Sales trend analysis via regression tree algorithm (case study: Veruby Store Pontianak)

Nengsy Verita[1*], Jimmy Tjen[2]

[1] Department of Digital Business, Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak, Indonesia.
[2] Department of Informatics, Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak, Indonesia.

*Email: [1] 22430172@widyadharma.ac.id, [2]jimmytjen@mathmods.eu

Abstract. This study aims to analyze sales trends in women's fashion stores in Pontianak using the regression tree method to predict the optimal timing for selling popular products. The data utilized comprises 1,200 transaction entries that record transaction dates, item names, categories, and sales amounts. The research focuses on the "tops" category, employing temporal variables such as day, date, and month as predictors to forecast sales quantities. The regression tree model was constructed using the fitrtree function, and the results were evaluated based on root means square error (RMSE) and normalized RMSE (NRMSE). The prediction results indicate an accuracy of 77.85%, with higher sales patterns observed on the 13th and specific days of the week, suggesting the influence of promotions or consumer shopping habits. This study provides important insights for inventory planning and promotional strategies in fashion stores, particularly in leveraging identified sales patterns to enhance efficiency and profitability. This analysis can be adapted for various product categories within the fashion industry to respond more timely to consumer demand.

Keywords: Regression tree, Fashion retail, Sales trends, Inventory planning, Promotional Strategies.

Introduction

The women's fashion retail sector in Pontianak is experiencing significant growth, with fashion serving as a medium for personal expression (Tripathi & Dubey, 2023). This expansion is fueled by a rising interest in contemporary clothing trends and styles. As the capital of West Kalimantan, Pontianak boasts a diverse array of fashion outlets that cater to various preferences, offering everything from casual and formal attire to modern Muslim fashion. The competitive landscape among these stores fosters innovation in product quality (Pei, Guo, Wu, Zhou, & Yeh, 2020), customer service, and overall shopping experiences. In response to increasing consumer demand, these fashion establishments exemplify the vibrant and promising local market (Perera & Ranaweera, 2023). The swift development of women's fashion stores in Pontianak aligns with the Sustainable Development Goals (SDGs), particularly in promoting economic growth (SDG 8) and encouraging sustainable consumption and production (SDG 12). As Pontianak flourishes, these fashion outlets have the potential to empower local communities by supporting small enterprises and creating job opportunities, especially for women. Moreover, with a growing consumer interest in contemporary trends (Kim, Cho, & Park, 2022), these exists an opportunity for the industry to advance sustainable fashion practices, thereby contributing to environmental sustainability while embodying the region's dynamic cultural identity.

To ascertain the optimal timing for selling popular products at women's fashion stores in Pontianak, the regression tree method serves as an effective analytical tool (Pavlyshenko, 2019). This approach analyzes historical sales data alongside variables such as seasonal trends, current fashion movements, national holidays, promotional events, and consumer behavior. By employing the decision tree, specific sales patterns can be discerned, highlighting peak demand periods for particular products, such as modern Muslim fashion in September, October, or November, and formal wear at the year's end. This analysis provides valuable

insights for women's fashion retail, enabling it to strategically plan inventory, promotions, and marketing initiatives, thereby enhancing profitability while satisfying market needs.

A systematic literature review of articles from Web of Science and Scopus examined predictive analytics encompassing data mining, big data, and machine learning and the decision tree method, highlighting their applications in predicting trends, enhancing corporate performance, and supporting data-driven decision-making in business analytics. (Lee, Cheang, & Moslehpour, 2022).

References to previous research on Climate change makes consumer behavior towards winter clothing searches difficult to predict. This study uses weather data and Google Trends for 12 years (2008-2012) with the classification and regression tree (CART) method to analyze the influence of weather, calendar, and promotions on consumer searches. This method was chosen because it is able to reveal hierarchical patterns between variables. The results show that searches increase when temperatures and *windchills* decrease, while promotions are only effective when temperatures are warmer than usual. These findings help retailers develop sales strategies and manage stock adaptively amidst climate change. (Oh, 2024).

