JOURNAL OF MODERN SCIENCE
SPECIAL ISSUE
3/57/2024
www.jomswsge.com

DOI: doi.org/10.13166/jms/191201

ZBIGNIEW SZYMAŃSKI
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0009-0009-0005-714X

MICHAŁ MAJ
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0002-7604-8559

JACEK PIWKOWSKI
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0009-0001-8095-5776

EWELINA JURCZAK
Graduate School of Business - National-Louis University, Poland
ORCID iD: orcid.org/0000-0001-6610-429X

TOMASZ CIEPLAK
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0002-2712-6098

DAMIAN PLISZCZUK
Netrix S.A., Poland
ORCID iD: orcid.org/0000-0002-5727-979X

OPTIMIZING DELIVERY TIME WITH AN INTELLIGENT FORECASTING MODEL: LEVERAGING AI AND MACHINE LEARNING FOR EFFICIENT LOGISTICS

OPTYMALIZACJA CZASU DOSTAWY ZA POMOCĄ INTELIGENTNEGO MODELU PROGNOZOWANIA: WYKORZYSTANIE SZTUCZNEJ INTELIGENCJI I UCZENIA MASZYNOWEGO W CELU ZAPEWNIENIA WYDAJNEJ LOGISTYKI

Abstract

Business analytics involves using various technologies to analyze data. Data mining focuses on the automated search for knowledge, patterns, or regularities in data. As a business analyst, it is essential to recognize the type of analytical technique appropriate for solving a specific problem. Exploratory Data Analysis (EDA) describes data using statistical and visualization techniques to highlight important aspects of that data for further analysis. This involves examining a data set from many angles, describing it, and summarizing it without making assumptions about its content. Exploratory data analysis is an essential step before diving into statistical modeling or machine learning to ensure that the data is really what it claims to be and that there are no apparent errors. This type of analysis should be part of data science projects in every organization. Visual analytics is sometimes confused with data visualization. Visual analysis is not simply a matter of graphically representing data. Modern, interactive visual analytics makes combining data from multiple sources easy and performs in-depth data analysis directly in the visualization. Additionally, artificial intelligence and machine learning algorithms can offer recommendations for exploration. Ultimately, visual analytics helps transform massive data sets into business insights that can positively impact an organization. Considering the previous comment about visual data analysis, it should be added that the system has extensive capabilities to create graphical dashboards containing reports and charts. It is essential that in the case of desktops, in addition to the visualizations included in the system itself, it is possible to embed reports from third-party tools. This approach makes it possible to combine existing reports and use other specialized tools to develop a coherent dashboard.: 900-1800 characters with spaces in English, below in the language of the article.

Streszczenie

Analityka biznesowa polega na zastosowaniu różnych technologii do analizy danych. Eksploracja danych koncentruje się na zautomatyzowanym poszukiwaniu w danych wiedzy, wzorców czy też prawidłowości. Dla analityka biznesowego istotna jest umiejętność rozpoznawania, jakiego rodzaju technika analityczna jest odpowiednia do rozwiązania konkretnego problemu. Eksploracyjna Analiza Danych (EDA) to proces opisywania danych za pomocą technik statystycznych i wizualizacyjnych w celu zwrócenia uwagi na ważne aspekty tych danych do dalszej analizy. Wiąże się to z badaniem zbioru danych pod wieloma kątami, opisywaniem go i podsumowywaniem bez żadnych założeń dotyczących jego zawartości. Eksploracyjna analiza danych jest ważnym krokiem do przed zanurzeniem się w modelowanie statystyczne lub uczenie maszynowe, aby upewnić się, że dane są naprawdę takie, za jakie się podają, i że nie ma oczywistych błędów. Tego rodzaju analiza powinna być częścią projektów data science w każdej organizacji. Analiza wizualna jest czasami mylona z wizualizacją danych. Analiza wizualna nie jest po prostu kwestią graficznego przedstawiania danych. Nowoczesna, interaktywna analityka wizualna ułatwia łączenie danych z wielu źródeł i dogłębną analizę danych bezpośrednio w samej wizualizacji. Ponadto algorytmy sztucznej inteligencji i uczenia maszynowego mogą oferować zalecenia, które pomogą w prowadzeniu eksploracji. Ostatecznie analiza wizualna pomaga przekształcić ogromne zbiory danych w spostrzeżenia biznesowe, które mogą mieć ogromny pozytywny wpływ na organizację. Biorąc pod uwagę poprzednią uwagę o analizę wizualnej danych, należy dopowiedzieć, że system posiada szerokie możliwości tworzenie pulpitów graficznych zawierających raporty oraz wykresy. Istotny jest fakt, że w przypadku pulpitów oprócz wizualizacji zawartych w samym systemie istnieje możliwość osadzania raportów z narzędzi firm trzecich. Zastosowanie takiego podejścia daje możliwości łączenia już istniejących raportów oraz możliwości wykorzystania innych specjalistycznych narzędzi do opracowania spójnego pulpitu.900-1800 znaków ze spacjami; w języku angielskim, poniżej w języku artykułu.

