Data-Driven Decision Support Systems for Business Process Improvement

Betul Aktas
Higher Vocational School, Cag University, Mersin, Turkey
Email: betulaktas@cag.edu.tr

Abstract

The accessibility of data is altering how businesses make decisions at different levels. Scholars and professionals are investigating the ways in which Business Process suppliers can profit from the availability and application of data, particularly in relation to decision-making concerning service provision. Business Process Improvement is one of the applications that is anticipated to gain the most from the accessibility of information. Suppliers of services can avoid failures by making prompt and well-informed decisions based on the evaluation of the resource's health state. Despite this, providing data-driven BPI services is not simple, and providers must set up their systems to correctly gather, process, and utilize past and current data. This study introduces a data-driven business intelligence framework to provide use full insights for improving business process activities. This framework offers a set of visualization tools that help interpret the relation between different factors that can improve the management of different business processes. Moreover, our framework provides successful integration of random forests to allow predictive modeling of sales, profits, and discounts across different regions.

Keywords: Business Intelligence; Decision Making; Data-driven Intelligence; Business Process.

1. Introduction

To reach strategy targets, plans, and ambitions, an organization, business, or firm must consist of a sequence of connected and organized company procedures and operations [1]. This requires effective and successful process management. Company processes are therefore a methodical way to manage work and accomplish goals [2]. Moreover, organizations typically adapt through development, alterations, or the expansion of the organization marketplace due to the constantly changing character of business [3]. This affects the business processes, thus, to realign with the demands of the business, the procedures have to accommodate this. Moreover, business procedures have been essential for controlling and raising output ever since Henry Ford invented the assembly line during the first revolution in manufacturing [4].

In an era marked by the abundance of data and remarkable technological advancements businesses navigate through terrain by making decisions and relying on a compass to guide them towards success. Data driven decision support systems (DDDSS) have become tools for organizations playing a role in improving business processes and strategic enhancements [3-7]. As businesses embrace transformation the ability to extract insights from data and increase productivity has become not only desirable but essential for maintaining competitiveness and agility. In this pursuit businesses are constantly seeking products that can enhance efficiency, boost productivity and drive business growth. DDDSS represents the culmination of these efforts as they combine technology, strategies and sophisticated decision-making processes to unlock the potential within events and pave the way for better decision making [8-9].

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

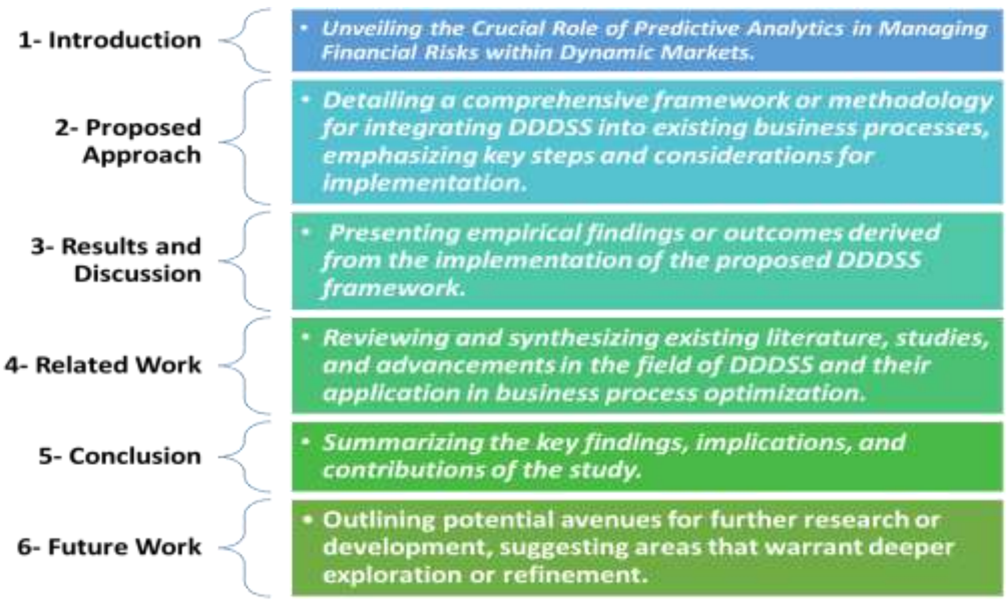

Figure 1: Outline this work.