Previous research utilized regression modelling through Decision Tree implemented with Python programming to identify and optimize activities or elements that have a significant impact on business processes, resulting in improved efficiency, enhance information accuracy, and better overall process quality. (Sishi & Telukdarie, 2021).

Earlier research employed decision techniques, and the findings of this study demonstrate that the model is capable of formulating strategies for sustainable site management moving forward, using data from new locations as it becomes available. (Li, Yi, Cundy, & Chen, 2022).
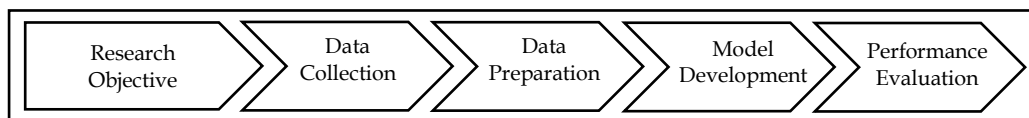
Methods

Figure 1. Workflow of research framework

This research employs decision tree regression to examine and forecast sales trends based on historical transaction data. The model was selected for its straightforwardness and capability to manage both categorical and numerical variables, with the intention of identifying trends and assessing predictive accuracy. The dataset comprises 1,200 transaction entries taken from the transaction history of a women's fashion retail named Veruby Store Pontianak, detailing transaction dates, item names, category and sales amounts. Focus was places on specific item category ("tops"), analyzing sales trends over time. Temporal variables, including the day of the week, date, and month, were utilized as predictors, while sales figures represented the target variable.

Temporal features were generated from transaction dates, and outliers were removed using a two-standard deviation method to exclude extreme values. The dataset was divided intro training (70%) and testing (30%) (Mansouri, Soui, & Abed, 2023) subsets for model development and evaluation.
A decision tree regression model was created using fitrtree function, treating the predictors as categorical variables. The tree structure was visualized to clarify the decision-making process.

Model accuracy was determined by comparing predicted sales figures with actual sales from the test dataset. Performance metrics included Root Mean Square Error (RMSE) and Normalized RMSE (NRMSE) (Ensafi, Amin, Zhang, & Shah, 2022), with accuracy calculated as Accuracy = (1 – NRMSE) x 100. Results were presented through plots comparing predicted and actual sales, demonstrating the model's effectiveness in capturing sales patterns. This methodology offered a clear and concise analysis of the model's performance.
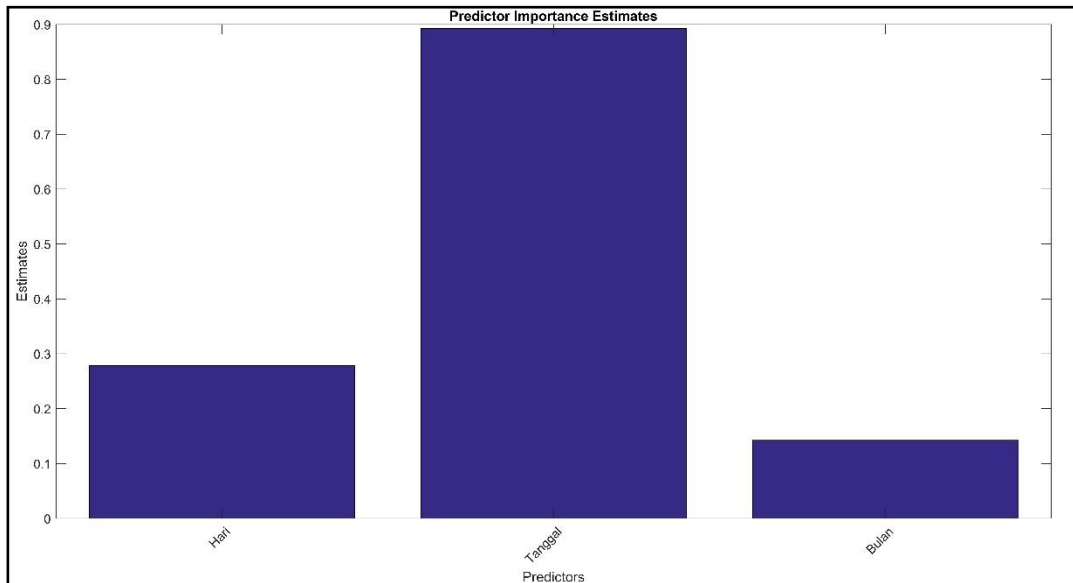
Result and Discussion