Keywords: *Data science, delivery time, CRM, artificial intelligence, machine learning*

Słowa kluczowe: *Badanie danych, czas dostaw, CRM, sztuczna inteligencja, uczenie maszynowe*

Introduction

Companies must pioneer ways to differentiate themselves from other market participants in today's complex business world by becoming more cooperative, efficient, precise, and flexible. They must be able to respond quickly to market needs and changes (Cieplak et al., 2021). Depending on the company's competitive advantage, which may be newness, price, excellent website content, or social media presence, specific online strategies must be used to reach the desired market. Many companies have noticed that the data they store and use in some way can build a market advantage. Data and information are becoming essential resources for many organizations (Dmowski, Wołowiec, Laskowska, Laskowski, 2023).

The customer profile on the market is constantly changing. This contributes to changing the business context and the conditions in which enterprises compete. Buyers are increasingly tired of the abundance of offers and are less and less loyal to the brand. They know perfectly well what other people think about products and services, and they constantly increase their requirements (Gołabek et al., 2023).

Measuring customer satisfaction, loyalty, contentment and value generates indicators critical to business results (such as profit or market share) (Sutanto et al., 2023). Therefore, the condition for the company to prosper is to meet the customer's expectations, i.e., adapting products to his expectations, taking care of the relationship with him before and after the transaction, and making every effort to anticipate his needs earlier than the competition. In addition to customer satisfaction, it is essential to know their opinions, attitudes, and preferences (the changes contributing to market transformations). Preference research allows us to determine the customer's profile, retain him in the company, and improve the quality of services and the company's efficiency (Zhao et al., 2023). *Knowing the profile of a customer (or a group of customers) means knowing, approximately, how he or she makes purchase decisions and being able to predict how he or she will react to sellers' proposals and behavior.*

Research on consumer preferences and needs answers several questions:

- How often does a customer buy a product from a given area, and where do they make purchases?
- What features are the priorities when purchasing and choosing a brand? How does the customer compare the features of different brands?
- How does the customer assess product availability?
- What does the customer think about the available sales channels? Which ones do they prefer, and why?
- Are your purchases spontaneous or planned?
- How, if at all, does the customer obtain information about products and services?

Correctly performed customer classification helps discover data characteristics and generalize or organize data consistent with the assumed knowledge-oriented structures. Data classification is an introduction to data analysis, and clustering algorithms are used to:

- data mining, including customer grouping,
- information extraction (document retrieval), i.e., simplifying access to information,
- image segmentation, including image division according to its specific properties, object, and character recognition.

However, in decision analysis, inconsistent information about decision-making situations is often encountered. Decision-makers often hesitate, express their preferential model imprecisely, and are unstable, resulting in the possibility of obtaining incomplete or uncertain information.

The built system allows for the analysis of distributed data and, based on it, research on customer behavior, business process efficiency, and marketing activities. This type of analysis results in the creation and availability of new, improved procedures and solutions that provide insight into the company's operation, which is the subject of the implementation.