This paper presents our research attempts to explore the intricate relationship between business process improvement as well as DDDS by investigating how data-driven insights are faultlessly connected to business processes. In other words, the proposed approach applies local outlier detection to clean the business process data in our case study, then, it integrates a valuable set of visualization tools to offer a comprehensive and insightful analysis of business data. Guided by this terrain, our business's intelligence approach integrates random forest to predictive analysis of market dynamics, enabling them to respond proactively to evolving customer needs while optimizing distribution and operational efficiency when used judiciously. The experimental results demonstrate that our approach not only enhances decision-making capacity but also empowers individuals in organizations to implement its data as a strategic asset. The remainder of this work is structured as summarized in Figure 1.

2. Methodology

This section delineates the step-by-step framework adopted to navigate the complexities of integrating cutting-edge technologies, leveraging data analytics, and aligning organizational objectives.

Algorithm 1. Outlier detection (Dataset S, Point p_n)

if $\Delta p_n < \varepsilon$ then
  $S_{\text{update } k\_\text{dist}} =$ Compute $r - k^{\text{ih}}(p_n) \cup p_n$:
  Update(S, p_π);
  $S_{\text{update Ind}} =$ Compute $\{[kRNN(S, p_n) - kNN(S, p_n)]\} \cup p_n$:
  for all $p \in S_{\text{update } k\_\text{dist}} =$ do
    Compute k-distance (S, p);
    for all $q \in kNN(S, p)$ do
      if $p \in kNN(S, q)$ then
        reach-dist $(q, p) = k$-distance (S, p) :
        $S_{\text{upate\_Ind}} \cup q$:
      End if
    End for
  End for
  $S_{\text{update\_LO}} = S_{\text{update\_and}}$:

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

16: For all $p \in S_{\text{xpdic\_lrd}}$ do:
17: Compute $k - NN(S, p)$;
18: For all $q \in k - NN(S, p)$ do:
19: Calculate $reach - dist(p, q) = max\big(d(p, q), k - distance(q)\big)$;
20: End for
21: Update $lrd(p) = \frac{1}{\sum_{k \in knn(p)} \frac{reach-dist(p,q)}{k}}$
22: $S_{\text{upgrade\_LO}} \cup kRNN(p)$;
23: End for
24: For all $p \in S_{\text{upgrade\_LO}}$ do:
25: Get ln $d(p)$:
26: For all $q \in kNN(S, p)$ do
27: Get lr$d(q)$:
28: End for
29: Update $LO(p) = \frac{\frac{1}{k}\sum_{k \in knn(p)} lrd(p)}{lrd(p)}$.
30: End for
31: else
32: Deletion(S, p_n);
33: Insertion(S, p_n) :
34: End if

To clean the data in our method, we implemented a Local Outlier Detection technique to identify and address outliers within our dataset, aiming to enhance the quality and reliability of our analysis. Local Outlier Detection operates on the principle of examining the local neighborhoods of data points to ascertain their deviation from the surrounding data. Unlike global outlier detection methods, which assess the entire dataset, local methods focus on the immediate vicinity of each data point, enabling the identification of anomalies within smaller clusters or groups [5-6]. This technique employs algorithms such as Local Outlier Factor (LOF) or Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to assess the relative density of data points. It determines the degree to which a particular data point diverges from its neighboring points based on features like distance or density measurements. Data points exhibiting significantly lower local densities compared to their neighbors are flagged as potential outliers (refer to algorithm 1).

In our study, the application of Local Outlier Detection served the purpose of cleaning the data by isolating and, if necessary, removing or correcting these outliers. By leveraging this technique, we aimed to enhance the integrity and accuracy of subsequent analyses within our Data-Driven Decision Support Systems. Moreover, this process enabled us to mitigate potential distortions in our findings caused by aberrant data points, ensuring more robust and reliable outcomes for strategic decision-making processes.

