Munich Personal RePEc Archive

From Discounts to Delivery: Decoding Customer Care Interactions in Warehousing

Leogrande, Angelo

LUM University Giuseppe Degennaro

2024

Online at https://mpra.ub.uni-muenchen.de/122693/
MPRA Paper No. 122693, posted 19 Nov 2024 14:22 UTC

From Discounts to Delivery: Decoding Customer Care Interactions in Warehousing

Angelo Leogrande, LUM University Giuseppe Degennaro, leogrande.cultore@lum.it

Abstract

The present research has delved deeper into the complex relationship of customer care calls with purchasing behavior in a WM system and has developed actionable insights to optimize operations. In this regard, the following critical factors have been considered: product attributes-cost, weight, and discount-on one hand, and delivery performance in terms of timeliness and reliability on the other, with a view to understand their impacts on customer satisfaction and interactions. Key takeaways are that high volumes of customer care calls reflect operational failure; there is a delay or expectation mismatch, and hence one needs strong process optimization. Also, heavy products, since perceived to be reliable, have fewer customer enquiries; lighter, cheap products cause more frequent queries since impulsive buying and lack of information occur. It further identifies timeliness of delivery as a main determinant of customer satisfaction while delays in delivery result in heightened discontent and rising demands for support. The study underlines the strategic relevance of advanced analytics, machine learning, and real-time monitoring to finally resolve the recurring inefficiencies. This may also be a good basis on which recommendations could be made concerning the use of predictive analytics for demand forecasting, effective logistical frameworks, and methods of customer service that would be in line with product-specific needs. Discounts become a two-edged factor: enhancing satisfaction but threatening brand value when used too frequently. In the end, strategies with discounts should be put into balance, proactive customer engagement should be there, with crystal clear communications with them, and the products to be more correctly described. The given study also identified how a warehouse clears the expectation from customers by applying data-driven strategies for better efficiency, customer satisfaction, and long-term loyalty. The above findings provide a comprehensive road map on how to integrate technology and customer-centric strategies in modern warehouse management.

JEL CODE: L9, L90, L91, L92, L93.

Keywords: e-Commerce, Warehouse, Logistics, Machine Learning, Tobit.

1. Introduction

This research paper goes deeply into the dynamics of calls about customer care and their purchase behavior in a warehouse management system. The study investigates how such issues as prior purchases, product cost, weight, discounts, and timely delivery bear on customer calls and their satisfaction level. The centrepiece of the paper is its connection of the efficiency of warehouse operations with customer behaviour, downplaying how advanced technologies, data analytics, and customer-oriented approaches are crucial elements for process optimisation that enhance customer experience. This paper outlines the growing importance of warehouse management in contemporary supply chains within the context of rapidly changing global markets. It is indicated that customer-driven behavior has shaped key operational strategies to meet the growing demands on timely delivery, order accuracy, and reliability of service. Business is encouraged to apply innovative technologies and data-oriented approaches in dealing with inherent complexities at both inbound and outbound levels. Some of the key variables analyzed in the paper are frequency of customer care calls, product characteristics including cost and weight, and delivery metrics including timeliness. It relates the variables to one another; for instance, an increased frequency of customer care calls could point out certain "hidden" inefficiencies in the warehouse operations, such as delays in delivery, misunderstanding about product information, or a shortfall in expectation. A lesser frequency of the calls, on the other hand, may indicate good customer relationship management, reliable service, and high satisfaction. The econometric models and statistical analyses that detail these relationships go into the methodology section, thus showing that these are not loose concepts or ideas, but rather a well-structured study of how things really are. This study extracts, through large-scale data and advanced analytics, important insights that enable warehouse optimization and improve customer satisfaction. Based on the findings, recommendations are made for the stated operational objectives, considering customers' needs in an integrated strategy touching on technological innovation, process smoothing, and customer-centric strategies.

2. Literature Review

Song and Hua (2022) say, "In Bonded warehouse operations, most of data mining technology has transformed and is continuously optimizing ". By this fact, their work testifies that with data-driven approaches, friction in the motion of goods will reduce, errors will fall off, and efficiency in trade will grow. The authors strongly raise arguments for predictive analytics and real-time monitoring when espousing the need to avail means of technological tools towards meeting the dynamic demands of international trade. More than anything else, the value of this paper lies in the fact that bonded warehouses comprise a niche yet substantial component of worldwide supply chains and really show what data mining can do in terms of precision related to customs clearances and/or inventory management. Odunjo (2020) draws on the presentation of an empirical evaluation of inbound and outbound logistics operations at Dangote Cement Industries, given that these operations have direct contact with organizational performance. This places the analysis in an industry with high stakes and logistical precision forming the very backbone of competitive advantage. The dissertation enumerates a few of the operational challenges faced, such as delays on the part of material delivery agencies and inefficiency in the distribution channels, while it advocates strategic investments in infrastructure and technology. Odunjo's (2020) study further supports the work of Song and Hua by adding weight to the improvement aspect of logistics through technology as the success factor for an organization. It adds another touch in underpinning challenges specific to the industry. Problems that Phan and Tran (2022) discuss as being covered in the development of a warehouse management information system underline that substantial attention has recently focused on how such practical software solutions will help improve operational transparency and efficiency. This is also very unique in giving a hands-on view of how the system was designed to actually work, both in theory and importantly in practice. It falls in the niche discussed by Song and Hua (2022) but narrows down into the development of the system, hence giving a prolific view of how technology can be tailored towards specific operational demands. This paper gives practical recommendations on what to do for any organization looking to implement or improve its WMS. The analysis by Hu and Weisel (2020) gave a more general strategic setting for optimizing inbound logistics. Indeed, their work probes how optimization in supplier relationships, transportation routes, or inventory could yield cost economies and customer satisfaction. Though their work has been less focused on technological interventions compared to the other articles mentioned above, their emphasis on process optimization bridges the gap from manual operations to digital transformations. This also extends the work of Phan and Tran (2022) by explaining in detail the areas within which technology can contribute to the traditional structure of logistics. Polim and Lestari (2023) have explained BPR in regards to achieving excellence in warehouse management, and their case study is about retail. The pragmatism inducted in the case study attests to the fact that reengineering the existing processes provides an insight into the estimation for running the warehouse operations in harmony with the organizational objectives. The present study represents a balanced view of the challenges and opportunities involved and brings to light the role of stakeholder involvement in implementing change initiatives. These series of articles put together give a comprehensive view of how modern logistics and warehouse management integrity blend with technology, process optimization, and industry-specific solutions to bring about operational excellence in the field. They recommend a strategic combination of technological innovation with process refinement to match demands that keep on growing from the supply chains of the world. Review articles allow reading innovation into warehouse management in all its diversity for industries that merge technological innovation, process optimization, and customer-oriented approaches. Put together, it reflects the state of evolution and modernization of logistics systems with their new demands determined by global supply chains. Praptodiyono et al. (2023) have conducted a study on the design of RFID-based WMS for AgroHUB Banten by concentrating on peculiarities related to the agricultural supply chain. This paper gives a glimpse into how RFID technologies can revolutionize inventory tracking and save manikins from operating inefficiencies within AgroHUB. The proposed system supports the concept of precision agriculture and smart chains through real-time capture and automation of processes. Thus, this is the real value of bespoke technological solutions: being able to attend to the needs of one industry in particular and, as such, standing perhaps as a beacon toward digitization in warehouses within the agrifood industry. Li et al. (2024) used Flexsim to simulate the process at the ZTO Distribution Center for both the inbound and outbound areas. In this regard, their work segregated in this succession by the connection of theory into practice through simulation to find bottlenecks and optimization of workflow for a complex logistics system. These findings are particularly crucial with respect to the fact that its insights drawn herein have practical significance in throughput improvement and lead time reduction, particularly for organizations operating high-volume distribution centers. Complementing the work of Praptodiyono et al. (2023) this manuscript will try to explore how the simulation tools could be adapted with RFID systems for complete end-to-end warehouse optimization. Münsberg et al. (2022) again take the customer-focused view by investigating the onboarding of new customers to a 3PL warehouse. The research responds to a call for process standardization through spread collaboration among stakeholders for onboard operations to happen seamlessly. Although unrelated to technological innovation, this research underlines the adaptable means in the context of the 3PL sector facing challenges related to diversity and customization across customers. Therefore, that brings another dimension into the discussion-that operational efficiency needs not be related to technology but equally to strategic process design. The discussion by Satish et al., (2023), was around warehouse managing and material handling practices at Polkart Logistics. This paper insists on the interaction between manual and automated operations, including the best combination thereof for better productivity with minimum cost. Again, this focuses on yet another very important but mostly discussed lesser area of the overall functions of the warehouse, thereby underlining the role of physical infrastructure and people management in pursuit of excellence in operations. A case study by Apolonio and Norona (2021) looks at inventory management automation in the FMCG sector, placing their discussion in the wider context of Supply Chain 4.0. Work such as that, for example, tested insight into how automation, IoT devices, and data analytics would prove to be highly transformative in inventory control related to high-demanding industries. This all befits complementary insights from RFID focus into wider implications for automation, such as inventory accuracy, replenishment cycles, and eventually customer satisfaction. In other words, for proper management to take place within the warehouse, the best technologies-ranging from RFID to simulation tools-must be integrated with strategic-level process enhancement and customer-centric approaches. The challenges thrown up by modern logistics are answered with applications related to industry-specific solutions, simulation-based planning, and integration of automation. The selected articles discuss various trends in developments related to warehouse management systems-from studying cloud-based architectures and machine learning applications, going further into the operational frameworks for leaner logistics operations. All these collectively indicate an increasing role of technology in causing changes in the process of warehousing and logistics across industries. Chen and Liu (2024) propose a state-of-the-art cloud-native architecture for AS/RS warehouses. Other key areas of the work involve developing how the cloud-based platform will facilitate the centralization of warehouse management and control systems for scalability, flexibility, and cost efficiency. This research work uses containerized services and microservice architecture to provide an effective means toward the delivery of real-time decision-making and operational adaptability. Also, with AS/RS-an automated system of warehousing-this automatically places the research at the bleeding edge of Industry 4.0 initiatives, showing just how traditional supply chain management can be changed through the integration of cloud-native technologies. To handle the demands imposed by FIFO and FEFO, Hietasari et al. (2024) proposed a conceptual model of WMS with auto-suggestions. Their work emphasizes lean warehousing through the removal of waste, intensification of inventory turnover, and timely use of stock. The framework automates decision-making processes for inventory allocation; hence, it contributes to solving some of the difficulties faced in perishable goods industries and others that require precision in stock management. This work thus complements the study of Chen and Liu (2024) by offering a more focused approach toward operational efficiency, targeting the aspect of inventory flow management. Albadrani et al. (2020) explore machine learning algorithms that can be utilized for the optimization of inbound logistics processes. Their work represents how predictive analytics and optimization models can mitigate costs of transportation, improve supplier selection, and develop smooth procurement processes. This study, through the amalgamation of machine learning, has revealed the potential of data-driven insights into transforming inbound logistics, a major but often neglected component of supply chain management. It goes well with the solutions of cloud-native approaches, with Chen and Liu illustrating the synergies between automation, advanced analytics, and real-time data processing. Dewi and Shofa (2023) introduce a WMS to manage the operation of warehouses effectively. This paper presents the design and implementation of an integrated system for addressing issues related to mismanagement of items in inventory and processing customer orders. Other than the general architectural or conceptual framework that many articles have been discussing, this study presents very handy insights into constructing user-friendly systems that can be implemented in medium to large-scale operations. Wahyuni et al. (2021) review logistics management in the contemporary retail market using operational modeling, taking Indonesia as a case study. Their work stresses how atypical such logistics challenges are in a retail context and thus require special solutions capable of balancing cost efficiency with customer satisfaction effectively. Although this study is less focused on technological interventions than the majority of other articles, it conveys an important perspective on the operational dynamics of logistics in a competitive market that complements the technological innovations discussed in the other articles. Entirely, these articles insist on the integration of heavyweight technologies like cloud computing, machine learning, and customized WMS frameworks within warehouse and logistic operations. This set of studies allows for further responsiveness, efficiency, and sustainability within the supply chain ecosystem through automation, data-driven decision-making, and operational modelling. It chooses diversified points of view-from digital transformation and frameworks for decision making to operational efficiency and logistics optimization. Put together, they help bring out a rich comprehension of the complications and development seen enlarged in the realm of warehousing practices. Kumar and Asthana (2023) have referred to the transformation of warehousing on Ekart from old-fashioned conventional godowns to state-of-the-art, technologically driven warehouse management systems. The above case study quite clearly describes how digitalization can help with inventory accuracy, operational efficiency, and scalability. The authors have amply explained on this platform how one of the e-commerce giants, eKart, put technology to use in satisfying consumer demand that is growing with each passing day. Such a transformation only modernized their operations but also aligned warehousing practices with the expected agility in fast-moving environments such as e-commerce. Gegeleso (2020) discusses the warehousing operation of the South Western Nigeria Fast-moving consumer goods industry performance talking about its inbound logistics. This paper realizes that good practices in warehousing will improve the relationship between supply and transportation efficiency among other activities believed to constitute inbound logistics. While doing this, those complications brought forth by deficiencies in infrastructure and resource constraints were region-specific; however, at the same turn, urged for strategic investments in those areas. This hence would complement the work done by Kumar and Asthana by taking into consideration some of the operational nuances of the work in the emerging markets since logistical inefficiencies are considered important in determining overall performance. Kara et al. (2023) also considered warehouse management in light of decision analysis-that is, through the suggested hybrid of CRITIC-MULTIMOORA with regard to choosing a warehouse manager. These authors have integrated single-valued neutrosophic sets into the provision of complex evaluation modeling about the candidates according to several criteria such as experience, leadership skill, and technical expertise. The comprehensive study shall find their significance in the fact that they are extending the scope of warehouse management in optimizing human resources-important but almost entirely ignored elements toward operational excellence. Rigor in methodology to make sure actionable insights are created for an organization on how to improve its leadership in the field of warehousing. Maheshwari et al. (2023) thought about a 'twin' digital version of the WMS and came up with the idea of a conceptual toolbox that can be used in future research. The related work this might involve could be done in the digital twin warehouse model for real-time simulation, monitoring, and optimization. It becomes capable of unparalleled efficiency and speed in managing the warehouse, having those advanced technologies like IoT, AI, and simulation models embedded; hence, the digital twin in that way. The present study lays a conceptual foundation in that regard; hence, it stands out among its peers of studies being prospective in this domain. This supports the themes of digital transformation discussed in the case by Kumar and Asthana. De Oliveira et al. made improvements to the logistics of receiving in warehouses in 2022. It involves data analysis and process mapping methodologies to identify bottlenecks and suggest improvements. Emphasizing that the receiving stage is so crucial and has been usually very much disregarded, the given study points to how smoothing initial logistics flows can make a difference in enhancing the general performance of the warehouse. It stands in concert with the theoretical insight of Maheshwari et al. (2023) in as much as concrete solutions have been provided in order to improve the concerns of operational flow. These papers finally reflect many-faceted dimensions of warehouse management, especially in driving home the imperative of technology, human capital, and operational efficiency in mobilizing large wheels of contemporary warehousing systems. This would, in turn, present an avenue to the readers for addressing theoretical and practical aspects that could hopefully be presented as a holistic roadmap on how best practices can be optimally enhanced and challenges heeded in the global dynamic supply chain. The selected articles explore diverse aspects of warehouse and inventory management, focusing on technology adoption, performance measurement, and process improvement. Collectively, they highlight how modern supply chain challenges can be addressed through strategic interventions, technological implementations, and operational redesigns. Corrêa (2023) examines the concept of the Supply Chain Control Tower, providing an analytical framework for its definition and application gaps. The study underscores the role of the control tower in offering real-time visibility and decision-making capabilities across supply chains. Despite its promise, the paper highlights significant gaps between theoretical concepts and practical implementation in industries, such as integration challenges and data standardization issues. This research emphasizes the need for technological advancements and cross-functional collaboration to unlock the full potential of control towers. By linking strategic oversight with operational execution, it lays the groundwork for future developments in supply chain management. Pereira et al. (2022) focus on the implementation of a Warehouse Management System (WMS) in a Danish logistics company. The study presents a detailed case of how WMS adoption improved inventory accuracy, workflow efficiency, and overall organizational performance. By leveraging automation and intelligent systems, the implementation addressed inefficiencies in warehouse operations, particularly in order processing and stock tracking. The practical insights from this study illustrate how companies can transition from manual to digital processes, aligning with broader trends in Industry 4.0. This complements Corrêa's (2023) research by demonstrating how operational-level interventions feed into the larger supply chain ecosystem. Karim et al. (2021) revisit warehouse productivity indicators, proposing ratio-based benchmarks to evaluate performance. Their study addresses a key challenge in warehouse operations: the lack of standardized metrics for productivity. By developing a comprehensive framework that considers space utilization, labor efficiency, and order fulfillment accuracy, the authors provide actionable tools for organizations to benchmark and improve their warehouse performance. This research is particularly valuable for managers seeking data-driven insights to enhance operational efficiency. Its focus on measurement aligns well with other articles discussing process improvements but adds a quantifiable layer to understanding warehouse effectiveness. Rector and Scott (2021) explore inbound inventory management system improvements at Hoffman. Their study identifies inefficiencies in the existing system, including delayed supplier communication and mismanaged stock levels. Through targeted interventions, such as enhanced supplier integration and automation of key processes, the study demonstrates measurable improvements in inventory turnover and accuracy. This case study offers a pragmatic approach to refining inbound logistics, bridging the gap between theoretical strategies and their practical application. Mane et al., (2020) focus on the implementation of the 5S methodology and an HTML-coded inventory control system. Their study demonstrates how structured workplace organization can streamline inventory processes, reduce waste, and improve productivity. The integration of a coded spare management system further illustrates how digital tools can complement lean practices to achieve operational excellence. This research showcases the value of combining traditional lean methods with modern technology to optimize warehouse operations. The selected articles introduce innovative means and unique challenges found in warehouse management, supply chain optimization, and inventory control. They provide a deeper understanding of how technological development, operational strategy, and context-specific factors affect the shaping of modern logistics systems. Indriyani (2020), in turn, examines the implementation of a WMS. She analyzes factors like how well it smooths out operational efficiency. Some of the identified key issues include inaccuracies in inventory, delayed shipment, and improper utilization of space. Based on such insufficient analysis, recommendations provided by the author include recommending the implementation of modern WMS tools and staff training to increase the efficiency of operations in a warehouse. This study shows practices within a warehouse must be re-aligned toward organizational objectives, with particular reference to the traditional logistic industry transiting into digitized systems. It therefore provides a better setting for further identification of how small-scale operations can adopt advanced WMS technologies in order to stay competitive. Herbe et al. (2024) develop the potentials of DLTs in managing the supply chain. This study explains how DLTs, such as blockchain, enable supply chains to adopt improved traceability and enhance the transparency and security of information. It is also being promoted as a means to enhance operational reliability by tacking common fraud and consistency problems in data. Meanwhile, the authors indicate implementation barriers due to a task's expensive cost and technological difficulty. The latter thus has specific significance for such industries, which require powerful authentication systems, and as one of the ways to prospectively view the integration of decentralized technologies within warehouse and logistic operations. Guo et al. (2024) discuss advanced machine learning applied to warehouse prediction tasks, such as using a CNN-BiLSTM-Attention model, indicating how AI is supposed to predict the move-ins and move-outs of silica powder into and out of a warehouse with very high accuracy. This represents a very good example of the growing role of AI in inventory management, especially for those industries concerned with special or high-value materials. The accuracy and efficiency of the model proposed herein represent how predictive analytics can reduce storage costs, increase order fulfillment, and optimize supply chain responsiveness. This research acts as a complement to traditional WMS studies by showing the capability of AI-based systems in improving operational forecasting. Bottani et al. (2020) present a simulation model to optimize storage space assignment in an e-commerce warehouse, particularly within the fashion supply chain. It also determines how to minimize the retrieval time by simulating various storage configurations of items in the warehouse, aiming to arrive at an efficient method of order processing. This hence creates the case that, in an e-commerce business, delivery speed coupled with correct order handling is vital to customer satisfaction and thus requires an industry-specific solution. Guided by the simulation approach, this work presents a model that can easily be replicated in other industries, by emphasizing scenario analysis as one of the key functions of warehouse optimization. Ahmed and Mohamed (2023) discuss the fragility of supply chains operating in a post-conflict environment using Somalia as a case experience. The authors have contextualized inbound and outbound disturbances, such as infrastructure deficiencies and security exposure to categorical disruptions within the operators of supply chains. It also focuses on resilient strategies that firms may adopt, such as local sourcing, risk-sharing agreements, or adaptive logistics practices. The research develops a critical stance towards contextual organizational dilemmas set in volatile regions and deepens the general discussion on warehouse and supply chain management (Table 1).

Table 1. References by methodologies, applications and results.

Authors	Methodologies	Applications	Results
Praptodiyono et al. (2023)	RFID technology implementation	AgroHUB warehouse management	Increased traceability and efficiency
Li et al.,(2024).	Simulation using Flexsim	Inbound/outbound process in ZTO Distribution Center	Improved process flow efficiency
Satish et al., (2023).	Operational analysis	Warehouse management at Polkart Logistics	Improved material handling processes
Hietasari et al., (2024)	Conceptual framework development	FIFO/FEFO in lean warehousing	Enhanced warehousing efficiency
Albadrani et al. (2024).	Machine learning algorithms	Inbound logistics optimization	Predicted future process improvements
Wahyuni et al., (2021).	Modeling and simulation	Logistics in retail markets	Improved logistical operations
Gegeleso (2020)	Impact assessment	Consumer goods industry logistics	Optimized inbound logistics practices
Kara et al., (2023).	CRITIC-MULTIMOORA hybrid method	Warehouse manager selection	Improved decision-making process
Maheshwari et al., (2023).	Digital twin concept	Future warehouse management	Theoretical guidance for innovation
de Oliveira et al., (2022)	Process improvement	Warehouse receiving processes	Enhanced efficiency
Corrêa (2023)	Literature review and gap analysis	Supply chain control towers	Identified practical application gaps

Pereira et al., (2022).	WMS implementation	Danish logistics company	Operational improvements
Karim et al., (2021).	Benchmark revision	Warehouse productivity measurement	Defined ratio-based indicators
Mane	5S implementation	Inventory control system	Improved inventory management
Indriyani (2020)	Operational analysis	Warehouse system at Pt. Pos Manado	Identified inefficiencies
Herbe et al., (2024).	Distributed ledger technology	Supply chain management	Secured and transparent operations
Guo et al., (2024).	CNN-BiLSTM-Attention model	Silica powder movement prediction	Improved accuracy
Bottani et al., (2024).	Simulation modeling	E-commerce fashion warehouse	Optimized storage assignment

3. Data and descriptive statistics

The database used in this study is a public database present on the Kaggle.com website. The database is entitled "E-Commerce Shipping Data" and takes into account some variables related to warehouse operations, logistics, with attention also to consumer responses in terms of satisfaction and ability to purchase the same products. It should be noted that this database does not offer information regarding the technical characteristics of the products that are stored, sold, shipped and in which consumers show interest. Therefore, this data has been used as an example to analyze some characteristics of warehouse management, logistics, shipping and customer satisfaction (Table 2).

Table 2. Variables of the model.

Variable	Definition	Source
Prior purchases	The number of times the customer has previously purchased the product or any related products.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024
Discount offered	The percentage or amount of discount provided on the product's original price for the customer.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024
Weight in gms	The weight of the product in grams.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024
Cost of the Product	The total price or cost of the product, often excluding any applied discounts or taxes.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024
Customer care calls	The number of times a customer has contacted customer support or customer care regarding inquiries or issues about the product.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024
Reached on time	Indicates whether the product was delivered to the customer within the expected or scheduled time frame.	Kaggle, E-Commerce Shipping Data, https://www.kaggle.com/datasets/prachi13/customer-analytics, accessed 10/10/2024

3.1 Descriptive Statistics

Customer Care Calls. The Customer care calls variable has a mean of approximately 4.05, with a median and mode of 4. This distribution suggests a relatively balanced trend around 4 calls per customer, with some customers contacting customer service more frequently than others. The minimum value of 2 and the maximum of 7 calls indicate that some customers might need more support, potentially signaling issues with product satisfaction or service complexity. The standard deviation of 1.14 is moderate, showing some variability but not extreme deviation from the mean.

The concentration around four calls may imply that customers typically reach out to customer care a few times, potentially to address common questions or minor concerns. Understanding the nature of these calls could help identify patterns and improve customer service operations, especially if high call volumes are linked to specific products or issues.

Cost of the Product. The Cost of the Product variable presents an average cost of about 210.2 units, with the median close to 214, and a mode of 245. The distribution shows a minimum of 144 and a maximum of 310, indicating a considerable range of product prices. The standard deviation is 48.34, highlighting some variability in pricing. This variance could be due to different product categories or customer segments within the dataset. The concentration around the median and mean values suggests a relatively stable pricing structure, though some products may fall into higher or lower price brackets. Analyzing the relationship between cost and customer satisfaction (perhaps through customer ratings or customer care calls) could provide insights into whether higher-priced items correlate with better or worse customer experiences. This information could guide pricing strategies and inform discount offerings.

Customer Rating. Customer rating has a mean of 2.09, with a median of 3 and a mode of 3, indicating a skew toward lower ratings. The minimum is 1, and the maximum is 5, with a standard deviation of 1.10. The ratings' concentration around the lower end (with a notable cluster at 3) may suggest customer dissatisfaction with certain aspects of the product or service. This distribution could indicate areas where improvements are needed, either in product quality, customer service, or value for money. Investigating the factors leading to lower ratings might reveal actionable insights. For instance, products with frequent customer care calls or delayed deliveries might correspond to lower ratings. A deeper dive into specific customer complaints could also help companies address pain points and enhance customer satisfaction.

Prior Purchases. The Prior purchases variable shows a mean of 3.59, a median of 3, and a mode of 3, indicating that most customers have purchased around three products previously. The range spans from a minimum of 2 to a maximum of 10, with a standard deviation of 1.02. This suggests a fair amount of customer retention or loyalty, as customers have purchased multiple products on average. However, there may be opportunities to further improve customer retention by examining the factors that encourage repeat purchases. For instance, analyzing if higher repeat purchases correlate with product quality, customer service, or discounts could inform strategies to increase repeat sales. Additionally, cross-referencing with customer ratings might show whether loyal customers are more likely to give positive ratings or if frequent buyers still experience similar pain points as one-time buyers.

Discount Offered. Discount offered has a mean of 13.37, with a median of 7 and a mode of 10, showing a right-skewed distribution. The maximum discount offered is 84, while the minimum is 1. The standard deviation is 13.68, indicating a substantial variability in discount rates. This variation suggests a differentiated discount strategy, potentially based on customer segments, product categories, or purchase behaviors. Higher discounts might be used to attract new customers or retain existing ones, while lower discounts could apply to premium products or high-demand items. It would be interesting to analyze the impact of discount levels on prior purchases and customer satisfaction. For instance, customers who received higher discounts might exhibit higher satisfaction or repeat purchase rates. On the other hand, significant discounts might sometimes correlate with lower ratings if they are tied to clearance items or products with known quality issues.

Weight in Grams. The Weight in gms variable has a mean weight of 3634 grams, with a median of 4149 and a mode of 4833, covering a range from 1001 to 7846 grams. The standard deviation of 1558.85 shows substantial variability, indicating that products vary widely in weight. This variability could be due to different product types, with lighter items potentially being accessories or smaller products, and heavier items being larger goods. Understanding the distribution of product weights and correlating them with other variables like customer satisfaction, delivery time, or cost could provide actionable insights. For instance, heavier products might face more logistical challenges, leading to delays or higher delivery costs. Examining the relationship between weight and customer ratings might reveal if heavier items are more prone to delivery issues or if they meet customer expectations differently than lighter items.

Reached on Time. The Reached on Time Y N variable indicates whether the product was delivered on time, with a mean of 0.597, suggesting that around 60% of deliveries were timely. The binary nature of this variable (0 = not on time, 1 = on time) allows us to quickly identify potential logistic issues. The mode and median are both 1, suggesting that most products did arrive on time, though a significant proportion did not. This could reflect operational inefficiencies or external factors affecting delivery times. Given that timely delivery is crucial to customer satisfaction, analyzing the impact of delayed deliveries on customer ratings and care calls could uncover areas for improvement. Timely delivery might correlate positively with higher customer ratings, while delays could lead to dissatisfaction and increased customer care interactions.

The dataset reveals several key insights and potential areas for improvement. Customer care calls show a moderate level of customer engagement, but there may be opportunities to reduce these calls by addressing recurring issues. Product cost appears fairly stable, but variability in customer ratings and the skew towards lower ratings suggest room for improvement in customer satisfaction. The discount strategy seems varied, which might effectively target different customer segments but could benefit from further optimization. The wide range in product weight highlights diversity in product offerings, with implications for logistics and customer satisfaction. Finally, ensuring timely delivery remains a challenge, as around 40% of deliveries appear to be delayed (Figure 1).

Figure 1. Descriptive Statistics.

Descriptive Statistics ▼

	Customer_care_calls	Cost_of_the_Product	Customer_rating	Prior_purchases	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
Valid	10999	10999	10999	10999	10999	10999	10999
Missing	0	0	0	0	0	0	0
Mode	4.000	245.000	3.000	3.000	10.000	4883.000	1.000
Median	4.000	214.000	3.000	3.000	7.000	4149.000	1.000
Mean	4.054	210.197	2.991	3.568	13.373	3634.017	0.597
Std. Error of Mean	0.011	0.458	0.013	0.015	0.155	15.593	0.005
95% CI Mean Upper	4.076	211.095	3.017	3.596	13.676	3664.579	0.606
95% CI Mean Lower	4.033	209.299	2.964	3.539	13.070	3603.454	0.588
Std. Deviation	1.141	48.063	1.414	1.523	16.206	1635.377	0.491
95% CI Std. Dev. Upper	1.156	48.479	1.424	1.558	16.496	1646.177	0.492
95% CI Std. Dev. Lower	1.128	47.594	1.403	1.486	15.895	1622.625	0.489
Coefficient of variation	0.282	0.229	0.473	0.427	1.212	0.450	0.822
MAD	1.000	40.000	1.000	1.000	3.000	1332.000	0.000
MAD robust	1.483	59.304	1.483	1.483	4.448	1974.823	0.000
IQR	2.000	82.000	2.000	1.000	6.000	3210.500	1.000
Variance	1.303	2310.078	1.998	2.319	262.619	$2.674\times10^{+6}$	0.241
95% CI Variance Upper	1.337	2350.213	2.029	2.427	272.114	$2.710\times10^{+6}$	0.242
95% CI Variance Lower	1.273	2265.215	1.967	2.209	252.647	$2.633\times10^{+6}$	0.239
Skewness	0.392	−0.157	0.004	1.682	1.799	−0.250	−0.394
Std. Error of Skewness	0.023	0.023	0.023	0.023	0.023	0.023	0.023
Kurtosis	−0.309	−0.972	−1.296	4.006	2.001	−1.448	−1.845
Std. Error of Kurtosis	0.047	0.047	0.047	0.047	0.047	0.047	0.047
Shapiro-Wilk	NaN	NaN	NaN	NaN	NaN	NaN	NaN
P-value of Shapiro-Wilk	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Range	5.000	214.000	4.000	8.000	64.000	6845.000	1.000
Minimum	2.000	96.000	1.000	2.000	1.000	1001.000	0.000
Maximum	7.000	310.000	5.000	10.000	65.000	7846.000	1.000
25th percentile	3.000	169.000	2.000	3.000	4.000	1839.500	0.000
50th percentile	4.000	214.000	3.000	3.000	7.000	4149.000	1.000
75th percentile	5.000	251.000	4.000	4.000	10.000	5050.000	1.000
25th percentile	3.000	169.000	2.000	3.000	4.000	1839.500	0.000
50th percentile	4.000	214.000	3.000	3.000	7.000	4149.000	1.000
75th percentile	5.000	251.000	4.000	4.000	10.000	5050.000	1.000
Sum	44595.000	$2.312\times10^{+6}$	32893.000	39240.000	147092.000	$3.997\times10^{+7}$	6563.000

3.2 Frequencies

Customer Care Calls. The Customer care calls variable indicates the number of times customers have contacted customer service. Here, the most common number of calls is 3 (29.25%), followed closely by 4 calls (32.34%) and then 5 calls (21.17%). A smaller percentage of customers made 6 or 7 calls (9.21% and 2.24%, respectively), while only 5.80% of customers contacted customer service twice. This distribution suggests that customers typically reach out multiple times, with the majority contacting customer care around 3-5 times. The relatively high frequency of calls may indicate recurring issues or a need for further assistance, either with product information or order support. This could signify potential areas of improvement in either the products themselves or the clarity of information available to customers pre-purchase. Companies may wish to investigate the nature of these calls to identify any common concerns or recurring issues, potentially reducing the need for repeated customer service interactions.

Customer Rating. The Customer Rating variable is spread fairly evenly across the five rating categories, with all ratings between 1 and 5 receiving close to 20% each. Specifically, 1-star ratings account for 20.32%, 2-stars for 19.68%, 3-stars for 20.36%, 4-stars for 19.90%, and 5-stars for 19.73%. This even distribution might indicate varied customer satisfaction levels, suggesting that while some customers are highly satisfied, an equally large portion is dissatisfied. The relatively high percentage of low ratings (1 and 2 stars) alongside the moderate number of high ratings (4 and 5 stars) points to inconsistency in customer experience or product quality. This could be due to variations in product types, different customer expectations, or quality control issues. Analyzing the factors associated with these ratings, such as product category, delivery timeliness, or frequency of customer care interactions, might help identify and address the root causes of dissatisfaction.

Prior Purchases. The Prior Purchases variable reveals that most customers have made relatively few previous purchases, with 2 prior purchases representing 23.63% and 3 prior purchases the most common with 35.96%. This gradually decreases with higher prior purchase counts, with only a small percentage of customers (less than 2%) having made 8 or more prior purchases. This trend suggests that a large portion of the customer base consists of relatively new or occasional buyers. The gradual drop in frequency as the number of prior purchases increases indicates that customer retention may be limited. Efforts to increase repeat purchases, perhaps through loyalty programs or improved post-purchase engagement, could help encourage more repeat customers and build a more loyal customer base. Understanding what drives repeat purchases, such as product satisfaction or effective customer support, could inform targeted retention strategies.

