The following table shows some important terms (with their definitions and in-text-references) in our target research paper (for big data understanding):

|  |  |  |
| --- | --- | --- |
| **Name of the term** | **Definition** | **In text reference** |
| Apache Drill | Apache Drill is a distributed system for interactive ad-hoc analysis of large-scale datasets. Designed to handle up to petabytes of data spread across thousands of servers, the goal of Drill is to respond to ad-hoc queries in a low-latency manner. | (Hausenblas & Nadeau, 2013) |
| Apache Hadoop | Open source software framework that enables distributed parallel processing of huge amounts of data across many inexpensive computers. | (Laudon and Laudon, 2018) |
| Artificial Intelligence | Artificial Intelligence (AI) is a branch of Computer Science, which is mainly concerned with automation of Intelligent behavior. This behavior we may consider from all domains—the human, animal world, and vegetation. | (Chowdhary, n.d.) |
| Autoencoder | An autoencoder is a type of artificial neural network used to learn efficient codings of unlabeled data (unsupervised learning). | (Kramer & Mark A., 1991) |
| Big Data | Data sets with volumes so huge that they are beyond the ability of typical relational DBMS to capture, store, and analyze. The data are often unstructured or semi-structured.  Size matters, but there are other important attributes of big data, namely data variety and data velocity. The three Vs of big data (volume, variety, and velocity) constitute a comprehensive definition, and they bust the myth that big data is only about data volume. In addition, each of the three Vs has its own ramifications for analytics. | (Laudon and Laudon, 2018), (SAS, 2011) |
| Big Data Analytics | Big data analytics is where advanced analytic techniques operate on big data. | (SAS, 2011) |
| Big Data Analytics Techniques | Big data analytics techniques involve the processing of data from different sources in different formats, with different data processing schemes (and correspondingly with different technological supporting platforms). | (Choi et al., 2018) |
| Big Data Analytics Strategies | Big data analytics faces various challenges which make them different from the typical data analytics. From the data side, these challenges include having a massive amount of data points (big data volume, high dimension), the presence of complex data (high variety of data with different classes and types), and the existence of high uncertainty. From the computing side, many existing methods are not flexible enough and “unscalable” to adapt to the requirements of big data. They also suffer the curse of dimensionality in which they cannot cope with huge-dimensional problems. To overcome these challenges, a few critical strategies are there which are called big data analytics strategies. | (Choi et al., 2018) |
| Big Data Architectures | Big Data cannot be solved effectively and approvingly if there is no good and proper architecture for the whole Big Data system. The proper architecture of the whole big data system is called big data architecture. | (Philip Chen & Zhang, 2014) |
| Big Data Methods | Big Data Methods include techniques, strategies and architectures for big data analytics. | (Choi et al., 2018) |
| Big Data Value Chain | The Big Data Value Chain is introduced to describe the information flow within a big data system as a series of steps needed to generate value and useful insights from data. | (Cavanillas et al., 2016) |
| Batch Processing | In batch processing, data is inserted into a database and polled at regular intervals for further analysis. | (Cavanillas et al., 2016) |
| Business Intelligence | Applications and technologies to help users make better business decisions. | (Laudon and Laudon, 2018) |
| Bayesian Statistics | Bayesian statistical methods are often presented in the form of an inference. The inference runs from a so-called prior probability distribution over statistical hypotheses, which expresses the degree of belief in the hypotheses before data has been collected, to a posterior probability distribution over the hypotheses, which expresses the beliefs after the data have been incorporated. The defining characteristic of Bayesian statistics is that it considers probability distributions over statistical hypotheses as well as over data. | (Jan-Willem, 2014) |
| Classification | It is a task of training a model based on labeled training data, such that the model can later be used to assign predefined class labels to new objects. | (Chowdhary, n.d.) |
| Cloud Computing | Model of computing in which computer processing, storage, software, and other services are provided as a shared pool of virtualized resources over a network, primarily the Internet. | (Laudon and Laudon, 2018) |
| Clustering | The process of clustering partitions a set of data, according to some similarity measure, into several groups such that “similar” records are in the same group, so that each group represents a similar subpopulation in the data. | (Chowdhary, n.d.) |
| Computer Vision | Computer Vision or Machine Version is concerned with analysis, modification, and understanding of images. The objective of this field is to understand what is happening in front of a camera, and use that understanding to control a computer or robotic system, or making use of these images provide people with new images that are more informative or aesthetically more pleasing than the original images. | (Chowdhary, n.d.) |
