ISSN 2661-2666 (Online) International Scientific Journal Monte (ISJM) DOI: 10.33807/monte.20233016 Volume 8, (No).1 (2023): December

THE IMPORTANCE OF DATA SCIENCE IN SUPPORTING BUSINESS STRATEGIES

Ejup Rustemi[1], Mefail Tahiri[2]*
[1]University of Tetova, North Macedonia
[2]University of Tetova, North Macedonia
[*]ejup.rustemi@yahoo.com

ABSTRACT

Our society is data-driven. The digital revolution, where enormous amounts of data are produced daily, is the result of the hunt for novel solutions and competitive advantage. The processing and analysis of this vast data set, gathered from various sources and comprising inconsistent, inaccurate, insufficient, and deceptive information, presents significant challenges for today's businesses. Due to the variety of data sources (such as the combination of text, graphics, audio, and video) and the speed at which real-time data is generated, gathered, and processed, these databases include a tremendous lot of complexity. In order to assess this massive amount of structured and unstructured data, new data analytics patterns and trends are required. Big Data is a vast collection of information-rich data that may be roughly categorized into five dimensions: I volume: describing the data's substantial size; I velocity: the rate at which new data is generated and analyzed in real time; (ii) variety: representing unstructured data from heterogeneous sources such as traditional databases, transactional systems, the Internet, and social media; (iii) veracity: the authenticity or origin of the data; and (iv) value: the value of the data, which depends on whether the data collected is old or recent. This paper will emphaisze the importance and challenges of data science inside the business landscape.

Keywords: *data, data science, business, strategy, big data.*

INTRODUCTION

The hardware and software requirements to handle such a vast volume of data from many sources are one of the main issues encountered during the implementation of Big Data in industrial processes and applications. Complex industrial applications still require more advanced technologies to handle real-time requirements, even though the development of advanced technologies such as cloud computing, virtualization of processes and storages, new programming and data processing frameworks like MapReduce and Hadoop has been somewhat successful.

Dual channel operation in a competitive environment not only necessitates service level improvement, but also collaborative service between channels, amplifying the rivalry between the manufacturer and the retailer and promoting cooperation throughout the entire supply chain. This paper builds the same services and differentiation service model based on the centralized decision and decentralized decision, through the theoretical model and numerical computation to explore the various service levels and network channels that effect on demands, based on the service level under the dual channel supply chain, considering that cooperative strategy between network direct channels and retail channels, to realize supply chain coordination (Xi Liu, 2017).

For a very long time, coordination in dual-channel supply chains has been studied. In order to make the best choice, Huang and Swaminathan take into account a supply chain that includes both traditional and Internet channels. The wholesale price contract can coordinate a dual-channel supply chain, according to Chen Zhang and Sun's analysis of the supplier's pricing strategy. David and Adida investigate competition and cooperation in a supply chain where a single supplier runs both a direct channel and sells its goods through a variety of shops with varying specialties.

In reality, the majority of providers sell through multiple retail channels. With a discount pricing policy, Sinha and Santanu optimize the supply chain of a single vendor and multiple buyers. Mateen and Chatterjee create analytical models for several strategies that can be used to coordinate a single supplier-multiple retailers system using vendor managed inventory.

The market's rapid globalization has resulted in a significant shift away from competition between governments toward competition between businesses, particularly brands. The competitiveness of an enterprise's core is determined by its brand. The United States typically holds 227 positions in the top 500 global brands, maintaining a position of considerable power, whereas Chinese businesses hold about 36 positions. It is clear that Chinese businesses need to increase their brand competitiveness if they want to stand out on the global market (Zhi Li, 2018).

THE DIGITALIZATION OF BUSINESS

E-commerce has advanced quickly in recent years, with a ten-fold growth in transaction volume over the previous ten years. In the interim, "Internet plus" has emerged as the industry's fastest-growing sector, having an impact on all facets of the sector, including the retail network, international e-commerce, online services, and Internet banking. These sectors have emerged as a new economic strength and development driver. Among these, network retail sales saw a significant uptick.