Reached on Time. The Reached on Time variable shows that 40.33% of deliveries did not reach the customer on time, while 59.67% did. This relatively high rate of delayed deliveries may impact overall customer satisfaction and could be a significant factor in the lower ratings observed in the Customer_rating variable. Timely delivery is often crucial for positive customer experiences, and delays can contribute to dissatisfaction and negative reviews. This delay rate suggests potential operational challenges in the delivery process, whether due to logistical issues, product availability, or external factors like shipping delays. Addressing these delays could be a high-impact area for improving customer satisfaction. Companies could consider optimizing inventory management, enhancing logistics, or setting more accurate delivery expectations to reduce the frequency of late deliveries.

In summary, this dataset reflects both strengths and areas for improvement. While there is a consistent frequency of customer care calls, the even distribution of customer ratings suggests mixed satisfaction levels. Many customers are new or infrequent buyers, with limited repeat purchases, and a significant percentage of orders are delayed. Addressing these challenges, particularly by enhancing delivery reliability and reducing customer service needs, could improve overall customer satisfaction, increase retention, and strengthen customer loyalty (Figure 2).

Figure 2. Frequencies.

Frequencies for Customer_care_calls

Customer_care_calls	Frequency	Percent	Valid Percent	Cumulative Percent
2	638	5.801	5.801	5.801
3	3217	29.248	29.248	35.049
4	3557	32.339	32.339	67.388
5	2328	21.166	21.166	88.554
6	1013	9.210	9.210	97.763
7	246	2.237	2.237	100.000
Missing	0	0.000		
Total	10999	100.000		

Note. The following variables have more than 10 distinct values and are omitted: Cost_of_the_Product, Discount_offered, Weight_in_gms.

Frequencies for Customer_rating

Customer_rating	Frequency	Percent	Valid Percent	Cumulative Percent
1	2235	20.320	20.320	20.320
2	2165	19.684	19.684	40.004
3	2239	20.356	20.356	60.360
4	2189	19.902	19.902	80.262
5	2171	19.738	19.738	100.000
Missing	0	0.000		
Total	10999	100.000		

Frequencies for Prior_purchases

Prior_purchases	Frequency	Percent	Valid Percent	Cumulative Percent
2	2599	23.629	23.629	23.629
3	3955	35.958	35.958	59.587
4	2155	19.593	19.593	79.180
5	1287	11.701	11.701	90.881
6	561	5.100	5.100	95.981
7	136	1.236	1.236	97.218
8	128	1.164	1.164	98.382
10	178	1.618	1.618	100.000
Missing	0	0.000		
Total	10999	100.000		

Frequencies for Reached.on.Time_Y.N

Reached.on.Time_Y.N	Frequency	Percent	Valid Percent	Cumulative Percent
0	4436	40.331	40.331	40.331
1	6563	59.669	59.669	100.000
Missing	0	0.000		
Total	10999	100.000		

3.3 Distributions

In the following paragraph the distributions of the variables are analyzed.

Customer Care Calls. The distribution of Customer care calls shows a peak around 3-4 calls, with fewer customers making 6 or 7 calls. This pattern suggests that most customers make a moderate number of calls to customer service, likely for standard inquiries or follow-ups. However, there are fewer extreme cases, indicating that, while some customers may have recurring issues, they are not widespread. Reducing the need for repeated calls through enhanced FAQs, proactive support, or improved product instructions could potentially decrease customer care calls and improve customer satisfaction.

Cost of the Product. The Cost of the Product distribution is relatively normal with a few peaks, indicating that most products fall within a similar price range, with a central tendency around 200-250 units. The distribution is slightly skewed, showing that while there are some higher-priced items, most products are priced near the average. This balanced pricing structure may reflect a standardized product range with limited luxury or budget options. Companies could explore offering products at different price points to appeal to a broader customer base or meet various budget requirements.

Customer Rating. The Customer Rating plot is almost uniform across all rating levels from 1 to 5. This distribution indicates that customer satisfaction is mixed, with each rating from 1 (lowest) to 5 (highest) receiving similar counts. This balanced yet varied distribution could mean that customer experiences are inconsistent, possibly due to variability in product quality, service, or delivery. Improving consistency in product quality and customer experience could help shift this distribution toward higher ratings, indicating better overall customer satisfaction.

Prior Purchases. The Prior Purchases plot shows a skewed distribution, with most customers having made only 2-3 prior purchases and significantly fewer customers making a high number of repeat purchases (6 or more). This suggests that customer retention may be a challenge, as few customers make frequent repeat purchases. Strategies like loyalty programs, targeted marketing for previous customers, or personalized recommendations could help increase repeat purchase rates and build a more loyal customer base.

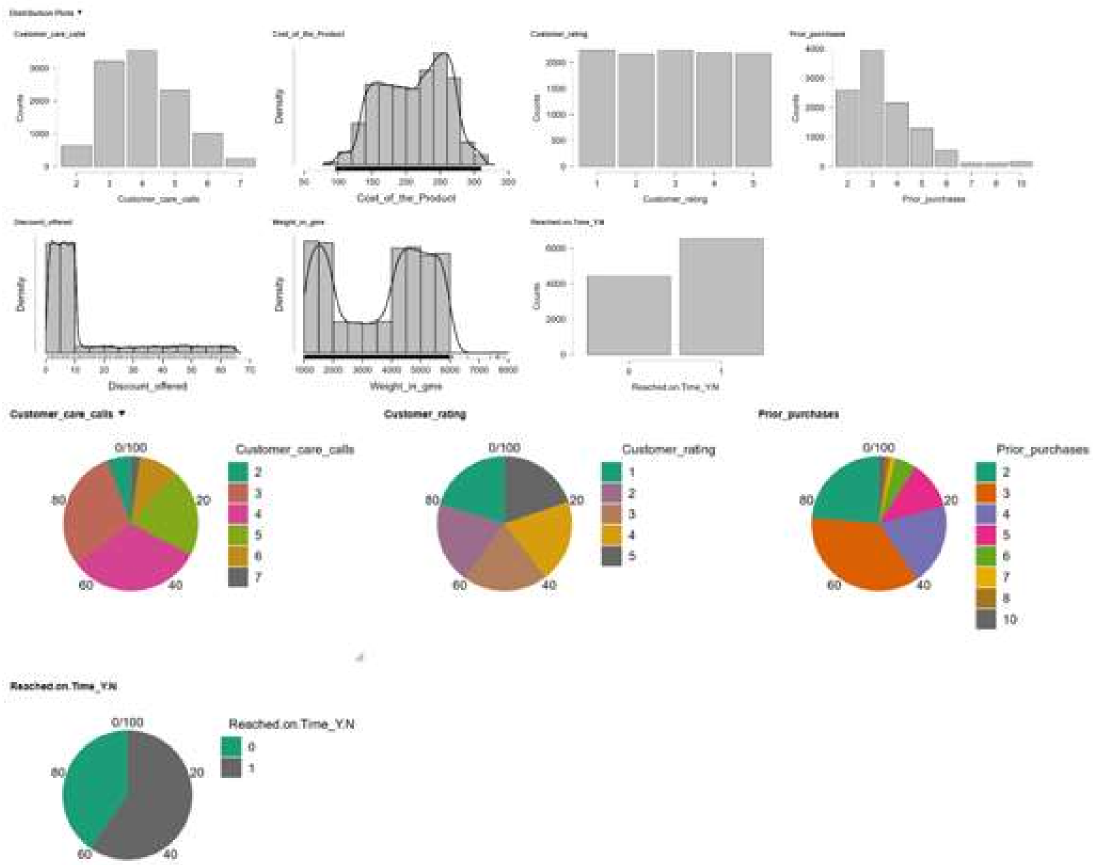
Discount Offered. The Discount_offered distribution is highly skewed, with most discounts clustered around the lower end, likely between 1 and 20%, with a few cases of higher discounts. This pattern indicates that while some discounts are offered, they are generally modest. Higher discounts may be offered selectively, perhaps to clear inventory or attract new customers. Analyzing the effectiveness of different discount levels on customer purchasing behavior could help refine the discount strategy, potentially leading to increased sales without needing significant markdowns.

Weight in Grams. The Weight in gms plot shows a bimodal distribution, indicating two main clusters of product weights. This could suggest two distinct types of products, one lighter and one heavier, possibly representing different product categories or types. For instance, the lighter products might be smaller items or accessories, while the heavier items could be more substantial goods. This distinction is important for logistical planning and could influence customer expectations, especially for delivery timelines and costs. Optimizing logistics based on product weight categories could improve efficiency and customer satisfaction.

Reached on Time. The Reached on Time variable is binary, indicating whether products were delivered on time (1) or not (0). The plot shows that about 60% of deliveries were on time, while 40% were delayed. This relatively high rate of delays could negatively impact customer satisfaction and is likely correlated with lower customer ratings. Addressing these delays, whether through better inventory management, streamlined logistics, or more accurate delivery estimates, could improve customer satisfaction and potentially reduce negative reviews.

The distribution plots reveal several patterns in customer behavior, product attributes, and operational efficiency. Customer satisfaction appears mixed, with uniform ratings distribution and a relatively high number of customer care calls. Limited repeat purchases indicate potential challenges with customer retention, while the high rate of delayed deliveries suggests logistical issues. Addressing these factors, such as by improving product consistency, enhancing delivery reliability, and encouraging repeat purchases, could strengthen customer loyalty and satisfaction, ultimately leading to better business performance (Figure 3).

Figure 3. Distribution plots.



3.4 Q-Q plots

In the following paragraph the Q-Q plots are analyzed.

Customer Care Calls. The Customer Care Calls Q-Q plot shows clear deviations from the normal line, with data points clustering at specific values along the y-axis, creating a stair-step pattern. This suggests that the variable is discrete rather than continuous and does not follow a normal distribution. This pattern is expected because customer care calls are typically limited to integer values (e.g., 2, 3, 4 calls), making it non-normal. The distribution is skewed, indicating that most customers make a similar number of calls, with fewer outliers.

Cost of the Product. The Cost of the Product Q-Q plot shows a closer fit to the normal line than some of the other variables, with points following the line reasonably well in the middle. However, deviations are seen at both ends, with the tails deviating from the line, indicating that the cost distribution has some skewness or kurtosis, with a few products priced significantly higher or lower than the central tendency. This suggests a relatively normal distribution with potential outliers on both ends.

Customer Rating. The Customer Rating plot exhibits a stair-step pattern similar to Customer Care Calls due to the discrete nature of this variable, with ratings typically limited to integer values from 1 to 5. The data points do not follow a normal distribution and deviate significantly from the red line, especially at the tails. This non-normality reflects the ordinal nature of customer ratings and the fact that ratings are not continuous values, thus making them inherently non-normal.

Prior Purchases. The Prior Purchases Q-Q plot also displays a non-normal, discrete distribution, with data points clustering in a stair-step fashion along certain quantile values. This pattern suggests that prior purchases are limited to integer counts, with most customers having a lower number of prior purchases and fewer with high counts. The clear deviation from the normal line confirms that this variable does not follow a normal distribution, as it is heavily skewed with more values concentrated at the lower end.

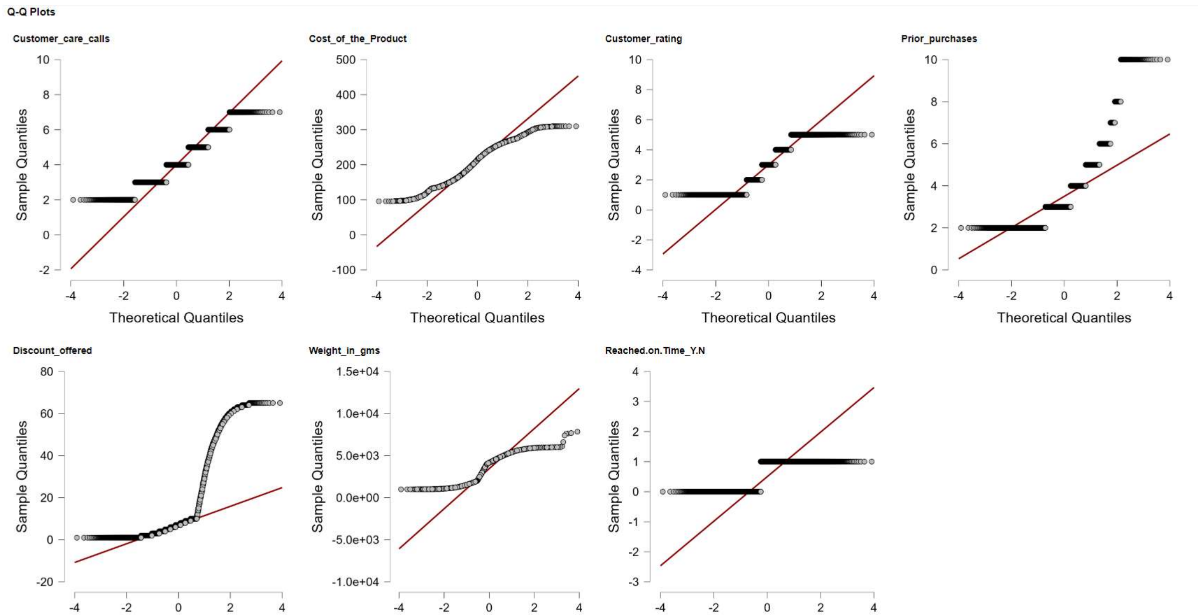
Discount Offered. The Discount Offered plot deviates from normality, with most points diverging from the line, especially in the tails. The points display a right-skewed pattern, indicating that most discounts are on the lower end, with a few outliers where high discounts are offered. This skewed distribution likely results from a discount strategy focused on smaller discounts for most products, with only occasional high discounts.

Weight in Grams. The Weight in Grams plot shows significant deviation from normality, particularly with heavy tails on both ends. The points do not follow the normal line closely and appear clustered in specific regions, suggesting the presence of different product types with distinct weight ranges. This bimodal or multimodal distribution indicates that product weights vary widely, likely due to different product categories (e.g., lighter versus heavier products), making the distribution far from normal.

Reached on Time. The Reached on Time plot demonstrates a clear departure from normality, with only two distinct levels (0 and 1), representing a binary outcome. This plot is expected to show a non-normal distribution as it is categorical. Points are clustered at two points on the y-axis, indicating that this variable does not fit a normal distribution and represents an outcome rather than a continuous metric.

In summary, these Q-Q plots reveal that most variables do not follow a normal distribution due to their discrete or categorical nature. Variables such as Customer care calls, Customer rating, Prior purchases, and Reached on Time are inherently non-normal as they are either ordinal or binary. Cost of the Product and Discount Offered show slight normality but have skewed distributions, while Weight in Grams appears bimodal, indicating distinct product groups. Understanding these distributions is essential for selecting appropriate statistical analyses, as many traditional tests assume normality, which is not the case for most of these variables (Figure 4).

Figure 4. Q-Q plots.



3.5 Correlation

The following paragraph analyses the correlations among the variables.

Customer Care Calls. The density plot for Customer Care Calls shows multiple peaks, suggesting that there are several common levels at which customers reach out, likely due to recurring issues or support needs. In the scatterplots involving Customer Care Calls, there doesn't appear to be a strong linear relationship with any other variables, as most of the points are scattered. This may imply that the number of customer care calls is influenced by factors not directly shown in this dataset or is impacted by multiple independent factors.

Cost of the Product. The density plot for Cost of the Product. shows a somewhat normal distribution but with slight skewness, indicating a typical pricing range with some higher-priced products. When examining its relationships with other variables, there are no significant trends in the scatterplots. This suggests that product cost does not strongly correlate with other variables like customer rating, prior purchases, or discount offered. It's possible that customer satisfaction, as represented by ratings or calls, is not strongly impacted by the price of the product alone.

Customer Rating. The Customer Rating density plot shows distinct peaks at each rating level, as expected for an ordinal variable. The scatterplots involving Customer Rating reveal no strong patterns or linear relationships with other variables. This lack of correlation could indicate that customer ratings are influenced by a wide range of factors beyond those captured in the dataset, such as customer expectations or external factors not related to the variables shown.

Prior Purchases. The Prior Purchases density plot has a peak at lower purchase counts, suggesting that most customers have only made a few prior purchases. This variable doesn't exhibit any clear relationship with other variables in the scatterplots. The lack of correlation may imply that the number of previous purchases does not significantly affect variables like customer rating, calls, or discount received, possibly due to diverse purchasing patterns or customer preferences.

Discount Offered. The Discount Offered density plot shows a high concentration at the lower end, with most discounts clustered around small values. In the scatterplots, Discount Offered does not exhibit any strong correlations with other variables, indicating that discounts are likely distributed independently of customer ratings, prior purchases, or product cost. This pattern might suggest a general discount strategy that is not tailored to individual customer profiles or purchase histories.

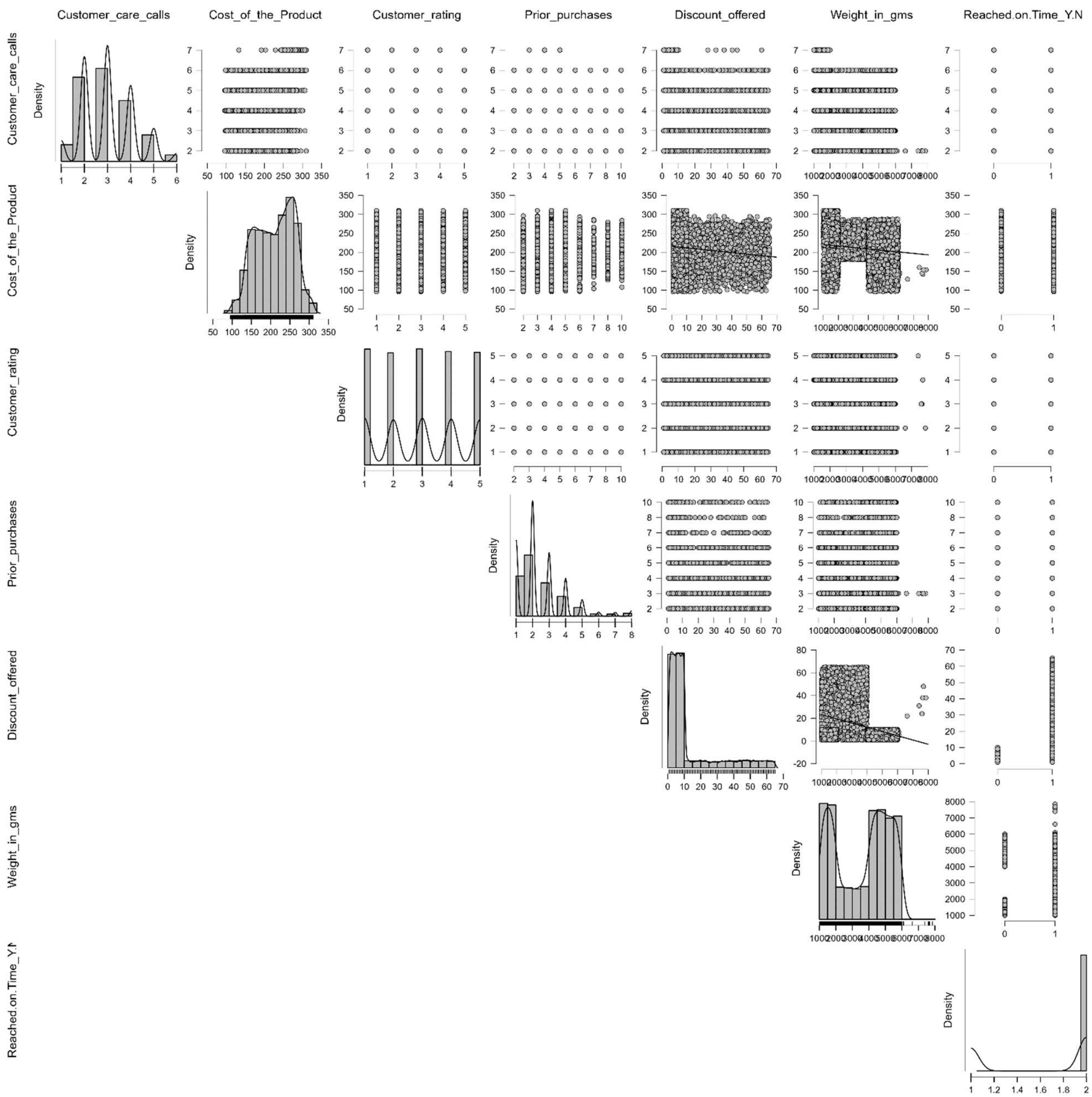
Weight in Grams. The Weight in Grams density plot is bimodal, reflecting two primary weight categories, which could indicate distinct types of products (e.g., lightweight items versus heavier products). However, scatterplots do not reveal any clear linear relationships between Weight in Grams and other variables. This lack of correlation could mean that product weight does not directly influence customer care calls, ratings, or discounts offered, although it may indirectly affect logistics or delivery times.

Reached on Time. Reached on Time is a binary variable with two possible values (0 and 1), representing whether the product was delivered on time. The density plot shows two distinct points, reflecting this binary nature. In the scatterplots, there is no strong relationship between Reached on Time and other variables. This lack of correlation may suggest that on-time delivery is influenced by external logistical factors rather than by customer behavior, product cost, or other characteristics within this dataset.

Overall, the pair plot matrix shows limited linear relationships or strong correlations between most variables, indicating that many of these factors operate independently or are influenced by external elements not captured in the dataset. Key takeaways include:

- Customer Care Calls and Customer Rating show no strong dependencies on factors like product cost, discount, or prior purchases, suggesting that customer satisfaction may be influenced by other factors.
- Discount Offered and Weight in Grams appear distributed independently of other variables, suggesting broad discounting practices and diverse product types.
- Reached on Time does not correlate strongly with other factors, which might reflect logistical or operational influences on delivery times rather than product or customer characteristics.

See Figure 5. Correlations.



3.6 Pareto distributions

The following paragraph analyses the characteristics of Pareto distributions.

Customer Care Calls:

- Description: This plot shows the count of customer care calls made by customers.
- Observation: Most customers made between 4 to 6 calls, with a sharp drop in the frequency for 7 or more calls.
- Cumulative Percentage: The cumulative line indicates that approximately 80% of calls are covered by up to 6 calls (marked by the orange dashed line).
- Implication: Limiting the number of customer care calls to 6 could address most cases, suggesting that focusing on customers who exceed this threshold might help in improving service efficiency.

Customer Rating:

- Description: This plot shows the distribution of customer satisfaction ratings.
- Observation: Ratings are clustered around 3 to 5, with each rating having a similar frequency.
- Cumulative Percentage: Around 80% of ratings are covered up to a score of 5 (again highlighted by the orange dashed line).
- Implication: Since most customers rate between 3 and 5, improving customer satisfaction efforts for this segment could enhance overall ratings. The plot also suggests limited occurrences of very low or very high ratings.

Prior Purchases:

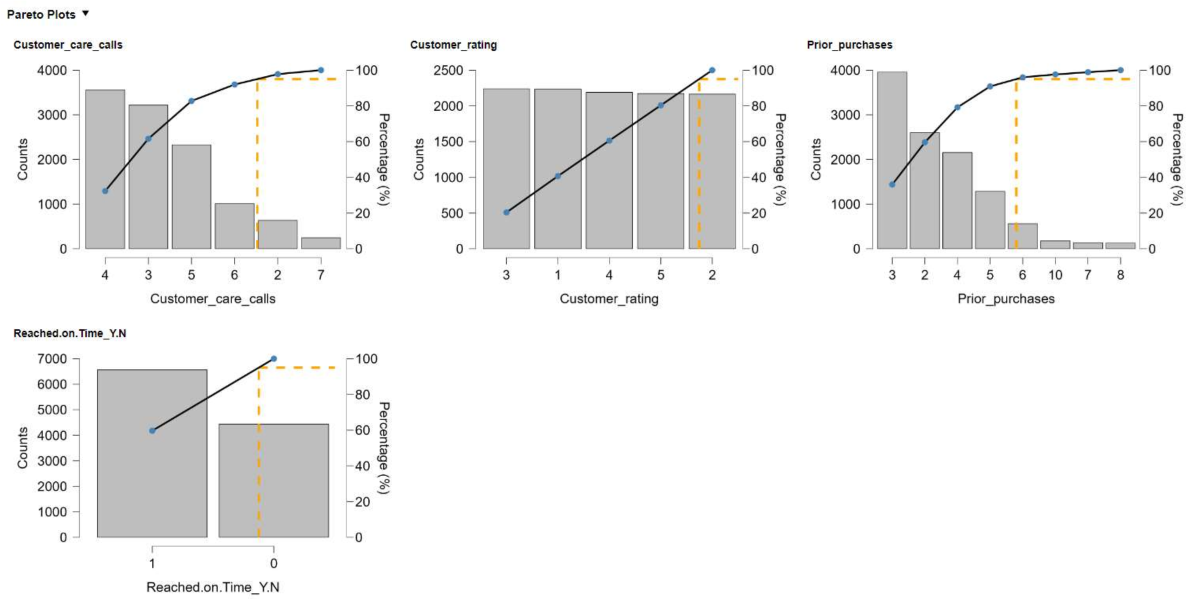
- Description: This plot displays the count of prior purchases made by customers.
- Observation: A high frequency is observed for 3 to 4 prior purchases, with counts diminishing rapidly for higher numbers of purchases.
- Cumulative Percentage: The cumulative line shows that up to 6 purchases account for around 80% of cases (indicated by the orange dashed line).
- Implication: Since most customers have made 3 to 4 prior purchases, strategies like loyalty rewards or targeted promotions for this group could help retain these high-frequency buyers.

Reached On Time:

- Description: This plot illustrates the count of on-time versus delayed deliveries (1 = on time, 0 = not on time).
- Observation: There is a higher count of deliveries marked as "1" (on time) than "0" (not on time).
- Cumulative Percentage: The orange dashed line indicates that around 80% of deliveries were made on time.
- Implication: Since the majority of deliveries are on time, the company performs well in terms of delivery timeliness. However, addressing the delayed deliveries could help improve customer satisfaction.

Each of these Pareto plots highlights areas where the 80-20 rule applies. For example, in customer care calls, focusing on those with more than 6 calls, improving satisfaction among mid-level ratings (3-5), engaging customers with frequent prior purchases, and ensuring timely delivery could all contribute to improving service quality and customer satisfaction. The cumulative lines and dashed thresholds in each plot are instrumental in identifying where 80% of the issues or opportunities lie (Figure 6)

Figure 6. Pareto Plots.



4. Econometric Models

We have estimated the following equation:

$$CustomerCareCalls = a_1 + b_1(PriorPurchases) + b_2(DiscountOffered) + b_3(WeightsInGms) + b_4(CostOfProduction)$$

A synthesis of the analysis is in the following table 3.

Table 3. Results of the econometric analysis.

	Heteroskedasticity-corrected, using observations 1-10999			LAD, using observations 1-10999		
	Coefficient	Std. Error	t-ratio	Coefficient	Std. Error	t-ratio
const	3.46645***	0.0682722	50.77	3.53223***	0.119979	29.44
Prior_purchases	0.0589729***	0.00651735	9.049	0.0640725***	0.00779334	8.221
Discount_offered	-0.0137017***	0.000696367	-19.68	-0.0154269***	0.00119056	-12.96
Weight_in_gms	-0.000204920***	6.47845e-06	-31.63	-0.000229838***	7.45585e-06	-30.83
Cost_of_the_Product	0.00604065***	0.000235451	25.66	0.00625005***	0.000437141	14.30
	WLS, using observations 1-10999 (n = 6563)			OLS, using observations 1-10999		
	Coefficient	Std. Error	t-ratio	Coefficient	Std. Error	t-ratio
const	3.23718***	0.0760699	42.56	3.59481***	0.0639107	56.25
Prior_purchases	0.0806730***	0.00849277	9.499	0.0612760***	0.00656531	9.333
Discount_offered	-0.0116684***	0.000735365	-15.87	-0.0145194***	0.000666951	-21.77
Weight_in_gms	-0.000175986***	8.83984e-06	-19.91	-0.000215029***	6.66901e-06	-32.24
Cost_of_the_Product	0.00610642***	0.000267495	22.83	0.00578807***	0.000208647	27.74

Positive relationship between customer care calls and prior purchases. The positive relationship of customer care calls and prior purchases shows that with higher purchases made by customers, they interact more with customer care. This dynamic might reveal a multitude of facets with regard to customer behavior and interaction. First, repeat customers are more acquainted with the company and probably feel their problems can be readily addressed to the customer support. Thus, with increased involvement with the brand, they may also develop higher expectations regarding service quality and product performance, possibly leading to more interactions with customer care. In addition, these very frequent buyers could feel more invested in the brand-financially and emotionally-and are, therefore, more apt to pursue assistance when something does not meet their expectations. This broader product experience may lead them to customer care for assistance in features, product care, or troubleshooting. Besides, if there is some inconsistency in product quality or consistency, frequent buyers would be more likely to come across these inconsistencies and hence feel the need to reach out for support. Another reason for this relationship could be the retention and loyalty efforts of the company. Customer service teams could provide more attentive support toward repeat buyers, thereby increasing interaction rates as part of a strategy for building loyalty. This proactive engagement may favorably affect the relationship in which customers will feel valued and bond with the brand. This would therefore establish a relationship between customer care calls and previous purchases, indicating that much valuable information can be obtained. It would also show that the customers are highly active, with a greater proportion of repeated buyers. On the other hand, it could also indicate sources of friction in product quality or unclear information. By delving deeper into exactly what such interactions entailed, the company would be in a position to determine with more specific detail if there was some kind of common problem with repeat buyers and take actionable steps to diminish unnecessary frequency of support. With these loyal customers in mind, more self-service resources can be developed through enhancements in the form of FAQs or tutorials and knowledge bases; these will definitely help customers solve on their own some of the very general issues. Indeed, personalized support for regular buyers could also enrich their experience and allow businesses to show appreciation for buyers' loyalty. Also, gathering feedback from these interactions may shed light on certain product consistency issues and serve as a driver for improvement efforts that make the experience seamless for loyal customers. In all, frequent customer care calls from repeat buyers may reflect the best engagement, but it also is an opportunity for the company to sharpen support resources and root out the problems. Once done, the company can refine the experience, thus fueling stronger loyalty and satisfaction with its most engaged customers (Khatoon et al., 2020; Yaqub et al., 2023; Sharma and Singh, 2023).

The negative relationship between customer care calls e discount offered. The negative association of customer care calls and discount offered: probably indicates for increased discounts; the frequency of customer care calls tends to go down. The interpretation for this may be varied and could draw some inferences related to customer behavior, expectation, and satisfaction. One probable explanation is that the availability of discounts has some kind of psychological effect, raising customer tolerance and putting them in a state where they feel satisfied. In certain occurrences, customers who receive a fat discount will be more tolerant to small issues or inconveniences, thinking they have gotten good value for money. In this way, discounts revolve around dissatisfaction by reducing cases where customers reach customer support over minor complaints. Additionally, deeper discounts would capture more price-conscious customers who may have lower expectations about premium service. These may weigh more on the deal they got as opposed to receiving faultless service or the best product quality, which could reduce the likelihood of their reaching out for support should minor issues arise. On the other hand, customers paying closer to full price may have higher expectations and thus be more willing to contact customer care if their experience doesn't meet those expectations. In addition, the discounts may also see quick purchasing decisions consequently potentially reducing customer care calls. Such customers, seeing that they are getting a really good deal, would likely be focused more on immediate benefits of having the discount rather than scrutiny into product details or specifications. Because they feel better value, this means they are less likely to call in customer support with questions or to complain. In that respect, the discounts can shift their attention to the deal and away from minor shortcomings. This might also mean that discounts are given for products that are simple and less likely to raise inquiries from customers. If the company selectively applies discounts to products that have fewer known issues or simpler use, there would naturally be fewer reasons for customers to call customer care. Whereas, complex or higher-priced products that do not usually have discounts may cause more questions or concerns; this would, in turn, increase the number of customer care interactions. Although discounts have a positive effect on customer satisfaction and decrease the need for customer care, they should be taken up cautiously. Where companies become dependent on discounts, they risk only attracting price-sensitive buyers, yet a factor which may degrade the brand's value over time. In other words, when customers frequently receive discounts, they will perceive value only if a discount is offered, thus depressing full-price sales. This is further supported by the negative relationship between customer care calls and discounts wherein, by offering a discount, it creates a value perception in the mind of the consumer whereby a person is unlikely to call if minor dissatisfaction occurs. Companies must avoid being too generous with discounts, though-they can have an unpredictable effect on customer expectations and behaviour (Yanchenko and Repilova, 2023; Troebs et al., 2021; Liu et al., 2021).

The negative relationship between Customer care calls and weight in gms. This negative correlation of Customer care calls and weight in gms might be interpreted as meaning there is a drop in the frequency of customer calls with increased weight of a particular product. This relationship is interesting with respect to expectations from customers regarding the nature of the product and the kind of issues that generate calls for support. Higher levels of weight could lead customers to perceive the products as more substantial or even reliable, probably lowering their need for support since they are less likely to have problems or questions about the product's quality or performance. The heavier products may also be those products in which customers have more confidence and understanding, usually because more research is conducted or much contemplation occurs over shipping costs or how much logistics it would take, as the costs associated with them are usually higher than other items. Taking a large appliance or furniture item as an example, these are typically big and heavy items in which lots of research is often done by the customer before buying, leaving them with only a few questions or misunderstandings that are normally left post-consumption. Such products themselves are more likely to have minor problems, breakages, or operational questions from customers who may call customer support more often. Lighter items might also be purchased more on impulse, with less prior research, thus leading to more questions or uncertainties afterward and higher customer care needs. Besides, logistics behind the transportation of heavier products may also be an issue. Heavier products are also sent out with special treatment, so to say, maybe by offering value-added services in the way of installation or assembling within a customer's home, which in turn might reduce the likelihood of confusion from the customer's end or troubleshooting that has specifically to be done. Again, these value-added services may avoid many frequently asked questions by the customers because customers get more support with heavier objects than with lighter ones. On the other hand, it is possible that the negative relationship between customer care calls and product weight may indicate selective customer behavior or a different set of expectations. The buying consumers who purchase heavier items may be more deliberate and patient, and they might expect minor difficulties arising from the nature of the product itself. For this reason, they are less likely to call customer support for small inconveniences or challenges, since they can undertake resolving such issues on their own. The negative relation of customer care calls to the product weight might be due to conceptual durability, better confidence among customers, extensive research before buying, or even extra support services for heavier products. On the other hand, this inversely related relationship indicates that heavier products contribute less to after-care needs of customers; however, companies are called upon to make sure a smaller, lighter weighted product satisfies the need of customers through smooth information and support resources to minimize any postpurchase assistance needs (Garver and Williams, 2020; Qin, et al., 2021; Du et al., 2022).

