DATA MINING ANALYSIS OF SHELL OIL SALES USING THE C4.5 ALGORITHM AT CV. HARAPAN KARYA MANDIRI

Seci Munika Sari[1)*], Muhammad Husni Rifqo[2)]

[1,2]Department of Informatics Engineering
[1,2]Faculty of Engineering
[1,2]University of Muhammadiyah Bengkulu
E-mail : Secimonika102@gmail.com[1)], mhrifqo@umb.ac.id[2)]

Abstract

This study focuses on the analysis of Shell oil sales at CV. Harapan Karya Mandiri (HKM) Bengkulu, which faces challenges in predicting consumer demand and managing stock efficiently. CV. HKM Bengkulu is an official distributor of PT. Shell Indonesia, competing in the vehicle lubricant industry. To address the challenges of competition and demand uncertainty, this study applies data mining methods, particularly the C4.5 algorithm, to analyze historical sales data and uncover significant patterns and trends. Data mining is a technique that helps identify hidden patterns and insights in large datasets to support decision-making. The C4.5 algorithm is employed to build a predictive model through a decision tree, which classifies data based on certain variables such as oil type, sales region, or time period. This model is expected to assist CV. HKM in predicting customer demand, optimizing sales strategies, and improving stock planning efficiency. Additionally, the results from the C4.5 algorithm provide practical benefits by enabling CV. HKM to optimize inventory management, target marketing efforts more effectively, and enhance operational efficiency. The insights derived from the model support data-driven decisions, improve business performance, and maximize profits by aligning stock levels with demand trends, thereby reducing wastage and improving profitability.

Keywords- *Analysis, Data Mining, Sales, C4.5 Algorithm*

1. INTRODUCTION

The advancement of Information Technology (IT) is currently growing very rapidly. This technology is also felt in all aspects of activities, particularly in the business or sales sector [1]. The ease of utilizing computerized systems greatly helps in analyzing, obtaining information, and making decisions in order to survive and develop capabilities to achieve goals [2].

CV. Harapan Karya Mandiri (HKM) Bengkulu is a lubricant distributor of PT. Shell Indonesia, where PT. Shell Indonesia is a global energy company engaged in the lubricant and fuel station industry operating in Indonesia since 1928. With the increasing competition in the lubricant industry, the company needs to optimize its sales strategies based on existing sales patterns and trends [3]. Historical sales data contain valuable information that can be used to understand customer behavior and demand trends [4].

The problems occurring at CV. Harapan Karya Mandiri (HKM) Bengkulu include difficulties in analyzing oil sales and predicting retail consumer demands, as well as uncertainty in the quantity of stock sold, which results in inefficient operations. Therefore, sales analysis is vitally important for CV. HKM Bengkulu to predict future sales, so that the company can estimate the amount of stock that needs to be prepared. Large and complex data cannot be analyzed manually. For this reason, data mining techniques are required to discover patterns that can help improve marketing effectiveness and stock planning [5].

Data mining is a process to discover patterns or interesting information from large datasets by using various specific techniques or methods. Its goal is to find meaningful patterns such as trends, relationships between data, or hidden anomalies that are difficult to identify manually [6]. The results of data mining should be easy to understand, novel, and useful for decision-making, providing added value by revealing hidden knowledge. Data mining also identifies relationships between relevant variables, such as the correlation between customer age and the type of product purchased, which can be used for various strategic purposes [2].

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

The C4.5 algorithm is one of the popular algorithms in data mining used to create decision tres [8]. This algorithm was developed by J. Ross Quinlan and is an extension of the ID3 algorithm. C4.5 is used for classification, namely mapping data into certain categories or classes based on existing attributes [9]. The C4.5 algorithm is one of the data mining methods that produces decision trees to predict categories or trends based on certain variables, such as oil type, sales region, or time period [10]. This research aims to implement the C4.5 algorithm to discover Shell oil sales patterns and produce predictive models that can assist management in decision-making [11]. C4.5 is an algorithm that breaks down large problems into smaller, more easily understood parts. To effectively analyze customer demand, various data classification algorithms can be applied. One such method is the C4.5 algorithm, which is commonly used to classify data containing both numerical and categorical variables [3].