Research Methodology

The first step will be to become familiar with existing delivery time forecasting methods, including both traditional and AI/ML-based ones (Saura, 2021). Delivery data will then be collected, including the order number, dates and times of submission, shipment and delivery, sender and recipient locations, product type and size, and courier company. This data will be obtained from courier companies' systems, parcel tracking websites, and customer surveys. The collected data will be prepared for analysis through cleaning, outlier removal, and normalization. Various AI/ML models, such as linear regression, decision trees, random forests and neural networks, will be used to predict delivery times (Hancock et al., 2023). The results will be compared to identify the best model. The best model will be validated on the test data set, and its accuracy will be assessed using metrics such as MAE and MSE (Akbarifard et al., 2021). The developed model will be implemented in a natural system, which will be used to predict delivery times for new orders.

Data analysis and delivery delays

There are many tools to provide access to data. These tools are typically SQL-based (Structured Query Language) or graphical user interface (GUI) tools that help formulate queries (e.g., Query By Example – QBE) (Zhang et al., 2023). For example, if an analyst can define something as *profitable* in operational terms that can be calculated from database entries, the query tool could answer the question: *Who are the most profitable customers in Warsaw?* The analyst can then run a query and receive a list of the most profitable customers ranked by profitability. This activity fundamentally differs from data mining because it does not involve discovering patterns or models.

Data warehouses collect and combine data from across the company, often from multiple transaction processing systems, each with its database. Analytical systems have access to data stores. Data warehousing can be viewed as a technology that supports data mining. Data warehousing is not always necessary because most data mining activities do not require access to data warehouses. Still, companies that invest in building data warehouses often use data mining much more widely within the organization. For example, if your data warehouse integrates records from your sales and billing departments and your human resources department, you can use it to find patterns that characterize successful salespeople (Saddad et al., 2020).

Customer behavior refers to an individual's purchasing habits, including social trends, frequency patterns, and background factors influencing their purchasing decision. Companies study customer behavior to understand their target audience and create more attractive products and service offerings. Customer behavior analysis is the qualitative and quantitative observation of customers' interactions with the company. Customers are first divided into buyer classes based on their common characteristics. Each group is then observed at different stages of the customer journey map to analyze how people interact with the company. Analyzing customer behavior provides insight into the various variables that influence your audience. It gives you an idea of the motives, priorities, and decision-making methods considered during the customer journey. This analysis helps you understand customers' feelings about your company and whether those perceptions align with their core values. Internet technologies implemented in e-commerce systems for many years provide excellent opportunities for tracking and analyzing customer behavior (Hendricks & Mwapwele, 2024).

The first step in analyzing customer behavior is categorizing your customer base. When doing this, it is important to use a wide range of features. Demographic characteristics such as gender, age, and location should be considered, and engagement trends such as online activity, preferred media channels, and online shopping habits should also be considered.

The most important thing is to determine the characteristics of customers who are most valuable to the company. One way to do this is to perform an RFM analysis, which determines how recently and often a customer is purchasing from a given company (Stormi et al., 2020). Another way is to calculate the customer's long-term value. Customer lifetime value takes metrics such as customer lifetime, purchase value, and frequency and determines how much revenue a company can expect from that customer. This information provides a quantitative picture of loyal customers' influence on a company.

Sales forecasting allows companies to act reasonably in advance; among other things, it supports decisions in planning marketing activities and determines the time of introducing and withdrawing individual products from the market. The forecasting result may be the sum of sales divided into product groups or distribution channels. Suppose a relationship exists between sales and the number of products or retail outlets. In that case, we will use the linear regression method to calculate the expected sales volume, which is the simplest way to make forecasting possible. However, when determining the sales volume depends on a more significant number of variables, we will use the multiple regression method for analysis (Helo & Hao, 2022).

The analysis results from the output data presented in charts generated by the implemented system. Forecasting affects the activities of all company areas and supports decision-making in determining the number of purchases, production planning, logistics, sales implementation, marketing campaigns, and after-sales services. If the forecasts significantly exceed the actual implementation, there may be an unnecessary freezing of cash, increased storage costs, price reductions, and business profitability. Underestimating forecasts, however, results in shortages of goods in warehouses, delays in logistics, and untimely service, which in turn causes customer dissatisfaction (Kalghatgi, 2023).