The positive relationship between customer care calls and cost of the product. The positive relation of customer care calls to cost of the product infers that when the price of a product increases, so does the number of customer care calls. This relationship describes a number of concepts in regard to customer expectations, the level of difficulty with a product, and the level of service customers are hoping for in relation to higher-priced products. Customers who purchase high-dollar products expect better quality, performance, and customer support. Spending a lot on a product makes customers more uptight for any issue or inconvenience; after all, they expect value from the product amounting to the price they paid for. This rise in expectation may make them more eager to finally contact customer support even on minor issues, since they don't want anything that will make the product not live up to their expectations regarding quality and functionality. The higher cost, in other words, gives a reason to feel entitled for premium service, encouraging customers to ask for help where and when needed. Second, the more expensive a product is, the more sophisticated it generally becomes: advanced features, installation requirements, and use by specialists. Sophistication and complexity mean that customers will most likely place themselves in situations where they have to seek assistance or advice, assuming that the product requires the customer to learn something new. This will probably lead to longer support interactions whereby customers seek the intervention of customer care for setting up, troubleshooting, or even locating certain features. They are usually higher-end electronics, appliances, and equipment where customers might require additional assistance because of how complex the product is. Apart from that, a customer buying expensive items may have more needs for investment protection. They would like to contact customer care about the proper usage of the product, warranty information, or maintenance advice for the product to last long and serve to its fullest capacity. The desire to maximize longevity and performance can therefore cause one to seek reassurance and advice through increased frequency of customer care contacts on how well their investments are protected. The positive relationship of the customer care calls with the cost of a product perhaps underlines the company's proactive support approach for high-value customers. Companies may consider such customers to be of higher financial value and offer them more personalized or attentive customer care, thereby encouraging frequent communication. This level of support is very conducive to relationships and customer satisfaction, reflecting upon the fact that a company is behind its high-value customers in helping them derive necessary assistance from what they have purchased. Although this relationship underlines the need for customer support for high-value items, it also gives a hint of areas where companies can optimize the customer experience. Clearer product documentation, comprehensive user guides, or instructional resources could help address common questions and reduce the need for frequent calls. Offering value products with premium support channels, such as a dedicated hotline for hotline service or priority service, will enhance customer satisfaction since the support accorded is in tandem with what customers perceive for that class of product or service (Supriyanto et al., 2021; Nguyen et al., 2020; Lim et al., 2021).

4.1 Estimation of operational efficiency in delivery and company customer care

Furthermore to have a measure of the efficiency in the delivery of the products towards consumers we have estimated the following equation:

$$\begin{aligned} ReachedOnTime &= a_1 + b_1(PriorPurchases) + b_2(DiscountOffered) + b_3(WeightInGms) \\ &+ b_4(CostOfTheProduction) + b_5(CustomerCareCalls) \end{aligned}$$

The results of the econometric analysis is showed in the following Table 4.

Table 4. Results of the econometric analysis for the estimation of the level of Reached on Time.

	Interval estimates, using observations 1-10999			Tobit, using observations 1-10999		
	Coefficient	Std. Error	z	Coefficient	Std. Error	z
const	0.878303***	0.0316364	27.76	0.865857***	0.0519022	16.68
Prior_purchases	-0.0146037***	0.00287517	-5.079	-0.0235274***	0.00483894	-4.862
Discount_offered	0.00953356***	0.000297134	32.09	0.0138165***	0.000479534	28.81
Weight_in_gms	-5.33283e-05***	3.04351e-06	-17.52	-8.70940e-05***	5.09204e-06	-17.10
Cost_of_the_Product	-0.000300506***	9.41451e-05	-3.192	-0.000416307***	0.000157158	-2.649
Customer_care_calls	-0.0246753***	0.00416022	-5.931	-0.0393050***	0.00696097	-5.646

The negative relationship between Reached on Time and Prior Purchases. Items being delivered on time-or "Reached on Time"-in respect to the volume of prior purchases made by the customer. Otherwise, the higher the count of a customer's past purchases, the less the probability of timely delivery. Several factors can be responsible for this unexpected fit, among them issues of operational complexity, demand expectations, and challenges in prioritizing customers. One explanation for this negative correlation exists in the operational complexities that often characterize frequent purchasers. High-value or frequent customers may also have orders that are larger, more varied, or more complicated, thus taking more time in processing and handling. For example, one simple order is easier to handle than multiple items that entail variable packaging and delivery needs. Added complexity will be certain to create delays, especially in the peak periods of a business cycle. Moreover, repeat customers may expect quicker and more personalized service, which could make them even more aware of delays and report them, therefore further skewing the delivery times of repeat customers. Another reason that explains an inverse association of on-time delivery with the previous purchases is that with more purchased items, the supply chain bears greater service stress. Companies often invest much of their resources in attracting new customers and delivering efficiently to them the first time, since the quality of the first impression is crucial. Repeat customers may therefore be given less priority for speeding up order fulfillment, especially during peak periods, which then affects on-time delivery rates. Although repeat customers are appreciated, the resource function sometimes unconsciously goes to the new customers with a view to expanding the customer base at the cost of marginally less reliable service for frequent purchasers. Moreover, high-frequency customers may start to be quite demanding because of the confidence built over time and might thus comment more likely when this expectation is not met. In other words, a repeat buyer familiar with on-time deliveries is more aware of-and apt to complain about-delays when those happen and reports them more frequently compared with casual buyers. This may create a perception-justified by complaint or feedback data-that returning customers are experiencing degraded service quality in delivery timeliness. This, therefore, can lead to a negative relationship between the frequency of past purchases and timely delivery rates, and can be a consequence of actual operational difficulties combined with resource allocation strategies and increased customer expectations. The solution might be in rationalizing the respective logistics strategies so that the service consistency balance is maintained between new and old customers, while the loyal purchasers get timely delivery, which they expect and deserve (Supriyanto et al., 2021; Nguyen et al., 2020; Lim et al., 2021).

The positive relationship between Reached on Time and Discount Offered. In the study of consumer behavior and business strategy, questions about timing and discount incentives are often interrelated in their capabilities for driving both customer satisfaction and brand loyalty. Perhaps one of the more interesting dynamics in this regard is indeed the positive relationship that exists between products or services arriving on time for the customer and the effectiveness of the discount offer. When companies can deliver on time, the perception of added value with discounts is furthered, ultimately contributing to customer trust and increasing the likelihood of making purchase decisions. The reliability of on-time delivery is a very important parameter that helps build trust among customers. Once the company promises to deliver on time and always meets this expectation, then there is a better chance for the customer to perceive the brand positively. Trust is one of the main aspects of consumer loyalty, and discounts given with such a feeling of trust will add value to the whole transaction. For example, if the product arrives with the customer in the estimated time, then the consumer is more likely to view the discount as a real discount that created a superior shopping experience. In this way, on-time delivery enhances a discount offer by making it valid through the company's commitment to satisfying the customer. Discount as Incentive to Loyalty and Repurchase Discounts on their own can be considered a means by which companies can rid themselves of unwanted stock or entice new customers. Discounts, coupled with timely delivery however, are used more strategically to ensure customer loyalty and repeat purchase. Customers will begin using the discounts offered by companies they learn to consider reliable, especially when these same companies demonstrate time and time again that they can keep their promises of good service. This can bring about a snowballing effect between timely delivery and discounts to ensure repeated purchases, since customers would always want to come back for more when businesses give them a quality service with sufficient and reasonable financial reasons. It has been evidenced by available studies, showing that customers are more open to promotional offers made through businesses which, in the past, have successfully met or exceeded customer delivery expectations. These are the companies that can always retain timely delivery with discounts in order to be a step ahead within the market compared to their competition. This is because such a combination enables them to attain a unique standing, different from the competitors that might be less reliable or less transparent in regards to the delivery process. Offers like these of discounts with ensured delivery create a competitive advantage by first meeting and then outpacing customer expectations. For instance, e-commerce platforms that provide swift delivery and discounts often see higher retention rates, as customers value the convenience and savings that such companies offer. In sum, the positive correlation between arriving on schedule and giving discounts manifests in building trust, enhancing customer satisfaction, and gaining an advantage over competitors. This dynamic-sixteenth century reflects an effective business strategy where the operational reliability and financial incentives work together to raise the level of customer experience and foster long-term loyalty (Yildiz and Savelsbergh, 2020; DÜNDAR and ÖZTÜRK, 2020; Chen et al., 2023).

The negative relationship between Reached on Time and Weight in Gms. In this regard, the scientific observation of transport and delivery information systems indicates that there is a negative relationship between "Reached on Time" and "Weight in Gms", which, out of logic, would suggest that the greater the weight, the less likelihood that these items would reach a destination on time. It is also generally expected that across logistics and supply chain processes, the physical characteristics of an item, in particular weight, would strongly relate to the delivery speed and the precision of timing. This inverse relationship is underpinned by a number of factors: mainly, the constraints of heavy weights on transportation modes, handling requirements, and the timeliness of delivery networks. The heavier the load, usually it means delivery speed suffers due to weight affecting the ease and cost of transport. Heavier parcels require handling processes that are more careful and often slower. For instance, heavier packages may require specialized equipment or a number of personnel in order for the packages to be loaded, unloaded, or transshipped from one mode of transportation to another. This in turn increases the time taken at every transit point and usually leads to delays. In a delivery system that is supposed to be very efficient and on time, especially when lighter packages can easily be moved in bulk and fast, heavier packages act as a bottle-neck, disturbing this flow. In addition, weight limits affect the very transportation vehicles. For example, in air transport, aircraft subject heavy items to strict weight limits. An intention to move heavier parcels will require proper planning in relation to fuel efficiency, balance, and space utilization. Heavy items in the cargo may cause increased fuel usage and hence affect not only cost-efficiency but also scheduling of transport. Such weighty logistical headaches have repercussions on on-time delivery since heavy-laden vehicles may have to travel at reduced speeds or be subject to additional safety checks. Moreover, heavier goods are regularly classified as falling into slower shipping classes that emphasize cost over speed, making them less likely to meet tight delivery windows. This relationship may again reflect decisions in supply chain management to then favor lighter packages, as these will still be processed faster and hence can deliver a high volume of shipments within stressed time frames. In such an environment, where throughput will be maximized, this would be the strategic choice where "Reached on Time" shows a negative correlation with "Weight in Gms.". Therefore, the overall negative relationship between "Reached on Time" and "Weight in Gms" may be because of physical and logistical bondage that relates with weightier packages. These definitely increase the time required to process and handle, decrease efficiency in transportation, and lower the chances of delivery on time (Ricardianto, 2023; Zhao et al., 2021; Brylla and Walsh, 2022).

The negative relationship between Reached on Time and Cost of the Product. The relationship between reaching on time and the cost of a product is typically negative, meaning that as the emphasis on timeliness increases, so does the cost associated with achieving it. This relationship is especially evident in industries where tight delivery schedules are crucial, such as manufacturing, retail, and supply chain logistics. When companies prioritize timely delivery, they must adopt practices that often drive up costs, such as expedited shipping, increased labor, and enhanced inventory management systems. Each of these factors requires substantial investment, which subsequently impacts the product’s overall cost. One of the primary reasons for the increase in cost is the reliance on expedited shipping methods. To ensure that products reach customers on time, companies may resort to air freight or express courier services, both of which are significantly more expensive than standard shipping options. By opting for quicker delivery methods, firms can meet customer expectations for timely delivery, but they incur higher costs that are often passed onto consumers. Additionally, to guarantee on-time delivery, companies might invest in more sophisticated logistics technology, such as real-time tracking systems and predictive analytics. While these tools increase the efficiency and accuracy of delivery estimates, they also add to the product’s cost due to the technology's implementation and maintenance expenses. Labor costs also play a significant role in the negative relationship between timeliness and cost. Companies that prioritize on-time delivery often need to maintain a more flexible workforce, with employees ready to work additional hours or adapt to changing schedules to meet tight deadlines. Such flexibility frequently involves overtime pay or hiring temporary workers, both of which increase operational costs. For businesses in sectors such as manufacturing, where tight schedules are common, this could mean higher labor expenses that are ultimately reflected in the final price of the product. Another contributing factor is inventory management. Companies aiming for timely delivery might invest in buffer stocks or safety inventory to mitigate delays. Maintaining this surplus inventory requires additional storage space, insurance, and handling, which adds to the product’s cost. Moreover, having excess inventory ties up working capital, making it a costly strategy despite its role in ensuring timeliness. These expenses directly impact the cost structure of products, making them more expensive for the end consumer. In conclusion, the negative relationship between reaching on time and the cost of a product reflects the significant financial investments required to meet tight deadlines. From expedited shipping and advanced logistics technology to flexible labor and safety inventory, the various factors that enable on-time delivery come at a high cost. Thus, as the demand for punctuality grows, so does the economic burden placed on companies and, subsequently, on consumers (Niemi et al., 2020; Ricardianto et al., 2023; Liu and Yang, 2022).

The negative relationship between Reached on Time and Customer Care Calls. This important attribute of customer satisfaction or operational efficiency in service-oriented businesses is supplemented as the frequency of calls for customer care by and large inversely correlates with timely service. Generally, where services have been delivered on time, customers are more satisfied, thus reducing their need to call the customer care. The motivation for this negative correlation follows logically from the psychology of customer expectations and service reliability: on-time delivery of services builds trust and dispels uncertainty, hence reduces motivational need for contact on the part of customers through complaints or inquiries. Delays are usually the leading cause of customer dissatisfaction in industries for which timeliness is a critical service metric, such as logistics, telecommunications, and public transportation. As is more often than not the case in the logistics industry, when deliveries do not turn up as promised, clients call the customer care lines to find out the status of their orders, to complain, or seek compensation. Every delayed delivery does not only add operational costs but also has a bearing on the reputation of a company. When companies manage to attain high levels of on-time delivery, the customer gets a much more seamless service, less frustration, therefore fewer calls into customer support. Lack of this very same expectation causes disruption in customers' experiences, which is generally marked by anxiety or disappointment, and heightens the chances of customer contact with support. On the business side, if a business meets or exceeds these expectations with on-time arrivals, customers are reassured and there is typically lower anxiety about their transactions, effectively reducing the need for further communication. This not only helps in reducing costs associated with customer care but also improves overall customer experience-maturing support resources for complex issues rather than issues related to time. It follows that the inverse relation between on-time reach and frequency of customer care calls clearly indicates timeliness is an important determinant of customer satisfaction and operational effectiveness. Indeed, reaching clients on time contributes to the expectations of customers and helps the cause of cost-saving by allowing for better resource allocation in customer service departments. Therefore, it stands to reason that this is one of the main factors driving customer loyalty and improvement in company reputation (Ricardianto et al., 2023; Candra et al., 2022; Ridho et al., 2021).

5. Machine Learning Regressions

5.1 Boosting

The model is built using 88 trees, a relatively modest number that balances the depth and complexity of the model, potentially reducing the risks of overfitting or underfitting. The learning rate, or shrinkage, is set to 0.1, controlling the impact of each tree in the model. Lower learning rates typically improve generalization but require more trees to achieve accurate results. The model employs a Gaussian loss function, which is well-suited for continuous targets in regression tasks and assumes that errors follow a normal distribution. The dataset is divided into three parts: a training set with 7040 samples, a validation set with 1760 samples, and a test set with 2199 samples. This split allocates about 64% of the data to training, 16% to validation, and 20% to testing, which is common in machine learning practice. The model's validation mean squared error (MSE) is 0.727, while the test MSE is 0.724, indicating similar performance on both validation and test sets. This consistency between the validation and test MSE suggests that the model is likely generalizing well to new data. Additionally, the note mentions that the model is optimized with respect to the "out-of-bag mean squared error," implying the use of ensemble techniques or resampling to enhance accuracy and reduce bias. In terms of evaluation metrics, the mean squared error (MSE) for the model is 0.724, representing the average squared difference between predicted and actual values. The root mean squared error (RMSE) is 0.851, providing an estimate of average prediction error on the same scale as the original data. The mean absolute error (MAE), or mean absolute deviation (MAD), is 0.69, indicating the average absolute difference between predictions and actual values. This is often more interpretable than MSE as it is less affected by large errors. The model also has a mean absolute percentage error (MAPE) of 238.86%, suggesting substantial deviation between predictions and actual values, possibly due to non-linear relationships not fully captured by the model. Finally, the R-squared value (R^2) is 0.285, indicating that the model explains only about 28.5% of the variance in the target variable. This relatively low R^2 suggests that the model may be underfitting and not fully capturing patterns in the data. The relative influence of features in the model reveals that "Cost of the Product" is the most significant predictor with an influence of 48.396, indicating it has the strongest impact on the target variable. "Prior Purchases" has the second-highest influence at 30.594, demonstrating the importance of purchase history in the model's predictions. "Weight in grams" has an influence of 16.316, and "Discount Offered" has the lowest influence at 4.695, making it the least impactful predictor. This ranking provides insight into which variables the model relies on most heavily to make predictions. In summary, this Boosting Regression model is moderately complex with consistent performance across validation and test sets but limited predictive power, as reflected in the low R^2 and high MAPE. The main drivers of the model's predictions are the cost of the product and prior purchases, highlighting these factors as key to the target outcome. Although the model performs consistently across different data subsets, the high MAPE suggests areas for improvement in capturing the underlying relationships within the data (Chicco et al., 2021; Kaliappan et al., 2021; Ali, 2020) (Figure 7).

Figure 7. Boosting Regression.

Boosting Regression

Trees	Shrinkage	Loss function	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
88	0.100	Gaussian	7040	1760	2199	0.727	0.724

Note. The model is optimized with respect to the *out-of-bag mean squared error.*

Data Split

Evaluation Metrics

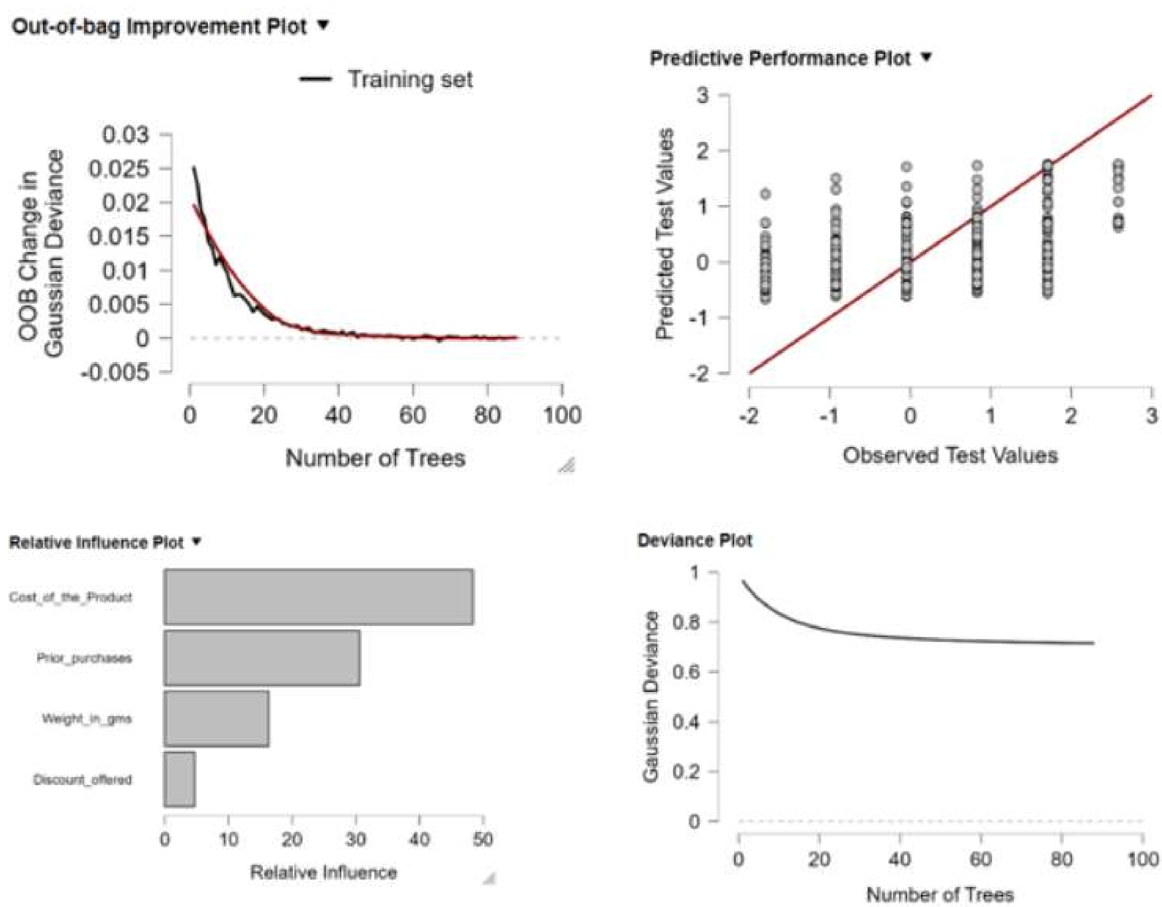
	Value
MSE	0.724
RMSE	0.851
MAE / MAD	0.69
MAPE	238.86%
R^2	0.285

Relative Influence

	Relative Influence
Cost_of_the_Product	48.396
Prior_purchases	30.594
Weight_in_gms	16.316
Discount_offered	4.695

The series of plots in Figure 8 provides insights into the performance and behavior of the Boosting Regression model. The "Out-of-bag Improvement Plot" in the upper left shows the change in Gaussian deviance across the number of trees for the training set. Initially, there is a significant improvement in out-of-bag (OOB) deviance as more trees are added, but this improvement tapers off around 50 trees, suggesting that additional trees have a diminishing impact on reducing deviance. The "Predictive Performance Plot" in the upper right compares predicted test values against observed test values. The red line represents the ideal scenario where predictions perfectly match actual values. However, the dispersion of points around the line indicates that there is considerable variability in the model's predictions, with predictions often deviating from the actual test values. This variability suggests some limitations in the model's predictive accuracy. The "Relative Influence Plot" in the lower left illustrates the relative importance of each predictor variable in the model. "Cost of the Product" is the most influential feature, followed by "Prior Purchases," which also plays a significant role. "Weight in grams" and "Discount Offered" have much lower relative influence scores, indicating they contribute less to the model's predictions. This ranking of feature importance aligns with the model summary, highlighting which features the model relies on most heavily. Finally, the "Deviance Plot" in the lower right tracks Gaussian deviance against the number of trees. The plot shows a gradual decline in deviance as more trees are added, though the rate of improvement slows over time. This curve reinforces the findings from the out-of-bag improvement plot, confirming that additional trees have a limited effect on improving model performance after a certain point. Overall, these plots collectively indicate that while the model does improve with more trees, its predictive performance has limitations, particularly in terms of accuracy and consistency with the test data. The high relative influence of "Cost of the Product" and "Prior Purchases" suggests that these features are key drivers in the model's predictions, even though the model itself may not fully capture the underlying complexity of the data (Plevris et al., 2022; Correndo et al., 2022; Suh et al., 2021).

Figure 8. Insights into the performance and behavior of the Boosting Regression model.



5.2 Decision Tree

The model was constructed with 44 splits and was trained on 8800 samples, with an additional 2199 samples allocated for testing. The test mean squared error (MSE) is reported as 0.635, indicating the average squared error of predictions on the test set. The data split section confirms that the majority of the data (80%) is used for training, with the remaining 20% used for testing, a common practice to ensure robust evaluation. In terms of evaluation metrics, the model's performance is characterized by an MSE of 0.635 and a root mean squared error (RMSE) of 0.797, both of which measure the error in prediction with RMSE providing a more interpretable scale. The mean absolute error (MAE), also called mean absolute deviation (MAD), is 0.646, which represents the average absolute difference between predicted and actual values. The mean absolute percentage error (MAPE) is 212.21%, suggesting that, on average, predictions differ significantly from actual values in relative terms. The R-squared (R^2) value of 0.339 indicates that the model explains about 33.9% of the variance in the target variable. This relatively low R^2 suggests that while the model has captured some predictive patterns, it may have limitations in fully explaining the data. The feature importance section provides insight into the variables that most influence the model's predictions. "Cost of the Product" is the most significant feature with a relative importance score of 46.396, indicating it plays the largest role in determining the target variable. "Prior Purchases" is the second most influential feature with a score of 28.593, followed by "Weight in grams" at 20.497. These three features together account for the bulk of the predictive power in the model. "Discount Offered" has a much lower importance at 3.824, while "Reached on Time (Y/N)" and "Customer Rating" contribute minimally with scores of 0.636 and 0.053, respectively. This distribution of feature importance suggests that the model's predictions rely heavily on product cost, purchase history, and weight, while factors like delivery timeliness and customer ratings have little impact. Overall, this Decision Tree Regression model shows moderate predictive capability with reasonably low error metrics but limited explanatory power as reflected by the R^2 value. The feature importance analysis highlights which attributes are most influential, offering insights into areas that may be key for improving model performance or refining business strategies based on these influential factors (Wong et al., 2023; Kaliappan et al., 2021).

Figure 9. Decision Tree Regression.

Decision Tree Regression ▼

Decision Tree Regression

Splits	n(Train)	n(Test)	Test MSE
44	8800	2199	0.635

Data Split

Evaluation Metrics

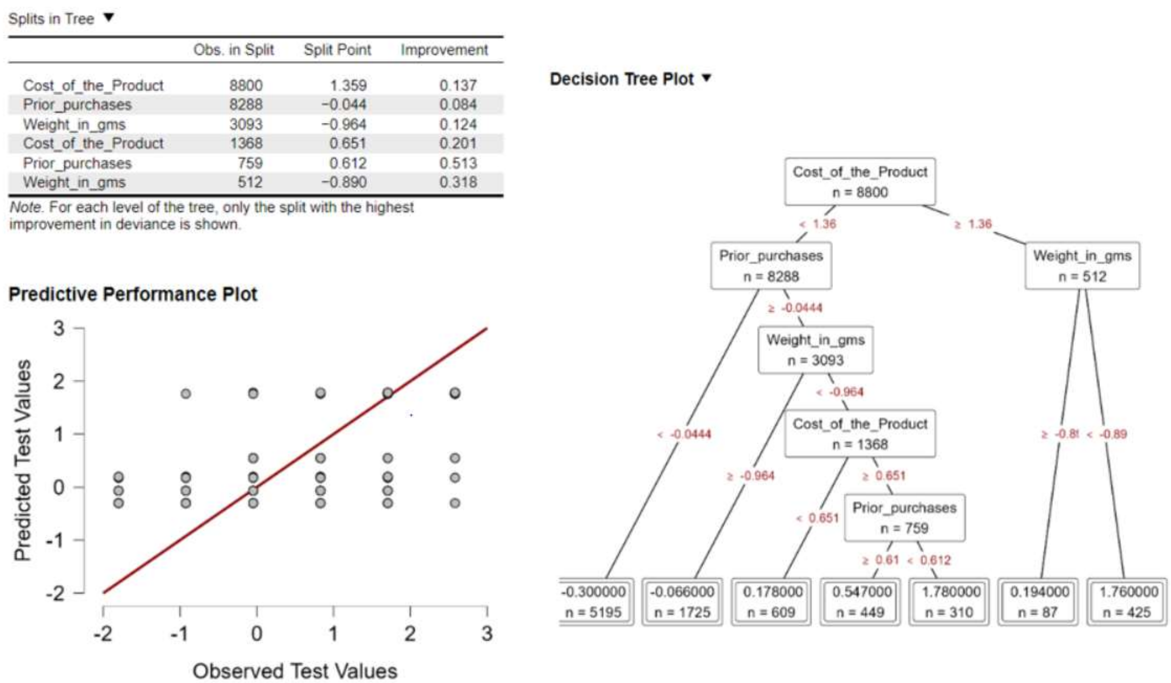
	Value
MSE	0.635
RMSE	0.797
MAE / MAD	0.646
MAPE	212.21%
R^2	0.339

Feature Importance

	Relative Importance
Cost_of_the_Product	46.396
Prior_purchases	28.593
Weight_in_gms	20.497
Discount_offered	3.824
Reached.on.Time_Y.N	0.636
Customer_rating	0.053

In Figure 10, At the top left, the "Splits in Tree" table outlines the primary splits made by the model to improve predictive accuracy. Each row lists a feature, the number of observations at that split, the specific split point, and the associated improvement in deviance. For instance, the first split on "Cost of the Product" with a split point at 1.359 yields a deviance improvement of 0.137. As the tree progresses, further splits are made based on features such as "Prior Purchases" and "Weight in grams," with each split contributing varying levels of improvement. This table shows that the tree prioritizes splits that maximize reduction in deviance at each step. The "Decision Tree Plot" on the top right visually represents the structure of the tree, starting with the root node split on "Cost of the Product." Each subsequent split directs observations through nodes that further divide based on conditions related to "Prior Purchases" and "Weight in grams." For example, after the initial split on "Cost of the Product," the tree splits based on "Prior Purchases" for a large subset of observations and then further splits on "Weight in grams." The nodes at the end of each branch represent final predictions made by the model, with each terminal node displaying the predicted value and the number of observations it contains. This structure illustrates the hierarchical approach taken by the decision tree to classify and predict values based on feature thresholds. The "Predictive Performance Plot" at the bottom left compares predicted test values to observed test values. The diagonal red line represents an ideal match between predicted and observed values, where all points would lie along the line in a perfectly accurate model. In this case, points are scattered around the line, indicating variability and some deviation between predictions and actual values. This dispersion suggests that the model is not perfectly accurate, with some predictions differing considerably from observed values, especially for extreme values. Overall, these visualizations indicate how the Decision Tree Regression model relies on specific features like "Cost of the Product," "Prior Purchases," and "Weight in grams" to make predictions, with each split aiming to maximize reduction in prediction error. The structure of the tree and the predictive performance plot together reveal both the logic of the model's decisions and the areas where predictive accuracy may be limited. The model's ability to segment the data into smaller groups based on the most impactful variables demonstrates its interpretability, though the spread in the predictive performance plot suggests that it may not fully capture all complexities in the data (Osojnik et al., 2016; Nissa et al., 2024; Zhang and Gionis, 2023).

Figure 10. Split in Tree.



Splits in Tree

	Obs. in Split	Split Point	Improvement
Cost_of_the_Product	8800	1.359	0.137
Prior_purchases	8288	−0.044	0.084
Weight_in_gms	3093	−0.964	0.124
Cost_of_the_Product	1368	0.651	0.201
Prior_purchases	759	0.612	0.513
Weight_in_gms	512	−0.890	0.318

Note. For each level of the tree, only the split with the highest improvement in deviance is shown.

5.3 K-Nearest Neighbours

The model showed in Figure 11 uses 10 nearest neighbors and applies the Epanechnikov kernel as a weighting function. The Epanechnikov function assigns higher weights to closer neighbors, with weights decreasing as the distance approaches the maximum. This weighting function helps prioritize nearby points, which is advantageous for local predictions. The model employs the Manhattan (or "city block") distance metric to measure proximity between data points, a common choice in KNN that sums the absolute differences across each dimension. The data is split into training, validation, and test sets, with 7040 samples allocated for training, 1760 for validation, and 2199 for testing. This split aligns with standard practices, reserving a substantial portion for training while ensuring enough data for validation and testing. The model was optimized based on the mean squared error (MSE) in the validation set, as noted in the text. The Validation MSE is 0.742, while the Test MSE is slightly higher at 0.756, indicating a small increase in error when applied to unseen data, suggesting reasonably stable performance but some room for further improvement. The "Epanechnikov Weight Function" plot below illustrates the decay in weight as the distance from a target point increases. The relative weight starts at 1.0 for points very close to the target and gradually decreases, reaching zero at the maximum distance. This characteristic of the Epanechnikov kernel helps limit the influence of more distant neighbors, making the model more sensitive to local patterns and reducing the impact of potentially irrelevant or far-off data points. Overall, this KNN Regression model is configured to prioritize closer neighbors with diminishing weights for more distant ones. The modest increase in error from the validation to the test set suggests reasonable generalization, though the performance metrics indicate that further tuning could potentially enhance accuracy. The choice of the Epanechnikov kernel with Manhattan distance reflects an approach focused on emphasizing the influence of proximity in the model's predictions, aiming for local rather than global accuracy.

Figure 11. K-Nearest Neighbors Regression with Epanechnikov kernel as a weighting function

K-Nearest Neighbors Regression

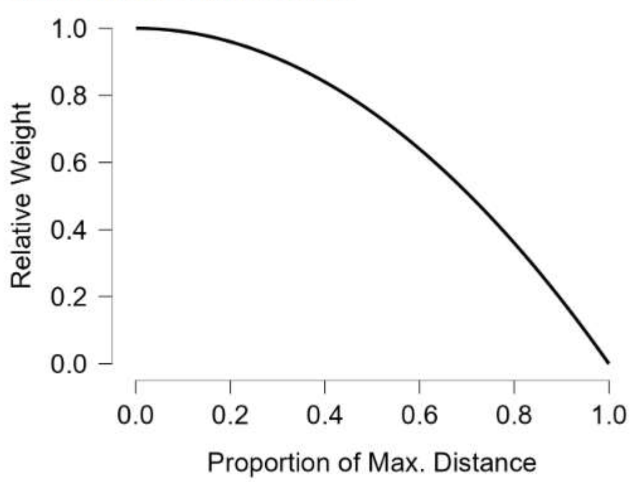
K-Nearest Neighbors Regression

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
10	epanechnikov	Manhattan	7040	1760	2199	0.742	0.756

Note. The model is optimized with respect to the *validation set mean squared error*.

Data Split

Epanechnikov Weight Function



The "Mean Squared Error Plot" on the left illustrates the mean squared error (MSE) across different values of "Number of Nearest Neighbors" for both the training set (dotted line) and the validation set (solid line). As the number of neighbors increases, the training error rises slightly, while the validation error decreases up to around 10 neighbors, which is marked with a red dot. This trade-off between training and validation error reflects the typical behavior in KNN models, where increasing the number of neighbors smoothens predictions, reducing variance but increasing bias. The "Predictive Performance Plot" on the right compares predicted test values with observed test values. The red diagonal line represents an ideal scenario where predictions perfectly match actual values. However, the dispersion of points around this line indicates variability and notable deviations, showing that the model's predictions are not highly accurate, particularly for extreme values. The clustering of points away from the line suggests that the model struggles with capturing certain patterns or outliers in the data. The evaluation metrics below further quantify the model's performance. The mean squared error (MSE) is 0.756, and the root mean squared error (RMSE) is 0.869, both measuring prediction accuracy, with RMSE providing an interpretable scale. The mean absolute error (MAE or MAD) is 0.694, indicating the average absolute difference between predictions and actual values. However, the mean absolute percentage error (MAPE) is quite high at 302.61%, suggesting that the model has substantial relative errors, particularly for smaller values in the target range. The R-squared (R^2) value of 0.253 indicates that the model explains only about 25.3% of the variance in the target variable, reflecting a limited ability to capture the underlying structure of the data. In summary, while the KNN model demonstrates a trend of decreasing validation error with more neighbors, the predictive performance plot and evaluation metrics suggest that it has limited accuracy and generalizability. The high MAPE and low R^2 values indicate that the model may not be fully capturing complex relationships in the data, leading to inconsistent predictions, especially for values that deviate from the mean (Tigga et al., 2023; Nedel'Ko Victor, 2023).