| Correlation | ‘‘Correlation’’ often is used loosely to mean data items that provide information about other data items—specifically, known quantities that reduce our uncertainty about unknown quantities. | (Provost and Fawcett, 2013) |
| Data | Data, in contrast, are streams of raw facts representing events occurring in organizations or the physical environment before they have been organized and arranged into a form that people can understand and use. | (Laudon and Laudon, 2018) |
| Data Analysis | Data Analysis is concerned with making the raw data acquired amenable to use in decision-making as well as domain-specific usage. Data analysis involves exploring, transforming, and modelling data with the goal of highlighting relevant data, synthesising and extracting useful hidden information with high potential from a business point of view. | (Cavanillas et al., 2016) |
| Data Analytics | Data analytics is defined as the application of computer systems to the analysis of large data sets for the support of decisions. | (Runkler, 2021) |
| Data Acquisition | Data Acquisition is the process of gathering, filtering, and cleaning data before it is put in a data warehouse or any other storage solution on which data analysis can be carried out. | (Cavanillas et al., 2016) |
| Data Driven Decision Making | Data-driven decision making (DDD)3 refers to the practice of basing decisions on the analysis of data rather than purely on intuition. | (Provost and Fawcett, 2013) |
| Data Envelopment Analysis (DEA) | DEA offers an insight into the relative efficiency of comparable ‘decision-making units' such schools. It offers a conservative estimate of comparative efficiency in situations in which multiple inputs and multiple outputs are found. The technique exploits limited data to its full extent, and can be used to examine technical efficiency, allocative efficiency and scale efficiency. | (Charnes et al., 1997) |
| Data Format | Data comes in many forms and these forms are called data formats. | (Cavanillas et al., 2016) |
| Data Mining | Data mining, or knowledge discovery in databases, provides the tools to sift through the vast data stores to find the trends, patterns, and correlations that can guide strategic decision-making. | (Chowdhary, n.d.) |
| Data Privacy | When talking about processing, integrating, or sharing medical data, a strong emphasis must be put on data privacy. | (Cavanillas et al., 2016) |
| Data Science | Data science is a set of fundamental principles that support and guide the principled extraction of information and knowledge from data. Possibly the most closely related concept to data science is data mining—the actual extraction of knowledge from data via technologies that incorporate these principles. Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data. | (Provost and Fawcett, 2013) |
| Data Storage | Data Storage is the persistence and management of data in a scalable way that satisfies the needs of applications that require fast access to the data. | (Cavanillas et al., 2016) |
| Data Usage | Data Usage covers the data-driven business activities that need access to data, its analysis, and the tools needed to integrate the data analysis within the business activity. | (Cavanillas et al., 2016) |
| Database | A collection of data organized to service many applications at the same time by storing and managing data so that they appear to be in one location. | (Laudon and Laudon, 2018) |
| Decision Support System | Information systems at the organization’s management level that combine data and sophisticated analytical models or data analysis tools to support semi-structured and unstructured decision making. | (Laudon and Laudon, 2018) |
| Deep machine learning | Deep learning is concerned with simulation of ANNs that “learn” gradually. The basic technique relies on ANNs, which do not precisely mimic how the actual neurons are working. Instead of this they are based on the general principles of mathematics that allow them to learn from examples e.g. to recognize people or objects in a photograph, and translate the spoken language from one to another etc. | (Chowdhary, n.d.) |
| Distributed Machine Learning | Distributing the machine learning workload across multiple machines, and turning the centralized into a distributed system. | (Verbraeken et al., 2020) |
| Dremel by Google | Dremel is a scalable, interactive ad-hoc query system for analysis of read-only nested data. By combining multi-level execution trees and columnar data layout, it is capable of running aggregation queries over trillion-row tables in seconds. | (Melnik et al., 2010) |
| Dryad | Dryad is a programming model used for data flow graph processing. | (Godara & Poonia, 2018) |
| Dynamic Programming | Dynamic Programming is a recursive method for solving sequential decision problems (hereafter abbreviated as SDP). Also known as backward induction. | (Rust, 2008) |