Application-focused institutions will inevitably encounter fresh prospects for change in the e-commerce major they teach. Every undergraduate institution used to provide similar e-commerce-related courses in the past, and the subject matter was comparatively dated. At the same time, instructional technology was rather antiquated, and teaching techniques were likewise. Additionally, there were not enough supporting materials for teaching materials that were up to date with the times. When it comes to hiring, many e-commerce businesses frequently choose to hire seasoned professionals over graduate students. The primary practical ability of e-commerce is much less than what businesses require, which is the cause.

The difficulties with practical instruction prevent e-commerce from developing more quickly and effectively. the initiative to introduce the e-commerce industry to construct a school for students to carry out training, practicing, and school productive training foundation The reform of teaching electronic commerce majors to serve local economic growth as the aim, with skill training as the important point. Incorporate the notion that "education helps the enterprise." In order to fulfill the needs of the enterprise, the existing teaching method must be modified and the teaching materials must be organized in accordance with the actual combat items used by the enterprise. Establish a training center with an entrepreneurial culture to boost students' professional success. One key component of cutting-edge vocational education is the integration of production and instruction. Encourage students to acquire real business skills for e-commerce companies, including application proficiency and the capacity to manage a functional platform for conducting business (Wang, 2017).

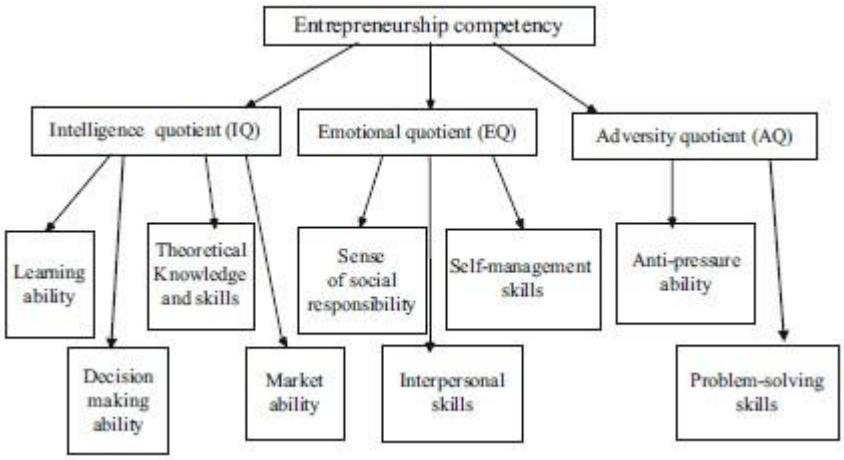

Fig 1. *Enterpreneurship competency model*[1]

To assess and enhance organizational performance, organizations might utilize the American Customer Satisfaction (ACSI) evaluation model. The six structure variables mentioned above make up the ACSI structural model. Consumer expectations, customer expectations, and perceived value are the causation variables, consumer complaints, and customer loyalty are the outcome variables, and customer happiness is the end target variable. Each of the six structural variables comprises one or more observed variables, and the information about the observed variables is typically gathered through actual research.

The ACSI technique may be used to evaluate customer satisfaction across all business types, industries, and departments, and the findings are equivalent. It allows for comparison of customer satisfaction across industries, as well as between various customers utilizing the same product.

As a result, the measurement of customer satisfaction can be expanded from the fields of engineering and environmental quality to those of product and service quality (Tong, Jia, 2017).

The stock market uses a significant amount of raw data that changes over time. Additionally, judgments about stock investments must take into account a variety of aspects, including the state of the economy on a national, regional, and international level, among others. Future investments will particularly benefit from analytics on pertinent historical data. Despite the fact that the stock market system is very challenging to accurately predict due to its unique data complexity features. As a result, forecasting stock investment returns is crucial in the financial industry.

Techniques for data analytics and data visualization have been used in stock market analysis. Predictive analytics rely on data processing because it has the pre/post processing capabilities and involves changing raw data. Its features include association analysis, cleaning, and the identification of data attribute descriptions, among others. On the other side, data visualization techniques give complex data visual representations, which is one of the most effective ways to help investors have a clear understanding of stock market movements (Zreika, Hua, Wang, 2018).

