Stock Resupplying Strategy via Decision Tree Learning Algorithm

Jose Justin[1,*]; Genrawan Hoendarto[2]
[1]Department of Digital Business, [2]Department of Informatics
[1,2]Universitas Widya Dharma Pontianak
[1,2]Pontianak, Indonesia
[1]21430103_jose_j@widyadharma.ac.id ; [2]genrawan@widyadharma.ac.id

Corresponding Author

The swift development of economic advancement has caused demand for better lead-time management to sharply increase profit. Fuel stations require precision and reliable resupply timetable to avoid congested traffic, which may potentially slowing down transactions, which is why the author suggests using Decision Tree Learning method to deal with the problem. PT. Perkasa Makmur, a distributor of fuel in Sekadau, West Kalimantan, Indonesia, has never analyzed sales data to gain insights for a potentially better resupply timetable. This research can be categorized as experimental qualitative research, with a population-spanning dataset collected within 36 months of 2021 to 2023 workyear. Sampling technique involves up to 720 data points using probability sampling. This study uses the last 5 months of data from 36 months of the collected dataset, starting in February 2024 and ending in June 2024. This research produces an optimal date for the fuel station to resupply its fuel stock within the weekly and monthly timeframes. The graph generated from leaf and branch nodes provides for the suggested date to resupply, with the accuracy rate of the method up to 76.5%.

Keywords—Decision Tree; Machine Learning; Purchase Pattern; Stock Resupplying

I. INTRODUCTION

The ever-increasing development of technology and business causes customers to demand better and better service, product quality, and service time. This indirectly forces companies to constantly outperform their opponents and perform better in order to keep the company growing. All companies that fail to deliver the best to their customers will eventually lose business. Companies can solve this problem by reengineering the value of their products or improving their internal efficiency. The first option is not the right answer for public companies due to strict government restrictions on business processes, so the second option is the only appropriate method. Preventing conflicting schedules between stock provisioning and crowded customer buying times will give the company the stock stability and flexibility it needs to handle emergency situations. An accurate forecast for resupply rescheduling also creates a basis for companies to design the best strategy to make profit and stay in business [1]. To develop such a strategy, sufficient data is required to generate such schedules. Companies often archive their sales history and just use it to generate monthly or yearly sales reports to measure their business' performance [2]. To utilize this information effectively, existing sales history data can be processed using machine learning algorithms.

The Decision Tree Learning method is a technique to achieve a boolean final result at the leaf nodes by constantly breaking down the data sequence and splitting it at the root. The method contains nodes and connections that are used to classify the dataset and ultimately to generate an analysis model in which all possible splits of the data are analyzed [3]. Decision tree learning is split into two processes. The first step is to convert most datasets into training data, then classify the rest as test data [4]. Decision tree learning consists of decision nodes and leaf nodes. Decision nodes will produce results and branch to other nodes. Leaf nodes are the end result of decision nodes and cannot branch anymore. Decision tree Learning is so called because of its resemblance to a tree structure [5]. Previous research has used decision tree methods to detect breast cancer [6][7], distinguish spam emails [8][9], detect ionization events [10], predict business failures [11], and predict service industry performance [12][13][14]. But overall, there is still a lack of analysis on the theme of development economics, specifically the creation of accurate and reliable schedules for corporate resupply activities.

PT Perkasa Makmur, a fuel distributor (gas station) located in Sekadau, West Kalimantan, benefited greatly from the application of decision tree learning to generate an appropriate replenishment schedule. However, PT. Perkasa Makmur has not utilized its sales data for such a purpose. The company places detailed orders with fuel suppliers controlled by the central government that will be delivered within two days. The delivery process has the potential to cause problems if it is done during customers' rush hours. Therefore, we took the initiative to process and analyze sales data to overcome the shortcomings.

II. RESEARCH METHODS

This research uses an experimental qualitative approach to assess the daily rate of fuel cubication sold to determine the appropriate day to resupply the fuel depot. Primary data was collected and analyzed from the historical sales data of PT Perkasa Makmur, which includes historical sales data from 2021 to 2023. The population of this study consists of 1,095 daily sales transactions from January 1, 2021, to December 31, 2023. The sample of this study is Pertalite transactions, which are used as test data. PT Perkasa Makmur, a fuel distributor located in Sekadau, West Kalimantan, Indonesia, is the subject of the study. The appropriate day to resupply was considered a dependent variable, while the daily sales rate of fuel cubic liters was treated as an independent variable.