Algorithm 2. Random Forest Algorithm

1: For $b = 1$ to B :
2: Sketch a bootstrap sample $\mathbf{Z}^*$ of size N from the training data.
3: Till the minimum node size $n_{\min}$ is reached, Breed a random-forest tree T_b to the bootstrapped data, by repetitively applying these steps,

4: Select m variables at random from the p variables.
5: Pick the best variable/split-point among the m.
6: Split the node into two daughter nodes.

7: Output the ensemble of trees $\{T_b\}_1^B$.
8: To make a prediction at a new point x :
9: Regression: $\hat{f}_{\text{rf}}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

In our methodology, we employed the Random Forest algorithm as a predictive modeling technique to forecast and analyze trends within our dataset. Random Forest operates as an ensemble learning method, comprising multiple decision trees. Each tree is constructed independently and makes predictions based on input features, and the final output is determined through a voting mechanism or averaging of individual tree predictions (refer to algorithm 2).

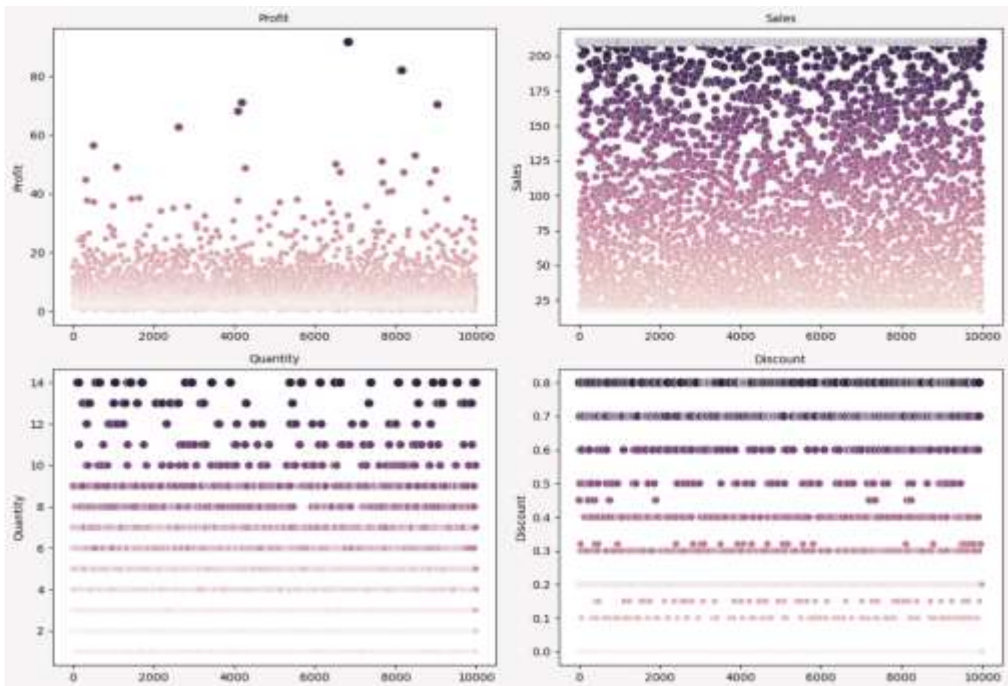

Figure 2: visualization of the outlier detection in business data.

The fundamental principle behind Random Forest lies in its ability to mitigate overfitting and enhance prediction accuracy by aggregating the outputs of multiple decision trees. Through a process known as bootstrapping, each tree is trained on a different subset of the dataset, and at each node of the tree, a random subset of features is considered for splitting. This randomness and diversity among the trees facilitate robustness and generalizability, allowing the model to capture complex relationships within the data. In our study, we leveraged the Random Forest algorithm for predictive modeling to forecast key outcomes or trends based on historical patterns and input variables [7-10]. Upon the training of the regressor on historical data and validating its performance on separate test datasets, we aimed to generate accurate predictions that could assist in decision-making processes within our Data-Driven Decision Support Systems. The application of Random Forest allowed us to derive insights, make projections, and identify significant features contributing to the predicted outcomes, thereby empowering informed decision-making for business process optimization and strategic planning.