Based on the background described above, the author will conduct a study entitled "Data Mining Analysis of Shell Oil Sales Using the C4.5 Algorithm at CV. Harapan Karya Mandiri."

The purpose of this research is to analyze Shell oil sales patterns at CV. Harapan Karya Mandiri (HKM) Bengkulu using data mining techniques and the C4.5 algorithm to classify Shell oil sales at CV. Harapan Karya Mandiri. This research aims to build a predictive model that is expected to help the company in estimating customer demand, optimizing sales strategies, and improving the efficiency of stock planning.

This research is urgently needed because the company is currently facing serious challenges in anticipating fluctuating market demands and managing inventory levels effectively. Without a systematic analytical approach, the risk of overstocking or stockouts increases, which can lead to financial losses, reduced customer satisfaction, and missed sales opportunities. The competitive nature of the lubricant industry further demands quick and accurate decision-making based on reliable data insights. By implementing a data mining solution such as the C4.5 algorithm, the company can shift from intuition-based to data-driven strategies, enabling faster responses to market trends and better operational control.

Thus, this research is expected to overcome difficulties in analyzing large and complex data, as well as assist management in making more effective decisions [13]. Shell oil sales analysis using data mining explores hidden patterns and trends in sales data by applying statistical techniques, algorithms, and artificial intelligence to provide in-depth insights that can support the company in decision-making, improve business performance, and maximize profits [14].

2. METHODOLOGY

The methods used in this research consist of two main parts: the use of the Confusion Matrix to evaluate the performance of the classification model and the application of the C4.5 algorithm to analyze Shell oil sales patterns.

1. Confusion Matrix

The Confusion Matrix is a method commonly used to evaluate the performance of a model in classification or grouping tasks. This method works by comparing the predicted data from the model with the actual data, and the results are broken down into four key categories: True Positive, True Negative, False Positive, and False Negative [15].

a. True Positive (TP): The model's prediction is correct when the actual data is positive.
b. True Negative (TN): The model's prediction is correct when the actual data is negative.
c. False Positive (FP): The model predicts positive, but the actual data is negative.
d. False Negative (FN): The model predicts negative, but the actual data is positive.

The Confusion Matrix allows for a deeper understanding of prediction errors, such as false positives or false negatives, and helps to evaluate accuracy, precision, recall, and various other performance metrics of the model [4].

2. C4.5 Algorithm

In addition to the Confusion Matrix, this research also uses the C4.5 algorithm to build a classification model. The C4.5 algorithm is a

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

method used to create decision trees based on the available data. This algorithm works by splitting the data into subsets based on the attribute that has the best information gain, which is calculated using entropy and information gain. Each branch in the decision tree represents a decision or condition that separates the data based on the attribute values.

a. The steps taken in applying the C4.5 algorithm in this research are as follows:
b. Data Preparation: The Shell oil sales data obtained from CV. Harapan Karya Mandiri (HKM) Bengkulu is processed and prepared for analysis. This data includes information about sales volume, price, and product type.
c. Data Splitting: The data is divided into two sets: the training set and the testing set. The training set is used to build the model, while the testing set is used to evaluate the model's performance.
d. Decision Tree Construction: The C4.5 algorithm is applied to the training data to build the decision tree. Each branch in the tree reflects the separation of data based on the attribute with the highest information gain, and this process continues until all data can be classified.
e. Model Evaluation: The model that is built is then tested using the testing data and evaluated using the Confusion Matrix. This allows the model to predict Shell oil sales patterns, supporting better decision-making in inventory planning and sales strategies.
f. Result Analysis: The evaluation results of the model using the Confusion Matrix are calculated to obtain accuracy, precision, recall, and other metrics to assess how effective the C4.5 model is in predicting sales patterns.