| Enterprise Resources Planning (ERP) | Enterprise Resource Planning (ERP) systems are used to integrate business processes in manufacturing and production, finance and accounting, sales and marketing, and human resources into a single software system. Information that was previously fragmented in many different systems is stored in a single comprehensive data repository where it can be used by many different parts of the business. | (Laudon and Laudon, 2018) |
| Fast Learning | In Fast Learning, input weights and hidden layer biases are randomly generated, and the weight values of the connection between the output layer and the input layer and the weight values connecting the output node and the input nodes are analytically determined based on least squares methods. | (Li, Niu, Duan and Zhang, 2013) |
| Feature Selection | Feature selection process helps in identifying the most effective subset of the original features to use. | (Chowdhary, n.d.) |
| Heuristics | The heuristics is defined in the form of a function, say f , which somehow represents the mapping to the total distance between the start node and the goal node. For any given node n, the total distance between start and goal node is f (n), such that f (n) = g(n) + h(n), (9.1) where g(n) is distance between start node and the node n, and h(n) is the distance between node n and the goal node. There exist two basic approaches to heuristics, “Divide and conquer” and “iterative improvement”. | (Chowdhary, n.d.) |
| Incremental Learning | When the training set is large, all the examples cannot be loaded into memory at one go. One approach to overcome this constraint is to train the classifier (machine learning model) using an incremental learning technique, whereby only subsets of the data are to be considered at any one time and results subsequently combined. Incremental learning techniques are one possible solution to the scalability problem, where data is processed in parts, and the results combined so as to use less memory. | (Syed et al., 1999) |
| Internet of Things (IoT) | Pervasive web in which each object or machine has a unique identity and is able to use the Internet to link with other machines or send data. Also known as the Industrial Internet. | (Laudon and Laudon, 2018) |
| Machine Learning | Machine learning aims to create theories and procedures—learning algorithms—that allow machines to learn. | (Chowdhary, n.d.) |
| MapReduce | MapReduce is a popular framework for data-intensive distributed computing of batch jobs. MapReduce has emerged as a popular way to harness the power of large clusters of computers. MapReduce allows programmers to think in a data-centric fashion: they focus on applying transformations to sets of data records, and allow the details of distributed execution, network communication and fault tolerance to be handled by the MapReduce framework. | (Condie et al., 2010) |
| Multivariate Statistical Analysis | Multivariate analysis generally refers to a range of statistical techniques/methods which primarily involves data with several variables, with the objective of investigating the dependence structure or relations between the involved variables. | (Chen, 2005) |
| Neural Networks | Hardware or software that attempts to emulate the processing patterns of the biological brain. | (Laudon and Laudon, 2018) |
| NP hard problems | A problem H is NP-hard (where, NP (nondeterministic polynomial time) is a [complexity class](https://en.wikipedia.org/wiki/Complexity_class) used to classify [decision problems](https://en.wikipedia.org/wiki/Decision_problem)) when every problem L in NP can be reduced in polynomial time to H; that is, assuming a solution for H takes 1 unit time, H‎'s solution can be used to solve L in polynomial time. | (Leeuwen, 1990) |
| Numerical Methods | Methods designed for the constructive solution of mathematical problems requiring particular numerical results, usually on a computer. A [numerical method](https://www.encyclopedia.com/earth-and-environment/ecology-and-environmentalism/environmental-studies/numerical-method) is a complete and unambiguous set of procedures for the solution of a problem, together with computable error estimates. | ("numerical methods", 2021) |
| Ontology Learning | Ontology Learning aims at the integration of a multitude of disciplines in order to facilitate the construction of ontologies (Ontologies constitute a formal conceptualization of a particular domain of interest that is shared by a group of people), in particular ontology engineering and machine learning. | (Staab et al., 2004) |
| Parallel Support Vector Machines | An algorithm for support vector machines (SVM) that can be parallelized efficiently and scales to very large problems with hundreds of thousands of training vectors. Instead of analyzing the whole training set in one optimization step, the data are split into subsets and optimized separately with multiple SVMs. The partial results are combined and filtered again in a ‘Cascade’ of SVMs, until the global optimum is reached. | (Parallel support vector machines: The cascade svm, 2004) |
| Pentaho | Pentaho is a business intelligence software used to generate reports for huge volumes of unstructured and structured data. It uses a big data platform to provide business services in the context of data integration, virtualization and exploration of data. | (Godara & Poonia, 2018) |