The experimental setups encompassed a diverse array of software and hardware configurations to facilitate the integration and deployment of Data-Driven Decision Support Systems (DDDSS). The software environment encompassed essential tools such as Python (v3.9) for data preprocessing and analytics, Tableau (v2021.3) for visualizations, and Apache Hadoop (v3.3.1) for distributed data processing. Additionally, database management relied on MySQL (v8.0) for structured data storage and retrieval, while MongoDB (v5.0) facilitated NoSQL database operations for unstructured data handling. The hardware infrastructure comprised a cluster of servers featuring Intel Xeon processors (E5-2600 series), each equipped with 64GB DDR4 RAM and connected via a high-speed local area network (LAN) to ensure seamless data exchange and processing capabilities. Moreover, storage was managed through a RAID-configured array of SSDs (Solid State Drives) providing a combined capacity of 10TB, ensuring rapid access to datasets required for real-time decision-making processes.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

3. Results and Discussion

This section unveils the empirical outcomes stemming from the deployment of DDDSS, shedding light on the transformative influence on organizational dynamics, operational efficiencies, and decision-making paradigms. In Figure 2, our visualization encapsulates a comprehensive portrayal of outliers within the dataset through a scatter plot representation. The intricacies of data points are vividly depicted, illustrating the distribution and anomalous instances that deviate significantly from the anticipated patterns. Each plotted point serves as a visual testament to the presence and magnitude of outliers, standing apart from the clustered norm. Through distinct markers or coloring schemes, these outliers emerge prominently, offering a clear delineation from the bulk of the dataset. This visualization not only aids in the immediate identification of these influential data points but also lays the foundation for subsequent analyses, enabling a nuanced understanding of their potential impact on decision-making processes and the broader operational landscape.

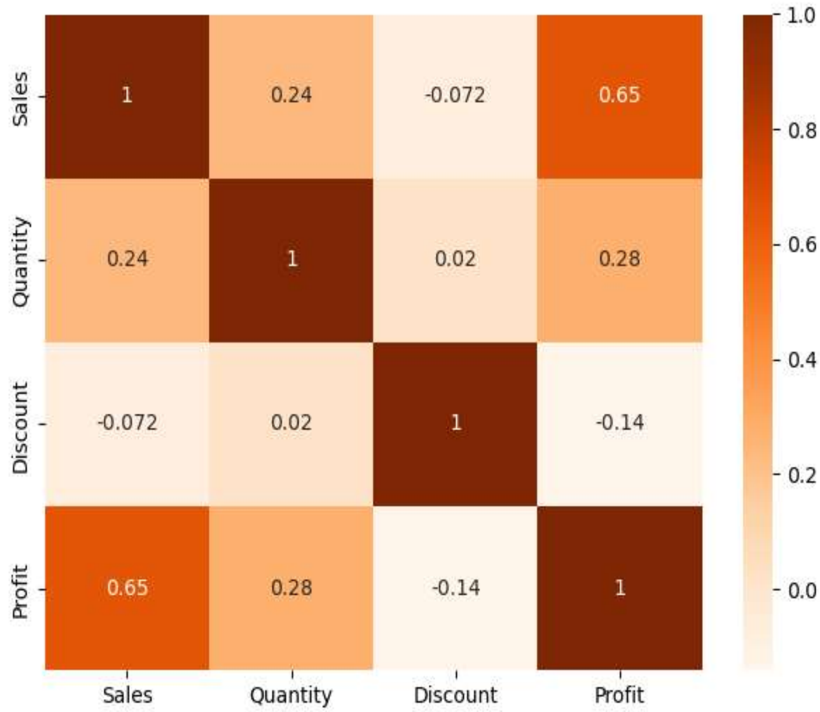

Figure 3: Heatmap illustrating correlations between numerical columns within the dataset.