3. RESULTS AND DISCUSSIONS

3.1 Calculation of the C4.5 Algorithm

In the design of the analysis using data mining, the application of the C4.5 Algorithm in this study produces a decision tree through a process of transforming data that was initially in tabular form into a decision tree [18]. The first step in creating a decision tree is to calculate the total number of best-selling and non-best-selling products based on the attributes used, namely brands, type, area, and sandal in the Shell oil sales data for the last three months at CV. Harapan Karya Mandiri. Next is to calculate the entropy value by computing the entropy and the highest gain for each value within the variables. In this study, since the gain ratio is used, the first step is to find the gain ratio in accordance with the flow of the C4.5 Algorithm, which is to determine entropy and gain first.

From the results of manual calculations to determine the root or the root node of the decision tree, the variable SANDAL has the highest gain ratio value of 0.75072863 and is therefore chosen as the root [19]. Within the SANDAL attribute, the values are: "PREMIUM ADVANCE, MATIC SERIES, HELIX 6, HELIX 5, HELIX 7, PRODUCT FOCUS, SPIRAX, HELIX CITY."

3.2 Results of C4.5 Algorithm Measurement Using RapidMiner

The processed data, which had been transformed into a data sheet ready for analysis, was tested in the RapidMiner application by applying the C4.5 algorithm. The operator used was X-Cross Validation with gain ratio calculation. The reason for using this operator model was the expectation that the testing would yield maximum accuracy. From the testing results by applying the C4.5 algorithm in RapidMiner, an accuracy of 79.29% was obtained, with the following confusion matrix results.

Table 1 . Confusion Matrix Results

	true HELIX	**true ADVANCE**	**true SPIRAX**	**true RIMULA**	**true COOLANT**
pred. HELIX	669	3	0	37	1
pred. ADVANCE	236	1303	0	235	13
pred. SPIRAX	0	0	3	1	0
pred. RIMULA	13	0	0	25	0
pred. COOLANT	0	0	0	0	0

The testing results of the data produced by RapidMiner using the C4.5 Algorithm model show the obtained accuracy.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

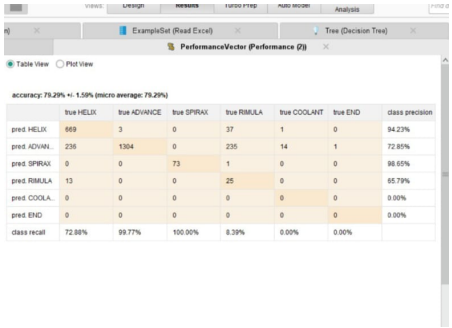

Figure 1. Accuracy Results

Figure 1 shows the accuracy results obtained from the C4.5 classification process. Based on the evaluation, the overall accuracy reached 79.29%. The highest number of correct predictions was for the product True Advance, totaling 1,303 predictions, indicating it as the best-selling oil. In contrast, True Coolant received 0 predictions, suggesting it had the lowest or no sales during the observed period.

Table 2. Confusion Matrix Formula

	true HELIX	**true ADVANCE**	**true SPIRAX**	**true RIMULA**	**true COOLANT**
pred. HELIX	TP	FP	FP	FP	FP
pred. ADVANCE	FN	TP	TN	FN	FN
pred. SPIRAX	FN	TN	TP	FN	TN
pred. RIMULA	FN	TN	TN	TP	TN
pred. COOLANT	FN	TN	TN	TN	TP

TP = 669+1303+73+25+0 = 2.071
TN = 0
FP = 3+0+37+1= 31
FN = 236+0+13+0+235+1+13 = 399
Manual Accuracy Calculation Using the Accuracy Formula

