts, choropleth maps, dot distribution maps, histograms, line graphs, pie charts, population pyramids, proportional symbol maps, scatter plots, stacked bar charts and tree maps.

When working with data sets that include big data points, automation of the data visualization process makes the process much easier. Therefore, a large variety of data visualization tools are developed to create visual representations of large data sets, including Tableau Software 2022.4 [\[14\]](#), Microsoft Power BI [\[15\]](#), Excel, FusionCharts, Sisense, etc. In addition to the visualization tools as software, there are many online visualization tools such as Infogram, RAWGraphs, Sovit, etc.

3.1.2. Data Analysis

Data analysis is to analyze the collected data and derive various quantitative characteristics reflecting objective phenomena. In addition to the traditional statistical methods of data concentration trend analysis, data dispersion analysis and data frequency distribution analysis, advanced data mining techniques probe more deeply into the underlying characteristics of data. Association and cluster analysis are two typical data analysis methods used in descriptive analytics.

Association analysis

Association analysis, also called association rule mining, is an unsupervised algorithm that is used to mine potential association relationships from data. There are two classical algorithms in association analysis: Apriori Algorithm and Frequent Pattern tree (FP-tree). Apriori Algorithm uses an iterative method of searching the database level by level to find the relationships of item sets to form rules. Its process consists of concatenation and pruning. To improve the Apriori algorithm, many improvement methods are proposed including the Direct Hashing and Pruning (DHP) algorithm [\[16\]](#), Dynamic Itemset Counting (DIC) [\[17\]](#), Parallel Apriori algorithms based on various frameworks such as MapReduce [\[18\]\[19\]\[20\]](#), Spark [\[21\]\[22\]](#) and Flink [\[23\]](#) and adaptive Apriori algorithms [\[24\]](#). Compared to the Apriori algorithm, the FP-tree algorithm only requires two scans of the database when performing frequent pattern mining and does not generate candidate item sets. There are various improvement algorithms based on FP-tree, such as QFP-growth [\[25\]](#), fuzzy FP-tree [\[26\]](#), PFP [\[27\]](#), balanced parallel FP-tree (BPFP) [\[28\]](#) and tree partition based parallel FP-tree [\[29\]](#).

Cluster analysis

Cluster analysis is a multivariate statistical analysis method for classifying samples or indicators. The clustering algorithms can be divided into five categories: partitioning-based, hierarchical-based, density-based, grid-based and model-based.

Partitioning-based algorithms include K-means ^[30], Fuzzy C-means (FCM) ^[31], K-medoids ^[32], CLARA (Clustering Large Applications) ^[32], K-modes ^[33], and CLARANS (Clustering Large Applications based on a RANDOMized Search) ^[34]. Hierarchical clustering algorithms include BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) ^[35], (CURE) ^[36], ROCK (Robust Clustering using Links) ^[37] and Chameleon (clustering using interconnectivity) ^[38]. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is the first density-based clustering algorithm ^[39]. In addition, DENCLU (DENSITY-based CLUstEring) ^[40] and OPTICS (Ordering Points to Identify the Clustering Structure) ^[41] are both widely used in cluster analysis. The typical grid-based algorithms include STING (Statistical Information Grid) ^[42], CLIQU (Clustering in Quest) ^[43] and WaveCluster ^[44]. There are usually two attempted ideas in the model-based algorithm: statistical methods and neural network methods. Among them, statistical methods are the COBWEB algorithm ^[45], GMM (Gaussian Mixture Model) ^[46] and the neural network algorithm is the SOM (Self Organized Maps) algorithm ^[47].

3.2. Predictive Analytics Techniques

In general, predictive analytics techniques can be divided into statistical techniques and machine learning techniques. Statistical methods to predict mainly refer to building suitable forecasting models and estimating model parameters, listing forecasting formulas and thus making extrapolated forecasts. In machine learning techniques, systems are trained to use specialized algorithms to study, learn and make predictions and recommendations based on large amounts of data.

3.2.1. Statistical Techniques

In statistical predictive techniques, statistical theories and methods are used for prediction by building statistical models and fitting the model parameters with past data. There are two groups of methods in statistical predictive techniques: regression models and time series models.

Regression model

Regression model is one of the most famous statistical techniques used to predict. The linear regression model is the basic model which is represented as an equation that finds specific weights for the input variables, which in turn describe a straight line that best fits the relationship between the input variables and the output variables ^[48]. When the output variable is a categorical variable, a classification model such as the logistic regression model ^[49] is needed. Meanwhile, polynomial regression models are used to fit nonlinear relationships between variables.