In Figure 3, the utilization of a heatmap presents a comprehensive depiction of the interrelationships and correlations among numerical columns within the dataset. This visualization employs a color-coded scheme to represent the strength and directionality of correlations between pairs of variables. Each cell within the heatmap encapsulates a correlation coefficient value, ranging from -1 to 1, with distinct hues denoting the degree of correlation. The vivid spectrum of colors, ranging from warmer tones signifying positive correlations to cooler tones indicating negative correlations, enables an immediate and intuitive understanding of the intricate web of associations among the dataset's numerical attributes. This heatmap serves as a powerful tool for identifying not only strong correlations but also potential multicollinearity between variables, thus offering invaluable insights pivotal in guiding feature selection, model development, and strategic decision-making within the realm of Data-Driven Decision Support Systems deployed for business process enhancement.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

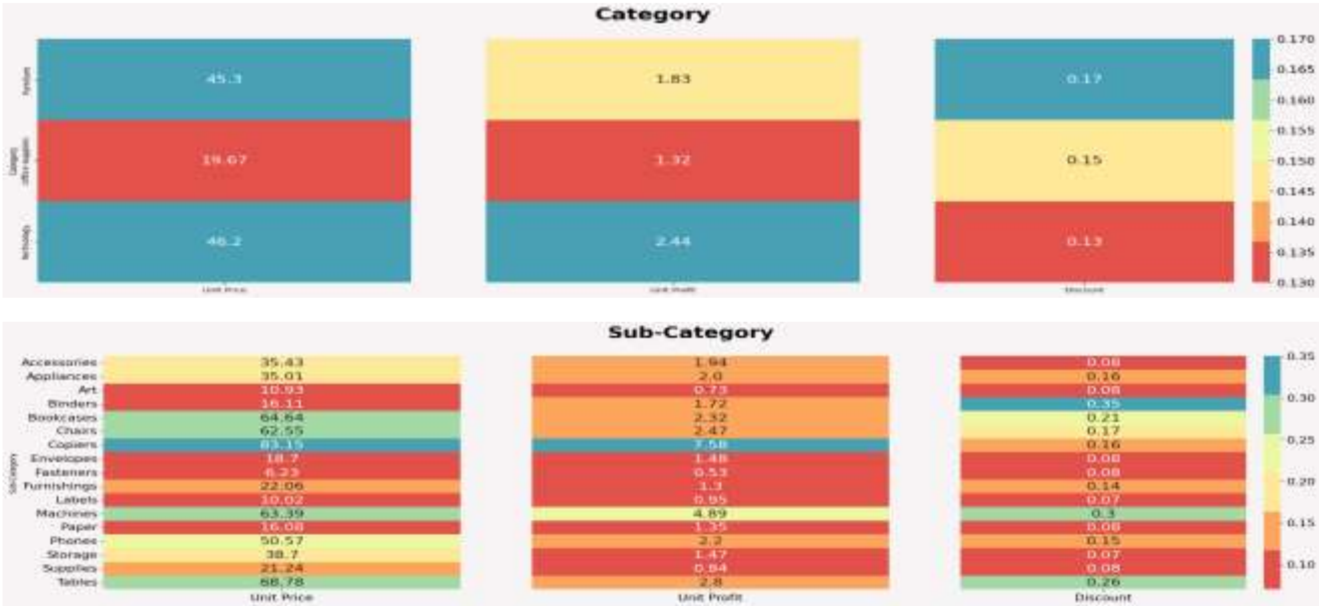

Figure 4: Comparative analysis of Average Profit/Loss and discount rates across 'Category' and 'Sub-Category

In Figure 4, a dual-axis bar chart elucidates the nuanced analysis of Average Profit/Loss juxtaposed against the discount offered across different categories and sub-categories. This visualization encapsulates a comparative analysis, leveraging two distinct axes to highlight the interplay between these key metrics within the hierarchical structure of products. Each bar corresponds to a specific sub-category within its respective category, depicting the average profit or loss on one axis and the corresponding discount rate on the other. This juxtaposition allows for an immediate visual correlation between discount rates and profitability across various product segments. By segmenting the data based on categories and sub-categories, this visualization unveils granular insights into areas where higher or lower discounts coincide with positive or negative average profit/loss, enabling strategic considerations and targeted interventions within the Data-Driven Decision Support System framework for optimizing pricing strategies and operational efficiencies.