| Real-time Stream Processing | Big data processing was initially carried out in batches of historical data. In recent years, stream processing systems such as Apache Storm have become available and enable new application capabilities. This technology is relatively new and needs to be developed further. | (Cavanillas et al., 2016) |
| Regression | The regression is the same as classification, except that the dependent variable values take real values. | (Chowdhary, n.d.) |
| S4 | S4 is a general-purpose, distributed, scalable, partially fault-tolerant, pluggable platform that allows programmers to easily develop applications for processing continuous unbounded streams of data. | (Neumeyer et al., 2010) |
| SAP HANA | SAP HANA is a pioneering, and one of the best performing, data platform designed from the grounds up to heavily exploit modern hardware capabilities, including SIMD, and large memory and CPU footprints. As a comprehensive data management solution, SAP HANA supports the complete data life cycle encompassing modeling, provisioning, and consumption. SAP HANA has triggered a major shift in the database industry from the classical disk-centric database system design to a groundbreaking main-memory centric system design. | (Sikka et al., 2013) |
| Sentiment Analysis | Mining text comments in an email message, blog, social media conversation, or survey form to detect favorable and unfavorable opinions about specific subjects. | (Laudon and Laudon, 2018) |
| Social Analytics | Social actors' everyday use and reflections on analytics, that is any digital tool that measures them and their presence in the world of online presence is called social analytics. | (Hanquinet & Savage, 2015) |
| Social Media | Internet-based, disentrained, and persistent channels of masspersonal communication facilitating perceptions of interactions among users, deriving value primarily from user-generated content. | (Carr & Hayes, 2015) |
| Statistics | Statistics is a mathematical and conceptual discipline that focuses on the relation between data and hypotheses. Statistical methods provide the mathematical and conceptual means to evaluate statistical hypotheses in the light of a sample. A statistical hypothesis is a general statement that can be expressed by a probability distribution over sample space, i.e., it determines a probability for each of the possible samples. | (Jan-Willem, 2014) |
| Statistical Machine Learning | Statistical machine learning helps the programs automatically learn from the data. This is an attractive option as an alternative to manually coding every rule in a program. | (Chowdhary, n.d.) |
| Structured Data | Structured data, which constitutes only 5% of all existing data (Cukier, 2010), refers to the tabular data found in spreadsheets or relational databases. | (Jan-Willem, 2014) |
| Supervised Learning | A supervised learning method uses machine learning techniques to induce a classifier from sense-annotated data sets. In general, a learning method is called supervised learning if the learning process requires some intervention from the user. Some approaches in supervised learning require the user to provide training examples. | (Chowdhary, n.d.) |
| Support Vector Machine | An SVM is a model where examples are nothing but representation of points in space that are mapped into two classes, such that the examples of the separate categories are divided by a clear gap that should be as wide as possible. Once the learning has taken place, any amount of new examples are then mapped into that same space and predicted to belong to a category depending on which side of the gap they fall on. | (Laudon and Laudon, 2018) |
| Unstructured Data | Unstructured data lack the structural organization required by machines for analysis e.g. text, images, audio, and video etc. | (Jan-Willem, 2014) |
| Unsupervised learning | This learning eliminates the need of a teacher, and the learner is solely responsible to form his own concepts and evaluate these for learning. In fact, the unsupervised methods have to discover the concept classes to which the given examples belong. The discovery-based learning is also a category of unsupervised learning. | (Chowdhary, n.d.) |
| Veracity | Represents the unreliability inherent in some sources of data. For example, customer sentiments in social media are uncertain in nature, since they entail human judgment. Yet they contain valuable information. | (Jan-Willem, 2014) |
| Visualization Analysis | Analysis for helping users see patterns and relationships in large amounts of data by presenting the data in graphical form. | (Laudon and Laudon, 2018) |
| Web Analytics | Basic and advanced web analytics are collectively we analytics. Basic Web metrics are generally accepted as the start of the Web analytics concept. Advanced web analytics aims to measure and understand the relationship between the customer and the Web site. | (Phippen et al., 2004) |
| Wireless Sensor Networks | A Wireless Sensor Network is simply what it says: a network of sensing devices connected together, and usually to some kind of a base station, by wireless means. What is sensed, how often, how much preprocessing is done, and how often the values are communicated with the base station are all variables which have to be traded off against the use of power. | (Guy, 2006) |