5.4 Neural Networks

The model in Figure 12 consists of 3 hidden layers with a total of 17 nodes, designed to capture complex, non-linear relationships in the data. The training, validation, and test sets contain 7040, 1760, and 2199 samples, respectively. The data split allocation follows a typical setup, with a large portion reserved for training and smaller portions for validation and testing, to allow robust evaluation of the model's generalization ability. The model is optimized based on the mean squared error (MSE) on the validation set. The validation MSE is 0.780, while the test MSE is 0.762, indicating that the model performs similarly across these sets, with a slight improvement in error on the test data. This suggests a relatively stable model with minimal overfitting. The evaluation metrics show an MSE of 0.762 and a root mean squared error (RMSE) of 0.873, which measure the magnitude of prediction errors, with RMSE providing an interpretable scale. The mean absolute error (MAE or MAD) is 0.689, representing the average absolute difference between predictions and actual values. The mean absolute percentage error (MAPE) is 133.97%, indicating that the model's predictions vary significantly relative to the true values, especially for smaller target values. The R-squared (R^2) value is 0.334, which means that the model explains about 33.4% of the variance in the target variable. This R^2 value suggests limited explanatory power, indicating that the model does not capture all underlying patterns in the data. In summary, the Neural Network Regression model exhibits moderate prediction accuracy with stable performance across validation and test sets. However, the high MAPE and relatively low R^2 indicate that the model may still have limitations in capturing complex relationships or outliers in the data. This performance profile suggests that while the neural network can handle non-linearity to some extent, further tuning or additional data might be needed to improve accuracy and explanatory power (Rahman and Asadujjaman, 2021; Caballero et al., 2020).

Figure 12. Neural Network Regression.

Neural Network Regression

Hidden Layers	Nodes	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
3	17	7040	1760	2199	0.780	0.762

Note. The model is optimized with respect to the *validation set mean squared error*.

Data Split ▼

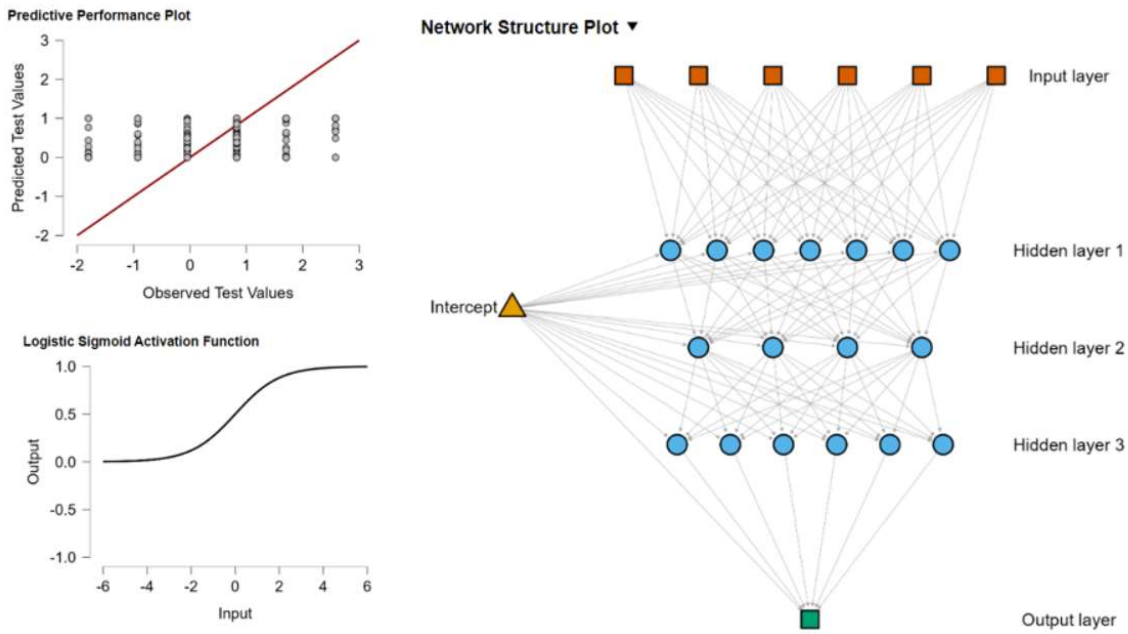
Train: 7040	Validation: 1760	Test: 2199	Total: 10999

Evaluation Metrics

	Value
MSE	0.762
RMSE	0.873
MAE / MAD	0.689
MAPE	133.97%
R^2	0.334

In Figure 13, on the right, the "Network Structure Plot" shows the neural network architecture. It consists of an input layer with multiple nodes corresponding to input features, three hidden layers with several interconnected nodes, and a single output layer node. Each hidden layer node is densely connected to nodes in the next layer, illustrating the fully connected nature of this neural network. An intercept node is also included, which feeds into each hidden layer to adjust biases and improve the network's ability to fit complex patterns in the data. The plot at the bottom left shows the "Logistic Sigmoid Activation Function," which is used within the network's nodes. This function maps input values into a range between 0 and 1, helping the model introduce non-linearity, which is crucial for capturing complex relationships. The sigmoid function's characteristic S-shape allows the model to smoothly transition between different activation levels, but it can also lead to vanishing gradients for extreme input values, potentially affecting training efficiency. The "Predictive Performance Plot" in the top left displays predicted test values against observed test values. Ideally, all points would lie along the red diagonal line, representing perfect prediction accuracy. However, the scattering of points around this line indicates some deviations between predictions and actual values, especially for extreme observed values, suggesting that the model struggles with certain data points. This spread highlights areas where the model may have difficulty capturing all underlying patterns in the data, despite its complex architecture. In summary, this neural network model, with three hidden layers and a sigmoid activation function, is designed to capture non-linear relationships. The network structure plot emphasizes its depth and connectivity, supporting its capacity for modelling complexity. However, the predictive performance plot indicates variability in prediction accuracy, pointing to possible limitations in fully capturing the range of data patterns or handling outliers effectively. The model's use of the sigmoid function aids in non-linearity but may also contribute to challenges with gradient-based optimization, especially in deeper layers (Pratiwi et al., 2020; Langer, 2021; Mulindwa and Du, 2023).

Figure 13. Neural Network Resutls.



5.5 Random Forest

The dataset is divided into training, validation, and test sets, with 7040 samples in training, 1760 in validation, and 2199 in testing. The model is optimized based on the out-of-bag (OOB) mean squared error, which allows it to estimate generalization error without requiring a separate validation set. The OOB error is reported as 0.658, close to the test mean squared error (MSE) of 0.671, indicating good consistency in performance. The evaluation metrics further detail the model's accuracy. The test MSE of 0.671 and root mean squared error (RMSE) of 0.819 reflect the average squared and root-squared deviations of predictions from actual values. The mean absolute error (MAE) of 0.661 shows the average absolute difference between predicted and actual values, while the mean absolute percentage error (MAPE) of 238.31% suggests substantial relative prediction errors, likely due to extreme values in the data. The R-squared (R^2) value of 0.312 indicates that the model explains approximately 31.2% of the variance in the target variable, suggesting limited explanatory power and room for improvement in capturing the full range of patterns within the data. The feature importance section provides insights into which variables most strongly influence the model’s predictions. "Cost of the Product" has the highest importance based on node purity (906.706), followed by "Weight in grams" (826.336) and "Prior Purchases" (495.563). "Discount Offered" also has a notable contribution (362.878). These variables collectively hold substantial predictive power, while "Customer Rating" and "Reached on Time (Y/N)" contribute minimally, indicating they are less influential in the model’s structure. The mean decrease in accuracy for each feature aligns with these rankings, confirming the importance of cost, weight, and prior purchases as primary drivers in prediction accuracy. In summary, the Random Forest Regression model demonstrates stable performance across OOB, validation, and test sets but has limitations in fully explaining variance, as shown by the relatively low R^2 and high MAPE. The feature importance metrics highlight which attributes the model relies on most heavily, with product cost, weight, and purchase history being the key predictors, while customer rating and timeliness have minimal impact. This suggests that further refinement or additional feature engineering could be beneficial for improving model accuracy and interpretability (Kaliappan et al., 2021; Agarwal et al., 2023; Han and Kim, 2021) (Figure 14).

Figure 14. Random Forest Regression.

Random Forest Regression

Random Forest Regression

Trees	Features per split	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE	OOB Error
100	2	7040	1760	2199	0.653	0.671	0.658

Note. The model is optimized with respect to the *out-of-bag mean squared error*.

Data Split

Evaluation Metrics

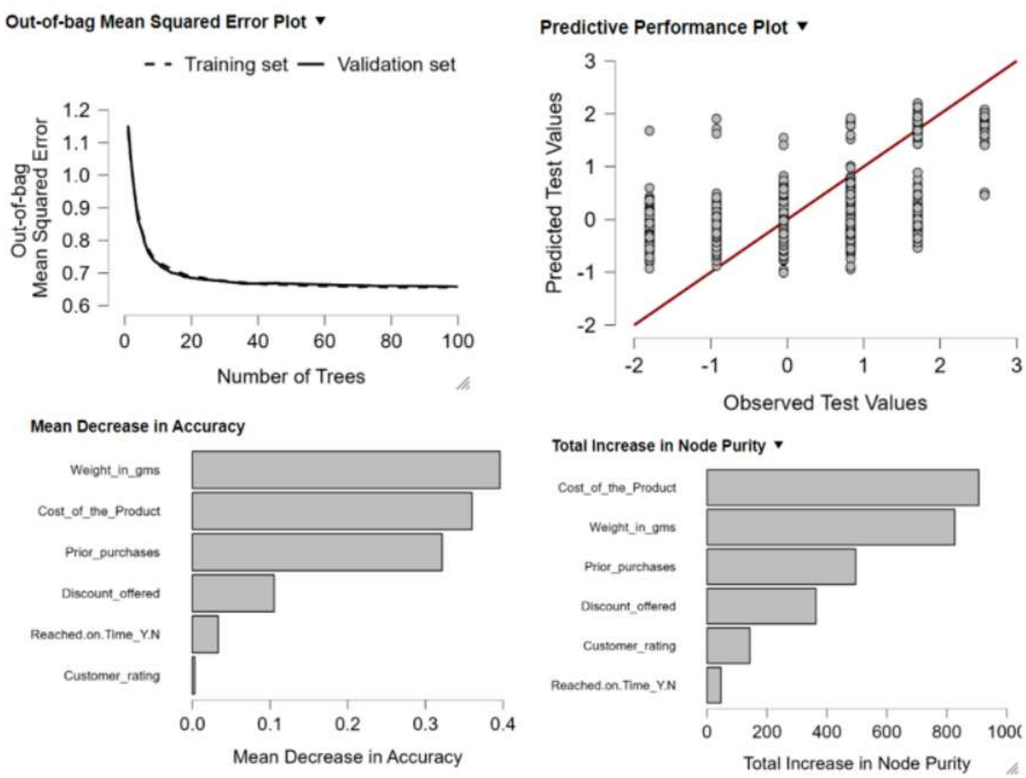
	Value
MSE	0.671
RMSE	0.819
MAE / MAD	0.661
MAPE	238.31%
R^2	0.312

Feature Importance

	Mean decrease in accuracy	Total increase in node purity
Cost_of_the_Product	0.360	906.706
Weight_in_gms	0.396	826.336
Prior_purchases	0.321	495.563
Discount_offered	0.105	362.878
Customer_rating	0.003	142.506
Reached.on.Time_Y.N	0.033	45.978

This set of visualizations in Figure 15 provides detailed insights into the performance and feature importance of a Random Forest Regression model. In the upper left, the "Out-of-bag Mean Squared Error Plot" shows the out-of-bag (OOB) mean squared error as the number of trees in the model increases. The plot indicates a steep decline in error with the addition of the first 20-30 trees, after which the error stabilizes around 0.65. This trend suggests that the model reaches optimal performance with around 30 trees, and adding more trees provides diminishing returns in terms of reducing the OOB error. The "Predictive Performance Plot" in the upper right compares the predicted test values to the observed test values, with the red diagonal line representing an ideal fit where predictions would match the actual values perfectly. However, the spread of points around this line, especially for extreme values, indicates variability and errors in predictions. This dispersion suggests that while the model is capturing some patterns in the data, it still struggles with precise predictions, particularly for certain ranges of values. The two bottom plots display the feature importance metrics. The "Mean Decrease in Accuracy" plot on the bottom left shows how each feature contributes to model accuracy. "Weight in grams" has the highest importance in terms of mean decrease in accuracy, followed closely by "Cost of the Product" and "Prior Purchases." "Discount Offered" has a moderate impact, while "Reached on Time (Y/N)" and "Customer Rating" contribute minimally. This plot highlights which features the model relies on most for making accurate predictions. The "Total Increase in Node Purity" plot on the bottom right further confirms feature importance based on how much each feature increases node purity. "Cost of the Product" and "Weight in grams" are again the most influential features, followed by "Prior Purchases" and "Discount Offered." "Customer Rating" and "Reached on Time (Y/N)" have minimal impact on node purity, underscoring their lower predictive relevance in this model. In summary, these plots reveal that while the Random Forest model stabilizes in performance after around 30 trees, there are still limitations in prediction accuracy, especially for extreme values. The feature importance metrics consistently show that "Cost of the Product," "Weight in grams," and "Prior Purchases" are the primary drivers in the model's predictions, while "Customer Rating" and "Reached on Time (Y/N)" play a negligible role. This analysis suggests that focusing on the most impactful features could further refine the model's predictive power (Agarwal et al., 2023).

Figure 15 provides detailed insights into the performance and feature importance of a Random Forest Regression model.



5.6 Regularized Linear

Figure 16 shows Regularized Linear Regression model uses Lasso (L1) regularization. The data is split into 7040 training samples, 1760 validation samples, and 2199 test samples. The model is optimized based on the mean squared error (MSE) on the validation set, with the validation MSE at 0.787 and the test MSE at 0.780, indicating consistent performance across different data subsets. The evaluation metrics provide further insight into the model's accuracy. The MSE is 0.78, and the root mean squared error (RMSE) is 0.883, both of which measure the average deviation of predictions from actual values, with RMSE on the same scale as the original data. The mean absolute error (MAE) is 0.718, representing the average absolute error between predictions and actual values. The mean absolute percentage error (MAPE) is 256.93%, suggesting that the model's relative prediction error is high, likely due to variation in the target values. The R-squared (R^2) value is 0.2, which means that the model explains only 20% of the variance in the target variable. This low R^2 indicates that the model has limited predictive power and may not fully capture the underlying relationships in the data. The regression coefficients provide insight into the contribution of each predictor variable. "Cost of the Product" has a positive coefficient (0.246), meaning that higher product costs are associated with higher predicted values. "Prior Purchases" also contributes positively with a coefficient of 0.072, suggesting that a higher purchase history is linked to an increase in the target. In contrast, "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)" have negative coefficients (-0.174, -0.312, and -0.053, respectively), indicating that increases in these variables lead to a decrease in the target prediction. "Customer Rating" has a coefficient of 0.000, meaning it has no impact on the prediction due to the regularization effect of Lasso, which shrinks insignificant coefficients to zero. In summary, this Regularized Linear Regression model shows stable but limited performance, as evidenced by its low R^2 and high MAPE. The Lasso regularization has effectively reduced the influence of less significant variables, with "Customer Rating" having no effect on the model's predictions. The most influential predictors are "Cost of the Product," "Weight in grams," and "Discount Offered," with the latter two having a negative impact on the target variable. This suggests that while Lasso regularization has simplified the model by eliminating irrelevant features, the model may still need further refinement or additional features to improve its predictive accuracy and explanatory power (Guo et al., 2021; Omer, 2022; Takada et al., 2020).

Figure 16. Regularized Linear Regression.

Regularized Linear Regression

Regularized Linear Regression

Penalty	λ	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
L1 (Lasso)	0.004	7040	1760	2199	0.787	0.780

Note. The model is optimized with respect to the *validation set mean squared error*.

Data Split

Evaluation Metrics

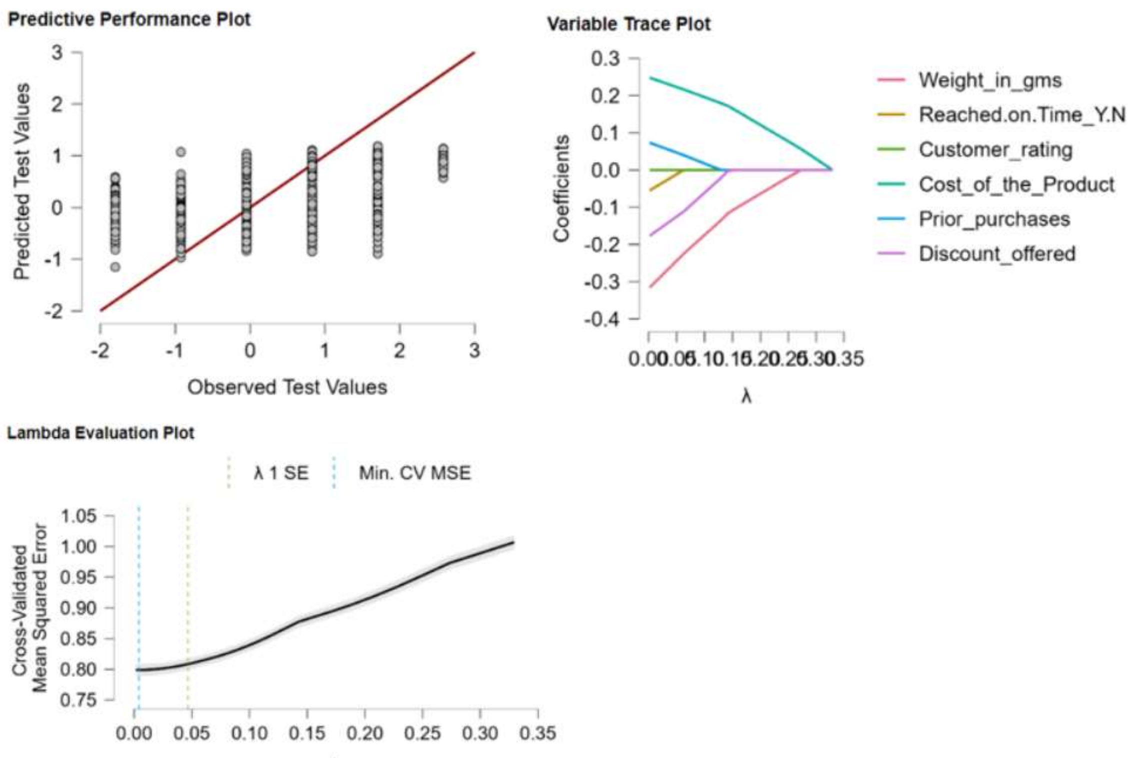
	Value
MSE	0.78
RMSE	0.883
MAE / MAD	0.718
MAPE	256.93%
R^2	0.2

Regression Coefficients

	Coefficient (β)
(Intercept)	−0.007
Customer_rating	0.000
Cost_of_the_Product	0.246
Prior_purchases	0.072
Discount_offered	−0.174
Weight_in_gms	−0.312
Reached.on.Time_Y.N	−0.053

The Figure in 17 shows "Predictive Performance Plot" on the top left shows predicted test values plotted against observed test values, with a red line representing perfect prediction accuracy. The dispersion of points around this line indicates variability in the model's predictions, suggesting that while it captures general trends, there are errors, especially for extreme values. This spread suggests limitations in the model's ability to fully capture the underlying data patterns. The "Variable Trace Plot" on the top right displays how each variable's coefficient changes as the regularization parameter λ increases. Lasso regularization gradually shrinks coefficients towards zero, with less influential variables dropping off as λ increases. For example, "Customer Rating" and "Reached on Time (Y/N)" quickly shrink towards zero as λ grows, indicating minimal influence on the model. In contrast, "Cost of the Product" and "Weight in grams" retain significant coefficient values even at higher levels of regularization, highlighting their importance in the model. This plot illustrates Lasso's ability to simplify the model by eliminating or minimizing the impact of less important features. The "Lambda Evaluation Plot" at the bottom shows the cross-validated mean squared error (MSE) across a range of λ values. The MSE initially decreases as λ increases, reaching a minimum around 0.04, where the model achieves optimal balance between bias and variance. Beyond this point, as λ increases further, the MSE begins to rise, indicating that the model is becoming too biased as more coefficients are reduced, ultimately leading to underfitting. The vertical lines indicate key values for λ: the minimum cross-validated MSE (Min. CV MSE) and one standard error above it (λ 1 SE), often used to select a simpler model with minimal loss of accuracy. In summary, these plots illustrate that while the Lasso regularization effectively reduces the influence of less important features, the model's prediction accuracy is moderate, with some variability in performance. The Variable Trace Plot confirms that "Cost of the Product" and "Weight in grams" are the most impactful features, while other variables have little to no influence at optimal regularization levels. The Lambda Evaluation Plot helps identify the ideal λ value for balancing model complexity and predictive accuracy, reinforcing the benefits of regularization in preventing overfitting while maintaining important predictors (Machkour et al., 2020; Kayanan and Wijekoon, 2020; Iparragirre et al., 2023).

Figure 17. Regularized Linear Regression.



5.7 Support Vector Machines

The model showed in Figure 18 was trained using 8800 support vectors, with the data split into 8800 samples for training and 2199 samples for testing, totalling 10,999 data points. The model's performance is measured with a test mean squared error (MSE) of 611,702.097, which is exceptionally high and suggests poor prediction accuracy. The evaluation metrics provide further insights into the model's limitations. The root mean squared error (RMSE) is 782.114, indicating substantial error magnitude in the same units as the target variable. The mean absolute error (MAE) is 639.141, showing that, on average, predictions deviate by a large amount from actual values. The mean absolute percentage error (MAPE) is an extremely high 478,256.69%, which implies that the model's predictions are highly inaccurate relative to the actual values, with prediction errors far exceeding the target values in proportional terms. Finally, the R-squared (R^2) value is 0.002, which means that the model explains only 0.2% of the variance in the target variable. This extremely low R^2 value indicates that the model fails to capture any meaningful patterns or relationships in the data. In summary, this SVM Regression model exhibits very poor performance, with high error metrics and an almost negligible R^2 value, indicating that it is not a suitable fit for the data. The high error values suggest that the model's predictions are far off from the actual values, likely due to inappropriate parameter tuning, a poor choice of kernel, or the inherent unsuitability of SVM for this particular dataset (Doganer et al., 2020; Xu, 2022).

Figure 18. Support Vector Regression.

Support Vector Machine Regression

Support Vector Machine Regression

Support Vectors	n(Train)	n(Test)	Test MSE
8800	8800	2199	611702.097

Data Split

Evaluation Metrics

	Value
MSE	611702.097
RMSE	782.114
MAE / MAD	639.141
MAPE	478256.69%
R^2	0.002

6. Machine Learning Classification

6.1 Decision Tree

The model showed in Figure 19 was built with 28 splits and trained on 8800 samples, with 2199 samples used for testing, resulting in a total of 10,999 data points. The model's test accuracy is 0.151, or 15.1%, which is notably low, indicating poor performance in correctly classifying the data. The confusion matrix further illustrates the model's performance by showing the distribution of predicted versus actual classifications across categories labeled 2, 3, 4, 5, and 6. Each row represents the observed class, while each column shows the predicted class. For example, in the row for observed class 3, only 21% of instances were correctly predicted as class 3, while the rest were misclassified across other classes. Similarly, the correct classification rates for other observed classes (4, 5, 6, etc.) are also low, with the model frequently misclassifying instances into incorrect categories. The dispersion across the matrix highlights the model's struggle with accurately differentiating between classes, as it frequently assigns incorrect labels. In summary, this Decision Tree Classification model has a very low classification accuracy of 15.1%, with substantial misclassification observed across all classes in the confusion matrix. This poor performance suggests that the model is not well-suited to the dataset or that the data may require additional preprocessing or feature engineering. Adjustments to the model's parameters, such as increasing the depth or experimenting with alternative classification algorithms, could potentially improve its performance (Krstinić et al., 2020; Zhao et al., 2021).

Figure 19. Decision Tree Classification.

Decision Tree Classification ▼

Decision Tree Classification ▼

Splits	n(Train)	n(Test)	Test Accuracy
28	8800	2199	0.151

Data Split

Confusion Matrix

		Predicted			
		3	4	5	6
Observed	2	0.04	0.01	0	0.01
	3	0.21	0.07	0.01	0.01
	4	0.2	0.06	0.03	0.01
	5	0.12	0.05	0.05	0.01
	6	0.01	0.02	0.01	0.06
	7	0	0	0	0.02

In the Figure 20 the Evaluation Metrics section, performance metrics are calculated separately for each class (labeled 2, 3, 4, 5, 6, and 7) across 2199 test samples. Each class has specific metrics for accuracy, precision, recall, F1 score, and more. For instance, the precision for class 3 is 0.355, indicating that 35.5% of the instances classified as class 3 are true positives, while the recall for this class is 0.714, showing that 71.4% of actual class 3 instances were correctly identified. The average F1 score across all classes is 0.337, reflecting a generally low balance between precision and recall. Additionally, the Area Under Curve (AUC) varies across classes, with class 6 achieving the highest AUC of 0.794, suggesting better discrimination for this class compared to others. The Average/Total column provides an overview across all classes, with overall accuracy at 0.793 and average precision and recall at 0.363 and 0.380, respectively. The low Matthews Correlation Coefficient (MCC) and varying precision and recall scores indicate that the model may be struggling to consistently classify instances across all classes. In the Feature Importance section, the relative importance of each feature in the model is displayed. "Prior Purchases" has the highest importance with a score of 44.926, followed closely by "Cost of the Product" at 39.020. "Weight in grams" has moderate importance at 15.860, while "Discount Offered" has minimal impact at 0.194. These scores indicate that the model heavily relies on purchase history and product cost when making classification decisions. The Splits in Tree section provides information on the most impactful splits in the decision tree. For example, the first split is based on "Cost of the Product" with a split point at 1.359, which leads to an improvement in deviance of 206.554, indicating a substantial reduction in error. Subsequent splits based on "Prior Purchases" (split point -0.044) and "Weight in grams" (split point -0.972) contribute further, though with diminishing improvements. This table highlights the model’s decision-making process, with initial splits based on high-impact features, followed by additional splits that refine the model’s classifications. In summary, this classification model has varying performance across classes, with generally low precision, recall, and F1 scores. The model’s reliance on "Prior Purchases" and "Cost of the Product" as key predictors is evident in both the feature importance rankings and the choice of initial splits. However, the low scores in key metrics like precision and recall indicate that the model may need further tuning or adjustments to handle classification more effectively across all classes (Dai et al., 2023; Riyanto et al., 2023; Obi, 2023).

Figure 20. Evaluation Metrics of Decision Tree Classification.

Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	132	637	673	506	207	44	2199
Accuracy	0.940	0.542	0.611	0.775	0.912	0.980	0.793
Precision (Positive Predictive Value)	NaN	0.355	0.293	0.524	0.526	NaN	0.363
Recall (True Positive Rate)	0.000	0.714	0.192	0.233	0.643	0.000	0.380
False Positive Rate	0.000	0.528	0.204	0.063	0.060	0.000	0.143
False Discovery Rate	NaN	0.645	0.707	0.476	0.474	NaN	0.575
F1 Score	NaN	0.475	0.232	0.323	0.578	NaN	0.337
Matthews Correlation Coefficient	NaN					NaN	NaN
Area Under Curve (AUC)	0.500	0.500	0.500	0.581	0.794	0.500	0.563
Negative Predictive Value	0.940	0.802	0.691	0.803	0.962	0.980	0.863
True Negative Rate	1.000	0.472	0.796	0.937	0.940	1.000	0.857
False Negative Rate	1.000	0.286	0.808	0.767	0.357	1.000	0.703
False Omission Rate	0.060	0.198	0.309	0.197	0.038	0.020	0.137
Threat Score	0.000	0.248	0.110	0.196	0.424	0.000	0.163
Statistical Parity	0.000	0.582	0.201	0.102	0.115	0.000	1.000

Note. All metrics are calculated for every class against all other classes.

Feature Importance

	Relative Importance
Prior_purchases	44.926
Cost_of_the_Product	39.020
Weight_in_gms	15.860
Discount_offered	0.194

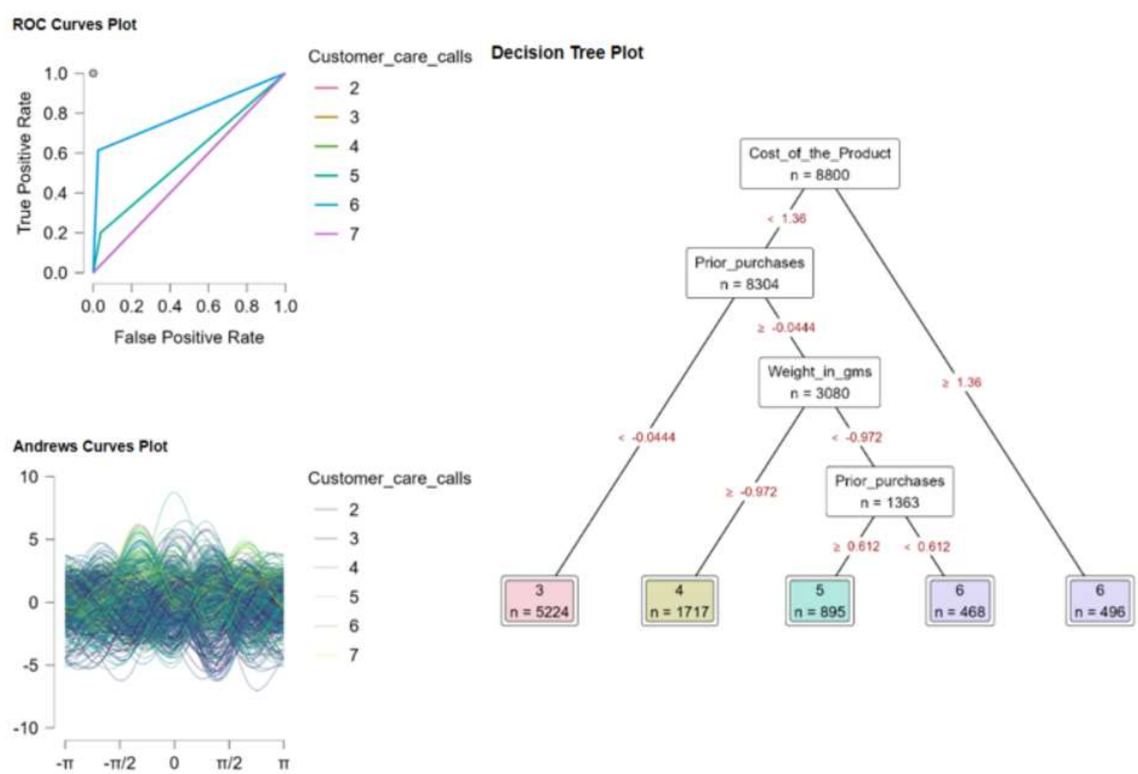
Splits in Tree

	Obs. in Split	Split Point	Improvement
Cost_of_the_Product	8800	1.359	206.554
Prior_purchases	8304	−0.044	132.151
Weight_in_gms	3080	−0.972	69.363
Prior_purchases	1363	0.612	125.502

Note. For each level of the tree, only the split with the highest improvement in deviance is shown.