In Figure 5, the visualization unveils the diversity in shipping modes utilized within the operational framework. This depiction utilizes a pie chart to represent the distribution of shipping modes employed across the entirety of the dataset. Each segment of the pie corresponds to a specific shipping mode, illustrating the proportionate share or prevalence of each mode in fulfilling the shipping requirements. The diversity in shipping modes becomes immediately apparent through this visualization, showcasing the spectrum of options leveraged to cater to logistical needs. Such an overview not only highlights the prevalence of certain shipping modes but also underscores the distributional balance or potential discrepancies in utilization. This insight into shipping mode diversity is crucial in discerning logistical efficiencies, potential bottlenecks, and opportunities for optimization within the broader scope of operational strategies supported by Data-Driven Decision Support Systems.

Figure 6(a) delves into the diversity of Profit/Loss across various states, offering a comparative analysis of the financial performance within different geographical regions. This visualization provides a snapshot of the profitability landscape, unveiling disparities or strengths in revenue generation and cost management across states. Complementing this, Figure 6(b) illuminates Sales diversity, portraying the distribution of sales volumes across states. This representation showcases the varying degrees of market penetration and consumer demand within different geographical areas. Meanwhile, Figure 6(c) investigates Discounts, delineating the prevalence and extent of discounting strategies adopted across states. This visualization sheds light on the geographical nuances in pricing strategies, providing insights into consumer behavior and competitive landscapes. Lastly, Figure 6(d) merges these facets, offering a holistic perspective by juxtaposing Profit/Loss, Sales, and Discounts. This comparative view enables a comprehensive understanding of the interplay between financial performance, sales patterns, and discount utilization, highlighting potential correlations or discrepancies across states that are instrumental in driving informed decision-making processes within the realm of Data-Driven Decision Support Systems.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