## 

## 

## **References:**

Allodi, L., & Massacci, F. (2017). Security Events and Vulnerability Data for Cybersecurity Risk Estimation. *Risk Analysis*, *37*(8), 1606-1627. https://doi.org/10.1111/risa.12864

Aral, S., & Walker, D. (2014). Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment. *Management Science*, *60*(6), 1352-1370. https://doi.org/10.1287/mnsc.2014.1936

Badiezadeh, T., Saen, R., & Samavati, T. (2018). Assessing sustainability of supply chains by double frontier network DEA: A big data approach. *Computers & Operations Research*, *98*, 284-290. https://doi.org/10.1016/j.cor.2017.06.003

Baughman, A., Bogdany, R., Harrison, B., O’Connell, B., Pearthree, H., & Frankel, B. et al. (2016). IBM Predicts Cloud Computing Demand for Sports Tournaments. *Interfaces*, *46*(1), 33-48. https://doi.org/10.1287/inte.2015.0820

Bertsimas, D., Kallus, N., & Hussain, A. (2016). Inventory Management in the Era of Big Data. *Production And Operations Management*, *25*(12), 2006-2009. https://doi.org/10.1111/poms.2\_12637

Biffis, E., & Chavez, E. (2017). Satellite Data and Machine Learning for Weather Risk Management and Food Security. *Risk Analysis*, *37*(8), 1508-1521. https://doi.org/10.1111/risa.12847

Carr, C., & Hayes, R. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal Of Communication*, *23*(1), 46-65. https://doi.org/10.1080/15456870.2015.972282

Cavanillas, J., Curry, E., & Wahlster, W. (2016). *New Horizons for a Data-Driven Economy*. Springer.

Charnes, A., Cooper, W., Lewin, A., & Seiford, L. (1997). Data Envelopment Analysis Theory, Methodology and Applications. *Journal Of The Operational Research Society*, *48*(3), 332-333. https://doi.org/10.1057/palgrave.jors.2600342

Chen, K. (2005). *Multivariate Statistical Analysis*.

Choi, T., Wallace, S., & Wang, Y. (2018). Big Data Analytics in Operations Management. *Production And Operations Management*, *27*(10), 1868-1883. https://doi.org/10.1111/poms.12838

Chong, A., Ch’ng, E., Liu, M., & Li, B. (2015). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal Of Production Research*, *55*(17), 5142-5156. https://doi.org/10.1080/00207543.2015.1066519

Chowdhary, K. *Fundamentals of artificial intelligence*.

Chung, S., Ma, H., & Chan, H. (2016). Cascading Delay Risk of Airline Workforce Deployments with Crew Pairing and Schedule Optimization. *Risk Analysis*, *37*(8), 1443-1458. https://doi.org/10.1111/risa.12746

Condie, T., Conway, N., Alvaro, P., & Hellerstein, J. (2010). MapReduce online. In *NASDI'10*. San Jose, California. Retrieved 7 July 2021.

Cui, R., Gallino, S., Moreno, A., & Zhang, D. (2017). The Operational Value of Social Media Information. *Production And Operations Management*, *27*(10), 1749-1769. https://doi.org/10.1111/poms.12707

Culotta, A., & Cutler, J. (2016). Mining Brand Perceptions from Twitter Social Networks. *Marketing Science*, *35*(3), 343-362. https://doi.org/10.1287/mksc.2015.0968

Ferreira, K., Lee, B., & Simchi-Levi, D. (2016). Analytics for an Online Retailer: Demand Forecasting and Price Optimization. *Manufacturing & Service Operations Management*, *18*(1), 69-88. https://doi.org/10.1287/msom.2015.0561

Godara, J., & Poonia, R. (2018). Big Data Analytics: A Review. *International Journal Of Research And Analytical Reviews (IJRAR)*, *5*(4), 933-937. Retrieved 7 July 2021, from.