Time series model

Time series models can be divided into two groups: exponential smoothing models and ARIMA series models [50]. The exponential smoothing model decomposes the time series into components and uses an additive or multiplicative structure to reassemble the smoothed components to predict future values [51]. Typical exponential smoothing models include simple exponential smoothing, Holt's exponential smoothing and Holt-Winters' seasonal exponential smoothing [52]. The ARIMA series model mainly includes AR (AutoRegressive) model, MA (Moving Average) model, ARMA (AutoRegressive Moving Average) model, ARIMA (AutoRegressive Integrated Moving Average) model and SARIMA (seasonal ARIMA) model [53].

3.2.2. Machine Learning and Artificial Intelligence Techniques

With the advent of the big data era, machine learning to guide predictive analytics has become a widely used approach. There are many classic machine learning prediction algorithms, such as support vector machine, nearest neighbor, decision tree, ensemble learning and artificial neural network and more advanced deep learning techniques.

Support vector machine

In machine learning, the support vector machine (SVM) is a supervised learning model for analyzing data in classification and regression analysis with associated learning algorithms [54]. Given a set of training instances, each labeled as belonging to one or the other of two classes, the SVM training algorithm builds a model that assigns new instances to one of the two classes. Thus, it is usually used to predict binary classification problems.

Nearest neighbor

The k-nearest neighbor algorithm, known as KNN, is a non-parametric supervised learning classifier [55]. It can be applied to classification problems or regression problems. In a classification problem, the output is a member of a category, and in a regression problem, the output is the value of an object's attributes. The nearest neighbor is considered the simplest type of machine learning algorithm [56].

Decision tree

Decision tree is a non-parametric supervised learning algorithm for classification and regression tasks. It is a hierarchical tree structure consisting of a root node, branches, internal nodes and leaf nodes. There are three typical decision tree algorithms: ID3, C4.5 and CART (Classification and Regression Tree). Iterative Dichotomiser 3 (ID3) uses information entropy and information gain as metrics to evaluate candidate splitting [57]. C4.5 is an improved version of ID3, which does not use the information gain directly but introduces the information gain ratio metric as the basis for feature selection [58].

Ensemble learning

The basic idea of the Ensemble Learning algorithm is combining multiple classifiers to achieve an integrated classifier with better prediction. Ensemble learning includes the bagging method, boosting method and stacking method.

Artificial Neural network

Artificial Neural network (ANN) is a model that mimics the structure and function of biological neural networks, especially in the brain [59]. According to the connectionism of networks, ANN can be divided into feed-forward neural networks and feedback neural networks. Feedforward neural networks (FNN) divide each neuron into different groups according to the order of receiving information, and each group can be considered as a neural layer [60]. The neurons in each layer receive the output of the neurons in the previous layer and output to the neurons in the next layer. FNN has two categories depending on the number of layers: single-layer and multi-layer networks [61]. Single-layer FNN is also known as fully connected feedforward neural networks (FC), and a typical multi-layer network is the convolutional neural network (CNN). In feedback neural networks, neurons can receive signals from other neurons and their own feedback signals. Compared with feedforward neural networks, the neurons in feedback neural networks have a memory function and have different states at different moments. Common feedback neural networks include recurrent neural networks (RNN) [62], Hopfield networks [63] and Boltzmann machines [64].

Deep learning

The concept of deep learning originates from the study of artificial neural networks, and a multilayer perceptron with multiple hidden layers is a deep learning structure. Recently, deep learning has been widely used in predictive analytics, including RNN, CNN, Transformer and Nbeats. LSTM is a well-known RNN algorithm used in prediction [65]. DeepAR employs a classical RNN model to solve the time series forecasting problem [66], and Deep state space model is proposed to improve DeepAR limitations [67]. Since DeepAR and Deep state space model are both one-horizon forecast models, MQRNN (multi-horizon forecast model) is designed to simultaneously predict for multiple future time steps [68]. The CNN-LSTM algorithm, which combines CNN and LSTM, has been applied in many predictive analyses [69][70][71].