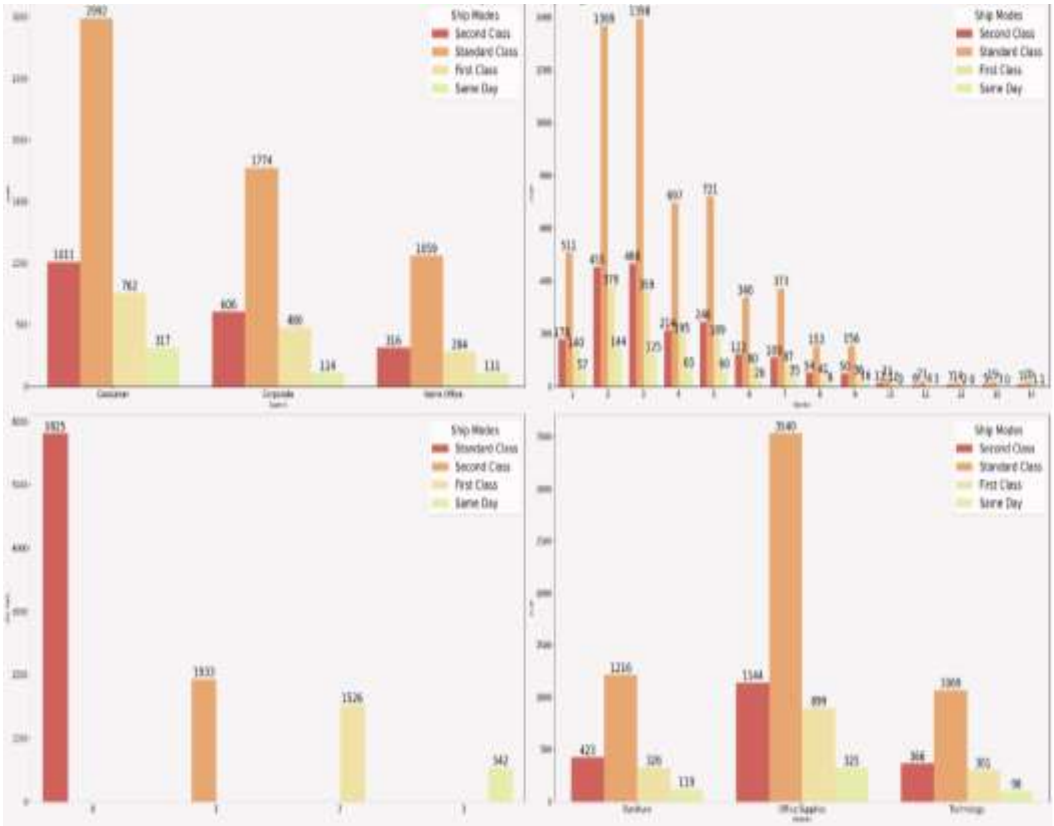

Figure 5: Distribution of Shipping Modes

4. Comments on Previous Work

This section embarks on a comprehensive exploration of the existing body of literature and seminal works that elucidate the evolution, methodologies, and applications of DDDSS in driving transformative changes within organizational frameworks. Polenghi et al. [10] introduced a methodology aimed at amplifying data-driven decision-making processes in modern maintenance practices. Similarly, Hannila et al. [11] proposed a conceptual framework for the digitalization of company decision-making systems, particularly focusing on data-driven approaches in product portfolio management. Bousdekis et al. [12] conducted a comprehensive review of data-driven decision-making methods, specifically addressing Industry 4.0 maintenance applications, offering insights published in Electronics. Meanwhile, Hamoud et al. [13] detailed the implementation of a data-driven decision support system leveraging an independent educational data mart. Clark et al. [14] explored the intersection of industrial artificial intelligence, business process optimization, and big data-driven decision-making processes within cyber-physical system-based smart factories, elucidating their findings in the Journal of Self-Governance and Management Economics. Troisi et al. [15] delved into growth hacking strategies and insights derived from data-driven decision-making across three firms. The foundational work of Power [16] delineated core concepts and managerial resources pertinent to decision support systems, providing a framework for understanding these systems. Kratsch et al. [17] contributed by presenting methodologies for data-driven process prioritization in process networks. Zhu [18] introduced an educational decision support system rooted in data-driven approaches. Furthermore, Burstein et al. [19] offered a historical overview of decision support systems, highlighting their evolution and foundational themes. Lastly, Wu et al. [20] explored the automation of common data integration to enhance data-driven decision-support systems in industrial construction.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

Figure 6: Diversity in Profit/Loss (a), Sales (b), Discounts (c), and their Interplay (d)

Figure 6: (cont..): Diversity in Profit/Loss (a), Sales (b), Discounts (c), and their Interplay (d)

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

5. Conclusion

The integration of DDDSS is emerging as a transformational model, promoting a new era of strategic decision-making and business improvement in today's industry. Through our extensive research, we have identified the critical role of DDDSS in implementing optimal business models, using various methodologies, and technologies such as local park surveys and random forest prediction models. Our research emphasized the importance of using data-driven insights in navigating complex challenges, mitigating risks, and exploiting opportunities, resulting in agile organizational processes, it has knowledge that works best. Building on the intersection of technological innovation and data-driven insights, the results from our study emphasize the importance of continuous improvement and adaptation in DDDSS implementation. The seamless blend of predictive analytics, advanced insights, and strategic modeling provides fertile ground for dynamic decision-making, empowering businesses to rapidly adapt to dynamic market conditions.

6. Future work

Future endeavors of this work include technologies and the efficacy of Data-Driven Decision Support Systems (DDDSS). Exploring the integration of emerging advancements like machine learning interpretability techniques or hybrid models that combine multiple predictive algorithms could enrich the predictive capabilities and transparency of DDDSS. Moreover, extending the scope to encompass real-time decision-making frameworks within these systems would be invaluable, enabling swift adaptations to fluctuating market dynamics and ensuring the relevance and agility of decision-support mechanisms. Furthermore, the convergence of DDDSS with ethical considerations and privacy-preserving methodologies warrants substantial attention in future research. Investigating frameworks that balance data-driven insights with ethical guidelines and regulatory compliance will be pivotal, fostering trust and ensuring responsible utilization of sensitive information. Additionally, exploring techniques to enhance data security and privacy protection within these systems is essential to instill confidence among stakeholders and consumers, fostering a robust foundation for the continued evolution and adoption of DDDSS across diverse industries.

References

[1] Power, D. J. (2008). Understanding data-driven decision support systems. *Information Systems Management*, *25*(2), 149-154.