Forecasting is based on historical data, which may be data from sales systems, but also analyzes the effectiveness and efficiency of other systems, which, when monitored, will generate data describing the operation or its course (Cadden et al., 2022). Global shipping volumes are increasing, and virtually all consumers recognize that delivery times affect their loyalty to specific companies. Although these statistics mainly pertain to B2C trends, B2B supply chain and logistics increasingly mirror B2C expectations, with enterprise customers seeking an Amazon-like experience no matter where they are. Nevertheless, delayed shipments are inevitable, and managing complex logistics remains challenging. Yet, there's a growing emphasis among logistics providers and supply chain participants on minimizing disruptions and improving on-time deliveries. Understanding the nature of delays and strategies to mitigate them is crucial to achieving these improvements (Han et al., 2023).

The primary challenges in shipping are often beyond the control of planners. Major weather events like hurricanes, tropical storms, or even minor disturbances like heavy rain or wind can quickly interfere with shipping routes. Similarly, road accidents can lead to traffic jams that disrupt truck schedules, potentially causing a chain reaction of missed transfers and lengthy delays. While you can't control the weather, you can manage your anticipation and response. If logistics workflows consistently follow static shipping routes, they may lack the flexibility needed in case of unexpected disruptions. However, simulating alternative routes for each new order based on various parameters and forecasts can allow replanning whenever a disruption is expected. This proactive approach enables supply chain managers to mitigate the impact of uncontrollable weather events (Osborn & Nault, 2012).

Logistics chains need to be highly visible from carriers and 3PLs down to the truck and container levels to implement the above repackaging and analysis effectively. Advanced analytics and digital twins can be used for ad hoc shipment optimization. This visibility allows for faster route planning and more precise tracking, thus preventing another common cause of delays: lost orders (Bag et al., 2023). In essence, timely deliveries can't happen if you don't know where shipments are. Logistics can be more efficiently managed by incorporating

IoT devices and RFID chips into the value chain, coupled with supply chain management software that enables item tracking (Tan & Sidhu, 2022).

Logistics errors can often lead to missing or mistakenly being shipped to the wrong destination. This typically happens due to human error, and rectifying such mistakes can be complex. If a shipment is wholly lost, restarting the order fulfillment process is the most straightforward approach. However, if its location is known but it's heading in the wrong direction, swift action is needed to get it back on track (Maj et al., 2023). Addressing this issue requires a mix of strategies: real-time planning analytics for visibility into different options and meticulous tracking throughout the supply chain. This way, you can quickly explore ways to rectify errors, like redirecting shipments to the nearest hub for consolidation or restarting the order while the consignment is rerouted. Harnessing supply chain transparency will determine whether delays can be effectively managed (Maheshwari et al., 2023).

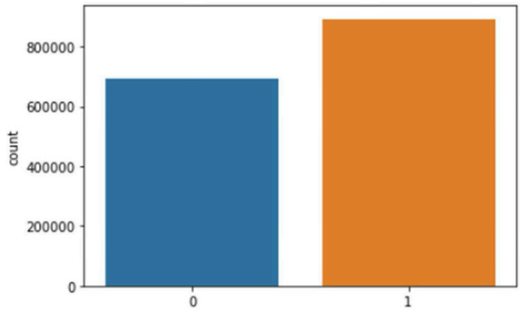
Development of a module for forecasting goods delivery times

The analyzed data set contains 1,595,196 observations. There are no missing values or duplicates in the dataset. The exact order document number may appear in the database several times, with differences in the goods, values, etc. This occurs when the available part of the goods is issued first, and other goods are issued later, after delivery. Based on the repeated numbers of order documents, a binary variable, partial, was created containing information about the partial execution of the order:

- 1 – a given order is partially implemented (in several sub-orders)
- 0 – the order is fully processed (entirely)

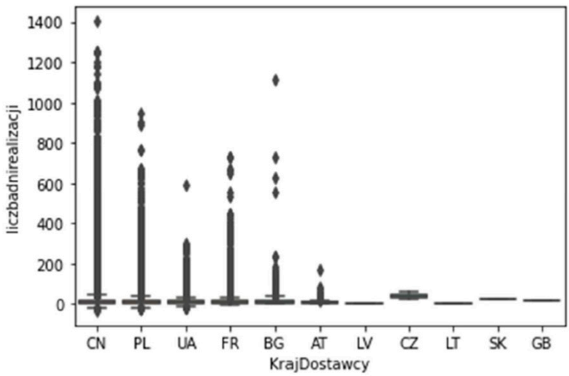
There are slightly more orders fulfilled partially, but considering the size of the entire collection, this is not a big difference compared to orders fulfilled in full.