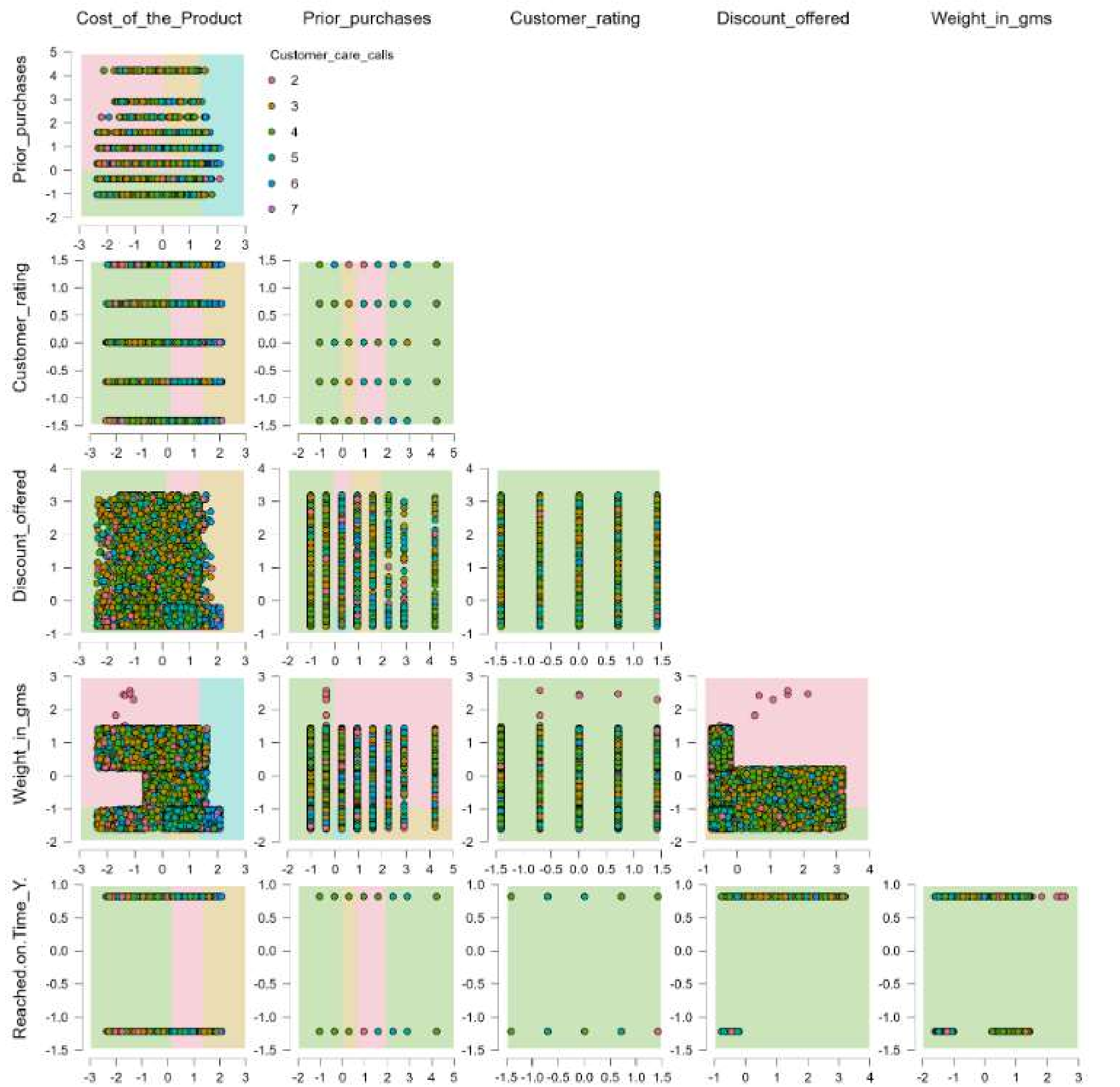
The Figure 21 shows ROC Curves Plot in the upper left shows the Receiver Operating Characteristic (ROC) curves for each class of "Customer Care Calls" labeled 2 through 7. Each curve plots the true positive rate (sensitivity) against the false positive rate, providing a measure of the model's ability to distinguish between classes. A curve closer to the top-left corner would indicate better performance. In this case, the curves for classes 2 and 3 show a somewhat stronger performance, while other classes, such as class 7, have ROC curves closer to the diagonal, indicating weaker discrimination and suggesting that the model struggles to accurately classify instances of those classes. The Decision Tree Plot in the upper right visualizes the structure of the decision tree used in the model. The tree starts with a split on "Cost of the Product" with a threshold of 1.36, dividing the dataset based on this value. Subsequent splits are made on "Prior Purchases" and "Weight in grams," with the tree further branching according to these features and specific thresholds (e.g., -0.0444 for "Prior Purchases" and -0.972 for "Weight in grams"). The nodes at the bottom of the tree represent final classifications, with each leaf node labeled by the predicted class and the number of samples it contains. This hierarchical structure shows how the model makes decisions based on key features, with "Cost of the Product" and "Prior Purchases" playing central roles in the classification process. The Andrews Curves Plot in the lower left visualizes patterns in the dataset for each class of "Customer Care Calls" using a continuous curve for each instance. Andrews curves map data to functions, enabling visualization of high-dimensional relationships. Here, each class (2 through 7) is represented by different colors. Although there is considerable overlap, some differences in the curves for each class may indicate subtle separations among classes, though the overlap suggests that classes may not be highly distinct. This visualization reinforces the idea that the classes are not easily separable, which could explain the classification challenges observed in the ROC curves. In summary, these plots provide insights into the classification model's limitations and decision-making process. The ROC curves reveal variability in performance across classes, with some classes, like 7, being particularly challenging to distinguish. The Decision Tree Plot highlights the central role of "Cost of the Product" and "Prior Purchases" in the classification process, while the Andrews Curves Plot suggests that, despite some separation, there is significant overlap between classes, making classification more difficult. These visualizations indicate that the model may benefit from further refinement, feature engineering, or exploration of alternative classification approaches to improve accuracy (Verbakel et al., 2020; Gajowniczek and Ząbkowski, 2021; Pendrill et al., 2023).

Figure 21. Decision Tree Classification Analysis.



The pair plot in Figure 22 visualizes the relationships between six variables—"Cost of the Product," "Prior Purchases," "Customer Rating," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)"—with data points color-coded by "Customer Care Calls" classes (2 through 7). Each scatterplot cell illustrates the pairwise relationship between two variables, while the diagonal cells display the distribution of individual variables. The plot of "Cost of the Product" against "Prior Purchases" shows some separation between classes, with specific clusters for classes 2 and 3 within certain ranges, though there remains significant overlap, indicating that these features alone may not be highly effective for distinguishing customer care call classes. Similarly, the relationship between "Weight in grams" and "Discount Offered" reveals dense overlap among all classes, though some specific regions show slightly higher densities for certain classes, hinting at weak class trends. The plots involving "Customer Rating" suggest limited variability, as ratings are clustered around a few discrete values, which implies that "Customer Rating" likely does not play a substantial role in predicting customer care call classes due to the lack of clear separation between them. Additionally, the binary nature of "Reached on Time (Y/N)" results in minimal spread across the scatterplots, with substantial class overlap, further indicating that this feature alone has limited predictive power in distinguishing between customer care call categories. Overall, while certain pairs, such as "Cost of the Product" and "Prior Purchases," show slight separation among some classes, the pervasive overlap across all features suggests that none of these variables alone can distinctly classify the different levels of customer care calls. This visualization implies that complex interactions between features likely drive customer care call classes, and further modeling techniques or feature engineering may be needed to improve predictive accuracy and class differentiation (Bernard et al., 2021; Ma and Maciejewski, 2020; Luque et al., 2022).

Figure 22. Relationships among variables.



6.2 k-Nearest Neighbors

The model showed in Figure 23 uses 10 nearest neighbors with a rectangular weighting function and the Euclidean distance metric. The data is split into 7040 training samples, 1760 validation samples, and 2199 test samples. The model's performance, optimized for validation set accuracy, achieved a validation accuracy of 36.8% and a test accuracy of 38.8%, indicating relatively low classification effectiveness. The confusion matrix further illustrates the model's performance by showing the distribution of predictions across classes. The observed classes (rows) are compared to the predicted classes (columns) for each category, labeled 2 through 7. The matrix indicates a high level of misclassification, with low accuracy across all classes. For example, in the row for observed class 3, only 12% of instances were correctly classified as class 3, with the remainder misclassified into other classes. Similarly, other classes (e.g., 4, 5, and 6) also show low correct classification rates, with predictions frequently scattered across multiple incorrect classes. Overall, the model's low validation and test accuracies, along with the high level of misclassification in the confusion matrix, suggest that this KNN model is not well-suited for distinguishing between the classes in this dataset. This underperformance may be due to overlapping feature distributions among classes or insufficiently distinct data points for effective classification by the KNN algorithm. To improve results, additional tuning of the number of neighbors, exploration of alternative distance metrics, or the use of different classification algorithms may be needed (Sugriyono and Siregar, 2020; Hinterreiter et al., 2020; Bajpai and He, 2020).

Figure 23. K-Nearest Neighbors Classification.

K-Nearest Neighbors Classification

K-Nearest Neighbors Classification

Nearest neighbors	Weights	Distance	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
10	rectangular	Euclidean	7040	1760	2199	0.368	0.388

Note. The model is optimized with respect to the *validation set accuracy*.

Data Split

Confusion Matrix

		Predicted					
		2	3	4	5	6	7
Observed	2	0	0.03	0.03	0	0	0
	3	0	0.12	0.15	0.01	0	0
	4	0	0.13	0.19	0.03	0	0
	5	0	0.06	0.11	0.03	0.01	0
	6	0	0.01	0.02	0.01	0.05	0
	7	0	0	0	0	0.02	0

In the Class Proportions section, each class (labeled 2 through 7) is represented by its proportion in the full dataset as well as in the training, validation, and test sets. The proportions are fairly consistent across all subsets, with class 4 being the largest at approximately 32%, followed by class 3 at around 29%, while class 7 is the smallest, making up only about 2% of the dataset. These class distributions indicate an imbalanced dataset, where certain classes have significantly fewer instances than others, potentially affecting the model's performance for minority classes. The Evaluation Metrics section provides detailed performance metrics for each class individually, as well as overall averages. Metrics include accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC), Area Under Curve (AUC), and others. For class 2, the accuracy is high at 94.2%, but both precision and recall are zero, indicating that while the model rarely misclassifies other classes as class 2, it struggles to identify true class 2 instances. In contrast, class 6 has a high accuracy of 97.9% with moderate precision (0.200) and recall (0.559), reflecting better classification but still significant room for improvement in identifying all true positives. The average F1 score across all classes is 0.362, which reflects the balance between precision and recall and highlights the model's overall struggle to maintain high performance for each class. The AUC values vary across classes, with class 7 achieving the highest AUC at 0.885, indicating relatively good discrimination between true and false instances for this class, while class 2 has a lower AUC of 0.603. The MCC, which measures the quality of classifications, is low across most classes, with an average of 0.028, showing weak correlation between predicted and actual classifications. Additional metrics such as the False Discovery Rate, False Negative Rate, and Statistical Parity provide further insight into the model's error rates and potential bias. Notably, the overall high false discovery rate and variable precision and recall values indicate that the model has difficulty consistently classifying all classes correctly, likely due to the class imbalance. In summary, this classification model shows varying performance across different classes, with generally low precision, recall, and F1 scores indicating limited effectiveness in accurately identifying each class. The imbalanced class distribution and low MCC values further highlight the model's challenges in delivering robust classification performance across all categories. Addressing the class imbalance, possibly through resampling or adjusting class weights, and refining the model could improve its overall accuracy and reliability (Leevy et al., 2022; Hancock et al., 2022; Riyanto et al., 2023), (Figure 24).

Figure 24. Evaluation Metrics of K-Nearest Neighbors Classification.

Class Proportions

	Data Set	Training Set	Validation Set	Test Set
2	0.058	0.059	0.055	0.057
3	0.292	0.294	0.310	0.274
4	0.323	0.323	0.299	0.344
5	0.212	0.209	0.223	0.212
6	0.092	0.092	0.091	0.093
7	0.022	0.023	0.023	0.020

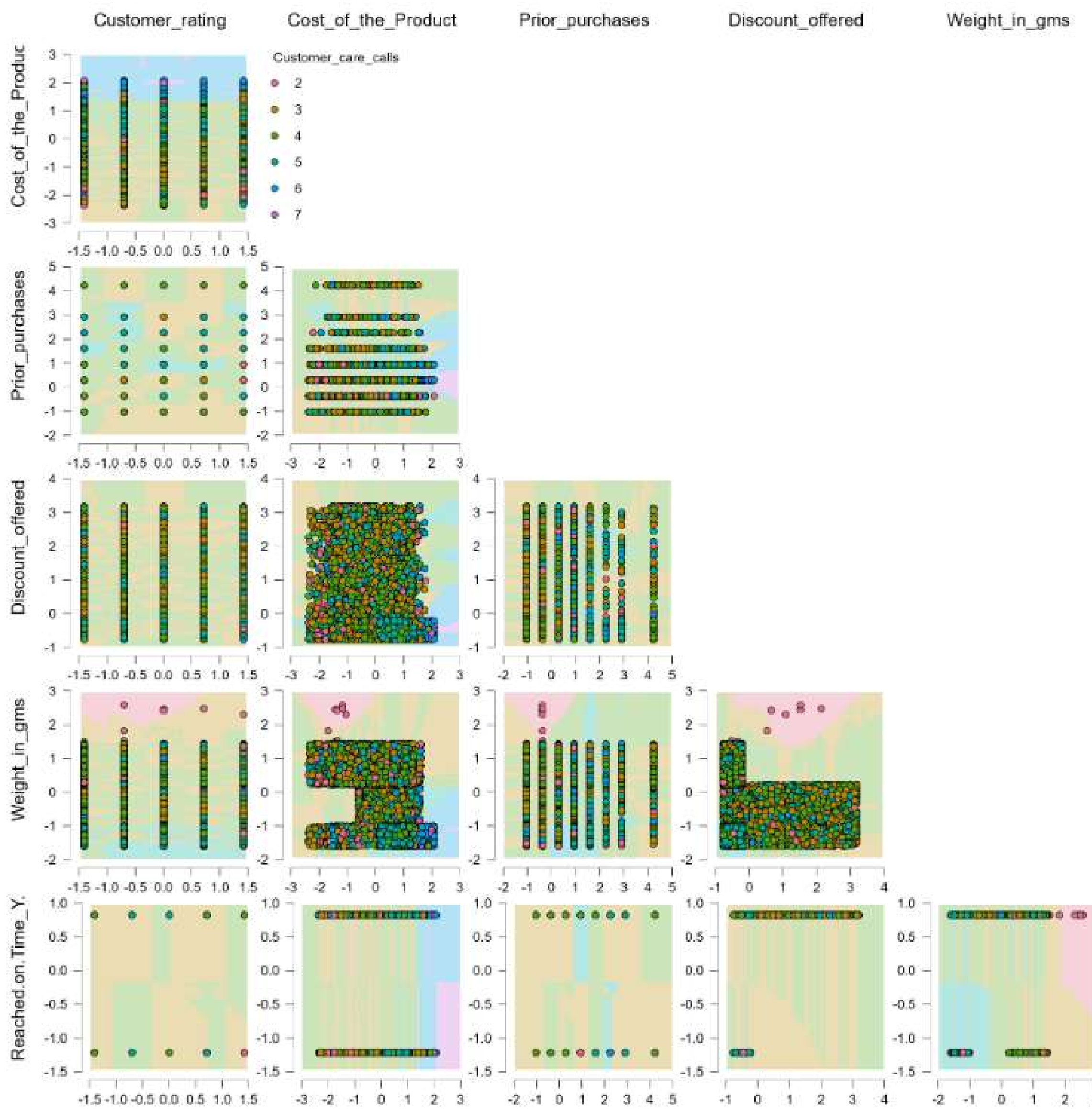
Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	126	602	757	466	204	44	2199
Accuracy	0.942	0.620	0.525	0.776	0.935	0.979	0.796
Precision (Positive Predictive Value)	0.000	0.342	0.370	0.427	0.679	0.200	0.379
Recall (True Positive Rate)	0.000	0.422	0.540	0.163	0.559	0.023	0.388
False Positive Rate	4.824×10^{-4}	0.306	0.483	0.059	0.027	0.002	0.146
False Discovery Rate	1.000	0.658	0.630	0.573	0.321	0.800	0.664
F1 Score	NaN	0.378	0.439	0.236	0.613	0.041	0.362
Matthews Correlation Coefficient	−0.005					0.061	0.028
Area Under Curve (AUC)	0.603	0.636	0.593	0.606	0.820	0.885	0.690
Negative Predictive Value	0.943	0.761	0.682	0.807	0.956	0.980	0.855
True Negative Rate	1.000	0.694	0.517	0.941	0.973	0.998	0.854
False Negative Rate	1.000	0.578	0.460	0.837	0.441	0.977	0.716
False Omission Rate	0.057	0.239	0.318	0.193	0.044	0.020	0.145
Threat Score	0.000	0.192	0.235	0.128	0.576	0.020	0.192
Statistical Parity	4.548×10^{-4}	0.337	0.503	0.081	0.076	0.002	1.000

Note. All metrics are calculated for every class against all other classes.

This pair plot in Figure 25 visualizes the relationships between six features—"Customer Rating," "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)"—with points color-coded by "Customer Care Calls" classes (2 through 7). Each scatterplot cell displays the pairwise relationship between two features, while the diagonal cells show the distribution of individual features. The background shading represents class regions to indicate where different "Customer Care Calls" classes tend to cluster within the feature space. Examining the relationships, we see that "Cost of the Product" and "Prior Purchases" show some mild separation for specific classes, though there is substantial overlap, particularly in regions with high densities of classes 3, 4, and 5, which suggests these features alone are insufficient for strong class discrimination. Similarly, the relationship between "Discount Offered" and "Weight in grams" reveals considerable overlap among all classes, with dense regions suggesting that these features also lack clear class separation, although certain class clusters, such as classes 2 and 7, are more distinct in some localized regions. The "Customer Rating" and "Reached on Time " features display limited variability, with data points clustered at a few specific values, indicating that these features are not likely to provide strong predictive power for distinguishing between customer care call classes. For instance, "Reached on Time (Y/N)" is binary, creating limited spread and class distinction in plots involving this feature. Overall, this pair plot indicates significant overlap across all classes within the feature space, suggesting that no single feature or pair of features provides a clear boundary for distinguishing customer care call classes. The color-coded scatter points show that while there are some areas with mild class clustering, the extensive overlap implies that more complex interactions between features are likely needed to achieve effective classification. This visualization suggests that additional feature engineering, dimensionality reduction, or more sophisticated modeling approaches may be required to improve class separability and predictive accuracy (Ma and Maciejewski, 2020; Bernard et al., 2021).

Figure 25. Relationships among the variables.



6.3 Linear Discriminant

The model showed in Figure 26 uses five linear discriminants with the "Moment" method for feature extraction and dimensionality reduction. The dataset is split into 8800 training samples and 2199 test samples, totaling 10,999 observations. The model achieves a test accuracy of 35.4%, indicating relatively low effectiveness in correctly classifying instances across the test set. The Confusion Matrix provides a detailed view of the model's classification performance for each class (labeled 2 through 7). The rows represent the observed classes, while the columns represent the predicted classes. Each cell shows the proportion of observations in each class that were classified as a specific class. For example, in the row for observed class 3, only 12% of instances were correctly classified as class 3, while the remaining instances were misclassified into other classes, with predictions scattered across classes 2, 4, and 5. Similarly, for observed class 4, only 18% were correctly classified, with a high proportion misclassified into other categories. This widespread misclassification across classes suggests that the LDA model struggles to differentiate between them effectively, likely due to overlapping feature distributions and insufficient class separation in the feature space. The overall test accuracy of 35.4%, combined with the high level of misclassification seen in the confusion matrix, indicates that LDA may not be a suitable choice for this dataset. The model's performance could potentially be improved by trying alternative classification algorithms that capture non-linear relationships or by performing additional feature engineering to enhance class separability (Li et al., 2023; Zorarpacı, 2021; Chandrasekar and Geetha, 2021).

Figure 26. Linear Discriminant Classification.

Linear Discriminant Classification

Linear Discriminant Classification

Linear Discriminants	Method	n(Train)	n(Test)	Test Accuracy
5	Moment	8800	2199	0.354

Data Split

Confusion Matrix

		Predicted					
		2	3	4	5	6	7
Observed	2	0	0.02	0.03	0	0	0
	3	0	0.12	0.16	0.01	0	0
	4	0	0.1	0.18	0.01	0.03	0
	5	0	0.06	0.1	0.01	0.05	0
	6	0	0.01	0.03	0	0.05	0
	7	0	0	0	0	0.02	0

In the Class Proportions section, the distribution of each class (labeled 2 through 7) is shown for the entire dataset as well as for the training and test sets. The proportions are consistent across splits, with class 4 being the largest (32.3%) and class 7 the smallest (2.2%). This class imbalance suggests that the model may struggle with accuracy for minority classes due to limited representation in the data. The Evaluation Metrics section provides detailed performance metrics for each class and an overall average. Accuracy varies widely across classes, with class 7 achieving the highest accuracy at 97.9% and class 5 having a relatively high accuracy at 85.3%. In contrast, class 4 has low accuracy (53.1%)

and recall (0.403), indicating difficulty in correctly identifying instances of this class. The overall average F1 score of 0.306 reflects the balance between precision and recall across classes and indicates that the model has limited effectiveness in achieving accurate and balanced classifications. The Area Under Curve (AUC) values also vary, with class 7 having the highest AUC of 0.953, showing better discrimination for this class, while class 3 has a lower AUC of 0.546. The Matthews Correlation Coefficient (MCC) is also generally low, averaging 0.078, further indicating weak correlations between observed and predicted classifications. In the Linear Discriminant Coefficients section, coefficients for each feature across the five linear discriminants (LD1 to LD5) are presented. These coefficients indicate the contribution of each feature to each discriminant function. For instance, "Cost of the Product" has negative coefficients in LD1, LD2, and LD5, suggesting a negative relationship with these discriminants, while "Discount Offered" shows high positive coefficients in LD1 and LD4, indicating a stronger influence on these discriminants. The varying signs and magnitudes of these coefficients reflect how each feature contributes differently to the linear discriminants, which are used to separate classes. Overall, this LDA model demonstrates inconsistent performance across classes, with low precision, recall, and F1 scores for many classes, indicating limited classification effectiveness. The class imbalance may be affecting performance, especially for minority classes like class 7. The discriminant coefficients show that certain features, such as "Discount Offered" and "Cost of the Product," have a more substantial impact on classification, while other features, like "Customer Rating," exhibit smaller coefficients and are less influential. These results suggest that the LDA model may benefit from further tuning, addressing class imbalance, or considering non-linear classification methods to improve performance (Chandrasekar and Geetha, 2021; Thölke et al., 2023; Mirza et al., 2021), (Figure 27).

Figure 27. Metrics of Linear Discriminant Classification.

Class Proportions

	Data Set	Training Set	Test Set
2	0.058	0.058	0.058
3	0.292	0.291	0.299
4	0.323	0.323	0.323
5	0.212	0.213	0.207
6	0.092	0.092	0.092
7	0.022	0.023	0.021

Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	128	657	711	455	202	46	2199
Accuracy	0.942	0.633	0.531	0.771	0.853	0.979	0.785
Precision (Positive Predictive Value)	NaN	0.389	0.354	0.184	0.323	NaN	0.299
Recall (True Positive Rate)	0.000	0.403	0.549	0.031	0.545	0.000	0.354
False Positive Rate	0.000	0.270	0.478	0.036	0.116	0.000	0.150
False Discovery Rate	NaN	0.611	0.646	0.816	0.677	NaN	0.687
F1 Score	NaN	0.396	0.430	0.053	0.405	NaN	0.306
Matthews Correlation Coefficient	NaN					NaN	NaN
Area Under Curve (AUC)	0.632	0.630	0.546	0.573	0.771	0.953	0.684
Negative Predictive Value	0.942	0.742	0.708	0.792	0.950	0.979	0.852
True Negative Rate	1.000	0.730	0.522	0.964	0.884	1.000	0.850
False Negative Rate	1.000	0.597	0.451	0.969	0.455	1.000	0.745
False Omission Rate	0.058	0.258	0.292	0.208	0.050	0.021	0.148
Threat Score	0.000	0.217	0.224	0.025	0.199	0.000	0.111
Statistical Parity	0.000	0.310	0.501	0.035	0.155	0.000	1.000

Note. All metrics are calculated for every class against all other classes.

Linear Discriminant Coefficients

	LD1	LD2	LD3	LD4	LD5
(Constant)	6.173×10^{-5}	−0.003	−0.004	0.002	−0.011
Customer_rating	−0.022	−0.055	0.042	0.665	−0.426
Cost_of_the_Product	−0.593	0.587	−0.464	−0.268	−0.430
Prior_purchases	−0.144	−0.857	−0.452	−0.221	−0.220
Discount_offered	0.460	4.496×10^{-4}	0.175	−0.274	−0.981
Weight_in_gms	0.869	0.148	−0.788	−0.144	−0.281
Reached.on.Time_Y.N	0.152	0.057	−0.064	−0.613	0.302

The Prior and Posterior Class Probabilities section lists the probabilities for each class (labeled 2 through 7) before and after considering the data. The prior and posterior probabilities are nearly identical for each class, indicating that the class distribution in the training data aligns closely with the model's classifications. Class 4 has the highest probability (32%), while class 7 has the lowest (2.3%), reflecting the class imbalance in the dataset. The Class Means in Training Data section shows the mean values of each feature for each class within the training data. This table helps to highlight any distinct patterns or differences in feature values across classes. For instance, "Cost of the Product" varies significantly between classes, with class 7 having a mean of 1.416 compared to -0.413 for class 2, indicating that this feature might be an important differentiator between classes. Similarly, "Weight in grams" also shows a noticeable difference, with a mean of -1.315 for class 7 compared to near-zero values for classes 2 and 4, suggesting potential class-specific distinctions in this feature. The Tests of Equality of Class Means section presents the results of F-tests to determine whether the means of each feature significantly differ across classes. High F-values and low p-values ($p < .001$) for most features, including "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)," indicate statistically significant differences in these features among classes. This suggests that these features contribute to distinguishing between classes. However, "Customer Rating" has a much lower F-value and a non-significant p-value ($p = 0.200$), implying that it does not differ significantly across classes and is therefore less useful for classification. The Tests of Equality of Covariance Matrices section reports Box's M test, which evaluates whether the covariance matrices are equal across classes, an assumption of LDA. The extremely high chi-square value (3522.523) and significant p-value ($p < .001$) indicate that the null hypothesis of equal covariance matrices is rejected. This suggests that the covariance structures differ significantly across classes, violating a key assumption of LDA and potentially impacting the model's performance. In summary, this analysis shows that features like "Cost of the Product," "Prior Purchases," and "Weight in grams" have significant mean differences across classes, making them valuable for classification. However, the unequal covariance matrices, as indicated by Box's M test, violate LDA assumptions, potentially limiting the model's effectiveness. The non-significance of "Customer Rating" further suggests that this feature may not provide meaningful separation between classes in this model. Addressing the class imbalance and potentially exploring alternative models that do not rely on equal covariance assumptions could improve classification performance (Mirza et al., 2021; Brzezinski et al., 2019; Thölke et al., 2023), (Figure 28).

Figure 28. Tests and Probabilities.

Prior and Posterior Class Probabilities

	Prior	Posterior
2	0.058	0.058
3	0.291	0.292
4	0.323	0.320
5	0.213	0.205
6	0.092	0.099
7	0.023	0.026

Class Means in Training Data

	Customer_rating	Cost_of_the_Product	Prior_purchases	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
2	0.023	−0.413	−0.079	0.222	−0.002	0.091
3	−0.020	−0.234	−0.200	0.125	0.245	0.064
4	0.011	−0.078	−0.054	−0.002	0.148	−0.017
5	−0.010	0.079	0.181	−0.043	−0.091	−0.019
6	0.069	0.774	0.426	−0.273	−0.740	−0.171
7	0.010	1.416	0.366	−0.407	−1.315	−0.177

Tests of Equality of Class Means

	F	df1	df2	p
Customer_rating	1.639	1	10997	0.200
Cost_of_the_Product	1282.557	1	10997	< .001
Prior_purchases	371.502	1	10997	< .001
Discount_offered	191.270	1	10997	< .001
Weight_in_gms	911.165	1	10997	< .001
Reached.on.Time_Y.N	49.775	1	10997	< .001

Note. The null hypothesis specifies equal class means.

Tests of Equality of Covariance Matrices

	χ^2	df	p
Box's M	3552.523	105	< .001

Note. The null hypothesis specifies equal covariance matrices.

The Pooled Within-Class Matrices Correlations section shows the correlations between features after pooling across classes, offering insights into the relationships between variables within the context of each class. "Cost of the Product" has a strong negative correlation with "Weight in grams" (-0.793) and a moderate negative correlation with "Discount Offered" (-0.322), suggesting that higher product costs tend to be associated with lighter products and lower discounts. "Discount Offered" and "Weight in grams" also have a substantial negative correlation (-0.672), indicating that higher discounts tend to be given to lighter products. Meanwhile, "Customer Rating" and "Prior Purchases" have relatively low correlations with other variables, suggesting limited relationships between these features and the rest of the dataset. The ROC Curves Plot in the bottom left evaluates the model's ability to discriminate between classes of "Customer Care Calls" (labeled 2 through 7). Each curve represents a class, plotting the true positive rate against the false positive rate. The ROC curve for class 7 stands out as it is closer to the top-left corner, indicating better discriminability for this class, while other classes, such as 4 and 5, have ROC curves closer to the diagonal, suggesting lower discrimination and more classification overlap. This indicates that the model performs better for some classes, like class 7, than others. The Andrews Curves Plot in the bottom right visualizes patterns in each "Customer Care Calls" class by mapping instances to continuous functions, with each class represented by a different color. The curves overlap considerably, indicating limited separation between classes in the feature space. This overlap suggests that the classes are not easily distinguishable based on the given features, making it challenging for the model to achieve high accuracy. In summary, the correlation matrix reveals notable relationships, such as the negative correlation between "Cost of the Product" and "Weight in grams," which may influence classification. The ROC curves highlight variability in classification performance, with class 7 showing better discrimination than others. The significant overlap in the Andrews curves plot underscores the difficulty in separating classes based on the current feature set, indicating that additional feature engineering or alternative modeling techniques may be necessary to improve classification accuracy and separability among classes (Ruisánchez et al., 2021; Aceved et al., 2022; Feng and Tian, 2021), (Figure 29).

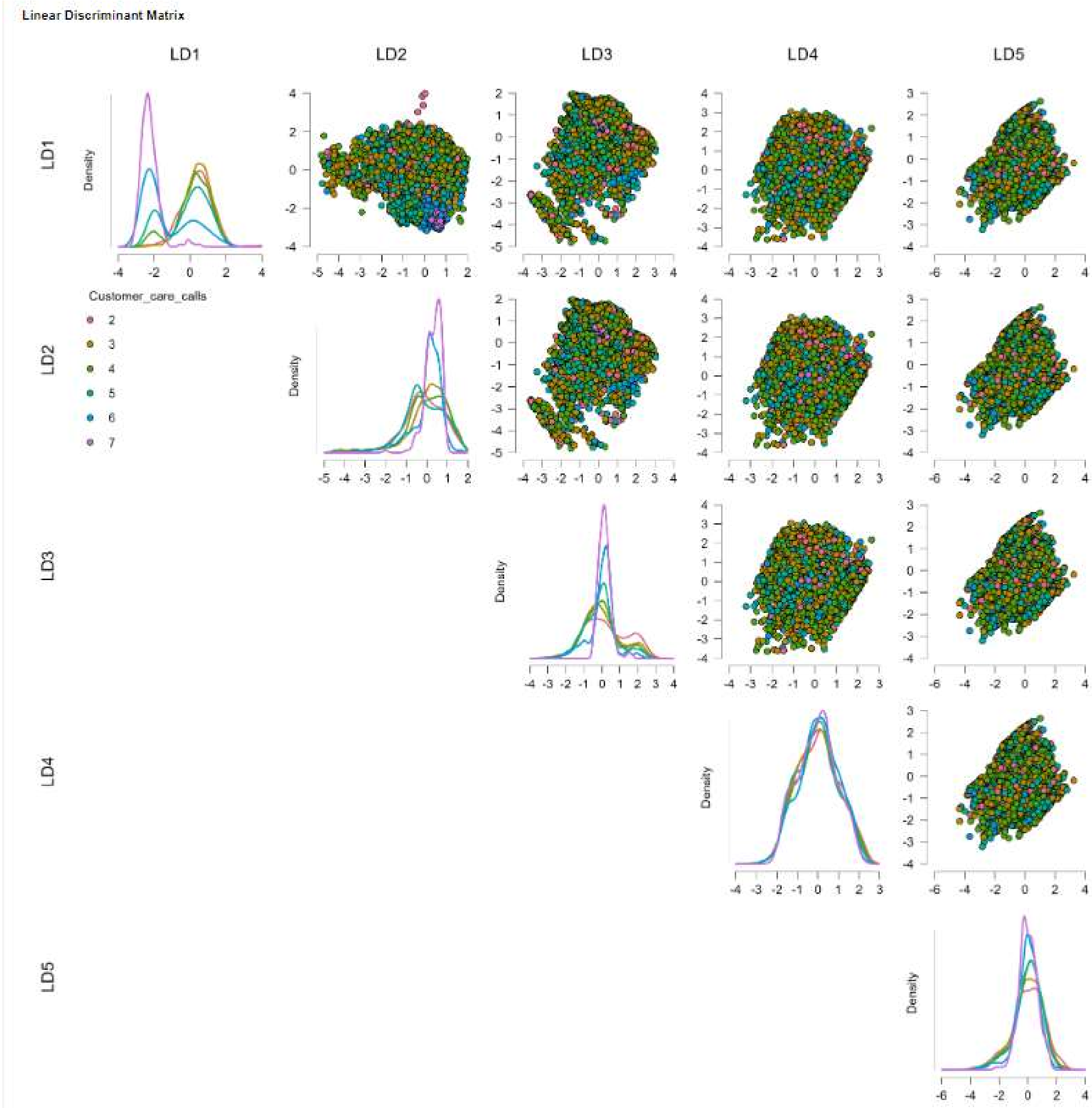
Figure 29. Correlations, ROC and Andrews curve.