3.3. Prescriptive Analytics

Prescriptive analytics is the final step of business analytics. Prescriptive analytics mainly refers to the use of operations research methods such as mathematical programming models and intelligent optimization algorithms to give recommendations on the optimal actions that an enterprise should take. Compared to the traditional decision methods which rely too much on human experience, prescriptive analytics gives more reliable and reasonable decisions through scientific approaches including traditional optimization algorithms and heuristic algorithms.

3.3.1. Traditional Optimization Algorithm

Based on the features of the objective function, constraints and decision variables, mathematical programs can be divided into linear programming, nonlinear programming, integer programming, stochastic programming, dynamic programming and so on [72]. In order to solve these problems, many traditional optimization algorithms are proposed. For constrained programming, Simplex algorithm is a well-known linear programming algorithm [73], and penalty-series methods are proposed for nonlinear programming. Gradient Descent Method [74], Quasi-Newton Method [75] and Conjugate gradient method [76] are classical iteration algorithms for unconstrained optimizations.

3.3.2. Heuristic Algorithm

Simple Heuristic Algorithms

Simple heuristic algorithms mainly contain greedy algorithms, local search algorithms and hill-climbing algorithms. The greedy algorithm is an algorithm that takes the optimal choice in the current state at each step of the selection process, thereby hopefully leading to the best or optimal outcome [77]. The local search algorithm is based on the greedy idea of starting with a candidate solution and continuously searching in its neighborhood until there are no better solutions in the neighborhood [78]. The hill-climbing algorithm is a simple greedy search algorithm that selects one optimal solution at a time as the current solution from the proximity solution space of the current solution until a local optimal solution is reached [79].

Meta-heuristic algorithms

Meta-Heuristic algorithms are improvements of simple heuristic algorithms, usually using randomized search techniques, and can be applied to a wide range of problems. Meta-heuristic algorithms include Evolutionary Algorithms, Swarm Intelligence algorithms, Simulated Annealing algorithms and Tabu Search algorithms. Evolutionary algorithms are inspired by the evolutionary mechanisms of living organisms and simulate the evolutionary processes to conduct evolutionary calculations on the candidate solutions of optimization problems. Typical evolutionary algorithms are Genetic Algorithm (GA), Differential Evolution (DE) and Immune Algorithm (IM). Swarm intelligence refers to the property of unintelligent subjects to exhibit intelligent behavior through cooperation and is a computational technique based on the behavioral laws of biological groups. Two representative swarm intelligence algorithms are Particle Swarm Optimization (PSO) [80] and ACO (ant colony optimization) [81]. Simulated Annealing is an algorithm that solves the global optimum by finding states with relatively small objective values in the neighborhood [82]. Tabu search algorithm searches for the optimal solution of the target by searching for a better solution in the solution neighborhood and puts the search history into a Tabu List during the search process to avoid duplicate searches [83].

Hyper-Heuristic algorithms

Hyper-Heuristic algorithms provide a high-level heuristic by managing or manipulating a set of Low-Level Heuristics (LLH) to generate new heuristics. These new heuristics are used to solve various combinatorial optimization problems.

4. Business Analytics Applications

4.1. Applications in Functional Areas

Supply chain management is a representative application of business analytics in the business area. Business analytics has a strong impact on the supply chain performance in the plan, source, make and deliver area [84][85][86]. For example, descriptive analysis helps to identify demand patterns and predict analysis forecasts customer demand in the future through statistical and machine learning algorithms. Based on the predictions, optimization algorithms are used to make pricing and inventory management decisions to maximize retailers' profit.

In the area of marketing management, business analytics integrates market and customer-related data and uses analysis algorithms to provide managers with a variety of relevant perspectives for better optimization decisions. Among the various areas of marketing, customer relationship management (CRM) is a key area that uses business analytics to analyze, integrate and utilize information resources and customer feedback to support CRM technology, such as acquiring and retaining customers [87].

Risk management is an essential area of company management, and business analytics techniques are widely used in the process of risk management. Predict analysis techniques such as artificial neural networks and support vector machines are applied to establish the early warning system [88][89] and risk evaluation [90][91]. Optimization tools of prescriptive analysis are used to make better risk-based decisions [92].

Strategic management plays an important role in the business area to create or sustain competitive advantages of an enterprise, which consists of analyses, decisions and actions undertaken. Business analytics helps firms to reveal their strengths and weaknesses by identifying business units, activities and processes [93].