[1] Chen, Ch., Wang, J. (2018). The Construction of University Students' Entrepreneurship Competency Model in Application-Oriented Universities (100-111. Recent Developments in Data Science and Business Analytics (ICDSB – 2017)

ISSN 2661-2666 (Online) International Scientific Journal Monte (ISJM) DOI:
10.33807/monte.20233016 Volume 8, (No).1 (2023): December

THE ERA OF MASS INFORMATION

The need for mass information processing and analysis has outgrown the capabilities of the traditional model of data analysis software. As a result, it is increasingly crucial to develop new data models and conduct data analysis and processing. This is because as data volumes and data growth increase, redundancy will also rise (Wang, Li, 2017). The following four categories of data-quality issues are present in data throughout its lifecycle, through human contact, computational processes, and communication methods, each of which can add errors and produce anomalous data: Data input and update, measurement, simplification, and integration are the first four steps.

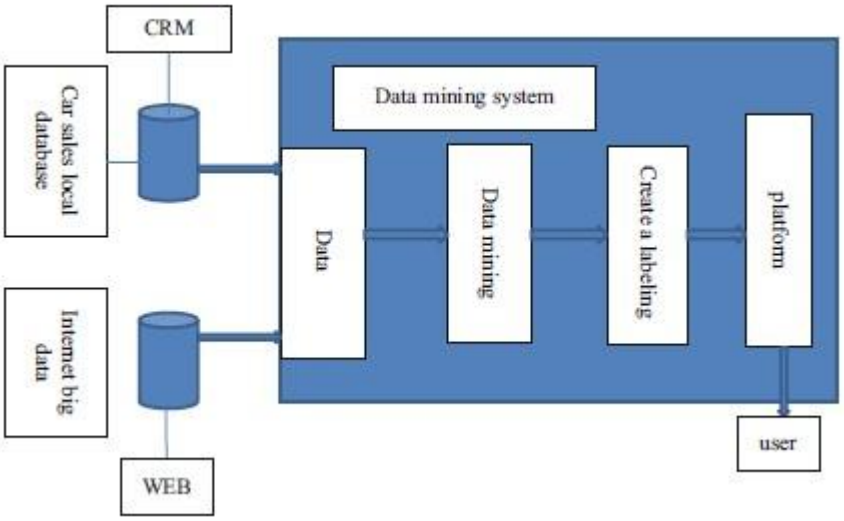

Fig 2. *Automotive industry intelligent sales platform*[2]

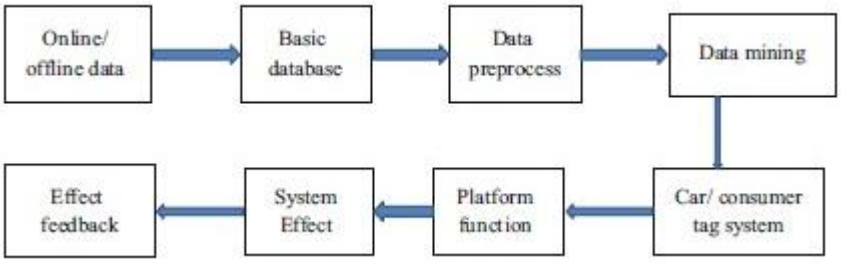

Fig 3. *Automotive industry intelligent sales platform chart*[3]

[2] Lee, J. (2018). Research on Intelligent Sales Platform of Automobile Industry Based on Large Data Mining (299-310. Recent Developments in Data Science and Business Analytics (ICDSB – 2017)

[3] Lee, J. (2018). Research on Intelligent Sales Platform of Automobile Industry Based on Large Data Mining (299-310. Recent Developments in Data Science and Business Analytics (ICDSB – 2017)

A relational database is constructed using the database's relational model. To deal with the data in the database, it makes use of mathematical ideas and techniques. Data in the form of images, videos, and other types of data storage underwent tremendous expansion in recent years with the rise of web2.0. The system's requirements have not been able to be satisfied by the current data storage techniques. Because of its scalability, NoSQL's distributed massive data storage technology is commonly used in cloud computing. Unstructured data storage issues can be resolved, and the quality of unstructured data storage services, by using cloud storage of unstructured storage technology. The significance and value of cloud databases are becoming more apparent as a result of the development of cloud computing.

Customers who purchase software as a service (SaaS) can access the program but not the hardware, operating system, or operational network infrastructure. This is the foundation of a service concept; rather than offering a notion of purchasing customer service, the software service provider offers the concept of leasing customer service. The more typical approach is to offer a set of account passwords.