Pooled Within-Class Matrices Correlations

	Customer_rating	Cost_of_the_Product	Prior_purchases	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
Customer_rating	1.000					
Cost_of_the_Product	−0.170	1.000				
Prior_purchases	−0.148	0.023	1.000			
Discount_offered	−0.138	−0.322	−0.202	1.000		
Weight_in_gms	−0.009	0.006	−0.186	−0.793	1.000	
Reached.on.Time_Y.N	−0.159	−0.305	−0.232	0.676	−0.672	1.000

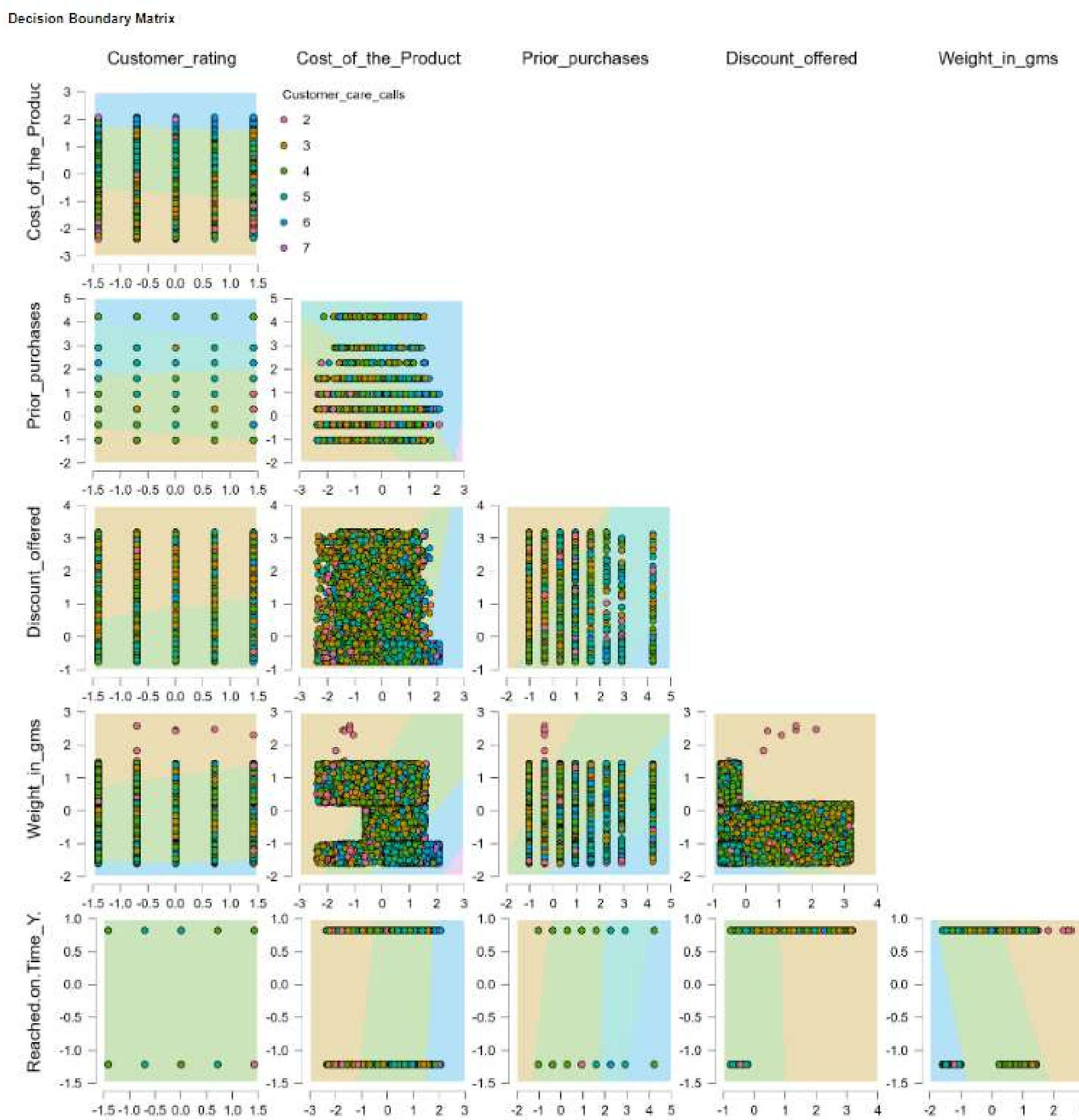
In Figure 30 Each point represents an instance, color-coded by "Customer Care Calls" classes (2 through 7), while the density plots along the diagonal illustrate the distribution of each class within individual discriminants. In the Density Plots on the diagonal, each linear discriminant (LD1, LD2, etc.) has its own density distribution for each class. LD1 shows some separation, with distinct peaks for different classes, particularly for classes 2 and 7, which appear to have their own density peaks. This suggests that LD1 may capture meaningful class separation. However, for other discriminants like LD3, LD4, and LD5, the distributions of different classes overlap substantially, indicating limited class separation within these discriminants. The Scatter Plots for each pair of discriminants (off-diagonal) reveal how classes are spread across the feature space defined by these discriminants. Most scatter plots, such as those between LD2 and LD3 or LD4 and LD5, show dense overlap among classes, with points from all classes densely clustered together. This overlap indicates that the linear discriminants do not provide clear separation between classes when combined, making it challenging for the LDA model to effectively distinguish among the customer care call categories based on these features. Overall, this plot suggests that while LD1 might contain some discriminatory power, the remaining discriminants (LD2 to LD5) do not contribute significantly to class separation. The substantial overlap in the scatter plots and density distributions indicates that the LDA model struggles to separate classes based on the current feature set. This result implies that additional feature engineering, alternative dimensionality reduction methods, or a non-linear classification approach may be needed to improve class separability and predictive performance (Ghiasi-Shirazi, 2022; Liu et al., 2023; Yan et al., 2020).

Figure 30. Relationship among variables.



This Decision Boundary Matrix shows pairwise decision boundaries between six features—"Customer Rating," "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)"—for a classification model that aims to predict "Customer Care Calls" classes (2 through 7). Each scatter plot cell represents a pair of features, with the background color representing the decision boundaries of the model, while each data point is color-coded according to its class. From the pairwise boundaries, we can observe that in most feature pairs, the decision boundaries are not well-separated, and there is considerable overlap in data points from different classes within each region. For instance, in the plot of "Cost of the Product" against "Prior Purchases," while there are some decision boundary regions colored to indicate different classes, the data points from classes 3, 4, and 5 are densely clustered together with significant overlap, suggesting that these features do not effectively separate the classes. Similarly, in the plot of "Discount Offered" versus "Weight in grams," data points from all classes cluster densely within a few regions, making it difficult for the model to distinguish between them based solely on these two features. The feature "Reached on Time (Y/N)" also shows minimal variability, as it is binary, leading to simple vertical or horizontal decision boundaries that divide classes into regions without substantial separation. In plots involving this feature, classes remain largely overlapping, indicating that it provides limited discriminatory power. Overall, this Decision Boundary Matrix highlights that no single pair of features provides clear separability between classes. The decision boundaries show that the model has defined regions for each class; however, the extensive overlap of data points across these boundaries suggests that the classes are not easily distinguishable with the current feature set. This extensive overlap implies that additional feature engineering, the introduction of non-linear decision boundaries, or the use of more complex classification algorithms may be necessary to improve class separability and the predictive accuracy of the model (Del Moral et al., 2022; Bemporad, 2022; Firdausanti et al., 2022), (Figure 31).

Figure 31. Decision Boundary Matrix.



6.4 Neural Network

The Neural Network Classification section outlines that the model has 5 hidden layers with 5 nodes each, trained on 8800 samples and tested on 2199 samples. The model's accuracy on the test set is only 5.8%, indicating very poor classification performance. This low accuracy suggests that the model struggles significantly to distinguish between classes in this dataset. The Confusion Matrix provides further insight into the model's classification errors. The matrix shows that, regardless of the true class, the model consistently predicts class 4 for almost all samples. For example, 0.06 (or 6%) of class 2 instances and 0.3 (or 30%) of class 3 instances were predicted as class 4. This pattern is repeated across all observed classes, indicating that the model has failed to learn meaningful patterns and instead defaults to a single class prediction, which is a common symptom of a model failing to generalize, possibly due to inadequate training, insufficient data, or issues with model complexity. The Class Proportions section shows the distribution of each class in the dataset, training set, and test set. The proportions are consistent across splits, with class 4 being the most prevalent (32.3%) and class 7 the least prevalent (2.2%). The class imbalance, particularly the dominance of class 4, might be contributing to the model's tendency to predict class 4 for most samples, as it may be biased toward the majority class. In summary, this Neural Network Classification model shows extremely poor performance, with a test accuracy of only 5.8% and a tendency to predict class 4 regardless of the input. This failure may be due to the model's structure, an imbalance in class representation, or inadequate training. To improve performance, it would be necessary to address class imbalance, adjust the model's architecture, or explore additional training techniques (Bemporad, 2022; Firdausanti et al., 2022; Del Moral et al., 2022), (Figure 32).

Figure 32. Neural Network Classification.

Neural Network Classification

Neural Network Classification

Hidden Layers	Nodes	n(Train)	n(Test)	Test Accuracy
5	5	8800	2199	0.058

Note. The model is optimized with respect to the *sum of squares*.

Data Split

Confusion Matrix

		Predicted
		4
Observed	2	0.06
	3	0.3
	4	0.33
	5	0.2
	6	0.09
	7	0.02

Class Proportions

	Data Set	Training Set	Test Set
2	0.058	0.058	0.058
3	0.292	0.290	0.302
4	0.323	0.322	0.330
5	0.212	0.215	0.198
6	0.092	0.093	0.088
7	0.022	0.022	0.024

The Evaluation Metrics section summarizes various performance metrics for each class (2 through 7) as well as overall averages across the test set, which includes 2199 samples. The metrics show considerable variation across classes, with class 2 achieving high accuracy (94.2%) and class 7 also performing well (97.6%). However, other classes, such as class 4, have very low accuracy (33.0%), and many metrics are labeled as "NaN," indicating they may not have been computed due to limited or absent predictions for those classes. Overall, the average precision, recall, and F1 scores are low, suggesting that the model struggles to balance performance across all classes. This is further highlighted by the Matthews Correlation Coefficient (MCC), which is "NaN" for most classes, indicating weak correlations between predicted and actual classifications, likely due to the model's over-reliance on certain classes and underperformance on others. The Network Weights section lists the weights between nodes in the neural network. The model has an intercept node and six input features: "Customer Rating," "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)," which connect to the nodes in the hidden layers. The weights vary significantly, with some being highly positive (e.g., 3.277 for "Intercept" to Hidden Node 2) and others highly negative (e.g., -3.268 from "Prior Purchases" to Hidden Node 1). These weights indicate the relative influence of each feature on different hidden nodes. For instance, "Discount Offered" has a positive weight of 0.701 to Hidden Node 1, suggesting it contributes positively to this node's activation, while "Prior Purchases" has a large negative weight to the same node (-3.268), indicating a strong negative influence. The weights from the hidden nodes to the output layer also vary, with some being highly positive (e.g., 3.008 from Hidden Node 1 to Output) and others highly negative (e.g., -1.199 from Hidden Node 7 to Output). These weights determine the final classification decision based on the activations from the hidden layers. Overall, this model exhibits poor and inconsistent performance across classes, as indicated by the evaluation metrics. The network weights show the structure and strength of connections within the model, but the varied and sometimes extreme weights suggest that the model may be overfitting certain features or struggling with class balance. Improvements could involve adjusting the network architecture, rebalancing the data, or tuning hyperparameters to enhance the model's generalizability across all classes (Barulina, et al., 2023; Sun et al., 2023; GONZALEZ-RAMIREZ et al., 2021).

Figure 33. Evaluation Metrics.

Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	127	665	725	435	194	53	2199
Accuracy	0.942	0.698	0.330	0.802	0.912	0.976	0.777
Precision (Positive Predictive Value)	NaN	NaN	0.330	NaN	NaN	NaN	0.109
Recall (True Positive Rate)	0.000	0.000	1.000	0.000	0.000	0.000	0.330
False Positive Rate	0.000	0.000	1.000	0.000	0.000	0.000	0.167
False Discovery Rate	NaN	NaN	0.670	NaN	NaN	NaN	0.670
F1 Score	NaN	NaN	0.496	NaN	NaN	NaN	0.163
Matthews Correlation Coefficient	NaN	NaN		NaN	NaN	NaN	NaN
Area Under Curve (AUC)	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Negative Predictive Value	0.942	0.698	NaN	0.802	0.912	0.976	0.866
True Negative Rate	1.000	1.000	0.000	1.000	1.000	1.000	0.833
False Negative Rate	1.000	1.000	0.000	1.000	1.000	1.000	0.833
False Omission Rate	0.058	0.302	NaN	0.198	0.088	0.024	0.134
Threat Score	0.000	0.000	0.246	0.000	0.000	0.000	0.041
Statistical Parity	0.000	0.000	1.000	0.000	0.000	0.000	1.000

Note. All metrics are calculated for every class against all other classes.

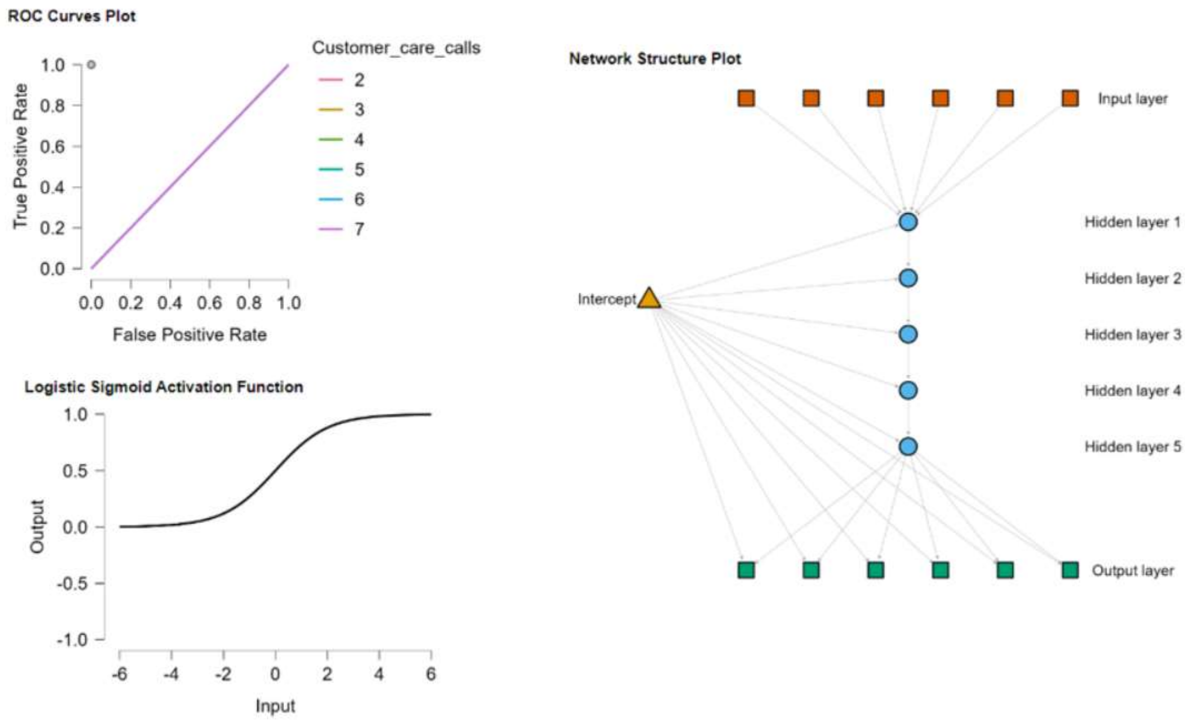
Network Weights

Node	Layer		Node	Layer	Weight
Intercept		→	Hidden 1	1	0.563
Customer_rating	input	→	Hidden 1	1	0.326
Cost_of_the_Product	input	→	Hidden 1	1	0.338
Prior_purchases	input	→	Hidden 1	1	−3.268
Discount_offered	input	→	Hidden 1	1	0.701
Weight_in_gms	input	→	Hidden 1	1	0.539
Reached.on.Time_Y.N	input	→	Hidden 1	1	0.035
Intercept		→	Hidden 1	2	−0.715
Hidden 1	1	→	Hidden 1	2	3.277
Intercept		→	Hidden 1	3	0.599
Hidden 1	2	→	Hidden 1	3	0.365
Intercept		→	Hidden 1	4	−0.130
Hidden 1	3	→	Hidden 1	4	−0.521
Intercept		→	Hidden 1	5	0.517
Hidden 1	4	→	Hidden 1	5	0.538
Intercept		→	4	4	−2.199
Hidden 1	3	→	4	4	−0.891
Intercept		→	5	5	−0.088
Hidden 1	4	→	5	5	−1.197
Intercept		→	6	6	−0.807
Hidden 1	5	→	6	6	0.092
Intercept		→	3	3	−1.722
Hidden 1	2	→	3	3	0.636
Intercept		→	7	output	−1.541
Hidden 1	6	→	7	output	−1.099
Intercept		→	2	2	−1.769
Hidden 1	1	→	2	2	−3.008

Note. The weights are input for the logistic sigmoid activation function.

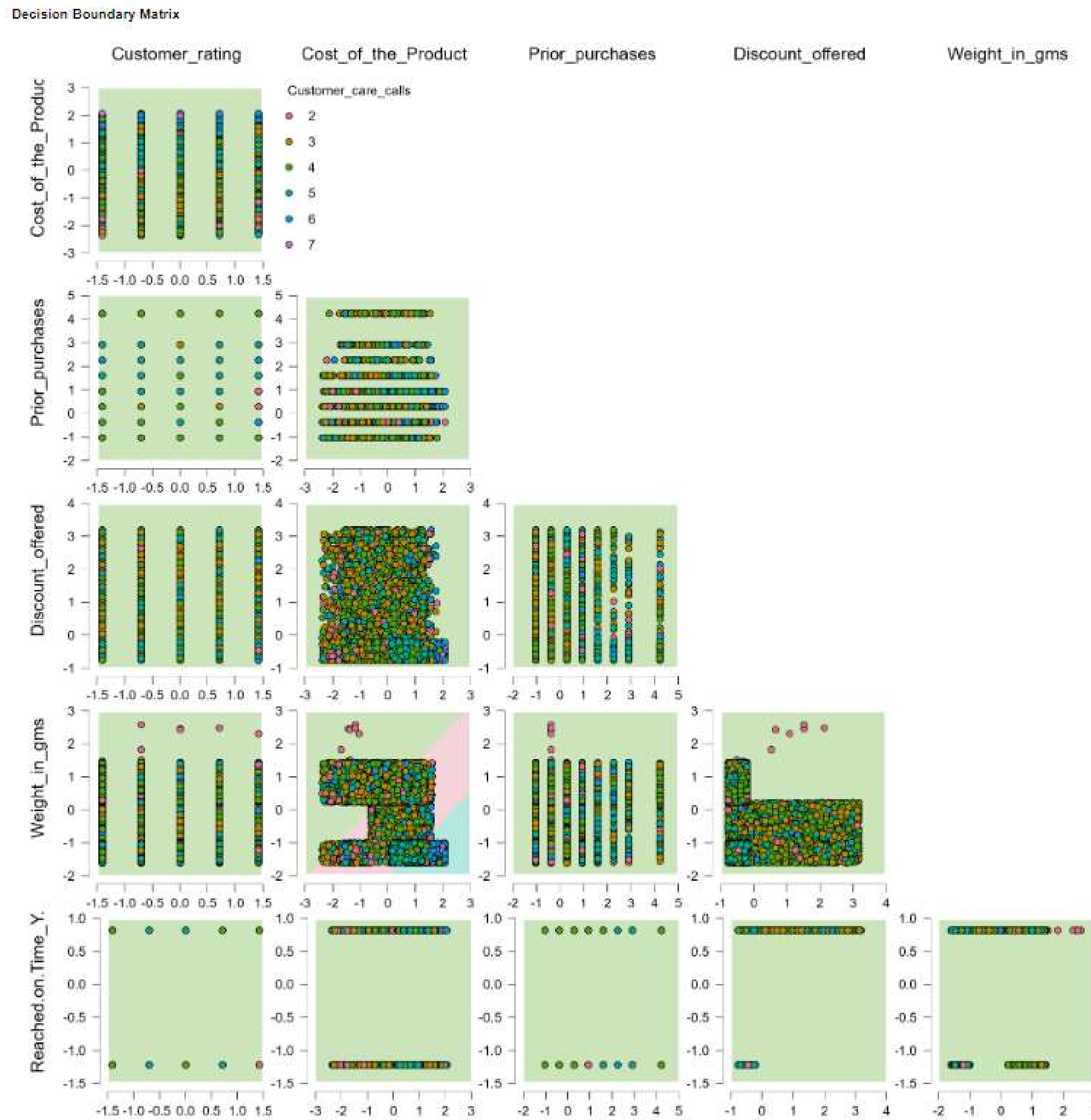
The ROC Curves Plot (top left) displays the true positive rate against the false positive rate for each class (2 through 7). The single diagonal line suggests that the model's performance is equivalent to random guessing, as there is no area under the curve for any specific class. This indicates that the model is unable to effectively distinguish between the different classes, achieving no better than a random classifier. This poor performance is consistent with low test accuracy and suggests that the model is not capturing any meaningful patterns in the data. The Logistic Sigmoid Activation Function plot (bottom left) illustrates the behavior of the activation function used in the neural network. The sigmoid function maps input values to an output range between 0 and 1, which is suitable for binary classification and is often used in hidden layers to introduce non-linearity. However, while the sigmoid function is shown here, the model's poor performance suggests that the activation function alone is not sufficient to enable effective classification, likely due to limitations in the network's structure or the feature set. The Network Structure Plot (right) visualizes the architecture of the neural network. The network has an input layer with six nodes (one for each feature), an intercept node, and five hidden layers with a single node each, connected sequentially. The final layer is an output layer with six nodes, representing the possible classes (2 through 7) for "Customer Care Calls." This simplistic structure with only one node per hidden layer is unusual and may be a key reason for the model's poor performance, as it lacks the complexity and capacity to capture patterns in the data. Neural networks typically require more nodes per hidden layer to learn intricate relationships, especially for multi-class classification problems. In summary, this neural network model shows severe limitations. The ROC curve indicates performance no better than random chance, while the simplistic structure of the network likely lacks the capacity to learn from the data. To improve this model, it would be necessary to increase the complexity by adding more nodes to each hidden layer, reconsider the activation function, or potentially redesign the architecture to better capture patterns in the data for effective multi-class classification (Aguilar-Ruiz and Michalak, 2022; Barulina, et al., 2023; Pawara et al., 2020), (Figure 34). .

Figure 34. Evaluation of Neural Networks.



This Decision Boundary Matrix visualizes pairwise decision boundaries between six features—"Customer Rating," "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)"—for a classification model that predicts "Customer Care Calls" classes (labeled 2 through 7). Each plot in the matrix represents a pair of features, where data points are color-coded by their actual class, and the background color indicates the model's predicted class for different regions in the feature space. From this matrix, it is evident that there is significant overlap among the data points from different classes within each decision boundary region. For example, in the scatter plots for "Cost of the Product" vs. "Prior Purchases" and "Discount Offered" vs. "Weight in grams," points from multiple classes (especially classes 3, 4, and 5) are densely clustered together in overlapping regions, making it difficult for the model to establish clear, distinct boundaries between classes. This overlap suggests that these features alone may not have strong predictive power for class differentiation. Additionally, some of the feature pairs, such as "Reached on Time (Y/N)" versus "Cost of the Product" or "Prior Purchases," show simple vertical or horizontal boundaries due to the binary nature of "Reached on Time (Y/N)." However, even in these cases, data points from different classes are mixed, indicating limited effectiveness in separating classes based on this feature. For example, "Reached on Time (Y/N)" creates clear-cut decision regions, but due to significant overlap, it does not provide a meaningful distinction among the classes. Overall, this Decision Boundary Matrix highlights that no single pair of features offers a strong basis for separating the classes clearly. The extensive overlap of data points across most feature pairs indicates that the model struggles to classify "Customer Care Calls" accurately with the current feature set. This visualization suggests that the model may benefit from additional feature engineering, more complex classification methods, or a different model that can handle non-linear relationships and overlapping class boundaries more effectively (Gong et al., 2020; Oliveira et al., 2022), (Figure 35)..

Figure 35. Decision Boundary Matric- Neural Networks.



6.5 Random Forest

The model has been trained with 86 trees and considers 2 features per split. The dataset includes 7040 training samples, 1760 validation samples, and 2199 test samples, totaling 10,999 observations. The model's test accuracy is 38.0%, with a validation accuracy of 38.2% and an out-of-bag (OOB) accuracy of 4.6%. The low accuracy values across validation and test sets indicate limited performance in correctly classifying instances across the test set.

Confusion Matrix. The Confusion Matrix shows that the model struggles to make accurate predictions, with predictions for each class largely scattered across the other classes. For example, only 15% of class 3 instances are correctly classified as class 3, while other class predictions are scattered across multiple categories, indicating poor class separation. This widespread misclassification suggests that the model is unable to capture the distinctions between classes effectively.

Class Proportions. The Class Proportions section details the distribution of each class in the full dataset, training set, validation set, and test set. The class proportions are consistent across these subsets, with class 4 being the largest (32.3%) and class 7 the smallest (2.2%). The class imbalance, especially the dominance of class 4, may contribute to the model's tendency to misclassify instances, as it may bias the model towards the majority class.

Evaluation Metrics. The Evaluation Metrics section provides a detailed breakdown of the model's performance for each class, including precision, recall, F1 score, accuracy, and other statistical measures. Notable observations include:

- The average precision is low at 0.398, and the average recall is 0.389, indicating poor performance in correctly identifying true positives across classes.
- The F1 scores are also low across all classes, with an average F1 score of 0.339, reflecting a weak balance between precision and recall.
- Area Under Curve (AUC) values vary across classes, with an average AUC of 0.689. This indicates that the model has some capability in distinguishing between positive and negative instances, but it is not highly effective for multi-class classification.
- The Matthews Correlation Coefficient (MCC) values are also low, with an average of 0.067, further indicating weak correlation between predicted and actual classes.

These metrics highlight the model's limited classification ability, with poor performance in precision, recall, and overall accuracy.

Feature Importance. The Feature Importance section shows the importance of each feature in the model. The values are measured in terms of mean decrease in accuracy and total increase in node purity:

- "Weight in grams" has the highest mean decrease in accuracy and node purity, suggesting it is the most influential feature in the model's decision-making process.
- "Prior Purchases" and "Cost of the Product" also contribute modestly to the model's performance.
- "Reached on Time (Y/N)" shows a very low importance, indicating it has minimal impact on the model's classification results.

In summary, this Random Forest model demonstrates weak classification performance, with low accuracy, poor recall, and scattered predictions across classes. The confusion matrix and low precision/recall scores suggest that the model struggles with class distinction, likely due to overlapping feature distributions and class imbalance. Feature importance indicates that only a few features, such as "Weight in grams," play a notable role in classification. To improve model performance, it may be beneficial to consider additional preprocessing steps, address class imbalance, or explore more complex classification methods (Shu et al., 2020; Salekshahrezaee et al., 2022; Wu et al., 2023) (Figure 36).

Figure 36. Random Forest Classification.

Random Forest Classification

Random Forest Classification

Trees	Features per split	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy	OOB Accuracy
66	2	7040	1760	2199	0.362	0.390	0.046

Note. The model is optimized with respect to the *out-of-bag accuracy*.

Data Split

Confusion Matrix

		Predicted					
		2	3	4	5	6	7
Observed	2	0	0.02	0.02	0.01	0	0
	3	0	0.15	0.11	0.01	0	0
	4	0	0.16	0.14	0.04	0	0
	5	0	0.08	0.08	0.05	0.01	0
	6	0	0.01	0.02	0.01	0.05	0
	7	0	0	0	0	0.02	0

Class Proportions

	Data Set	Training Set	Validation Set	Test Set
2	0.058	0.057	0.069	0.052
3	0.292	0.296	0.288	0.288
4	0.323	0.320	0.324	0.335
5	0.212	0.211	0.210	0.215
6	0.092	0.095	0.087	0.087
7	0.022	0.022	0.022	0.025

Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	115	628	736	473	192	55	2199
Accuracy	0.946	0.593	0.572	0.765	0.900	0.973	0.797
Precision (Positive Predictive Value)	0.300	0.359	0.373	0.416	0.611	0.143	0.389
Recall (True Positive Rate)	0.026	0.538	0.409	0.226	0.557	0.018	0.390
False Positive Rate	0.003	0.384	0.347	0.087	0.034	0.003	0.143
False Discovery Rate	0.700	0.641	0.627	0.584	0.389	0.857	0.633
F1 Score	0.048	0.431	0.390	0.293	0.583	0.032	0.371
Matthews Correlation Coefficient						0.043	0.043
Area Under Curve (AUC)	0.631	0.630	0.575	0.632	0.870	0.931	0.711
Negative Predictive Value	0.949	0.769	0.687	0.812	0.958	0.975	0.858
True Negative Rate	0.997	0.616	0.653	0.913	0.966	0.997	0.857
False Negative Rate	0.974	0.462	0.591	0.774	0.443	0.982	0.704
False Omission Rate	0.051	0.231	0.313	0.188	0.042	0.025	0.142
Threat Score	0.024	0.226	0.208	0.161	0.484	0.015	0.186
Statistical Parity	0.005	0.428	0.367	0.117	0.080	0.003	1.000

Note. All metrics are calculated for every class against all other classes.

Feature Importance

	Mean decrease in accuracy	Total increase in node purity
Weight_in_gms	-0.006	0.029
Prior_purchases	0.002	0.028
Cost_of_the_Product	0.008	0.022
Discount_offered	-0.007	0.008
Customer_rating	-0.002	0.004
Reached.on.Time_Y.N	-4.607×10^{-4}	0.002

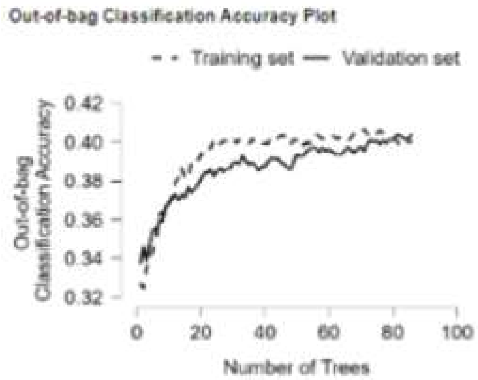
- Out-of-Bag Classification Accuracy Plot (top left): This plot tracks the model's out-of-bag accuracy against the number of trees in the forest. The accuracy plateaus around 38%, indicating that adding more trees does not significantly improve the model's classification accuracy. This relatively low out-of-bag accuracy suggests that the model may not be effectively capturing patterns in the data, possibly due to overlapping classes or limited feature differentiation.
- Mean Decrease in Accuracy and Total Increase in Node Purity** (top right): These bar charts illustrate the importance of each feature based on two metrics: mean decrease in accuracy and

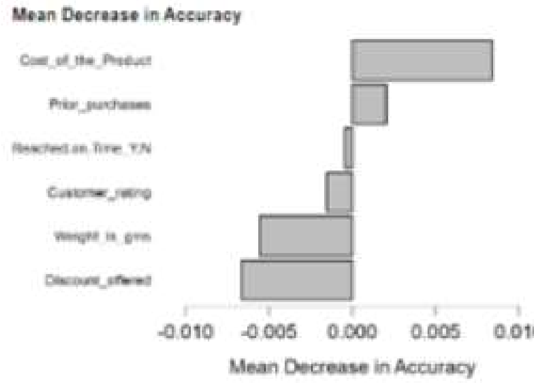
total increase in node purity. "Weight in grams" is the most important feature by both metrics, contributing significantly to the model's predictive power. "Prior purchases" and "Cost of the Product" are also influential, although to a lesser extent. "Reached on Time (Y/N)" and "Customer Rating" show minimal importance, indicating they have little impact on the model's ability to differentiate between classes.

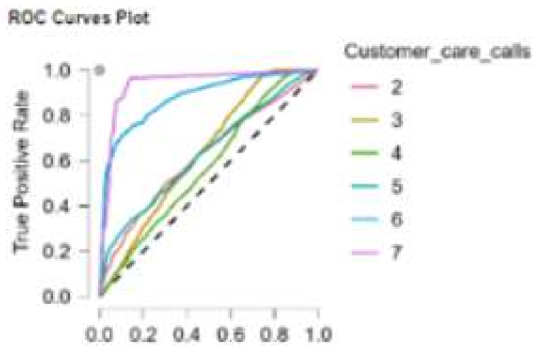
- ROC Curves Plot (middle left): This plot displays the ROC curves for each "Customer Care Calls" class (2 through 7), which show the model's ability to distinguish between positive and negative instances for each class. While some classes, such as class 7, achieve relatively high true positive rates, others, such as classes 4 and 5, follow a line close to the diagonal, indicating poor discrimination between these classes and others. The overall performance suggests that the model has difficulty separating certain classes, likely due to feature overlap.
- Andrews Curves Plot (bottom left): Andrews curves visualize class separability by representing each instance as a continuous curve. The significant overlap of curves across all classes indicates limited separability in the feature space. This overlap suggests that the classes are not well-differentiated based on the provided features, contributing to the model's low accuracy and difficulty in achieving clear class distinctions.

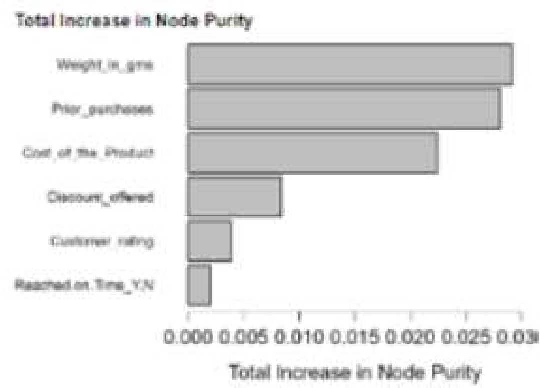
In summary, these plots indicate that while features like "Weight in grams" and "Prior purchases" contribute meaningfully to the model, overall class separability is poor. The out-of-bag accuracy remains low, and the ROC and Andrews curves confirm that the model struggles with class distinction, particularly for classes 4 and 5. The results suggest that further feature engineering, balancing class distributions, or exploring more complex models might be necessary to improve classification performance (Janitza and Hornung, 2018; Ouma et al., 2022; Loecher, 2022) (Figure 37).

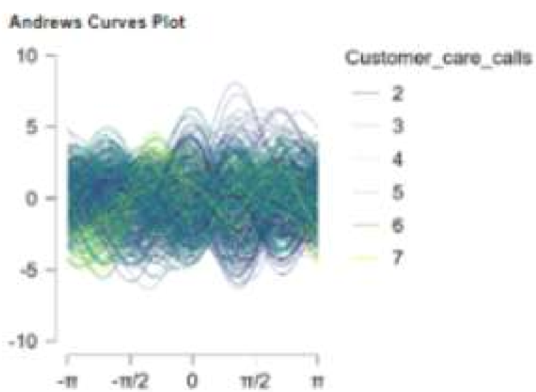
Figure 37. Characteristics of Random Forest.