The emergency of business analytics drives the development of data-driven human resources (HR) management [94]. Human resources management is progressively increasing its adoption of advanced data analytics, visualization models and techniques to strengthen strategic decision-making and serve the needs of decision-makers. Descriptive analytics uses internal and external organizational data and HR administrative information to generate ratios, metrics, dashboards and reports on HR. Predictive analytics can analyze process data and make predictions. Based on predictive analytics and the large and diverse HR data available, HR departments gain decision options to optimize performance and completely reshape the decision-making process [95].

4.2. Applications in Industry Sectors

Business analytics is widely used in the healthcare sector. Data visualization tools such as dashboards and control charts are used to monitor outcomes and look for variations in process [96]. Descriptive analytics techniques are used to mine genetic data to identify the relationships between human genes, diseases, variants, proteins, cells and biological pathways [97]. Predictive analytics methods help to forecast the emergency and development of diseases [98]. The application of prescriptive algorithms can increase efficiency and reduce costs in the healthcare industry [99].

The retail industry has various applications of business analytics. Retailers can collect customer demographics and behavior data to analyze customer preferences and shopping features through business analytics. The classical one is the market basket analysis using data mining methods to examine large transaction databases and determine which items are most frequently purchased [\[100\]](#)[\[101\]](#). Customer visit segments can be mined by data mining rules [\[102\]](#). Business analytics techniques are also used in the establishment of recommend systems, especially in the electric-commerce fields [\[103\]](#)[\[104\]](#).

I 5. Challenges in Business Analytics

5.1. Data Quality

With the advent of the Big Data era, the accessibility of data and the volume of data available have increased significantly compared to the past. However, the problem that arises is how to select useful and accurate data for analytics from the vast amount of information. Machine learning plays an important role in business analytics, which relies on data. Thus, business analytics can be considered a data-driven analytics process; so, data quality is very important for subsequent analysis and guidance. In business analytics, data quality challenges mainly include data completeness, consistency and accuracy.

Data accuracy refers to anomalies or errors in the information recorded in the data. Common data accuracy errors include garbled data and abnormally large or small data. There are various outlier detection algorithms, each with its advantages, disadvantages and scope of application, and it is difficult to directly determine which one is the best. In practical applications, an appropriate outlier detection algorithm is selected according to the characteristics of business operations, such as the requirements for computational volume and tolerance for outliers.

5.2. Data Security and Privacy

There is no completely secure data infrastructure unless it is isolated and disconnected from all other networks. However, this is impossible for business analytics, especially when cloud computing emerges [\[8\]](#). Throughout the data lifecycle, enterprises need to comply with stricter security standards and confidentiality regulations; therefore, the security requirements for data storage and use are increasingly high.

Meanwhile, the security needs of data are changing, and a new complete chain has been formed from data collection, data integration, data refinement, data mining, security analysis, security posture determination and security detection to threat discovery. In this chain, data may be lost, leaked, accessed by unauthorized access, tampered with, or even involved in user privacy and corporate secrets. Therefore, data security protection in the big data environment is a significant challenge for business analytics. From the perspective of customers, there are concerns about the privacy of individuals. The use of the personal data of customers, even within the limits of the law, should be avoided or scrutinized to prevent the organization from adverse effects and public condemnation.