The dataset's random nature will be measured by an equation called entropy. Its value always swings between 0 and 1, whereas by attaining a result close to 0, the produced decision tree will be more accurate [3]. Pi is the probability ratio of the sample number selected from an example in class i and the i-th attribute or class identifier, as shown in (1):

$$\text{Entropy } (S) = \sum \text{P}i\ c\ i\text{=1 log 2 Pi} \tag{1}$$

Gained information will be represented as Gain (S,A). This is a segmentation metric used to measure the amount of knowledge inside a random variable's value. This information can also be used to acquire an attribute's position in the tree itself [15]. Usually, a higher result is likely to generate a better model [3]. Range of attribute A value is $V(A)$, and Sv is the subset entropy of the dataset S equal to the value of attribute v (2):

$$\text{Gain (S , A)} = \sum |\text{Sv}|\ |\text{S}|\ \text{v} \in \text{V(A) Entropy(Sv)} \tag{2}$$

The pseudocode generally accepted in decision tree learning starts with ordering the system to maintain the best features of the input parameters at the root node of the tree. The dataset is trained by splitting it into several subsections. These split subsets generate a similar value for each input attribute. The process is then repeated subsequently until leaf nodes in the tree are generated, thus ending the process [6].

The study began by compiling the data collected from PT Perkasa Makmur into a table, with the first column being the date (DD-MM-YYYY), the second column being numbers in auto-increment to enable the system to recognize the date, while the remaining columns contained the fuel cubication sold in liters of Pertalite, Dexlite, Pertamax, Solar, and Pertamina Dex. A lag matrix is created to generate an interval between the training data and the predicted result. This data table is then processed and converted into a decision tree, with 90% of the samples as training data and the rest as test data. Apart from Pertalite, all types of fuel sold by PT Perkasa Makmur were not included in the decision tree because the sales were insignificant and fluctuated abnormally.

Sales data is then augmented with its average value to incorporate the predicted amount of average fuel sales to generate the prediction. A huge amount of leaf and branch nodes will then be produced to identify which day within a week or month, represented by a number, is the most appropriate to conduct refueling. Lastly, since identifying them through messy plotting is difficult, the decision tree is simplified by generating weekly and monthly graphs of fuel stock movement, and by finding its lowest point, the result will then be recognized as the most appropriate day to resupply PT. Perkasa Makmur's fuel stock.

III. RESULT AND DISCUSSION

In this study, sales data can be used to predict the future period when it is appropriate to resupply within a weekly or monthly timeframe. Initially, the Decision Tree is generated as in Figure 1. The accuracy value generated by this prediction model can be seen in Figure 2, which is 76.5%. The blue line represents the result generated by the model, while the red line represents the actual test data's value. The value of the accuracy result leaves room for improvement, possibly by using a more detailed dataset down to hourly data. The analysis revealed some optimal timescales for PT Perkasa Makmur to restock Pertalite, as seen in Figure 3.

For example, the model predicts there are four likely days to resupply fuel stock within the monthly timeframe of December 2024, which are on December 6, 10, 15, and 22. This corresponds to the current resupply schedule of once every week. From the analysis, we can also recognize two patterns that existed in the last month of 2024. First, sales tend to decline on the first days of the week, allowing the company to resupply during this period. In addition, pertalite sales also tend to increase in the middle of the week, which indicates that companies should avoid resupplying during this timeframe.