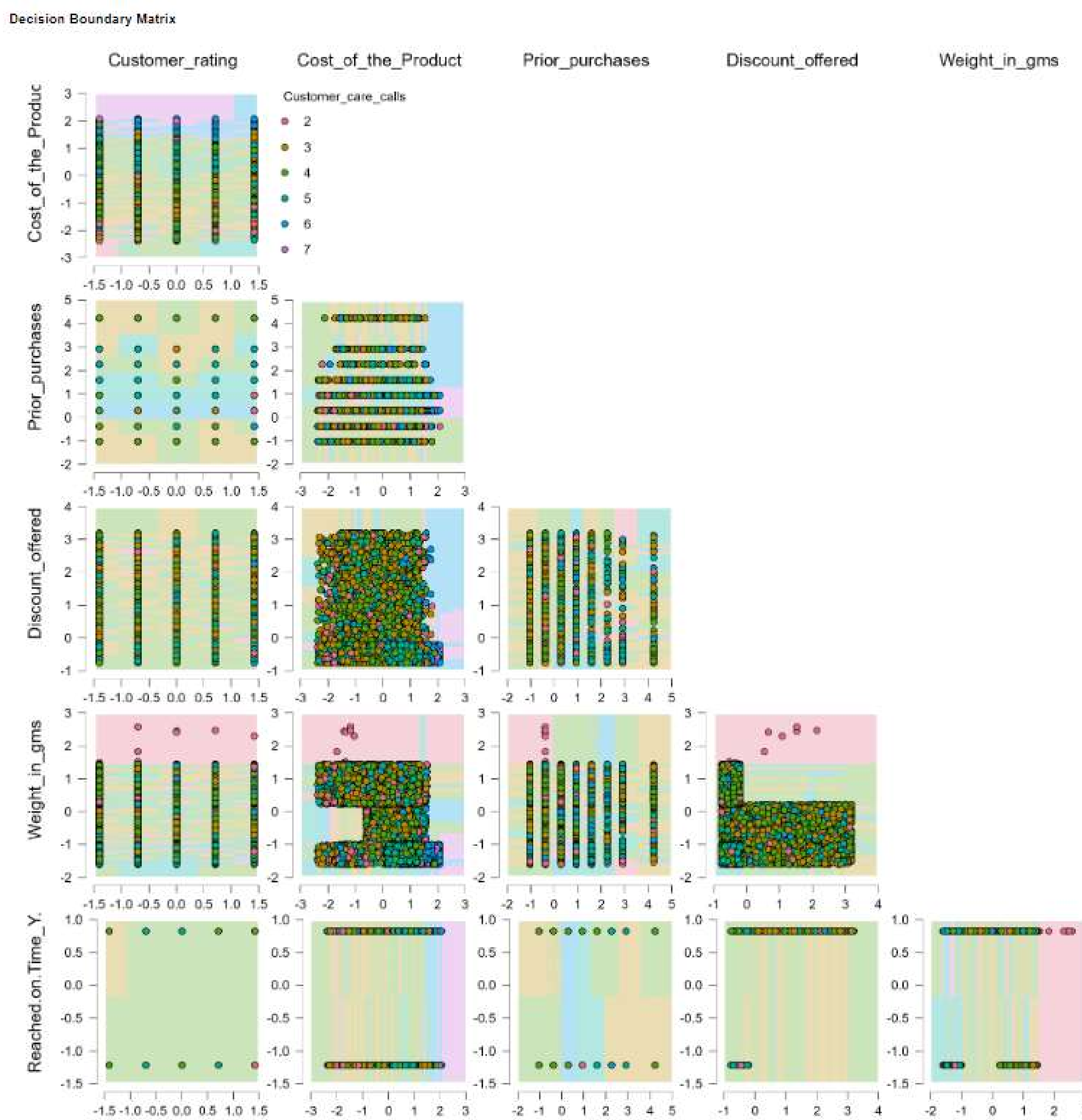






The Decision Boundary Matrix visualizes pairwise decision boundaries for a classification model applied to six features: "Customer Rating," "Cost of the Product," "Prior Purchases," "Discount Offered," "Weight in grams," and "Reached on Time (Y/N)." Each scatter plot in the matrix shows the decision boundaries for a specific pair of features, with data points color-coded according to "Customer Care Calls" classes (2 through 7). The background colour in each plot represents the regions where the model predicts each class, based on the distribution of feature values. The visualizations reveal substantial overlap between classes in almost every feature pair, indicating that the model struggles to distinguish between classes with the current feature set. For instance, in the plot for "Cost of the Product" vs. "Prior Purchases," classes are densely clustered together across most regions, with points from classes 3, 4, and 5 overlapping extensively. This overlap makes it difficult for the model to define clear boundaries, resulting in inaccurate or uncertain classifications. The binary feature "Reached on Time (Y/N)" also fails to provide substantial class separation. In plots involving this feature, vertical or horizontal boundaries are visible due to its binary nature, yet there is still a significant mix of classes on either side of the boundary, indicating that this feature alone does not offer a strong distinction between classes. In general, the background colors in each plot reflect the model's decision boundaries, but the overlapping data points show that these boundaries do not align well with class separations. This lack of clear separability suggests that these features may not be sufficiently informative for accurately predicting customer care call categories. The extensive overlap implies that the model may benefit from more informative features, additional feature engineering, or possibly a more complex model that can handle the non-linear and overlapping class boundaries present in this dataset (Gong et al., 2020; Ma and Maciejewski, 2020; Paranjape et al., 2021), (Figure 38).

Figure 38. Random Forest Classification.



6.6 Support Vector Machine

The SVM model is trained with 8,800 observations and uses 8,782 support vectors. It is tested on 2,199 samples, achieving a low test accuracy of 35.4%, indicating limited effectiveness in classifying instances across the test set.

Data Split. The data is divided into 8,800 training samples and 2,199 test samples out of a total of 10,999 observations. The training set is used to learn the decision boundaries, while the test set evaluates the model's generalization ability.

Confusion Matrix. The Confusion Matrix shows significant misclassification across classes, with many instances being incorrectly predicted as class 4. For example:

- Only 2% of true class 2 instances are correctly classified as class 2, while most are classified incorrectly, particularly as class 4.
- Similarly, for class 3, only 7% of instances are correctly classified, while 22% are incorrectly classified as class 4.
- Across other classes, there is a similar pattern where predictions are heavily biased towards class 4, indicating that the model is unable to separate the classes effectively and has a strong bias toward the majority class.

This misclassification pattern suggests that the model struggles with class distinction, likely due to overlapping feature distributions or insufficient class-specific information in the feature set.

Class Proportions. The Class Proportions table shows the distribution of each class across the data set, training set, and test set. The proportions are consistent across splits, with class 4 being the largest (32.3%) and class 7 the smallest (2.2%). The imbalance, particularly the high prevalence of class 4, may contribute to the model's bias toward predicting class 4, as the model tends to favor the majority class in the absence of strong distinguishing features.

In summary, the SVM model demonstrates poor classification performance with substantial misclassification and a strong bias toward class 4, likely due to class imbalance and overlapping feature distributions. The low test accuracy (35.4%) and misclassification patterns suggest that the model cannot capture meaningful class distinctions. To improve performance, addressing class imbalance, using more informative features, or tuning the SVM parameters may be necessary. Additionally, exploring more complex models or feature engineering may help achieve better class separation and overall accuracy (Choudhary and Shukla, 2022; Naboureh et al, 2020; Cao et al., 2020), (Figure 39).

Figure 39: Support Vector Machine

Data Split

Confusion Matrix

		Predicted					
		2	3	4	5	6	7
Observed	2	0	0.02	0.04	0	0	0
	3	0	0.07	0.22	0	0	0
	4	0	0.07	0.25	0	0	0
	5	0	0.04	0.17	0	0	0
	6	0	0.01	0.05	0	0.03	0
	7	0	0	0.01	0	0.02	0

Class Proportions

	Data Set	Training Set	Test Set
2	0.058	0.058	0.058
3	0.292	0.294	0.286
4	0.323	0.323	0.323
5	0.212	0.212	0.210
6	0.092	0.092	0.094
7	0.022	0.021	0.029

The following metrics include accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC), and other statistical measures, giving insight into the model's performance for each class individually.

Key Observations:

- Accuracy: The accuracy varies widely across classes, from 44.1% for class 4 to 97.1% for class 7, with an overall average accuracy of 78.5%. This variability suggests that the model performs well for some classes (e.g., classes 2 and 7) but struggles with others, particularly class 4.
- Precision and Recall: The precision and recall values are inconsistent across classes, with some metrics labeled as "NaN" (not a number), indicating they could not be computed, possibly due to limited or absent predictions for those classes. Precision is relatively low for most classes, with an overall average of 32.0%, indicating that when the model predicts a class, it is often incorrect. Recall averages at 35.4%, meaning that the model fails to capture a substantial portion of the true instances for each class.
- False Positive Rate (FPR) and False Discovery Rate (FDR): The FPR and FDR values are high for certain classes, such as class 3, where FPR is 18.4% and FDR is 66.7%. This indicates a high rate of false positives for this class, contributing to the model's poor precision.
- F1 Score: The F1 score, which balances precision and recall, is also low, averaging at 27.7%. Class 6 achieves the highest F1 score (45.5%), while other classes, like class 5, have particularly low F1 scores (1.3%), indicating that the model fails to consistently balance true positives against false positives and negatives across classes.
- Matthews Correlation Coefficient (MCC): MCC values, which reflect the quality of binary classifications for each class, are generally low or "NaN," indicating that there is little to no correlation between predicted and actual classes for many categories. This further supports the model's weak classification performance.

- Area Under Curve (AUC): The AUC values are all 0.500, which is equivalent to random guessing. This indicates that the model does not effectively distinguish between positive and negative instances for any of the classes.
- Negative Predictive Value (NPV) and True Negative Rate (TNR): The NPV and TNR are relatively high, with averages of 84.9% and 84.4%, respectively. This suggests that the model is generally better at identifying true negatives than true positives, possibly due to class imbalance.
- False Negative Rate (FNR) and False Omission Rate (FOR): The FNR is particularly high, with an average of 77.0%, indicating that the model frequently fails to detect true instances of each class. The FOR is also relatively high, averaging 15.1%, which suggests that many instances not predicted as a specific class are actually positive instances of that class.
- Threat Score and Statistical Parity: The Threat Score, which reflects the proportion of true positives relative to the sum of true positives, false negatives, and false positives, is low across all classes, averaging at 11.9%. Statistical Parity is maintained, with a value of 1.000, indicating balanced predictions across classes without bias toward any particular group.

In summary, this model demonstrates poor classification performance across most evaluation metrics. High false positive and false negative rates, low precision, recall, and F1 scores, and the random-guess AUC of 0.500 for all classes suggest that the model is largely ineffective in distinguishing between classes. Improvements could include addressing class imbalance, using more informative features, or exploring alternative models to improve classification accuracy and overall performance (Belinda et al., 2023; Sitarz, 2022; Carrington et al., 2022), (Figure 49).

Figure 49. Evaluation Metrics.

Evaluation Metrics

	2	3	4	5	6	7	Average / Total
Support	127	629	711	461	207	64	2199
Accuracy	0.942	0.648	0.441	0.788	0.918	0.971	0.785
Precision (Positive Predictive Value)	NaN	0.333	0.341	0.273	0.610	NaN	0.320
Recall (True Positive Rate)	0.000	0.229	0.783	0.007	0.362	0.000	0.354
False Positive Rate	0.000	0.184	0.722	0.005	0.024	0.000	0.156
False Discovery Rate	NaN	0.667	0.659	0.727	0.390	NaN	0.611
F1 Score	NaN	0.271	0.475	0.013	0.455	NaN	0.277
Matthews Correlation Coefficient	NaN					NaN	NaN
Area Under Curve (AUC)	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Negative Predictive Value	0.942	0.725	0.728	0.791	0.936	0.971	0.849
True Negative Rate	1.000	0.816	0.278	0.995	0.976	1.000	0.844
False Negative Rate	1.000	0.771	0.217	0.993	0.638	1.000	0.770
False Omission Rate	0.058	0.275	0.272	0.209	0.064	0.029	0.151
Threat Score	0.000	0.135	0.242	0.006	0.329	0.000	0.119
Statistical Parity	0.000	0.197	0.742	0.005	0.056	0.000	1.000

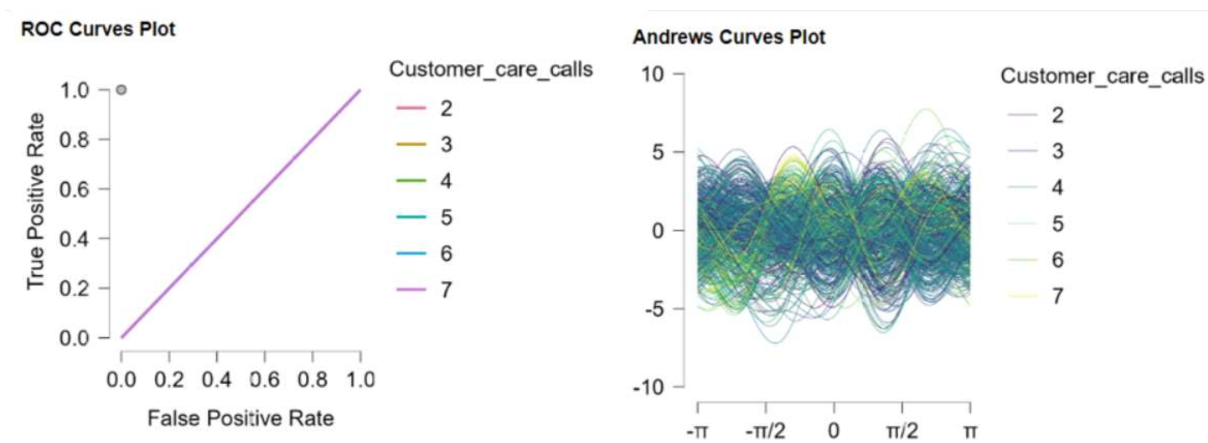
Note. All metrics are calculated for every class against all other classes.

ROC Curves Plot (Left). The ROC Curve Plot displays the true positive rate (TPR) against the false positive rate (FPR) for each class. However, the ROC curve for each class is essentially a diagonal line, indicating that the model's classification performance is no better than random guessing for all classes. A random classifier, or a model with no discriminatory ability, would produce an ROC curve along this diagonal, resulting in an area under the curve (AUC) of 0.5. This suggests that the model fails to distinguish between positive and negative instances for any of the classes effectively. Such a result typically indicates severe limitations in the model's capacity to learn class distinctions, possibly due to overlapping feature distributions or insufficient feature informativeness.

Andrews Curves Plot (Right). The Andrews Curves Plot is a way to visualize the separability of different classes by mapping each observation into a continuous curve. In this plot, each class (2 to 7) is represented by a different color, with each instance within a class mapped to a specific curve. However, there is substantial overlap between the curves of different classes, which indicates poor class separability. Ideally, distinct classes should have clearly separated curves with minimal overlap, which would indicate that the model could potentially distinguish between them. In this case, the high degree of overlap suggests that the feature set does not allow for clear differentiation between classes, leading to the model's difficulty in achieving accurate classification.

Overall, both plots indicate significant issues with the model's performance. The ROC Curve Plot demonstrates that the model's predictions are equivalent to random guessing, while the Andrews Curves Plot shows extensive overlap among classes, indicating poor class separability. These issues suggest that the model may need additional or more informative features, improved class balancing, or more complex modeling techniques to achieve better classification performance (Verbakel et al., 2020; Carrington et al., 2020; Namdar et al., 2021) (Figure 50)

Figure 50. Support Vector Machine.

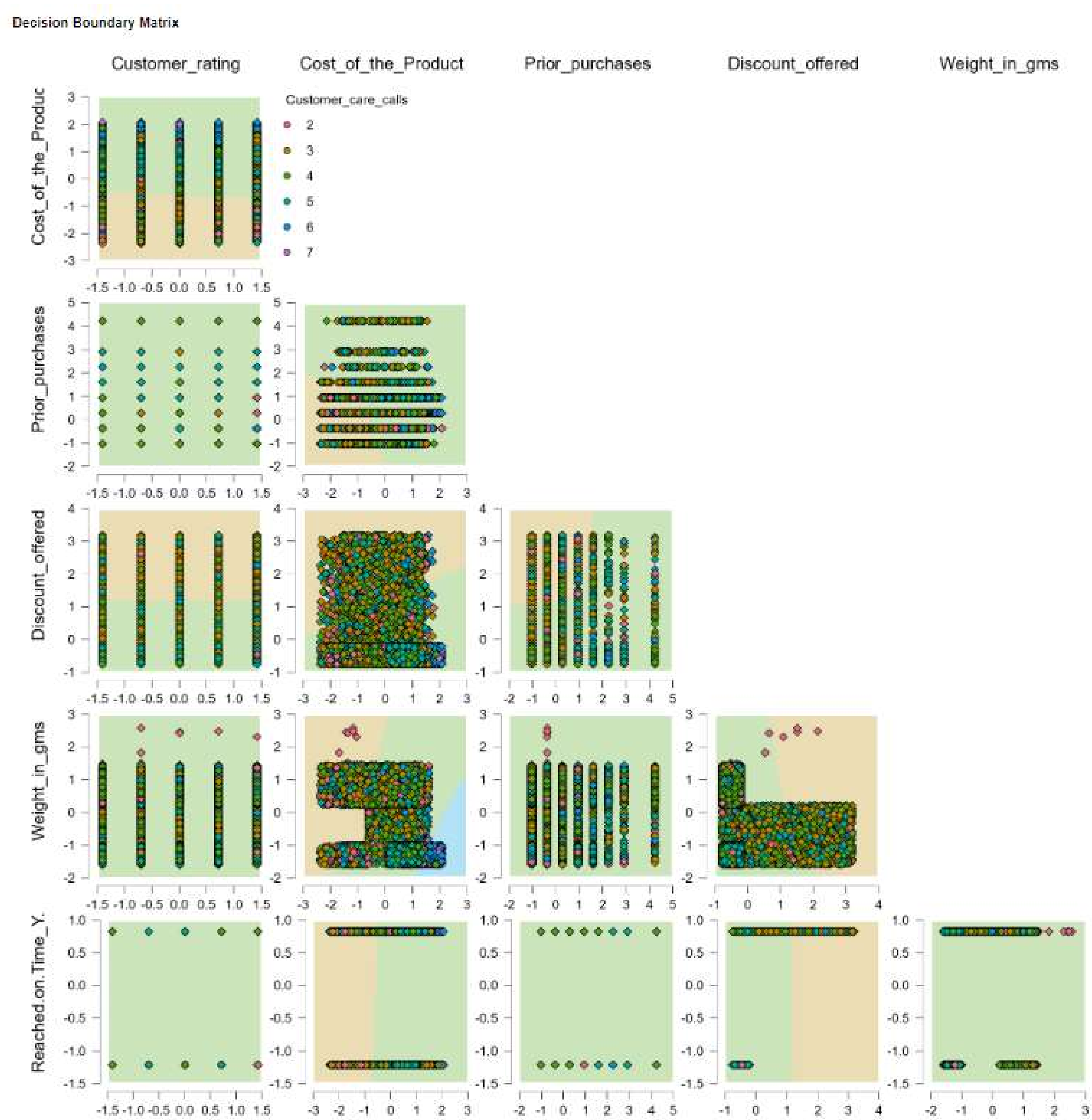


- High Class Overlap: The plots indicate a high degree of overlap among classes in most feature pair combinations. For example, in the plots involving "Cost of the Product" and "Prior Purchases," multiple classes are mixed within the same regions, with no clear separation between them. This suggests that the model struggles to find distinctive boundaries, likely due to similar feature distributions across classes.
- Limited Class Separability: For certain feature pairs, such as "Discount Offered" vs. "Weight in grams," the decision boundaries are complex and irregular. This complexity suggests that the model may be attempting to fit to noise or minor variations in the data, further limiting its effectiveness in drawing clear separations among classes. The plots do not show distinct regions where only a single class is dominant, indicating poor class separability.
- Binary Feature Patterns: The binary feature "Reached on Time (Y/N)" creates rigid, binary boundaries in plots involving this feature. However, these boundaries do not appear to contribute meaningful class separation; instances from multiple classes are distributed similarly on both sides of the binary split, indicating that this feature is not providing substantial information for differentiating classes.
- Class Imbalance Effects: Classes with higher frequencies appear more dominant in the decision boundary regions, while less frequent classes are scattered across the plots, often

without clear regions of their own. This dominance suggests that the model may be biased towards predicting the majority classes, possibly due to imbalanced class distributions.

Overall, the Decision Boundary Matrix reveals that the current feature set does not effectively differentiate the "Customer Care Calls" classes. The high degree of overlap and lack of distinct decision regions across most feature pairs suggest that the model struggles with class separability, likely due to insufficiently informative features. To improve model performance, additional feature engineering, such as creating new features that capture underlying patterns, or exploring alternative classification techniques, may be necessary to achieve clearer class boundaries and enhance predictive accuracy (Kim and Kim, 2020; Vuttipittayamongkol et al., 2021; Fu et al., 2020), (Figure 51.. Decision Boundary Matrix).

Figure 51. Decision Boundary Matrix.



7. Machine Learning Clustering

7.1 Hierarchical Clustering optimized with respect to the BIC (Bayesian Information Criterion) value

The model partitions the data into 10 clusters, with a total of 10,999 observations and an R^2 value of 0.838, indicating that 83.8% of the total variance is explained by the clustering solution. The AIC (Akaike Information Criterion) and BIC values for this model are 5419.010 and 5638.170, respectively, and the silhouette score is 0.380, suggesting moderate cluster cohesion and separation.

Cluster Information. Each of the 10 clusters varies significantly in size and internal structure:

- Size: Cluster sizes range from 6 observations in Cluster 10 to 3,880 observations in Cluster 9, showing a highly imbalanced distribution across clusters.
- Explained Proportion of Within-Cluster Heterogeneity: Cluster 9 explains the largest proportion of within-cluster heterogeneity (41.3%), while Cluster 10, with only 6 observations, explains virtually none of it (4.582×10^{-4}).
- Within Sum of Squares (WSS): This metric shows the variation within each cluster, with higher values indicating more dispersed clusters. Cluster 9 has the largest WSS (2212.018), suggesting it has the most internal variance, while Cluster 10, being extremely small, has the lowest WSS (2.456).
- -Silhouette Score: The silhouette score measures how similar each point is to its own cluster compared to other clusters. Cluster scores range from 0.046 in Cluster 4, indicating poor cohesion, to 0.624 in Cluster 10, suggesting better-defined boundaries.

Evaluation Metrics. The Evaluation Metrics provide additional insights into the clustering quality:

- Maximum Diameter: The largest distance within any cluster is 3.050, indicating the maximum spread of observations within a cluster.
- Minimum Separation: The smallest separation between clusters is 0.045, which indicates a close proximity between certain clusters, potentially leading to overlap.
- Pearson's γ (Gamma) and Dunn Index: Pearson's γ is 0.628, suggesting moderate cluster separation, while the Dunn index is very low at 0.015, indicating poor separation between the closest clusters relative to the within-cluster spread.
- Entropy: The entropy value of 1.834 measures the diversity within clusters, with lower values indicating more homogeneous clusters.
- Calinski-Harabasz Index: The Calinski-Harabasz index is high at 6296.379, generally indicating well-defined clusters, although other metrics suggest some issues with cluster separation.

In summary, this hierarchical clustering model demonstrates a moderate level of cohesion and separation, as shown by an average silhouette score of 0.380 and an R^2 of 0.838. The model has well-separated clusters based on the Calinski-Harabasz index, but low values for the Dunn index and some clusters with minimal separation indicate potential overlap. Cluster sizes are highly imbalanced, with some clusters containing thousands of observations and others containing very few, which could impact the interpretability and stability of the clustering solution. Adjusting the number of clusters or experimenting with different clustering algorithms might yield a more balanced and clearly defined structure (Da Silva et al., 2020; Laskowski and Tomiło, 2023; Liu, 2022), (Figure 51).

Figure 51. Hierarchical clustering.

Hierarchical Clustering ▼

Hierarchical Clustering ▼

Clusters	N	R^2	AIC	BIC	Silhouette
10	10999	0.838	5419.010	5638.170	0.380

Note. The model is optimized with respect to the *BIC* value.
Note. The optimum number of clusters is the maximum number of clusters. You might want to adjust the range of optimization.

Cluster Information

Cluster	1	2	3	4	5	6	7	8	9	10
Size	539	533	848	521	468	193	1896	2115	3880	6
Explained proportion within-cluster heterogeneity	0.041	0.061	0.114	0.070	0.046	0.016	0.127	0.112	0.413	4.582×10^{-4}
Within sum of squares	217.725	327.366	608.659	373.846	247.428	85.467	682.230	601.811	2212.018	2.456
Silhouette score	0.539	0.198	0.355	0.046	0.194	0.361	0.488	0.520	0.338	0.624

Note. The Between Sum of Squares of the 10 cluster model is 27634.99
Note. The Total Sum of Squares of the 10 cluster model is 32994

Evaluation Metrics

	Value
Maximum diameter	3.050
Minimum separation	0.045
Pearson's γ	0.628
Dunn index	0.015
Entropy	1.834
Calinski-Harabasz index	6296.379

Note. All metrics are based on the *euclidean* distance.

Elbow Method Plot (Top Left). The Elbow Method Plot shows within-cluster sum of squares (WSS) and its change as the number of clusters increases. The plot includes AIC and BIC metrics, with the lowest BIC identified at 10 clusters, suggesting this as the optimal number of clusters. The plot indicates a diminishing improvement in WSS beyond 10 clusters, supporting the choice of 10 clusters as a point where adding more clusters yields minimal additional benefit.

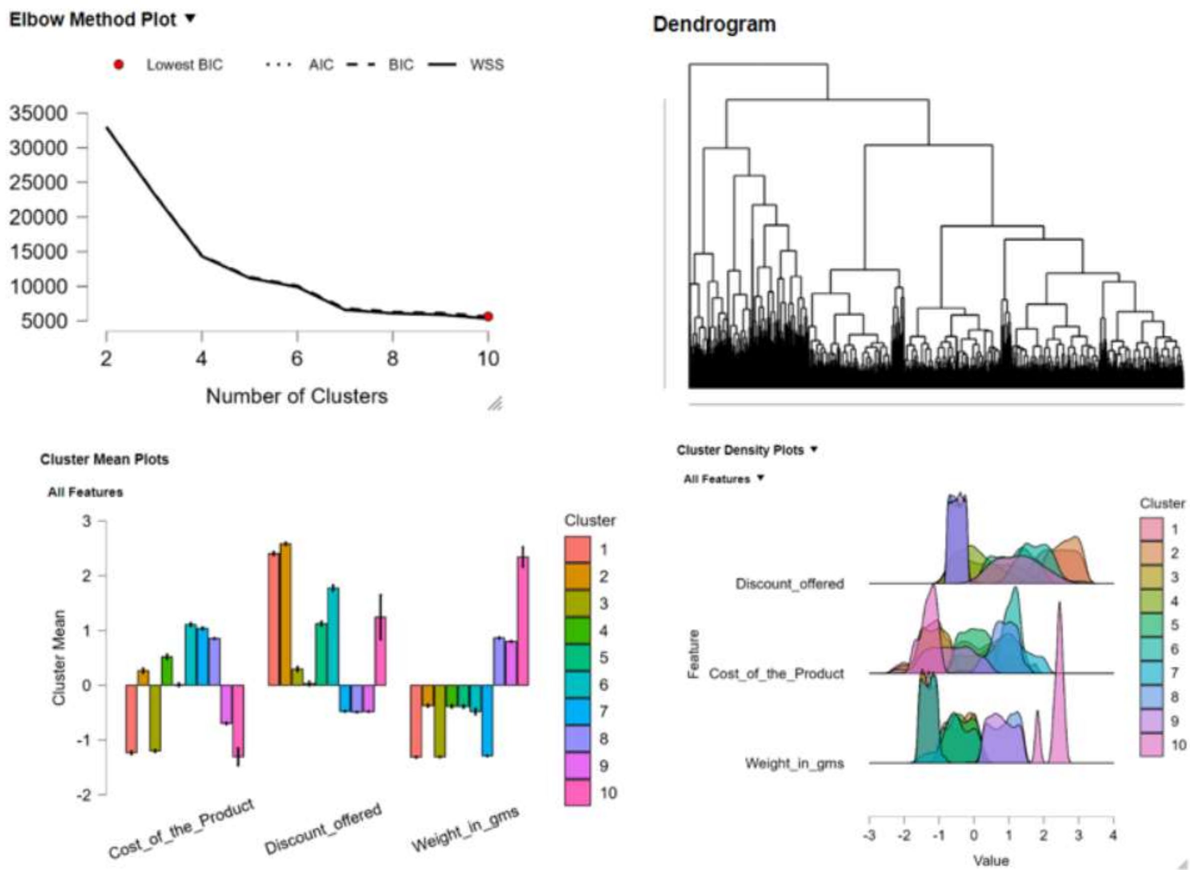
Dendrogram (Top Right). The Dendrogram illustrates the hierarchical structure of the clustering, showing how individual observations or smaller clusters merge into larger clusters as the threshold distance increases. The branching structure provides insight into the relationship and similarity between clusters. Observations that merge early in the dendrogram are more similar, while those merging at higher levels represent larger, more generalized groupings. This hierarchical view helps visualize the nested clustering structure and the relative distance between clusters.

Cluster Mean Plots (Bottom Left). The Cluster Mean Plots depict the mean values of each feature ("Cost of the Product," "Discount Offered," and "Weight in grams") within each of the 10 clusters. Different colors represent different clusters, and bars show mean values, highlighting the variation in feature distributions across clusters. For example, "Cost of the Product" has both positive and negative means across clusters, suggesting that this feature helps differentiate clusters. Similarly, "Discount Offered" and "Weight in grams" show distinct mean values in certain clusters, indicating these features also contribute to defining cluster characteristics.

Cluster Density Plots (Bottom Right). The Cluster Density Plots show the distribution of values for each feature ("Discount Offered," "Cost of the Product," and "Weight in grams") within each cluster. Each cluster's density curve is color-coded, representing the frequency of feature values within that cluster. This visualization helps identify overlaps and unique distributions, showing how well each feature separates clusters. For example, "Discount Offered" and "Cost of the Product" exhibit distinctive peaks for some clusters, while other clusters overlap significantly, indicating areas where clusters may not be well-separated.

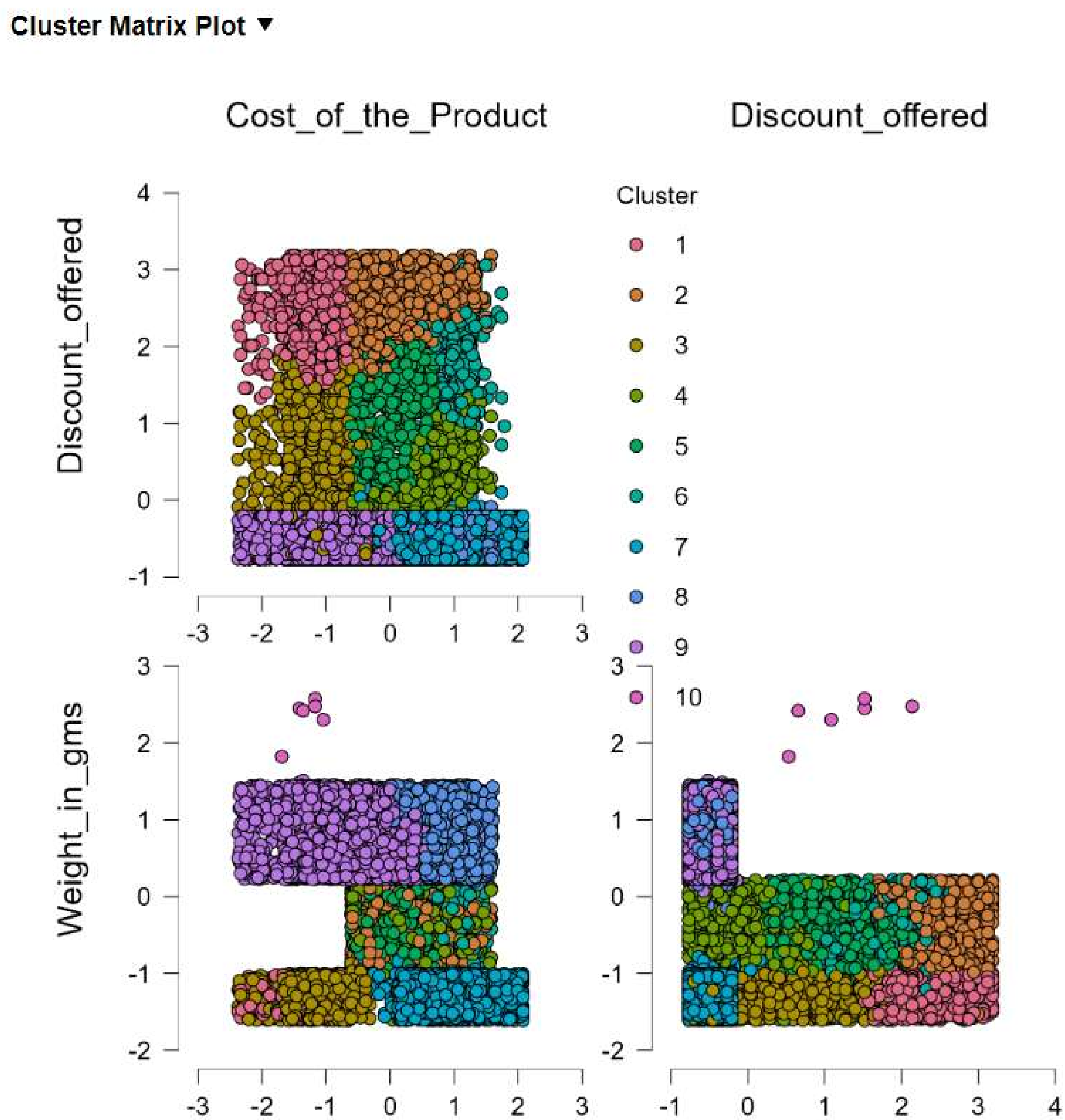
In summary, these plots provide a comprehensive view of the hierarchical clustering model with 10 clusters. The Elbow Method Plot and Dendrogram support the choice of 10 clusters as optimal based on WSS and BIC. The Cluster Mean and Density Plots show that the features "Cost of the Product," "Discount Offered," and "Weight in grams" contribute to cluster differentiation, though some clusters exhibit overlapping feature distributions, indicating potential areas where the clustering may benefit from additional or alternative features to improve separation (Aksan et al., 2021; Li and Wei, 2022; Onumanyi et al., 2022), (Figure 52).

Figure 52. Hierarchical Clusterting.



This Cluster Matrix Plot provides a visual representation of the hierarchical clustering across three features: "Cost of the Product," "Discount Offered," and "Weight in grams." Each point is color-coded according to its cluster (from clusters 1 to 10). In the plot for Cost of the Product vs. Discount Offered (top left), clusters occupy different vertical regions, suggesting that "Discount Offered" contributes to differentiating clusters. However, there is overlap among clusters along the "Cost of the Product" axis, indicating that this feature alone may not fully separate some clusters. The plot for Cost of the Product vs. Weight in grams (bottom left) shows a more stratified distribution, with clusters occupying distinct areas along the "Weight in grams" axis, indicating this feature's importance in separating clusters. Some clusters are closely packed, suggesting partial overlap in feature space. Lastly, the Discount Offered vs. Weight in grams plot (bottom right) highlights a clear separation of clusters, with most clusters occupying distinct blocks. However, some clusters, particularly 8 and 10, are more spread out, suggesting that within-cluster variance is high for these clusters, especially in the "Discount Offered" dimension. Overall, this Cluster Matrix Plot demonstrates how these three features contribute to distinguishing clusters, with "Weight in grams" and "Discount Offered" playing a significant role in cluster differentiation, while "Cost of the Product" provides limited separation. There is also evidence of overlap among clusters in certain regions, indicating potential improvements in clustering quality if additional distinguishing features were available (Wei, 2023; van Dongen, 2022; Guénard and Legendre, 2022), (Figure 53).

Figure 53. Cluster Matrix Plot.