References

1. Mortenson, M.J.; Doherty, N.F.; Robinson, S. Operational Research from Taylorism to Terabytes: A Research Agenda for the Analytics Age. *Eur. J. Oper. Res.* 2015, 241, 583–595.
2. Holsapple, C.; Lee-Post, A.; Pakath, R. A Unified Foundation for Business Analytics. *Decis. Support Syst.* 2014, 64, 130–141.
3. Bayrak, T. A Review of Business Analytics: A Business Enabler or Another Passing Fad. *Procedia-Soc. Behav. Sci.* 2015, 195, 230–239.
4. Duan, L.; Xiong, Y. Big Data Analytics and Business Analytics. *J. Manag. Anal.* 2015, 2, 1–21.
5. Chen, H.; Chiang, R.H.; Storey, V.C. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q.* 2012, 36, 1165–1188.
6. Delen, D.; Zolbanin, H.M. The Analytics Paradigm in Business Research. *J. Bus. Res.* 2018, 90, 186–195.
7. Watson, H.J. Tutorial: Business Intelligence—Past, Present, and Future. *CAIS* 2009, 25, 39.
8. Delen, D.; Ram, S. Research Challenges and Opportunities in Business Analytics. *J. Bus. Anal.* 2018, 1, 2–12.
9. INFORMS. Certified Analytics Professional Handbook; INFORMS: Catonsville, MD, USA, 2016.
10. Kraus, M.; Feuerriegel, S.; Oztekin, A. Deep Learning in Business Analytics and Operations Research: Models, Applications and Managerial Implications. *Eur. J. Oper. Res.* 2020, 281, 628–641.
11. Davenport, T.H.; Harris, J.G. *Competing on Analytics: The New Science of Winning*. Language 2007, 15, 24.
12. Lee, C.S.; Cheang, P.Y.S.; Moslehpour, M. Predictive Analytics in Business Analytics: Decision Tree. *Adv. Decis. Sci.* 2022, 26, 1–29.
13. Ware, C. *Information Visualization: Perception for Design*; Morgan Kaufmann: Burlington, MA, USA, 2019; ISBN 0-12-812876-3.
14. Batt, S.; Grealis, T.; Harmon, O.; Tomolonis, P. Learning Tableau: A Data Visualization Tool. *J. Econ. Educ.* 2020, 51, 317–328.
15. Becker, L.T.; Gould, E.M. Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data. *Ser. Rev.* 2019, 45, 184–188.
16. Park, J.S.; Chen, M.-S.; Yu, P.S. An Effective Hash-Based Algorithm for Mining Association Rules. *ACM Sigmod Rec.* 1995, 24, 175–186.

17. Brin, S.; Motwani, R.; Ullman, J.D.; Tsur, S. Dynamic Itemset Counting and Implication Rules for Market Basket Data. In Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, Tucson, AZ, USA, 13–15 May 1997; pp. 255–264.
18. Yang, X.Y.; Liu, Z.; Fu, Y. MapReduce as a Programming Model for Association Rules Algorithm on Hadoop. In Proceedings of the 3rd International Conference on Information Sciences and Interaction Sciences, Chengdu, China, 23–25 June 2010; pp. 99–102.
19. Li, N.; Zeng, L.; He, Q.; Shi, Z. Parallel Implementation of Apriori Algorithm Based on Mapreduce. In Proceedings of the 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Kyoto, Japan, 8–10 August 2012; pp. 236–241.
20. Sornalakshmi, M.; Balamurali, S.; Venkatesulu, M.; Krishnan, M.N.; Ramasamy, L.K.; Kadry, S.; Lim, S. An Efficient Apriori Algorithm for Frequent Pattern Mining Using Mapreduce in Healthcare Data. *Bull. Electr. Eng. Inform.* 2021, 10, 390–403.
21. Qiu, H.; Gu, R.; Yuan, C.; Huang, Y. Yafim: A Parallel Frequent Itemset Mining Algorithm with Spark. In Proceedings of the 2014 IEEE International Parallel & Distributed Processing Symposium Workshops, Phoenix, AZ, USA, 19–23 May 2014; pp. 1664–1671.
22. Rathee, S.; Kaul, M.; Kashyap, A. R-Apriori: An Efficient Apriori Based Algorithm on Spark. In PIKM '15 Proceedings of the 8th Workshop on Ph.D. Workshop in Information and Knowledge Management, Melbourne, Australia, 19 October 2015; ACM: New York, NY, USA, 2015; pp. 27–34.
23. Akil, B.; Zhou, Y.; Röhm, U. On the Usability of Hadoop MapReduce, Apache Spark & Apache Flink for Data Science. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 303–310.
24. Patil, S.D.; Deshmukh, R.R.; Kirange, D.K. Adaptive Apriori Algorithm for Frequent Itemset Mining. In Proceedings of the 2016 International Conference System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 25–27 November 2016; pp. 7–13.
25. Qiu, Y.; Lan, Y.-J.; Xie, Q.-S. An Improved Algorithm of Mining from FP-Tree. In Proceedings of the 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Shanghai, China, 26–29 August 2004; Volume 4, pp. 1665–1670.
26. Lin, C.-W.; Hong, T.-P.; Lu, W.-H. Linguistic Data Mining with Fuzzy FP-Trees. *Expert Syst. Appl.* 2010, 37, 4560–4567.
27. Li, H.; Wang, Y.; Zhang, D.; Zhang, M.; Chang, E.Y. Pfp: Parallel Fp-Growth for Query Recommendation. In RecSys '08 Proceedings of the 2008 ACM conference on Recommender systems, Lausanne, Switzerland, 23–25 October 2008; ACM Press: Lausanne, Switzerland, 2008; p. 107.