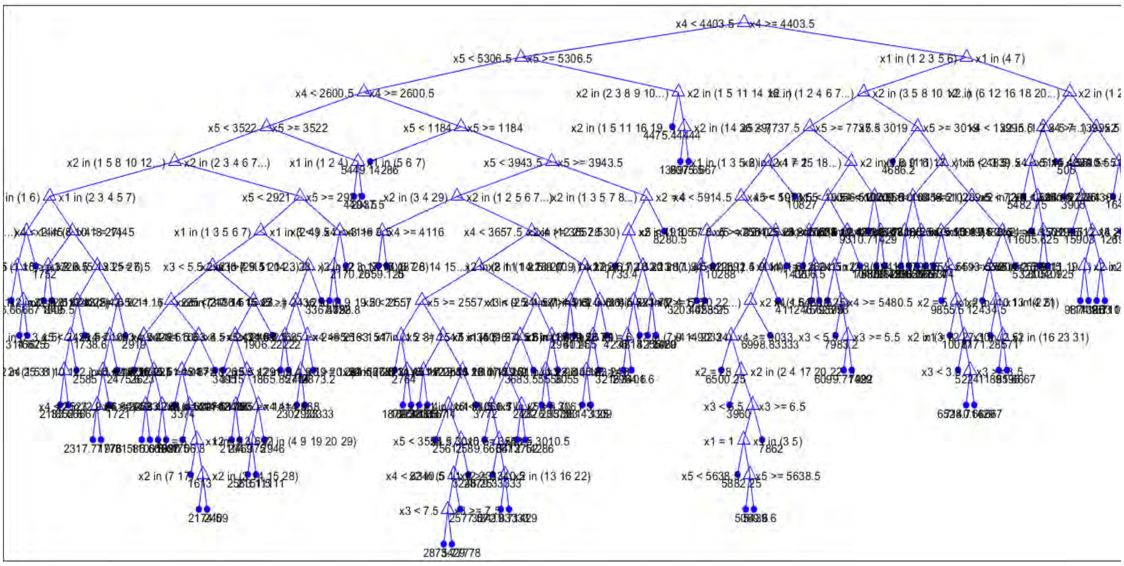

Figure 1. Decision Tree Calculation Results

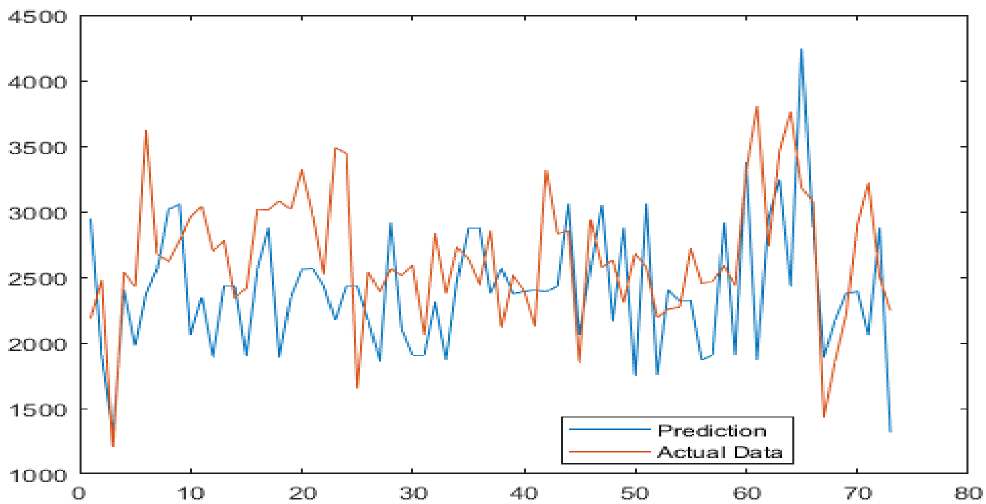

Figure 2. Accuracy Results

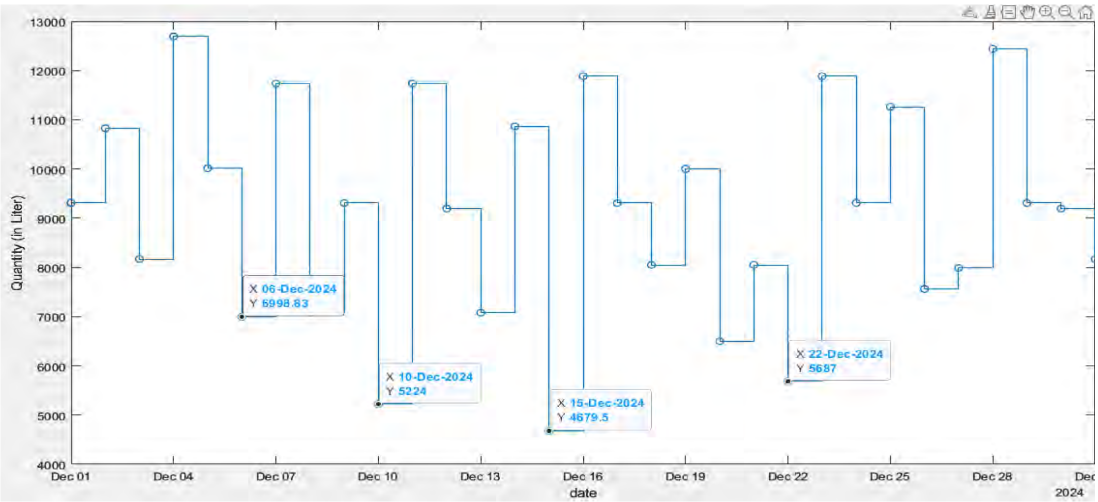

Figure 3. Prediction Results

The resulting schedule can be a valuable resource for the company in strategizing the right time for refueling. Thus allowing as little friction as possible between consumer transactions and refueling actions by the company. In addition, this schedule can also serve as a reference for implementing other timely activities, such as stock purchasing and delivery, to increase product sales and maintain customer satisfaction.