Figure 1. *Division of orders into fully completed (0) and partially completed (1)*



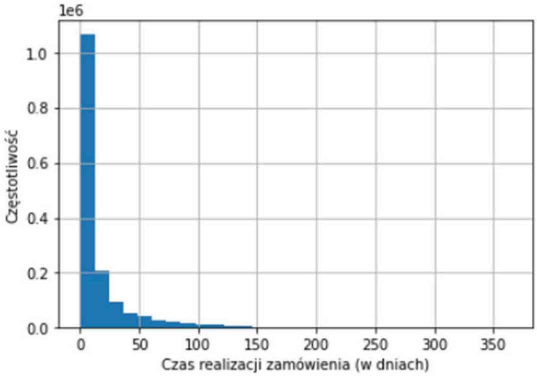
The box plot of the number of days of execution by supplier's country (Figure 2) shows outliers in the number of days of order execution (there are values even above 1000 days and negative ones). Observations for which the lead time was non-negative or exceeded 365 days were selected for further analysis. This results in a table of size (1591471, 16).

Figure 2. *Boxplots of the number of lead days by Supplier Country*



From the histogram of the number of execution days variable (Figure 3) and based on the statistics, in particular, the mean and median, where the mean is much larger than the median, we conclude that there is a significant right-sided asymmetry in the distribution of the number of order execution days. Most orders are completed within a short time. The median turnaround time is seven days, and the average is approximately 18 days – similar to the third quartile. Only 25% of observations concerned orders were completed over 18 days. The histogram shows very few orders with execution time exceeding 100 days, and orders exceeding 150 days are practically unnoticeable on the chart.

Figure 3. *Histogram of the number of days of order fulfillment*



Conclusions

The development of an intelligent goods delivery time model can significantly impact various stakeholder groups and contribute to the development of knowledge in the field of forecasting. AI allows you to analyze vast amounts of data and identify complex patterns that may be invisible to traditional forecasting methods. This enables the Intelligent to Delivery Time Model to consider many factors that impact delivery times, leading to more accurate estimates. Accurate delivery time forecasts allow courier companies to optimize routes, plan resources, and manage warehouses more efficiently. This can lead to faster delivery times, reduced costs, and improved customer satisfaction. AI can be used to adapt delivery time forecasts to individual customer needs. For example, customers can receive real-time updates on the status of their orders or personalized notifications about potential delays. AI enables analysis of delivery time data to identify trends and areas for improvement. This information can be used to optimize business processes, implement new strategies, and make better business decisions.

References

Akbarifard, S., Sharifi, M. R., Qaderi, K., Madadi, M. R. (2021). Optimal operation of multi-reservoir systems: Comparative study of three robust metaheuristic algorithms. Water Science and Technology: Water Supply, 21(2). https://doi.org/10.2166/ws.2020.368

Bag, S., Dhamija, P., Singh, R. K., Rahman, M. S., Sreedharan, V. R. (2023). Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain: An empirical study. *Journal of Business Research*, *154*. https://doi.org/10.1016/j.jbusres.2022.113315

Cadden, T., Dennehy, D., Mantymaki, M., Treacy, R. (2022). Understanding the influential and mediating role of cultural enablers of AI integration to supply chain. *International Journal of Production Research*, *60*(14). https://doi.org/10.1080/00207543.2021.1946614

Cieplak, T., Rymarczyk, T., Kłosowski, G., Maj, M., Pliszczuk, D., Rymarczyk, P. (2021). Methods of process mining and prediction using deep learning. *Przeglad Elektrotechniczny*, *97*(3). https://doi.org/10.15199/48.2021.03.28