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} x100\%$$

$$= \frac{2.071+0}{2.071+0+31+399} x100\%$$

$$= \frac{2.071}{2.611} x100\%$$

$$= 79{,}31\%$$

1. Decision Tree or C4.5
   From the testing results using RapidMiner by applying the C4.5 algorithm, a decision tree was obtained as shown in the figure below:

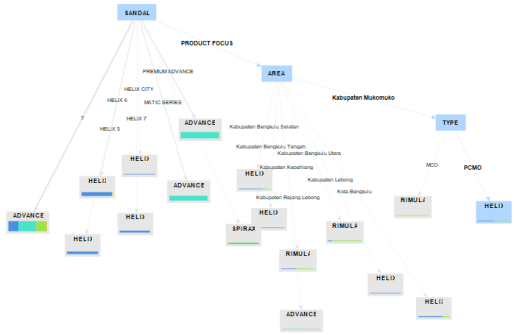

Figure 2. Decision Tree Results (C4.5)

Figure 2 displays the decision tree generated using the C4.5 algorithm in RapidMiner. From this tree structure, a set of classification rules was extracted to guide predictions based on various input attributes. The decision tree enables systematic classification of oil sales data and reveals key decision paths derived from historical patterns. These rules serve as a foundation for understanding the relationship between product categories and their corresponding sales behavior. From the decision tree above, the testing results in RapidMiner produced the following set of rules:

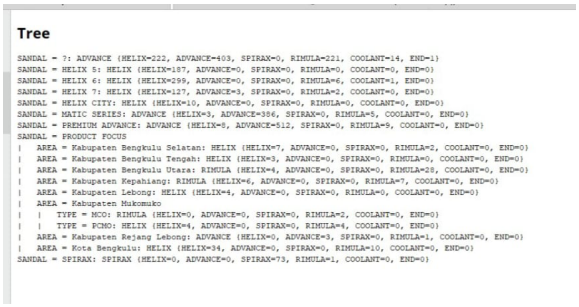

Figure 3. Decision Tree Rules

Figure 3 outlines several decision rules extracted from the decision tree generated by the C4.5 algorithm. These rules are derived based on the values of key attributes such as sandal, area, and type, which significantly influence the classification of best-selling oil products. The following rules summarize the most relevant patterns identified:

a. If the value of the sandal attribute is overall, then the best-selling brand is Advance with a value of 303.
b. If the value of the sandal attribute is Helix5, Helix6, or Helix7, then the best-selling

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

product is Helix with values of 187, 299, and 127 respectively.

c. If the value of the sandal attribute is Matic Series, then the best-selling brand is Advance with a value of 386.
d. If the sandal attribute is Premium Advance, then the best-selling brand is Advance with a value of 512.
e. If the value of the area attribute is Kabupaten Bengkulu Selatan and Kabupaten Bengkulu Tengah, it shows that the best-selling product is Helix, while in Bengkulu Utara and Kepahiang the best-selling product is Rimula.
f. According to the type attribute, the best-selling types are MCO and PCMO with the brands Rimula and Helix.

3.3 Data

The data used by the researcher in this study are the Shell oil sales data from the last three months, obtained from CV. Harapan Karya Mandiri through one of its employees.

3.3.1 Steps of the Process in RapidMiner

The first step is to prepare the data that will be tested, followed by the testing process in the RapidMiner application. The testing uses the X-Cross Validation operator model in order to obtain maximum accuracy.

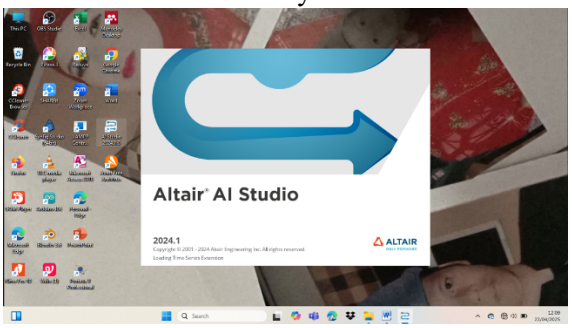

Figure 4. Initial Page of RapidMiner

Figure 4 shows the initial interface of Altair® AI Studio 2024.1, a software used for data science workflows including AI, machine learning, and predictive analytics. This screen appears upon launching the software and provides access to tools for importing datasets, building analytical pipelines, and evaluating model performance. Altair AI Studio offers visual programming capabilities similar to platforms like RapidMiner, making it suitable for users with minimal coding experience.