Figure 2. Parameter Importance

Based on the results of the Predictor Importance Estimates analysis as shown in (Figure 2), it is evident that the "Date" parameter plays the most significant role in influencing the model's predictions, as indicated by its highest importance value compared to other parameters. This suggests that date-related information has a dominant impact on determining the target variable. Meanwhile, the "Day" parameter shows a moderate contribution, while the "Month" parameter has the lowest importance score, indicating its relatively minor influence on the model's predictions. These results imply that the model relies heavily on date information for accurate predictions, while the impacts of the day and month are comparatively less significant.
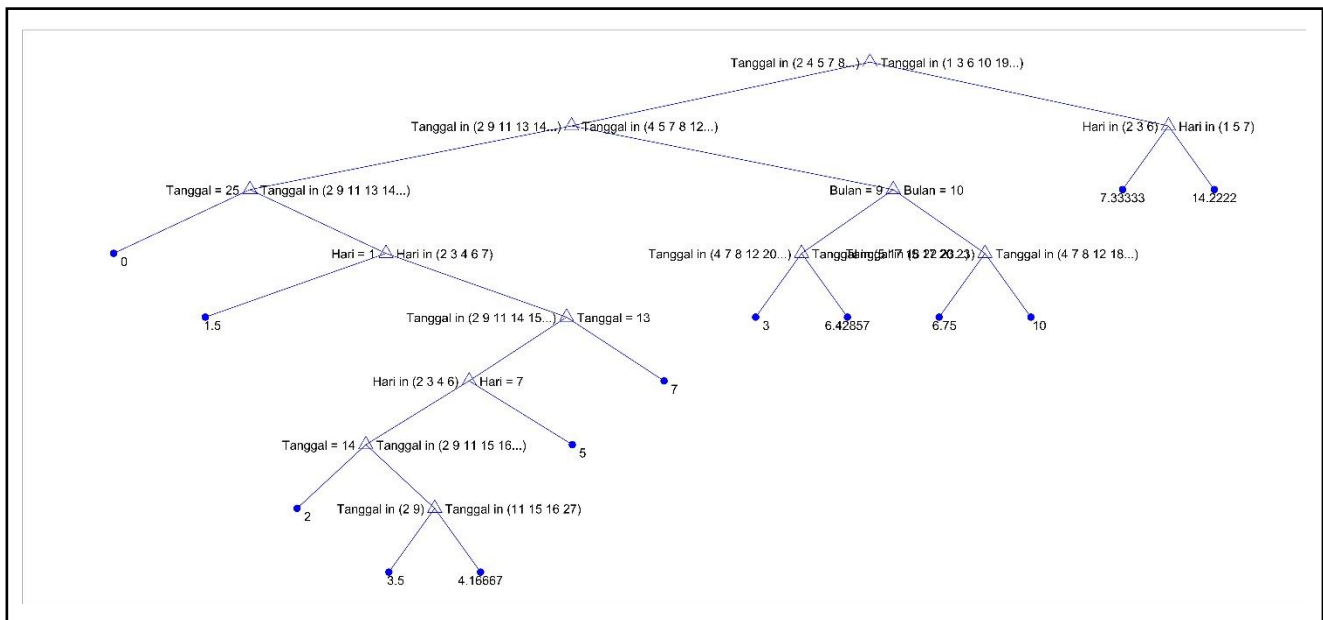

Figure 3. Decision Tree Calculation Results

In this study, the Decision Tree was constructed as shown in (Figure 3). The analysis of parameter importance indicates that the Date is the most influential factor in predicting sales for the "tops" category (Figure 2), followed by the Day of the week, while the Month has the least impact. The structure of the decision

tree reveals that certain dates, such as the 13th, and the specific weekdays (particularly weekdays excluding Thursday, with Day = {2, 3, 4, 6, 7}), significantly affect sales predictions. For example, in the branch where the Date = 13 and the Day includes {2, 3, 4, 6, 7}, the predicted sales reach 7 units.