Guy, C. (2006). Wireless sensor networks. *Sixth International Symposium On Instrumentation And Control Technology: Signal Analysis, Measurement Theory, Photo-Electronic Technology, And Artificial Intelligence*. https://doi.org/10.1117/12.716964

Hanquinet, L., & Savage, M. (2015). *Routledge International Handbook of the Sociology of Art and Culture*. Routledge

Hausenblas, M., & Nadeau, J. (2013). Apache Drill: Interactive Ad-Hoc Analysis at Scale. *Big Data*, *1*(2), 100-104. https://doi.org/10.1089/big.2013.0011

Huang, T., & Van Mieghem, J. (2013). Clickstream Data and Inventory Management: Model and Empirical Analysis. *Production And Operations Management*, *23*(3), 333-347. https://doi.org/10.1111/poms.12046

Jamshidi, A., Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., & Dollevoet, R. et al. (2017). A Big Data Analysis Approach for Rail Failure Risk Assessment. *Risk Analysis*, *37*(8), 1495-1507. https://doi.org/10.1111/risa.12836

Jan-Willem, R. (2014). Philosophy of Statistics. In *Stanford Encyclopedia of Philosophy* (5th ed.). Metaphysics Research Lab, Stanford University.

Kaur, H., & Singh, S. (2018). Heuristic modeling for sustainable procurement and logistics in a supply chain using big data. *Computers & Operations Research*, *98*, 301-321. https://doi.org/10.1016/j.cor.2017.05.008

Kramer, & Mark A. (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *Aiche Journal*, *37*(2), 233–243. https://doi.org/10.1002%2Faic.690370209

Lau, R., Zhang, W., & Xu, W. (2017). Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data. *Production And Operations Management*, *27*(10), 1775-1794. https://doi.org/10.1111/poms.12737

Laudon, K., & Laudon, J. (2018). *Management Information Systems* (5th ed.). Pearson.

Leeuwen, J. (1990). *Handbook of theoretical computer science* (1st ed.). Elsevier.

Li, G., Niu, P., Duan, X., & Zhang, X. (2013). Fast learning network: a novel artificial neural network with a fast learning speed. *Neural Computing And Applications*, *24*(7-8), 1683-1695. https://doi.org/10.1007/s00521-013-1398-7

Li, L., Chi, T., Hao, T., & Yu, T. (2016). Customer demand analysis of the electronic commerce supply chain using Big Data. *Annals Of Operations Research*, *268*(1-2), 113-128. https://doi.org/10.1007/s10479-016-2342-x

Liu, X., Singh, P., & Srinivasan, K. (2016). A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing. *Marketing Science*, *35*(3), 363-388. https://doi.org/10.1287/mksc.2015.0972

López‐Cuevas, A., Ramírez‐Márquez, J., Sanchez‐Ante, G., & Barker, K. (2017). A Community Perspective on Resilience Analytics: A Visual Analysis of Community Mood. *Risk Analysis*, *37*(8), 1566-1579. https://doi.org/10.1111/risa.12788

Lorca, Á., Çelik, M., Ergun, Ö., & Keskinocak, P. (2016). An Optimization‐Based Decision‐Support Tool for Post‐Disaster Debris Operations. *Production And Operations Management*, *26*(6), 1076-1091. https://doi.org/10.1111/poms.12643

Lu, S., Xiao, L., & Ding, M. (2016). A Video-Based Automated Recommender (VAR) System for Garments. *Marketing Science*, *35*(3), 484-510. https://doi.org/10.1287/mksc.2016.0984

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. (2014). Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions On Intelligent Transportation Systems*, 1-9. https://doi.org/10.1109/tits.2014.2345663

Melnik, S., Gubarev, A., Long, J., Romer, G., Shivakumar, S., Tolton, M., & Vassilakis, T. (2010). Dremel: interactive analysis of web-scale datasets. *Proceedings Of The VLDB Endowment*, *3*(1-2), 330-339. https://doi.org/10.14778/1920841.1920886