28. Zhou, L.; Zhong, Z.; Chang, J.; Li, J.; Huang, J.Z.; Feng, S. Balanced Parallel FP-Growth with MapReduce. In Proceedings of the 2010 IEEE Youth Conference on Information, Computing and Telecommunications, Beijing, China, 28–30 November 2010; pp. 243–246.
29. Chen, D.; Lai, C.; Hu, W.; Chen, W.; Zhang, Y.; Zheng, W. Tree Partition Based Parallel Frequent Pattern Mining on Shared Memory Systems. In Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium, Rhodes Island, Greece, 25–29 April 2006; p. 8.
30. MacQueen, J. Classification and Analysis of Multivariate Observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 21 June–18 July 1967; pp. 281–297.
31. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The Fuzzy c-Means Clustering Algorithm. *Comput. Geosci.* 1984, 10, 191–203.
32. Kaufman, L.; Rousseeuw, P.J. (Eds.) *Finding Groups in Data*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; ISBN 978-0-470-31680-1.
33. Huang, Z.; Ng, M.K. A Fuzzy K-Modes Algorithm for Clustering Categorical Data. *IEEE Trans. Fuzzy Syst.* 1999, 7, 446–452.
34. Ng, R.T.; Han, J. Efficient and Effective Clustering Methods for Spatial Data Mining. In VLDB'94 Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, Chile, 12–15 September 1994; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1994; pp. 144–155.
35. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *ACM Sigmod Rec.* 1996, 25, 103–114.
36. Guha, S.; Rastogi, R.; Shim, K. CURE: An Efficient Clustering Algorithm for Large Databases. *ACM Sigmod Rec.* 1998, 27, 73–84.
37. Guha, S.; Rastogi, R.; Shim, K. ROCK: A Robust Clustering Algorithm for Categorical Attributes. *Inf. Syst.* 2000, 25, 345–366.
38. Karypis, G.; Han, E.-H.; Kumar, V. Chameleon: Hierarchical Clustering Using Dynamic Modeling. *Computer* 1999, 32, 68–75.
39. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. Density-Based Spatial Clustering of Applications with Noise. In Proceedings of the Second International Conference Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; Volume 240.
40. Hinneburg, A.; Keim, D.A. *An Efficient Approach to Clustering in Large Multimedia Databases with Noise*; Bibliothek der Universität Konstanz: Konstanz, Germany, 1998; Volume 98.
41. Ankerst, M.; Breunig, M.M.; Kriegel, H.-P.; Sander, J. OPTICS: Ordering Points to Identify the Clustering Structure. *ACM Sigmod Rec.* 1999, 28, 49–60.

42. Wang, W.; Yang, J.; Muntz, R. STING: A Statistical Information Grid Approach to Spatial Data Mining. *Vldb* 1997, 97, 186–195.
43. Agrawal, R.; Gehrke, J.; Gunopulos, D.; Raghavan, P. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. In *SIGMOD '98 Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, Seattle, WA, USA, 1-4 June 1998; ACM: New York, NY, USA, 1998; pp. 94–105.
44. Sheikholeslami, G.; Chatterjee, S.; Zhang, A. WaveCluster: A Wavelet-Based Clustering Approach for Spatial Data in Very Large Databases. *VLDB J.* 2000, 8, 289–304.
45. Arifovic, J. Genetic Algorithm Learning and the Cobweb Model. *J. Econ. Dyn. Control.* 1994, 18, 3–28.
46. Reynolds, D.A. Gaussian Mixture Models. *Encycl. Biom.* 2009, 741, 659–663.
47. Kohonen, T. *Self-Organizing Maps*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 30, ISBN 3-642-56927-7.
48. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Wasserman, W. *Applied Linear Regression Models*; McGraw-Hill/Irwin: New York, NY, USA, 2004; Volume 4.
49. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398, ISBN 0-470-58247-2.
50. Jain, G.; Mallick, B. A Study of Time Series Models ARIMA and ETS. *SSRN J.* 2017.
51. Gardner, E.S. Exponential Smoothing: The State of the Art. *J. Forecast.* 1985, 4, 1–28.
52. Hyndman, R.J.; Koehler, A.B.; Snyder, R.D.; Grose, S. A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. *Int. J. Forecast.* 2002, 18, 439–454.
53. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*, 3rd ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1994; ISBN 978-0-13-060774-4.
54. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* 1995, 20, 273–297.
55. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* 1967, 13, 21–27.
56. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction. *Sci. Rep.* 2022, 12, 6256.
57. Quinlan, J.R. *Discovering Rules by Induction from Large Collections of Examples*. *Expert Syst. Micro Electron. Age* 1979.
58. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014; ISBN 0-08-050058-7.