Platform as a Service (PaaS): Users run applications on the host. Consumers have no control over the operating system, hardware, or operational network infrastructure, but they can modify the environment in which the application is executing. An application infrastructure typically makes up a platform.

Consumers employ fundamental computer resources like processor power, storage capacity, network components, or middleware when they use infrastructure as a service (IaaS). Customers can manage the operating system, storage capacity, running applications, and network parts (such load balancers, firewalls, etc.), but not the cloud architecture.

The future of cloud databases is highly bright in the current era of the data explosion. The study paper predicts that enterprise storage needs for structured data will keep rising in the future. In the case of small-scale applications, the system load fluctuates due to the redundant resources the system must manage, but in the case of large-scale applications, not only are there significant data storage requirements but also a dynamic demand for resources, which results in an increase or decrease in the number of virtual machines. The huge data storage, cost-effectiveness, flexible scalability, and other requirements of NOSQL databases have proven insurmountable for the conventional relational database. The cloud database is an unavoidable option because it can more effectively address the issue of data storage in the distributed database of scalability factors. Although cloud storage cannot yet fully meet our needs for storage, its advent has helped to lessen the pressure caused by the current explosion in data growth and opened our eyes to potential future developments in data storage (Zhang, Xu, 2017).

HIGH FREQUENCY TRADING

In recent years, high frequency trading (HFT) has dominated the financial sector. It accounts for over 80% of trading volumes in foreign exchange futures and about 55% of trading volumes in the U.S. equity market. In contrast to traditional trading, it uses computer algorithms to carry out trading automatically in a high frequency mode, where a transaction can be completed in a few seconds or even milliseconds.

Challenges in finance are brought about by the advent of high frequency trading in both theory and reality. It boosts market liquidity and favors trading the most liquid large-cap stocks. As a result, there is a difference between the bid and ask prices in trading for large-caps and small-caps, or the bid-ask spread. Due to its "high-frequency" character, it somewhat significantly raises

trading risks. The DOW lost 998.5 points (9.2%) in 7 minutes during the May 6, 2010, Flash Crash, and the S&P 500 lost 3% in 4 minutes. It was also cited as one of the "major" causes of the 2008 financial crisis.

HFT data are typical structured big data that only include a small number of features but a large number of observations. In actuality, it is a kind of highly structured streaming data that is produced quickly and in large quantities. For instance, an HFT profile of five stocks over the course of a month might require more than 80 Gigabytes of storage and contain more than 20 million transaction records. Traditional data analytics techniques are actually not very good at applying to such streaming huge data, let alone extracting trading expertise from it (Han, Li, 2018).

The advantages are too numerous to list, but to give you an idea of them, have a look at a few examples:

• Anomaly Detection: surveillance of production, continuos predictive maintenance, manufacturing optimization;
• Fraud detection procedures;
• Identity theft detection;
• Account and transaction issues;
• Customer analytics:
– Customer Relationship Management (*CRM*);
– Analysis and prevention;
– Customer Satisfaction;
– Pricing: promotional effects tracking, on-time competitive price responses;
• Competitive monitoring; as well as
• New Product Development Strategies.

All of these issues call for some sort of decision. The complex decision-making process is best shown by new product development. Throughout a pipeline for product development, decisions are made. From conception or conceptualization to product launch and post-launch tracking, there are several stages in this process. A pipeline has five steps, according to Paczkowski (2020): ideation, design, testing, launch, and post-launch tracking. Between each step, choices are made regarding whether to continue with the project or stop development or even production. A business case analysis that looks at the anticipated revenue and market share for the product is used to identify each decision point. For each business case evaluation, it is necessary to estimate expected sales, anticipated price points (which are adjusted as the product progresses through the pipeline), production and marketing costs, and competitive analyses that include current products, sales, pricing, and promotions, as well as competitive responses to the proposed new product. Any of these will be canceled and taken out of the pipeline if they have a negative impact on the concept. For each of the business case check points, information is required (Paczkowski, 2021).

In daily speech, the terms information and data are interchangeable. For instance, it is typical to hear a business manager state she has a lot of data in one context and then say she has a lot of information in another, conflating the meaning of the two phrases. The accompanying technology employed in those systems is known as Information Technology, and the computer systems that manage data are referred to as Information Systems (IS) (IT).