The high influence of the Date suggests that sales are affected by customer behavior, such as increased shopping activity on specific dates like the 13th, which falls within the early month when people typically receive their salary. In reality, on Mondays, Tuesdays, and Wednesdays, many individuals purchase clothing such as shirts for work, while on Fridays and Saturdays, people buy blouses to prepare their outfits for the weekend. This may explain why sales of "tops" tend to be higher on those specific dates, particularly on weekdays.

The low impact of the Month indicates that seasonal variations are not significant for this category. These findings can be leveraged to enhance promotional strategies, such as offering additional discounts at the beginning of the month until 13th, as well as developing marketing strategies for the latest "tops" or different "tops" models on specific days by creating engaging videos shared on social media to attract customer attention. Furthermore, ensuring the availability of "tops" such as shirts and blouses during weekdays can also be implemented. By understanding these patterns, inventory management and promotional strategies can be optimized to boost sales on high-potential dates and days.

Conclusion

In this research, the women's fashion retail sector in Pontianak is experiencing significant growth, driven by increasing interest in contemporary clothing trends and a diverse array of outlets offering various options, from casual to formal wear and modern Muslim fashion. The existing competition fosters innovation in product quality and overall shopping experience, aligning with the Sustainable Development Goals (SDGs) that focus on economic growth and sustainable consumption. This study employs the decision tree method to determine the optimal timing for selling "tops", with the regression model achieving an accuracy of 77.85% in predicting sales trends. The decision tree structure illustrates the influence of temporal parameters on sales predictions, providing insights that can be utilized to enhance sales strategies and inventory management. Specifically, the analysis facilitates targeted marketing efforts, enabling the store adjust inventory based on predicted demand and seasonality. For future development, expanding the scope of research by incorporating additional predictors such as weather data, discount periods, or online search trends could yield more comprehensive results. Integrating teal-time analysis from a larger dataset, including hourly data, would enable faster and more adaptive decision-making, supporting the optimization of sales strategies and inventory management across various product categories.

References

Ensafi, Y. a. (2022). Time-series forecasting of seasonal items sales using machine learning--A comparative analysis. *International Journal of Information Management Data Insights, 2*(1), 100058.

Kim, H. a. (2022). Analyzing genderless fashion trends of consumers' perceptions on social media: using unstructured big data analysis through Latent Dirichlet Allocation-based topic modeling. *Fashion and Textiles, 9*(1), 6.

Lee, C. S., Cheang, P. Y., & Moslehpour, M. (2022). Predictive analytics in business analytics: decision tree. *Advances in Decision Sciences, 26*(1), 1-29.

Li, X. a. (2022). Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms. *Journal of Cleaner Production, 371*, 133612.

Mansouri, N. a. (2023). SFS feature selection with decision tree classifier for massive open online courses (MOOCs) recommendation. *Journal of Computers in Education*, 1-22.

Oh, J. (2024). Classification and regression tree approach for the prediction of the seasonal apparel market: focused on weather factors. *Journal of Fashion Marketing and Management: An International Journal, 28*(5), 893-910.

Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *{Data, 4*(1), 15.

Pei, X.-L., Guo, J.-N., Wu, T.-J., Zhou, W.-X., & Yeh, S.-P. (2020). Does the effect of customer experience on customer satisfaction create a sustainable competitive advantage? A comparative study of different shopping situations. *Sustainability, 12*(18), 7436.

Perera, G. A., & Ranaweera, A. (2023). Fashion localism: evaluation and extensions of utility in Sri Lankan fashion brands. *Research Journal of Textile and Apparel, 27*(2), 264-280.

Sishi, M. a. (2021). The application of decision tree regression to optimize business processes. *Proceedings of the International Conference on Industrial Engineering and Operations Management Sao Paulo, Brazil*, 48-57.

Tripathi, D. a. (2023). {Online Shopping And Buying Behaviour: A Study On Female Students With Special Reference To Apparels And Fashion Industry. *Journal of Namibian Studies: History Politics Culture, 35*, 5385-5406.