1. Go to the main menu of RapidMiner

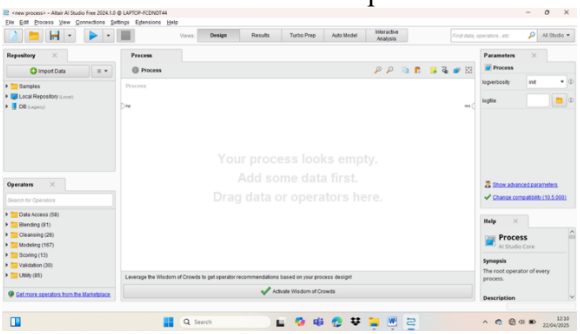

Figure 5. Main Menu of RapidMiner

Figure 5 displays the main interface of RapidMiner Studio, a data science platform commonly used for machine learning workflows. The central area labeled “Process” is where users can drag and drop operators to build analysis pipelines. The left panel includes the Repository (for accessing datasets and projects) and Operators (a categorized list of processing tools such as data transformation, modeling, and validation). The right-hand panel provides settings and parameters for each selected operator.

In the main menu, we can see the display of each tool that will be used in conducting the data testing. Then click Import Data. In the operators section, click Read Excel to import the data that will be tested.

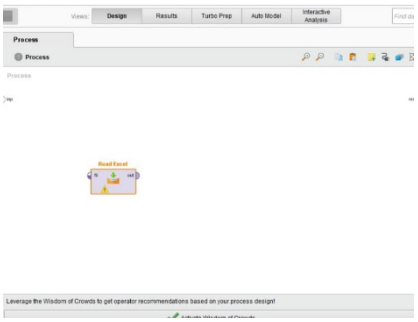

Figure 6. Process Menu

As shown in Figure 6, the Process Menu in RapidMiner allows users to design the analysis workflow. In this step, the operator “Read Excel” is used to import the dataset to be tested. This operator can be accessed from the Operators panel on the left. The imported data will serve as the main input for subsequent analysis tasks such as preprocessing, modeling, and evaluation.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

Then drag Read Excel from the operators section into the process view to process the data. Select Import Configuration Wizard and then select the training data that will be processed.

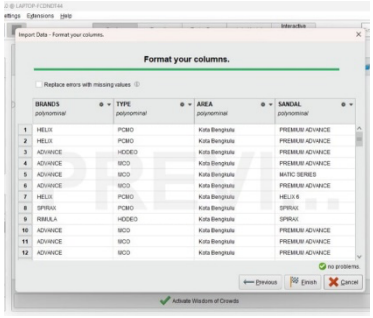

Figure 7. Data Import Wizard View

After dragging the Read Excel operator from the Operators section into the Process View, the next step is to configure the data import settings. As shown in Figure 7, by clicking the Import Configuration Wizard, users are guided to select the specific training data file (usually in .xlsx or .csv format) to be processed. This step ensures that the dataset is correctly structured and compatible with the analysis pipeline in RapidMiner.

Add the ID attribute with id and the description attribute with label and polynomial, then click Finish.