59. Jain, A.K.; Mao, J.; Mohiuddin, K.M. Artificial Neural Networks: A Tutorial. *Computer* 1996, 29, 31–44.
60. Sanger, T.D. Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network. *Neural Netw.* 1989, 2, 459–473.
61. Murat, H.S. A brief review of feed-forward neural networks. *Commun. Fac. Sci. Univ. Ank.* 2006, 50, 11–17.
62. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Phys. D Nonlinear Phenom.* 2020, 404, 132306.
63. Gopalsamy, K.; He, X. Stability in Asymmetric Hopfield Nets with Transmission Delays. *Phys. D Nonlinear Phenom.* 1994, 76, 344–358.
64. Ackley, D.H.; Hinton, G.E.; Sejnowski, T.J. A Learning Algorithm for Boltzmann Machines. *Cogn. Sci.* 1985, 9, 147–169.
65. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to Forget: Continual Prediction with LSTM. *Neural Comput.* 2000, 12, 2451–2471.
66. Salinas, D.; Flunkert, V.; Gasthaus, J. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *Int. J. Forecast.* 2020, 36, 1181–1191.
67. Rangapuram, S.S.; Seeger, M.W.; Gasthaus, J.; Stella, L.; Wang, Y.; Januschowski, T. Deep State Space Models for Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* 2018, 31, 7796–7805.
68. Wen, R.; Torkkola, K.; Narayanaswamy, B.; Madeka, D. A Multi-Horizon Quantile Recurrent Forecaster. *arXiv* 2017, arXiv:1711.11053.
69. Lu, W.; Li, J.; Li, Y.; Sun, A.; Wang, J. A CNN-LSTM-Based Model to Forecast Stock Prices. *Complexity* 2020, 2020, 1–10.
70. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities. *Sensors* 2018, 18, 2220.
71. Kim, T.-Y.; Cho, S.-B. Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks. *Energy* 2019, 182, 72–81.
72. Williams, H.P. *Model Building in Mathematical Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2013; ISBN 1-118-50618-9.
73. Klee, V.; Minty, G.J. How Good Is the Simplex Algorithm. *Inequalities* 1972, 3, 159–175.
74. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. *arXiv* 2016, arXiv:1609.04747.
75. Dennis, J.; Moré, J.J. Quasi-Newton Methods, Motivation and Theory. *SIAM Rev.* 1977, 19, 46–89.

76. Shewchuk, J.R. An Introduction to the Conjugate Gradient Method without the Agonizing Pain; Carnegie-Mellon University, Department of Computer Science Pittsburgh: Pittsburgh, PA, USA, 1994.
77. DeVore, R.A.; Temlyakov, V.N. Some Remarks on Greedy Algorithms. *Adv. Comput. Math.* 1996, 5, 173–187.
78. Johnson, D.S.; Papadimitriou, C.H.; Yannakakis, M. How Easy Is Local Search? *J. Comput. Syst. Sci.* 1988, 37, 79–100.
79. Tsamardinos, I.; Brown, L.E.; Aliferis, C.F. The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm. *Mach. Learn.* 2006, 65, 31–78.
80. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In *Proceedings of the ICNN'95-International Conference on Neural Networks*, Perth, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.
81. Dorigo, M.; Birattari, M.; Stutzle, T. Ant Colony Optimization. *IEEE Comput. Intell. Mag.* 2006, 1, 28–39.
82. Kirkpatrick, S.; Gelatt, C.D., Jr.; Vecchi, M.P. Optimization by Simulated Annealing. *Science* 1983, 220, 671–680.
83. Glover, F.; Laguna, M. Tabu Search. In *Handbook of Combinatorial Optimization*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 2093–2229.
84. de Oliveira, M.P.V.; McCormack, K.; Trkman, P. Business Analytics in Supply Chains—The Contingent Effect of Business Process Maturity. *Expert Syst. Appl.* 2012, 39, 5488–5498.
85. Wu, P.-J.; Huang, P.-C. Business Analytics for Systematically Investigating Sustainable Food Supply Chains. *J. Clean. Prod.* 2018, 203, 968–976.
86. Trkman, P.; McCormack, K.; de Oliveira, M.P.V.; Ladeira, M.B. The Impact of Business Analytics on Supply Chain Performance. *Decis. Support Syst.* 2010, 49, 318–327.
87. Nam, D.; Lee, J.; Lee, H. Business Analytics Use in CRM: A Nomological Net from IT Competence to CRM Performance. *Int. J. Inf. Manag.* 2019, 45, 233–245.
88. Zhang, Z.; Xiao, Y.; Fu, Z.; Zhong, K.; Niu, H. A Study on Early Warnings of Financial Crisis of Chinese Listed Companies Based on DEA–SVM Model. *Mathematics* 2022, 10, 2142.
89. Zhou, W.; Chen, M.; Yang, Z.; Song, X. Real Estate Risk Measurement and Early Warning Based on PSO-SVM. *Socio-Econ. Plan. Sci.* 2021, 77, 101001.
90. Jianying, F.; Bianyu, Y.; Xin, L.; Dong, T.; Weisong, M. Evaluation on Risks of Sustainable Supply Chain Based on Optimized BP Neural Networks in Fresh Grape Industry. *Comput. Electron. Agric.* 2021, 183, 105988.