IV. CONCLUSION

In this research, the Decision Tree Learning (DTL) method served as an analytical tool to uncover customer buying patterns for PT Perkasa Makmur's fuel products based on its Pertalite sales data. These insights were then used to uncover the company's rush hours in a weekly or monthly timeframe, which in turn were used to avoid selecting these times to conduct stock replenishment. This prediction model also provides information about unique sales patterns at PT Perkasa Makmur, i.e., information about off days that can be used for expected resupply and stock replenishment. In addition, Decision Tree Learning provides important information to guide the company in determining the required lead time strategy for purchasing and delivering goods and services, using this data as a reference point. Further research may be conducted to achieve an even more accurate result by using daily and hourly datasets.

Based on this analysis, DTL is capable of providing PT Perkasa Makmur with a specific and precise system to base their resupply schedule on. In addition, this analysis gives PT Perkasa Makmur a considerable advantage over other fuel stations in the region, whereas by aligning replenishment efforts with customer preferences, the company can effectively engage customers' rush hours and capitalize on maximal sales growth.

V. REFERENCES

[1] S. Seyedan, "Development of Predictive Analytics for Demand Forecasting and Inventory Management in Supply Chain using Machine Learning Techniques," 2023.

[2] A. R. Efrat, R. Gernowo, and Farikhin, "Consumer purchase patterns based on market basket analysis using apriori algorithms," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, June 2020. doi: 10.1088/1742-6596/1524/1/012109.

[3] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, March 2021, doi: 10.38094/jastt20165.

[4] M. Z. Arif, R. Ahmed, U. H. Sadia, M. S. I. Tultul, and R. Chakma, "Decision Tree Method Using for Fetal State Classification from Cardiotography Data," *Journal of Advanced Engineering and Computation*, vol. 4, no. 1, p. 64, March 2020, doi: 10.25073/jaec.202041.273.

[5] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine

learning," *Decision Analytics Journal*, vol. 3, p. 100071, June 2022, doi: 10.1016/j.dajour.2022.100071.

[6] H. Rajaguru and S. R. Sannasi Chakravarthy, "Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer," *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 12, pp. 3777–3781, 2019, doi: 10.31557/APJCP.2019.20.12.3777.

[7] P. Sathiyanarayanan, S. Pavithra, M. , S. Saranya, and M. Makeswari, "Identification of Breast Cancer using the Decision Tree Algorithm," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, Mar. 2019.

[8] S. S. I. Ismail, R. F. Mansour, R. M. Abd El-Aziz, and A. I. Taloba, "Efficient E-Mail Spam Detection Strategy Using Genetic Decision Tree Processing with NLP Features," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7710005.

[9] M. Jazzar, R. F. Yousef, and D. Eleyan, "Evaluation of Machine Learning Techniques for Email Spam Classification," *International Journal of Education and Management Engineering*, vol. 11, no. 4, pp. 35–42, August 2021, doi: 10.5815/ijeme.2021.04.04.

[10] N. Linty, A. Farasin, A. Favenza, and F. Dovis, "Detection of GNSS ionospheric scintillations based on machine learning decision tree," *IEEE Trans Aerosp Electron Syst*, vol. 55, no. 1, pp. 303–317, February 2019, doi: 10.1109/TAES.2018.2850385.

[11] S. Y. Kim and A. Upneja, "Majority voting ensemble with a decision trees for business failure prediction during economic downturns," *Journal of Innovation and Knowledge*, vol. 6, no. 2, pp. 112–123, April 2021, doi: 10.1016/j.jik.2021.01.001.

[12] J. Patalas-Maliszewska, H. Łosyk, and M. Rehm, "Decision-Tree Based Methodology Aid in Assessing the Sustainable Development of a Manufacturing Company," *Sustainability (Switzerland)*, vol. 14, no. 10, May 2022, doi: 10.3390/su14106362.

[13] C. S. Lee, P. Yeng, S. Cheang, and M. Moslehpour, "Predictive Analytics in Business Analytics: Decision Tree," 2022.

[14] B. Yeo and D. Grant, "Predicting Service Industry Performance Using Decision Tree Analysis," *Int J Inf Manage*, vol. 38, no. 1, pp. 288–300, 2018.

[15] M. I. Hoque, A. kalam Azad, M. Abu Hurayra Tuhin, and Z. U. Salehin, "University students result analysis and prediction system by decision tree algorithm," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 3, pp. 115–122, 2020, doi: 10.25046/aj050315.