Mukherjee, U., & Sinha, K. (2017). Product Recall Decisions in Medical Device Supply Chains: A Big Data Analytic Approach to Evaluating Judgment Bias. *Production And Operations Management*, *27*(10), 1816-1833. https://doi.org/10.1111/poms.12696

Neumeyer, L., Robbins, B., Nair, A., & Kesari, A. (2010). S4: Distributed Stream Computing Platform. *2010 IEEE International Conference On Data Mining Workshops*. https://doi.org/10.1109/icdmw.2010.172

Papadopoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S., & Fosso-Wamba, S. (2017). The role of Big Data in explaining disaster resilience in supply chains for sustainability. *Journal Of Cleaner Production*, *142*, 1108-1118. https://doi.org/10.1016/j.jclepro.2016.03.059

Parallel support vector machines: The cascade svm. (2004) (pp. 521-528). Retrieved 6 July 2021, from.

Philip Chen, C., & Zhang, C. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314-347. https://doi.org/10.1016/j.ins.2014.01.015

Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research*, *14*(4), 284-293. https://doi.org/10.1108/10662240410555306

Provost, F., & Fawcett, T. (2013). Comment on Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51-59. https://doi.org/10.1089/big.2013.1508

Runkler, T. (2021). *Data Analytics* (2nd ed.). Springer.

Rust, J. (2008). Dynamic Programming. *The new Palgrave dictionary of economics*. Retrieved 6 July 2021, from.

Romero Morales, D., & Wang, J. (2010). Forecasting cancellation rates for services booking revenue management using data mining. *European Journal Of Operational Research*, *202*(2), 554-562. https://doi.org/10.1016/j.ejor.2009.06.006

SAS. (2011). *BIG DATA ANALYTICS*. Retrieved from https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf

See-To, E., & Ngai, E. (2016). Customer reviews for demand distribution and sales nowcasting: a big data approach. *Annals Of Operations Research*, *270*(1-2), 415-431. https://doi.org/10.1007/s10479-016-2296-z

Sikka, V., Färber, F., Goel, A., & Lehner, W. (2013). SAP HANA: The Evolution from a Modern Main-Memory Data Platform to an Enterprise Application Platform. *Proceedings Of The VLDB Endowment*, *6*(11), 1184-1185. https://doi.org/10.14778/2536222.2536251

Staab, S., Alexander, & Maedche. (2004). *Handbook on Ontologies* (pp. 173-190). Springer.

Sagaert, Y., Aghezzaf, E., Kourentzes, N., & Desmet, B. (2018). Temporal Big Data for Tactical Sales Forecasting in the Tire Industry. *Interfaces*, *48*(2), 121-129. https://doi.org/10.1287/inte.2017.0901

Shang, Y., Dunson, D., & Song, J. (2017). Exploiting Big Data in Logistics Risk Assessment via Bayesian Nonparametrics. *Operations Research*, *65*(6), 1574-1588. https://doi.org/10.1287/opre.2017.1612

Syed, N., Liu, H., & Sung, K. (1999). Incremental Learning with Support Vector Machines. https://doi.org/10.1.1.46.6367

Van Jaarsveld, W., & Scheller-Wolf, A. (2015). Optimization of Industrial-Scale Assemble-to-Order Systems. *INFORMS Journal On Computing*, *27*(3), 544-560. https://doi.org/10.1287/ijoc.2015.0641

Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., & Rellermeyer, J. (2020). A Survey on Distributed Machine Learning. *ACM Computing Surveys*, *53*(2), 1-33. https://doi.org/10.1145/3377454

Wang, G., Gunasekaran, A., & Ngai, E. (2016). Distribution network design with big data: model and analysis. *Annals Of Operations Research*, *270*(1-2), 539-551. https://doi.org/10.1007/s10479-016-2263-8

Xie, K., Ozbay, K., Kurkcu, A., & Yang, H. (2017). Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots. *Risk Analysis*, *37*(8), 1459-1476. https://doi.org/10.1111/risa.12785

Xue, Z., Wang, Z., & Ettl, M. (2016). Pricing Personalized Bundles: A New Approach and An Empirical Study. *Manufacturing & Service Operations Management*, *18*(1), 51-68. https://doi.org/10.1287/msom.2015.0563