91. Jiang, H.; Ching, W.-K.; Yiu, K.F.C.; Qiu, Y. Stationary Mahalanobis Kernel SVM for Credit Risk Evaluation. *Appl. Soft Comput.* 2018, 71, 407–417.
92. Gerrard, M.; Gibbons, F.X.; Houlihan, A.E.; Stock, M.L.; Pomery, E.A. A Dual-Process Approach to Health Risk Decision Making: The Prototype Willingness Model. *Dev. Rev.* 2008, 28, 29–61.
93. Pröllochs, N.; Feuerriegel, S. Business Analytics for Strategic Management: Identifying and Assessing Corporate Challenges via Topic Modeling. *Inf. Manag.* 2020, 57, 103070.
94. van der Togt, J.; Rasmussen, T.H. Toward Evidence-Based HR. *JOEPP* 2017, 4, 127–132.
95. Margherita, A. Human Resources Analytics: A Systematization of Research Topics and Directions for Future Research. *Hum. Resour. Manag. Rev.* 2022, 32, 100795.
96. Stadler, J.G.; Donlon, K.; Siewert, J.D.; Franken, T.; Lewis, N.E. Improving the Efficiency and Ease of Healthcare Analysis Through Use of Data Visualization Dashboards. *Big Data* 2016, 4, 129–135.
97. Stelzer, G.; Rosen, N.; Plaschkes, I.; Zimmerman, S.; Twik, M.; Fishilevich, S.; Stein, T.I.; Nudel, R.; Lieder, I.; Mazor, Y.; et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinform.* 2016, 54, 1–30.
98. Fanelli, D.; Piazza, F. Analysis and Forecast of COVID-19 Spreading in China, Italy and France. *Chaos Solitons Fractals* 2020, 134, 109761.
99. Ward, M.J.; Marsolo, K.A.; Froehle, C.M. Applications of Business Analytics in Healthcare. *Bus. Horiz.* 2014, 57, 571–582.
100. Kaur, M.; Kang, S. Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Comput. Sci.* 2016, 85, 78–85.
101. Videla-Cavieres, I.F.; Ríos, S.A. Extending Market Basket Analysis with Graph Mining Techniques: A Real Case. *Expert Syst. Appl.* 2014, 41, 1928–1936.
102. Griva, A.; Bardaki, C.; Pramataris, K.; Papakiriakopoulos, D. Retail Business Analytics: Customer Visit Segmentation Using Market Basket Data. *Expert Syst. Appl.* 2018, 100, 1–16.
103. Hwangbo, H.; Kim, Y.S.; Cha, K.J. Recommendation System Development for Fashion Retail E-Commerce. *Electron. Commer. Res. Appl.* 2018, 28, 94–101.
104. Isinkaye, F.O.; Folajimi, Y.O.; Ojokoh, B.A. Recommendation Systems: Principles, Methods and Evaluation. *Egypt. Inform. J.* 2015, 16, 261–273.

Retrieved from <https://encyclopedia.pub/entry/history/show/94970>