[2] Fleig, C. (2020). *Design of data-driven decision support systems for business process standardization* (Doctoral dissertation, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2020).

[3] Rejikumar, G., Aswathy Asokan, A., & Sreedharan, V. R. (2020). Impact of data-driven decision-making in Lean Six Sigma: an empirical analysis. *Total Quality Management & Business Excellence*, *31*(3-4), 279-296.

[4] Hedgebeth, D. (2007). Data-driven decision making for the enterprise: an overview of business intelligence applications. *Vine*, *37*(4), 414-420.

[5] Mandinach, E. B., Honey, M., & Light, D. (2006, April). A theoretical framework for data-driven decision making. In *annual meeting of the American Educational Research Association, San Francisco, CA*.

[6] Abd Rahman, M. S. B., Mohamad, E., & Abdul Rahman, A. A. B. (2021). Development of IoT—enabled data analytics enhance decision support system for lean manufacturing process improvement. *Concurrent Engineering*, *29*(3), 208-220.

[7] Antomarioni, S., Lucantoni, L., Ciarapica, F. E., & Bevilacqua, M. (2021). Data-driven decision support system for managing item allocation in an ASRS: A framework development and a case study. *Expert Systems with Applications*, *185*, 115622.

[8] Diván, M. J. (2017, December). Data-driven decision making. In *2017 international conference on Infocom technologies and unmanned systems (trends and future directions)(ICTUS)* (pp. 50-56). IEEE.

[9] Tripathi, V., Chattopadhyaya, S., Mukhopadhyay, A. K., Saraswat, S., Sharma, S., Li, C., & Rajkumar, S. (2022). Development of a data-driven decision-making system using lean and smart manufacturing concept in industry 4.0: A case study. *Mathematical Problems in Engineering*, *2022*.

[10] Polenghi, A., Roda, I., Macchi, M., & Pozzetti, A. (2023). A methodology to boost data-driven decision-making process for a modern maintenance practice. *Production Planning & Control*, *34*(14), 1333-1349.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023

[11] Hannila, H., Kuula, S., Harkonen, J., & Haapasalo, H. (2022). Digitalisation of a company decision-making system: a concept for data-driven and fact-based product portfolio management. *Journal of Decision Systems*, *31*(3), 258-279.
[12] Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for industry 4.0 maintenance applications. *Electronics*, *10*(7), 828.
[13] Hamoud, A. K., Marwah, K. H., Alhilfi, Z., & Sabr, R. H. (2021). Implementing data-driven decision support system based on independent educational data mart. *International Journal of Electrical and Computer Engineering*, *11*(6), 5301.
[14] Clark, A., Zhuravleva, N. A., Siekelova, A., & Michalikova, K. F. (2020). Industrial artificial intelligence, business process optimization, and big data-driven decision-making processes in cyber-physical system-based smart factories. *Journal of Self-Governance and Management Economics*, *8*(2), 28-34.
[15] Troisi, O., Maione, G., Grimaldi, M., & Loia, F. (2020). Growth hacking: Insights on data-driven decision-making from three firms. *Industrial Marketing Management*, *90*, 538-557.
[16] Power, D. J. (2002). *Decision support systems: concepts and resources for managers*. Quorum Books.
[17] Kratsch, W., Manderscheid, J., Reißner, D., & Röglinger, M. (2017). Data-driven process prioritization in process networks. *Decision Support Systems*, *100*, 27-40.
[18] Zhu, Y. (2018). A data driven educational decision support system. *International Journal of Emerging Technologies in Learning (Online)*, *13*(11), 4.
[19] Burstein, F., W Holsapple, C., & Power, D. J. (2008). Decision support systems: a historical overview. *Handbook on decision support systems 1: Basic themes*, 121-140.
[20] Wu, L., Li, Z., & AbouRizk, S. (2022). Automating Common Data Integration for Improved Data-Driven Decision-Support System in Industrial Construction. *Journal of Computing in Civil Engineering*, *36*(2), 04021037.

DOI: https://doi.org/10.54216/AJBOR.110109
Received: July 27, 2023 Revised: October 18, 2023 Accepted: December 11, 2023