The Chief Information Officer is the C-Level executive in charge of this data and IT infrastructure (CIO). Take note of how often the phrase "information" appears. Despite the fact that individuals frequently use these two terms interchangeably, they do not necessarily mean the same thing. Data and information are separate concepts with a relationship. Simply put, data are facts, or objects that are true on their face, and they must be arranged and manipulated in order to provide insight

into previously undiscovered phenomena. They become information when they are controlled and manipulated. Organization and manipulation are mutually exclusive; neither can exist without the other. Your company's IT department organizes data, but it does not change it to create information. To use the information in a decision, it must be extracted from the latent information that is concealed inside the data.

Data has five characteristics or qualities that influence how they can be studied and by what instrument. There is a taxonomy that includes:

1. Source
2. Domain
3. Level
4. Continuity
5. Measurement Scale

Each of these poses unique issues. These characteristics make up a data taxonomy (Paczkowski, 2021).

CONCLUSION

Understanding linkages and applying them to forecast the most likely effects of business decisions is a key component of business data analytics. Price elasticity, which is used to reposition a current product's price or establish a new product's starting price, is an illustration of the former. In any scenario, the elasticity is determined using the price-quantity connection while accounting for other elements including income levels, the time of year, geographical regions, and so on. Except for the fact that it is always negative in economic theory and practice, the precise relationship between price and quantity can only be discovered by thoroughly analyzing data. It is nearly usually the case that you will need to know the outcome of a decision for a key performance measure (KPM), such as revenue, sales, shipments, and so forth, for the latter application of relationships, predicting. Predictions are difficult to create and of great importance, but they are still dependent on correlations between features or variables in a data collection, much like the main measurements.

As its name implies, hierarchical clustering creates a hierarchy of groups of things. The items are measured using a variety of continuous, ordinal, or nominal properties. The final two must be properly encoded, typically using dummy encoding. Character traits can be employed, but they must, of course, be properly coded. A dendrogram, which is a type of tree used to represent hierarchies, typically shows all the individual objects as independent leaves at one end of the tree and the entire collection of things at the other, or root, end. The hierarchical clustering dendrogram is comparable to a decision tree in this way. The dendrogram is made differently in this case.

Two types of hierarchical clustering exist:

1. Agglomerative, where each initial cluster only contains one object and the hierarchy is built up from there, with each object starting as its own cluster. Lower-level clusters serve as the foundation for higher-level clusters. Thus, the dendrogram is constructed using a bottoms-up strategy.
2. Divisive in which the initial cluster, which is the root, contains every object (i.e., the initial cluster contains every object). After that, the objects are gradually removed from this cluster to create other clusters beneath the high-level cluster. Thus, a top-down technique is used to construct the dendrogram.

The Divisive technique requires additional computation because each object must be examined individually to identify which cases to pluck away. The method that is most frequently utilized is agglomerative since it is more effective.

REFERENCES AND BIBLIOGRAPHY

- Brownlees, T., Cipollini, F., & Gallo, M. (2011). Intra-daily volume modelling and prediction for algorithmic trading. Journal of Financial Economics, 9(3), 489–518.
- Han, H., Li, M. (2018). Big Data Analytics for High Frequency Trading Volatility Estimation (349-357). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Li, Zh., Tao, M. (2018). Research on the Construction of Enterprise Brand Competitiveness Evaluation System Based on the Integration of SWOTand AHP Model (67-74). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Liu, X., Ji, G. (2017). A study on Cooperation Strategies of Dual Channel Supply Chain Based on Service level (29-43). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Paczkowski, W. (2021). Business Analytics – Data Science for Business Problems. Springer, Plainsboro.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117.
- Wang, M., Li, Zh. (2017). Research Status and Prospect of Data Extraction and Cleaning Technology in Large Environment (292-298). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Wang, Sh. (2017). Reflections on the Training Mode of E-Commerce Professionals with Improved Practical Exercises and Innovative Ability (75-81). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Zhang, X., Xu, F. (2017). Research on Data Storage Based on Cloud Platform (333-338. Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.
- Zreika, M., Hua, J., Wang, G. (2018). Applying Data Processing Method for Relationship Discovery in the Stock Market (249-255). Recent Developments in Data Science and Business Analytics (ICDSB – 2017), Springer.