8. Warehouse Management: Lesson Learned

This paper represents an in-depth analysis of the relationship developed between customer care calls and buying behavior in a warehouse management system. Such valuable insights from this research would go a long way in optimizing warehouse operations and help in documenting key drivers of customer satisfaction and operational efficiency. Several lessons can be identified from the study regarding the integration of advanced analytics and customer-centric strategies in order to bring change to the logistical processes, as well as customer experiences. One important conclusion seems to be that customer care calls have something to do with operational inefficiencies. In fact, analysis has been able to establish that more customers contact the company when there are delays or misunderstandings about products or when expectations go unmet. The generation of these inefficiencies through process optimization can improve service reliability and reduce the complaints received from customers substantially. It points out the need for finding robust means of data collection and analysis in finding and eliminating recurring problems within the warehouse systems. Businesses can also avoid major grievances among customers through proactive measures like clear communication channels, facilitation in processing orders, and accuracy in product descriptions. A second key insight is the nature and type of the underlying product attributes driving changes in customer behavior and operational performance. For example, the study indicates that heavier products tend to translate into fewer customer-care calls, either because customers are confident about substantial items or because installation or delivery support comes with big items. Lighter or inexpensive items could spur more questions and complaints because people are buying on impulse or with insufficient product information. It indicates that warehouses will be obliged, through inventory nature, to modify their handling and support strategies, further extending the gamut of tailored services to match the profile of customer expectations. It also emerges from the findings that timely delivery remains one of the most key determinants of customer satisfaction. Besides, delayed delivery tends to erode customer trust and may also increase the chances of customer calls to care. On the other hand, on-time delivery creates a positive attitude toward reliability with less post-purchase support. Again, this brings in the requirement for an effective logistical network, helped with predictive analytics in demand forecasting to avoid delays.Warehouse managers should put more focus on real-time tracking, ensuring efficient inventory management, and improving coordination with transporters to ensure timely deliveries that guarantee customer loyalty. Discounts come out as another critical determinant of customer service. This is evidenced by the study that the discount rate is inversely proportional to customer care interactions, probably because the value that customers perceive is well worth minor inconveniences. However, a high reliance on discounting can devalue a brand and ultimately eat away at profitability over time. A balance has to be achieved whereby strategic discounts are given out without compromising product quality or levels of service. Data analytics can be used in warehouses to personalize the discount strategy by customer segment; hence, they are winning price-sensitive buyers without decreasing the brand's value. Finally, this here shows how different data-driven decisions can actually be in business. The application of econometric models and statistical analyses gives an actionable understanding of customer behavior and holds up a mirror to operational bottlenecks. Therefore, it will be easy to have warehouses matching their strategies with the expectations of their customers, whereby all would be working toward efficiency and satisfaction. Predictive analytics, machine learning, and real-time monitoring systems are some of the key investments that can drive businesses to better anticipate challenges, smoothen operations, and offer superior customer experiences. In a nutshell, the article has much to teach on warehouse management. It identifies key data-driven strategies, customer-oriented approaches, and customized operational practices that are essential in negotiating the challenges of fragmented supply chains in the modern context. Efficiency improvement, use of technology, and customer satisfaction are three promoting elements, which, with time, will help businesses reach operational excellence and command long-term customer loyalty (Martínez et al., 2020; Sharma, et al., 2022; Luo et al., 2020).

9. Conclusion

This study investigates the complex link between customer care calls and purchase behavior in the context of a warehouse management system to deliver the major insights of the dynamics of operational efficiency and customer satisfaction. This research, by analyzing the interplay of product attributes, delivery performance, and customer engagement, underlines the role of advanced analytics and technology in modern logistical operations. It notices that product attributes like price have weight and discount have different impacts on said customer contact and satisfaction scores. Heavier products, seen as durable and lasting longer, tend to be marked by lower volumes of calls to the customer care whereas lighter and cheaper items, often purchased on impulse, tend to be reflective of more calls due to lack of information prior to buying or perceived faults at the time of delivery. Likewise, discounting again shows up as a two-edged sword-while value for money boosts satisfaction, it erodes the brand upon overuse. It points out that timely delivery has been a cornerstone in building customer satisfaction, while delays notably increase dissatisfaction and fire customer care interactions. These findings bring into light that these difficulties are increased by operational inefficiencies such as delays in delivery, misaligned customer expectations, or poor product information. This also tends to establish that the price of a product is positively related to frequency in customer care interactions, as a reflection of higher expectations by customers when products are priced high. These insights then suggest the need for businesses to apply differentiated strategies of customer engagement across different product types and levels of customer expectations. This begets the use of advanced analytics, machine learning, and real-time monitoring as a key recommendation that would solve the recurring efficiencies. Of these, predictive analytics offers great potential in demand forecasting, logistical optimization, and proactive customer care. Such tools would enable companies to proactively address delays, prevent overstocking or understocking of inventory, and time customer engagement for greater operational efficiency and higher levels of customer satisfaction. The research also advocates transparency in communication with customers about timelines of deliveries to avoid mismatched expectations leading to resultant calls to customer care for help. The findings imply, from a practical standpoint, the need to balance the discount strategy to attract customers with the long-run brand equity. Based on this, businesses are suggested to find data-driven approaches through which customer grouping would be done based on purchasing pattern and product characteristics so that restoration offers that result in customer loyalty can be designed without hurting profitability. Alternatively, strategic relevance of Warehouse Management Systems using technological innovations like RFID and simulation tools that could streamline operations and mitigate inefficiencies should be pursued. Although the research provides a sound structure for understanding customer behavior dynamics and operational efficiencies, it recognizes certain limitations. The lack of detailed product-specific attributes in the dataset restricts the depth of analysis, especially with respect to correlating technical features with customer satisfaction. Further research can expand this study to include variables such as customer demographics, geographic influences, and competitor benchmarks to better probe into customer behavior related to warehouse operations. In addition, it would be very valuable for future research to understand how customer care interactions influence retention and loyalty over an extended period of time. The study, in essence, maps out an all-inclusive route that combines technology, customer-focused approach, and operational enhancements in the management of a modern warehouse. It identifies areas where efficient processes are going on, further aligns customer expectations, and identifies predictive analytics to improve customer satisfaction and loyalty. The strategies, in this respect, are not only scheduled to make instantaneous impacts on customer experiences but also achieve long-term growth rates and competitive advantage in dynamic market settings.

10. References

Acevedo, A., Duran, C., Kuo, M. J., Ciucci, S., Schroeder, M., & Cannistraci, C. V. (2022). Measuring Group Separability in Geometrical Space for Evaluation of Pattern Recognition and Dimension Reduction Algorithms. IEEE Access, 10, 22441-22471.

Agarwal, A., Kenney, A. M., Tan, Y. S., Tang, T. M., & Yu, B. (2023). MDI+: A flexible random forest-based feature importance framework. arXiv preprint arXiv:2307.01932.

Aguilar-Ruiz, J. S., & Michalak, M. (2022). Multiclass classification performance curve. IEEE Access, 10, 68915-68921.

Ahmed, A. M., & Mohamed, A. A. (2023). Inbound, outbound disturbances and supply chain vulnerability of firms operating in post-conflict zones: the case of Somalia. Journal of System and Management Sciences, 13(3), 352-363.

Albadrani, A., Zohdy, M. A., & Olawoyin, R. (2020, July). An approach to optimize future inbound logistics processes using machine learning algorithms. In 2020 IEEE International Conference on Electro Information Technology (EIT) (pp. 402-406). IEEE.

Ali, A. (2020, October). Ensemble learning model for prediction of natural gas spot price based on least squares boosting algorithm. In 2020 International conference on data analytics for business and industry: way towards a sustainable economy (ICDABI) (pp. 1-6). IEEE.

Apolonio, S. D., & Norona, M. I. (2021). Automating inventory management in distribution centers of a leading fast-moving consumer goods (FMCG) industry player in the beverage market: A supply chain 4.0 journey. In Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management (pp. 1602-1614).

Bajpai, D., & He, L. (2020, September). Evaluating KNN performance on WESAD dataset. In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 60-62). IEEE.

Barulina, M., Okunkov, S., Ulitin, I., & Sanbaev, A. (2023). Sensitivity of Modern Deep Learning Neural Networks to Unbalanced Datasets in Multiclass Classification Problems. Applied Sciences, 13(15), 8614.

Belinda, L., Dahbia, I., & Zahira, C. (2023, October). A Hybrid Artificial Intelligence Approach for Early Stroke Prediction. In 2023 5th International Conference on Pattern Analysis and Intelligent Systems (PAIS) (pp. 1-8). IEEE.

Bemporad, A. (2022). A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems. IEEE Transactions on Automatic Control, 68(6), 3194-3209.

Bernard, J., Hutter, M., Zeppelzauer, M., Sedlmair, M., & Munzner, T. (2021). ProSeCo: Visual analysis of class separation measures and dataset characteristics. Computers & Graphics, 96, 48-60.

Bottani, E., Tebaldi, L., Rossi, M., & Casella, G. (2020, September). A storage assignment simulation model for optimizing processes in an e-commerce warehouse of a fashion supply chain. In International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation (pp. 1-7). CAL-TEK Srl.

Brylla, D., & Walsh, G. (2022). When Faster Online Delivery Backfires: Examining the Negative Consequences of Split Deliveries. International Journal of Electronic Commerce, 26(4), 497-525.

Brzezinski, D., Stefanowski, J., Susmaga, R., & Szczech, I. (2019). On the dynamics of classification measures for imbalanced and streaming data. IEEE transactions on neural networks and learning systems, 31(8), 2868-2878.

Caballero, L., Jojoa, M., & Percybrooks, W. S. (2020). Optimized neural networks in industrial data analysis. SN Applied Sciences, 2(2), 300.

Candra, F. A., Sutanto, J. E., & Sondak, M. R. (2022). The influence of product quality, service quality, and timeliness of delivery on repurchase intention at cake by tenggilis surabaya.

Cao, B., Liu, Y., Hou, C., Fan, J., Zheng, B., & Yin, J. (2020). Expediting the accuracy-improving process of svms for class imbalance learning. IEEE Transactions on Knowledge and Data Engineering, 33(11), 3550-3567.

Carrington, A. M., Fieguth, P. W., Qazi, H., Holzinger, A., Chen, H. H., Mayr, F., & Manuel, D. G. (2020). A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC medical informatics and decision making, 20, 1-12.

Chandrasekar, K. S., & Geetha, P. (2021). A new formation of supervised dimensionality reduction method for moving vehicle classification. Neural Computing and Applications, 33(13), 7839-7850.

Chen, C., & Liu, J. (2024, April). Research of Cloud-Native AS/RS Warehouse Management and Control Platform Architecture. In 2024 5th International Conference on Computer Engineering and Application (ICCEA) (pp. 277-285). IEEE.

Chen, S., Lei, J., & Moinzadeh, K. (2023). Inventory and supply chain management with auto-delivery subscription. Production and Operations Management, 32(12), 4068-4087.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, e623.

Choudhary, R., & Shukla, S. (2022). Reduced-Kernel weighted extreme learning machine using universum data in feature space (RKWELM-UFS) to handle binary class imbalanced dataset classification. Symmetry, 14(2), 379.

Corrêa, A. C. (2023). Supply Chain Control Tower: Definitions and Gaps Between Literature and Application in Industries.

Correndo, A. A., Rosso, L. H. M., Hernandez, C. H., Bastos, L. M., Nieto, L., Holzworth, D., & Ciampitti, I. A. (2022). metrica: an R package to evaluate prediction performance of regression and classification point-forecast models. Journal of Open Source Software, 7(79), 4655.

Da Silva, L. E. B., Melton, N. M., & Wunsch, D. C. (2020). Incremental cluster validity indices for online learning of hard partitions: Extensions and comparative study. IEEE Access, 8, 22025-22047.

Dai, Z., Li, P., Zhu, M., Zhu, H., Liu, J., Zhai, Y., & Fan, J. (2023). Dynamic prediction for attitude and position of shield machine in tunneling: A hybrid deep learning method considering dual attention. Advanced Engineering Informatics, 57, 102032.

de Oliveira, A. V., Pimentel, C. M. O., Godina, R., Matias, J. C. D. O., & Garrido, S. M. P. (2022). Improvement of the Logistics Flows in the Receiving Process of a Warehouse. Logistics, 6(1), 22.

Del Moral, P., Nowaczyk, S., & Pashami, S. (2022). Why is multiclass classification hard?. IEEE Access, 10, 80448-80462.

Dewi, I. K., & Shofa, R. N. (2023). Development of warehouse management system to manage warehouse operations. International Journal of Applied Information Systems and Informatics (JAISI), 1(1).

Doganer, A., Yaman, S., Eser, N., & Metin, T. O. (2020). Different machine learning methods based prediction of mild cognitive impairment.

Du, W., Peng, X., & Feng, D. (2022, June). Product Key Characteristics Mapping Transformation Method Based on Customer Demand. In Proceedings of the 2022 2nd International Conference on Control and Intelligent Robotics (pp. 349-353).

DÜNDAR, A. O., & ÖZTÜRK, R. (2020). The effect of on-time delivery on customer satisfaction and loyalty in channel integration. Business & Management Studies: An International Journal, 8(3), 2675-2693.

Feng, Y., & Tian, L. (2021). Issues and solutions in biomarker evaluation when subclasses are involved under binary classification. Statistical Methods in Medical Research, 30(1), 87-98.

Firdausanti, N. A., Fatyanosa, T. N., Data, M., Mendonça, I., & Aritsugi, M. (2022, December). Two-stage sampling: A framework for imbalanced classification with overlapped classes. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 271-280). IEEE.

Fu, G. H., Wu, Y. J., Zong, M. J., & Yi, L. Z. (2020). Feature selection and classification by minimizing overlap degree for class-imbalanced data in metabolomics. Chemometrics and Intelligent Laboratory Systems, 196, 103906.

Gajowniczek, K., & Ząbkowski, T. (2021). ImbTreeAUC: An R package for building classification trees using the area under the ROC curve (AUC) on imbalanced datasets. SoftwareX, 15, 100755.

Garver, M. S., & Williams, Z. (2020). Utilizing relative weight analysis in customer satisfaction research. International Journal of Market Research, 62(2), 158-175.

Gegeleso, O. M. (2020). Effect of Warehousing Operation on Inbound Logistics Practices of Consumer Goods Industries Performce in South Western Nigeria.

Ghiasi-Shirazi, S. K. (2022). Revisiting Classical Multiclass Linear Discriminant Analysis with a Novel Prototype-based Interpretable Solution. arXiv preprint arXiv:2205.00668.

Gong, N., Wang, Z., Chen, S., Liu, G., & Xin, S. (2020, November). Decision boundary extraction of classifiers. In Journal of Physics: Conference Series (Vol. 1651, No. 1, p. 012031). IOP Publishing.

GONZALEZ-RAMIREZ, A., LOPEZ, J., TORRES-ROMAN, D., & YAÑEZ-VARGAS, I. (2021). Analysis of multi-class classification performance metrics for remote sensing imagery imbalanced datasets. Journal of Quantitative and Statistical Analysis, 8-22.

Guénard, G., & Legendre, P. (2022). Hierarchical clustering with contiguity constraint in R. Journal of statistical software, 103, 1-26.

Guo, D., Duan, P., Yang, Z., Zhang, X., & Su, Y. (2024). Convolutional Neural Network and Bidirectional Long Short-Term Memory (CNN-BiLSTM)-Attention-Based Prediction of the Amount of Silica Powder Moving in and out of a Warehouse. Energies, 17(15), 3757.

Guo, Y., Wang, W., & Wang, X. (2021). A robust linear regression feature selection method for data sets with unknown noise. IEEE Transactions on Knowledge and Data Engineering, 35(1), 31-44.

Han, S., & Kim, H. (2021). Optimal feature set size in random forest regression. Applied Sciences, 11(8), 3428.

Hancock, J., Khoshgoftaar, T. M., & Johnson, J. M. (2022, December). Informative evaluation metrics for highly imbalanced big data classification. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1419-1426). IEEE.

Herbe,A., Estermann, Z., Holzwarth, V., & vom Brocke, J. (2024). How to effectively use distributed ledger technology in supply chain management?. International Journal of Production Research, 62(7), 2522-2547.

Hietasari, D. N., Subianto, D. I., & Adi, T. W. (2024, January). Conceptual framework of warehouse management system with auto suggesting features for FIFO/FEFO implementation towards lean warehousing. In AIP Conference Proceedings (Vol. 2951, No. 1). AIP Publishing.

Hinterreiter, A., Ruch, P., Stitz, H., Ennemoser, M., Bernard, J., Strobelt, H., & Streit, M. (2020). ConfusionFlow: A model-agnostic visualization for temporal analysis of classifier confusion. IEEE Transactions on Visualization and Computer Graphics, 28(2), 1222-1236.

Hu, X., & Weisel, E. (2020). Inbound Logistics Optimization.

Indriyani, S. (2020). Analyzing the Warehouse Management System at Pt. Pos Manado. Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi, 8(4).

Iparragirre, A., Lumley, T., Barrio, I., & Arostegui, I. (2023). Variable selection with LASSO regression for complex survey data. Stat, 12(1), e578.

Janitza, S., & Hornung, R. (2018). On the overestimation of random forest's out-of-bag error. PloS one, 13(8), e0201904.

Kaliappan, J., Srinivasan, K., Mian Qaisar, S., Sundararajan, K., & Chang, C. Y. (2021). Performance evaluation of regression models for the prediction of the COVID-19 reproduction rate. Frontiers in Public Health, 9, 729795.

Kara, K., Yalçın, G. C., & Edinsel, S. (2023). Warehouse manager selection by CRITIC-MULTIMOORA hybrid method based on single-valued neutrosophic sets. Journal of Maritime Transport and Logistics, 4(1), 48-64.

Karim, N. H., Abdul Rahman, N. S. F., Md Hanafiah, R., Abdul Hamid, S., Ismail, A., Abd Kader, A. S., & Muda, M. S. (2021). Revising the warehouse productivity measurement indicators: ratio-based benchmark. Maritime Business Review, 6(1), 49-71.

Kayanan, M., & Wijekoon, P. (2020). Stochastic Restricted LASSO-Type Estimator in the Linear Regression Model. Journal of Probability and Statistics, 2020(1), 7352097.

Khatoon, S., Zhengliang, X., & Hussain, H. (2020). The Mediating Effect of customer satisfaction on the relationship between Electronic banking service quality and customer Purchase intention: Evidence from the Qatar banking sector. Sage Open, 10(2), 2158244020935887.

Kim, B., & Kim, J. (2020). Adjusting decision boundary for class imbalanced learning. IEEE Access, 8, 81674-81685.

Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label classifier performance evaluation with confusion matrix. Computer Science & Information Technology, 1, 1-14.

Kumar, A., & Asthana, S. (2023). From Godown to Warehouse Management: Transformation at eKart, An Illustrative Case Study. Jindal Journal of Business Research, 12(1), 96-110.

Langer, S. (2021). Approximating smooth functions by deep neural networks with sigmoid activation function. Journal of Multivariate Analysis, 182, 104696.

Laskowski, J. F., & Tomiło, P. (2023). A new AI-based method for clustering survey responses. Journal of Modern Science, 54(5), 355-377. Aksan, F., Jasiński, M., Sikorski, T., Kaczorowska, D., Rezmer, J., Suresh, V., ... & Janik, P. (2021). Clustering methods for power quality measurements in virtual power plant. Energies, 14(18), 5902.

Leevy, J. L., Khoshgoftaar, T. M., & Hancock, J. (2022, October). Evaluating performance metrics for credit card fraud classification. In 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 1336-1341). IEEE.

Li, L., Pan, Y., Ma, X., & Liu, D. (2024, April). Simulation Research on the Inbound and Outbound Process of ZTO Distribution Center Warehousing System Based on Flexsim. In The International Conference on Artificial Intelligence and Logistics Engineering (pp. 281-289). Cham: Springer Nature Switzerland.

Li, Z., & Wei, W. (2022). Modular design for optimum granularity with auto-generated DSM and improved elbow assessment method. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 236(4), 413-426.

Li, Z., Nie, F., Wu, D., Wang, Z., & Li, X. (2023). Sparse trace ratio LDA for supervised feature selection. IEEE transactions on cybernetics, 54(4), 2420-2433.

Lim, Y., Lee, J., & Kim, H. (2021). Customer Satisfaction and Implied Cost of Equity: Moderating Effects of Product Market Conditions and Chief Marketing Officer.

Liu, G. (2022). A new index for clustering evaluation based on density estimation. arXiv preprint arXiv:2207.01294.

Liu, H., Lobschat, L., Verhoef, P. C., & Zhao, H. (2021). The effect of permanent product discounts and order coupons on purchase incidence, purchase quantity, and spending. Journal of Retailing, 97(3), 377-393.

Liu, J., Cai, X., & Niranjan, M. (2023). GO-LDA: Generalised optimal linear discriminant analysis. arXiv preprint arXiv:2305.14568.

Liu, L., & Yang, X. (2022). A multi-objective model and algorithms of aggregate production planning of multi-product with early and late delivery. Algorithms, 15(6), 182.

Loecher, M. (2022). Unbiased variable importance for random forests. Communications in Statistics-Theory and Methods, 51(5), 1413-1425.

Luo, H., Ma, S., Rao, J., Zheng, Z., Li, M., & Xu, H. (2020, November). Research on the influence of logistics service quality on post-purchase behavior in e-business. In 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT) (pp. 31-34). IEEE.

Luque, A., Mazzoleni, M., Carrasco, A., & Ferramosca, A. (2022). Visualizing Classification Results: Confusion Star and Confusion Gear.

Ma, Y., & Maciejewski, R. (2020). Visual analysis of class separations with locally linear segments. IEEE Transactions on Visualization and Computer Graphics, 27(1), 241-253.

Machkour, J., Muma, M., Alt, B., & Zoubir, A. M. (2020). A robust adaptive Lasso estimator for the independent contamination model. Signal Processing, 174, 107608.

Maheshwari, P., Kamble, S., Kumar, S., Belhadi, A., & Gupta, S. (2023). Digital twin-based warehouse management system: a theoretical toolbox for future research and applications. The International Journal of Logistics Management, (ahead-of-print).

Mane, S., Bhuva, J., & Patel, S. (2020). Implementation of 5S to Set Up Inventory Control System with HTML Coded Spare Management System. In Proceedings of International Conference on Intelligent Manufacturing and Automation: ICIMA 2020 (pp. 327-336). Springer Singapore.

Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, 281(3), 588-596.

Mirza, B., Haroon, D., Khan, B., Padhani, A., & Syed, T. Q. (2021). Deep generative models to counter class imbalance: A model-metric mapping with proportion calibration methodology. IEEE Access, 9, 55879-55897.

Mulindwa, D. B., & Du, S. (2023). An n-Sigmoid Activation Function to Improve the Squeeze-and-Excitation for 2D and 3D Deep Networks.

Münsberg, T. M., Hvam, L., & Tsintzou, L. (2022). Optimizing the Integration Process of New Customers in a 3PL Warehouse. In 10th International Conference on Mass Customization andPersonalization–Community of Europe (pp. 96-101). University of Novi Sad.

Naboureh, A., Ebrahimy, H., Azadbakht, M., Bian, J., & Amani, M. (2020). RUESVMs: An ensemble method to handle the class imbalance problem in land cover mapping using Google Earth Engine. Remote Sensing, 12(21), 3484.

Namdar, K., Haider, M. A., & Khalvati, F. (2021). A modified AUC for training convolutional neural networks: taking confidence into account. Frontiers in artificial intelligence, 4, 582928.

Nedel'Ko Victor, M. (2023). On the Properties of Bias-Variance Decomposition for kNN Regression. Известия Иркутского государственного университета. Серия: Математика, 43, 110-121.

Nguyen, D. T., Pham, V. T., Tran, D. M., & Pham, D. B. T. (2020). Impact of service quality, customer satisfaction and switching costs on customer loyalty. The Journal of Asian Finance, Economics and Business, 7(8), 395-405.

Niemi, T., Hameri, A. P., Kolesnyk, P., & Appelqvist, P. (2020). What is the value of delivering on time?. Journal of Advances in Management Research, 17(4), 473-503.

Nissa, N., Jamwal, S., & Neshat, M. (2024). A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques. Computation, 12(1), 15.

Obi, J. C. (2023). A comparative study of several classification metrics and their performances on data. World Journal of Advanced Engineering Technology and Sciences, 8(1), 308-314.

Odunjo, F. O. O. (2020). A study of the impact of logistic inbound and outbound operations in organizational performance at Dangote Cement Industries (Doctoral dissertation, Dublin, National College of Ireland).

Oliveira, A. A. A. D. M., Espadoto, M., Hirata Júnior, R., & Telea, A. C. (2022). SDBM: supervised decision boundary maps for machine learning classifiers. In Proceedings.

Omer, P. A. (2022). Improving Prediction Accuracy of Lasso and Ridge Regression as an Alternative to LS Regression to Identify Variable Selection Problems. DOI: https://doi. org/10.31972/ticma2022.

Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset. Applied Sciences, 12(15), 7515.

Osojnik, A., Džeroski, S., & Kocev, D. (2016). Option predictive clustering trees for multi-target regression. In Discovery Science: 19th International Conference, DS 2016, Bari, Italy, October 19–21, 2016, Proceedings 19 (pp. 118-133). Springer International Publishing.

Ouma, Y., Nkwae, B., Moalafhi, D., Odirile, P., Parida, B., Anderson, G., & Qi, J. (2022). Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 681-689.

Paranjape, P. N., Dhabu, M. M., & Deshpande, P. (2021). dSubSign: Classification of Instance-Feature Data Using Discriminative Subgraphs as Class Signatures. International Journal of Software Engineering and Knowledge Engineering, 31, 917-947

Pawara, P., Okafor, E., Groefsema, M., He, S., Schomaker, L. R., & Wiering, M. A. (2020). One-vs-One classification for deep neural networks. Pattern Recognition, 108, 107528.

Pendrill, L. R., Melin, J., Stavelin, A., & Nordin, G. (2023). Modernising Receiver Operating Characteristic (ROC) Curves. Algorithms, 16(5), 253.

Pereira, M. T., Santos, E., Sá, J. C., Oliveira, M., & Ferreira, L. P. (2022, June). Improvements in the Warehouse of a Danish Transport and Logistics Company Through a WMS Implementation. In International Conference on Flexible Automation and Intelligent Manufacturing (pp. 325-336). Cham: Springer International Publishing.

Phan, M. H., & Tran, A. T. (2022). Development a warehouse management information system. Applied Mechanics and Materials, 907, 131-143.

Plevris, V., Solorzano, G., Bakas, N. P., & Ben Seghier, M. E. A. (2022, November). Investigation of performance metrics in regression analysis and machine learning-based prediction models. In 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022). European Community on Computational Methods in Applied Sciences.

Polim, D. N., & Lestari, Y. D. (2023). Business process reengineering to excellence warehouse management system: a case study of retail industry. Jurnal Syntax Transformation, 4(1), 92-109.

Praptodiyono, S., Santoso, M. I., Sukarna, R. H., Muhammad, F., & Fuad, A. (2023, June). The design of an RFID-Based warehouse management system for AgroHUB banten. In AIP Conference Proceedings (Vol. 2760, No. 1). AIP Publishing.

Pratiwi, H., Windarto, A. P., Susliansyah, S., Aria, R. R., Susilowati, S., Rahayu, L. K., ... & Rahadjeng, I. R. (2020, February). Sigmoid activation function in selecting the best model of artificial neural networks. In Journal of Physics: Conference Series (Vol. 1471, No. 1, p. 012010). IOP Publishing.

Qin, Y., Zhaofa, Y., Xuzheng, L., Zufang, Z., Weijie, C., & Sheng, R. (2021). Product module attribute parameter configuration model considering customer requirements preferences. Mathematical Problems in Engineering, 2021(1), 6632057.

Rahman, M., & Asadujjaman, M. (2021, September). Implementation of artificial neural network on regression analysis. In 2021 5th Annual Systems Modelling Conference (SMC) (pp. 1-7). IEEE.

Rector, L. R., & Scott, B. P. (2021). Hoffman Inbound Inventory Management System Improvement.

Ricardianto, P., Kurniawan, I., Ikawati, I., Gutomo, T., Sijabat, E., Mahadita, A., ... & Endri, E. (2023). Service quality and timeliness: Empirical evidence on the parcel delivery service in Indonesia. Uncertain Supply Chain Management, 11(4), 1645-1656.

Ridho, F., Hartono, S., & Istiqomah, I. (2021). Easy, quality of service, and timeliness of delivery to customer satisfaction. Almana: Jurnal Manajemen dan Bisnis, 5(2), 184-191.

Riyanto, S., Imas, S. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. International Journal of Advanced Computer Science and Applications, 14(6).

Ruisánchez, I., Jiménez-Carvelo, A. M., & Callao, M. P. (2021). ROC curves for the optimization of one-class model parameters. A case study: Authenticating extra virgin olive oil from a Catalan protected designation of origin. Talanta, 222, 121564.

Salekshahrezaee, Z., Leevy, J. L., & Khoshgoftaar, T. M. (2022, August). A class-imbalanced study with feature extraction via pca and convolutional autoencoder. In 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 63-68). IEEE.

Satish, R., Manoj, K., & Vp, S. (2023). A study on warehouse management and material handling process at polkart logistics. Russian Law Journal, 11(11S), 16-22.

Sharma, A., Pratap, A., Vyas, K., & Mishra, S. (2022, December). Machine learning approach: Consumer buying behavior analysis. In 2022 IEEE Pune Section International Conference (PuneCon) (pp. 1-10). IEEE.

Sharma, D. R., & Singh, B. (2023). Understanding the relationship between customer satisfaction, customer engagement and repeat purchase behaviour. Vision, 27(4), 449-457.

Shu, Q., Hu, T., & Liu, S. (2020, May). Random Forest Algorithm based on GAN for imbalanced data classification. In Journal of Physics: Conference Series (Vol. 1544, No. 1, p. 012014). IOP Publishing.

Sitarz, M. (2022). Extending F1 metric, probabilistic approach. arXiv preprint arXiv:2210.11997.

Song, Y., & Hua, X. (2022). Implementation of data mining technology in bonded warehouse inbound and outbound goods Trade. Journal of Organizational and End User Computing (JOEUC), 34(3), 1-18.

Stepišnik, T., Osojnik, A., Džeroski, S., & Kocev, D. (2020). Option predictive clustering trees for multi-target regression. Computer Science and Information Systems, 17(2), 459-486.

Sugriyono, S., & Siregar, M. U. (2020). Preprocessing kNN algorithm classification using K-means and distance matrix with students' academic performance dataset.

Suh, A., Appleby, G., Anderson, E. W., Finelli, L., & Cashman, D. (2021, October). Communicating Performance of Regression Models Using Visualization in Pharmacovigilance. In 2021 IEEE Workshop on Visual Analytics in Healthcare (VAHC) (pp. 6-13). IEEE.

Sun, P. Z., Zuo, T. Y., Law, R., Wu, E. Q., & Song, A. (2023). Neural network ensemble with evolutionary algorithm for highly imbalanced classification. IEEE Transactions on Emerging Topics in Computational Intelligence, 7(5), 1394-1404.

Supriyanto, A., Sutrisna, S., & Listen, G. (2021). The Relation of Price, Product Quality And Brand Image To Customer Loyalty In Manufacturing Companies.

Takada, M., Suzuki, T., & Fujisawa, H. (2020). Independently Interpretable Lasso for Generalized Linear Models. Neural computation, 32(6), 1168-1221.

Thölke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., ... & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. NeuroImage, 277, 120253.

Tigga, O., Pal, J., & Mustafi, D. (2023, February). A Comparative Study of Multiple Linear Regression and K Nearest Neighbours using Machine Learning. In 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-5). IEEE.

Troebs, C. C., Wagner, T., & Herzog, W. (2021). Do Customer Discounts Affect Frontline Employees?. Journal of Service Research, 24(3), 390-404.

van Dongen, S. (2022). Fast multi-resolution consensus clustering. bioRxiv, 2022-10.

Verbakel, J. Y., Steyerberg, E. W., Uno, H., De Cock, B., Wynants, L., Collins, G. S., & Van Calster, B. (2020). ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. Journal of Clinical Epidemiology, 126, 207-216.

Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in imbalanced data classification. Knowledge-based systems, 212, 106631.

Wahyuni, D., Sembiring, M. T., Budiman, I., & Hutagalung, T. R. (2021, March). Modelling Operation of Logistics Management in Modern Market: A Case Study in Indonesian Retail Company. In IOP Conference Series: Earth and Environmental Science (Vol. 709, No. 1, p. 012052). IOP Publishing.

Wei, B. (2023). A brand-new clustering method and analysis system designed for revealing the truth of the high-dimension large data deciphered the complex composition structure of human brain endothelial cells from single-cell RNA sequence data. bioRxiv, 2023-06.

Wong, P., Thant, P., Yadav, P., Antaliya, R., & Woo, J. (2023). Using Spark Machine Learning Models to Perform Predictive Analysis on Flight Ticket Pricing Data. arXiv preprint arXiv:2310.07787.

Wu, J., Lin, X., Yang, X., Li, S., Zhang, B., & Gao, L. (2023, September). Research on random forest algorithm based on oversampling and feature selection. In 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC) (Vol. 7, pp. 1531-1537). IEEE.

Xu, H. (2022). Prediction of Students' Performance Based on the Hybrid IDA-SVR Model. Complexity, 2022(1), 1845571.

Yan, C., Chang, X., Luo, M., Zheng, Q., Zhang, X., Li, Z., & Nie, F. (2020). Self-weighted robust LDA for multiclass classification with edge classes. ACM Transactions on Intelligent Systems and Technology (TIST), 12(1), 1-19.

Yanchenko, A. Y., & Repilova, M. M. (2023). On the question of applying discounts in the marketing process. Entrepreneurâ€™ s Guide, 16(4).

Yaqub, R. M. S., Akram, H. M. Z., Ahmed, N., & Ali, M. Q. (2023). Moderating Role of Worth of Win-Back Offer on the Relationship of Post Purchase Regret, Social Capital, and Prior Experience with Switch Back Intentions among Prepaid Mobile Subscribers. Pakistan Journal of Humanities and Social Sciences, 11(2), 1013-1026.

Yildiz, B., & Savelsbergh, M. (2020). Pricing for delivery time flexibility. Transportation Research Part B: Methodological, 133, 230-256.

Zhang, G., & Gionis, A. (2023). Regularized impurity reduction: accurate decision trees with complexity guarantees. Data mining and knowledge discovery, 37(1), 434-475.

Zhao, L., Lee, S., & Jeong, S. P. (2021). Decision tree application to classification problems with boosting algorithm. Electronics, 10(16), 1903.

Zhao, M., Ma, Z., Zhou, Z., & Zheng, K. (2021, October). Wspeed: drone energy optimization for multiple-package delivery considering weight changes. In 2021 International Conference on Space-Air-Ground Computing (SAGC) (pp. 1-6). IEEE.

Zorarpacı, E. (2021, June). A hybrid dimension reduction based linear discriminant analysis for classification of high-dimensional data. In 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 1028-1036). IEEE.