Dmowski, A., Wołowiec, T., Laskowski, J., Laskowska, A. 2023. Economic issue of material reserves management taking into account probabilistic model. Journal of Modern Science, 54(5), p. 395-409. https://doi.org/10.13166/jms/176382

Gołabek, Ł., Pliszczuk, D., Maj, M., Bogacki, S., Rzemieniak, M. (2023). Artificial intelligence in a distributed supply chain control model for personalizing and identifying products in real time. In *Innovation in the Digital Economy* (pp. 85–97). Routledge. https://doi.org/10.4324/9781003384311-8

Han, M., Yang, T., Zhong, J., Zhong, Y. (2023). AI applications and supply chain concentration. *Applied Economics Letters*. https://doi.org/10.1080/13504851.2023.2210813

Hancock, J. T., Bauder, R. A., Wang, H., Khoshgoftaar, T. M. (2023). Explainable machine learning models for Medicare fraud detection. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00821-5

Helo, P., Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: an exploratory case study. *Production Planning and Control*, *33*(16). https://doi.org/10.1080/09537287.2021.1882690

Hendricks, S., Mwapwele, S. D. (2024). A systematic literature review on the factors influencing e-commerce adoption in developing countries. *Data and Information Management*, *8*(1). https://doi.org/10.1016/j.dim.2023.100045

Kalghatgi, U. S. (2023). Creating Value for Reliability Centered Maintenance (RCM) in Ship Machinery Maintenance from BIG Data and Artificial Intelligence. *Journal of The Institution of Engineers (India): Series C*, *104*(2). https://doi.org/10.1007/s40032-022-00900-1

Maheshwari, P., Kamble, S., Belhadi, A., Venkatesh, M., Abedin, M. Z. (2023). Digital twin-driven real-time planning, monitoring, and controlling in food supply chains. *Technological Forecasting and Social Change*, *195*. https://doi.org/10.1016/j.techfore.2023.122799

Maj, M., Rymarczyk, T., Maciura, Ł., Cieplak, T., Pliszczuk, D. (2023, October). Cross-Modal Perception for Customer Service. *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. https://doi.org/10.1145/3570361.3615751

Osborn, B. E., Nault, B. R. (2012). A Classification of Supply Chain Problems. *Engineering Management Research*, *1*(2). https://doi.org/10.5539/emr.v1n2p1

Saddad, E., El-Bastawissy, A., Mokhtar, H. M. O., Hazman, M. (2020). Lake data warehouse architecture for big data solutions. *International Journal of Advanced Computer Science and Applications*, *11*(8). https://doi.org/10.14569/IJACSA.2020.0110854

Saura, J. R. (2021). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation and Knowledge*, *6*(2). https://doi.org/10.1016/j.jik.2020.08.001

Stormi, K., Lindholm, A., Laine, T., Korhonen, T. (2020). RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers. *Journal of Management and Governance*, *24*(3). https://doi.org/10.1007/s10997-018-9447-3

Sutanto, J. E., Harianto, E., Balkan, N. (2023). The effect of supply chain organizational and supply agility on supply chain performance: The mediation role of supply chain strategy in retail shops. *Uncertain Supply Chain Management*, *11*(1). https://doi.org/10.5267/j.uscm.2022.11.002

Tan, W. C., Sidhu, M. S. (2022). Review of RFID and IoT integration in supply chain management. *Operations Research Perspectives*, *9*. https://doi.org/10.1016/j.orp.2022.100229

Zhang, F., Ge, W., Huang, L., Li, D., Liu, L., Dong, Z., Xu, L., Ding, X., Zhang, C., Sun, Y., Jun, A., Gao, J., Guo, T. (2023). A Comparative Analysis of Data Analysis Tools for Data-Independent Acquisition Mass Spectrometry. *Molecular and Cellular Proteomics*, *22*(9). https://doi.org/10.1016/j.mcpro.2023.100623

Zhao, N., Hong, J., Lau, K. H. (2023). Impact of supply chain digitalization on supply chain resilience and performance: A multi-mediation model. *International Journal of Production Economics*, *259*. https://doi.org/10.1016/j.ijpe.2023.108817