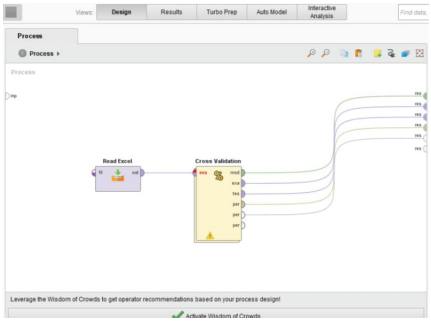

Figure 8. Read Excel and Validation Display in Main Process

After successfully importing the dataset, the next step involves setting the attributes required for processing. As illustrated in Figure 8, the ID attribute is set to id, and the label or target attribute (which will be predicted) is assigned and defined with the data type as polynomial. Once these configurations are completed, the Cross Validation operator is connected to the Read Excel operator. This step is crucial to evaluate the model performance through training and testing cycles using k-fold validation. After completing the setup, click Finish to proceed.

Type X-Validation in the search column in the operators section and drag it into the process as shown in the figure above. The reason for using the Cross Validation operator with polynomial validation is to obtain accurate results. In the validation box, double-click so that the process page for training and testing appears, as shown in the following figure.

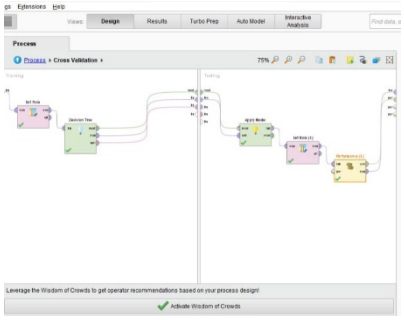

Figure 9. Validation Process View

As shown figure 9. After double-clicking the Cross Validation operator, the process editor splits into two sections: Training (left panel) and Testing (right panel). In the Training section, the Decision Tree operator is used to build the model using the labeled training data. In the Testing section, the model is applied to test data using the Apply Model operator, and the performance is measured with the Performance (Classification) operator.

This validation process helps evaluate the accuracy and effectiveness of the model. The configuration ensures the model is trained and tested properly across folds to avoid overfitting and to generalize better on unseen data.

For the final step, click the Run icon, and a display like the following will appear:

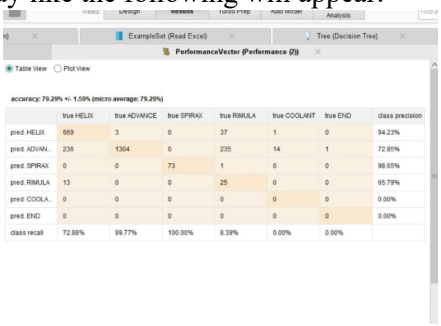

Figure 10. Accuracy Display in PerformanceVector

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

As shown in Figure 10, after clicking the Run icon, the system processes the data using the cross-validation method with the C4.5 algorithm. Figure 10 presents the PerformanceVector output, which includes the accuracy, error rate, and confusion matrix. The achieved accuracy is 79.29%, with an error rate of 11.87% across 7-fold validation.

The confusion matrix in Figure 10 also indicates the classification results for each class, showing how many instances were correctly or incorrectly predicted. This confirms that the C4.5 decision tree algorithm provides relatively good performance for the dataset used.

The following is the PerformanceVector Text View display:

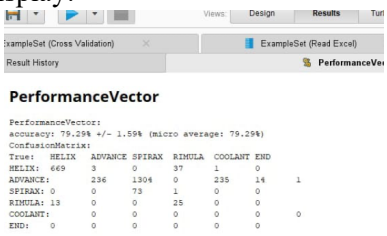

Figure 11. Text View Display in PerformanceVector

As shown in Figure 11, the PerformanceVector Text View displays the classification performance in a textual format, including accuracy and confusion matrix results.

3.3.2 Calculation of the C4.5 Algorithm and RapidMiner Testing

The use of the gain ratio method overcomes the weakness of the previous gain method, which could not be applied to continuous variables and to handle missing values in data, which later became known as the C4.5 algorithm [20].

Based on testing with manual calculations and RapidMiner software, different results were produced in determining the root of the decision tree. From the manual calculation results, the attribute SANDAL was obtained with a gain ratio of 0.75072863 [20].

3.3.3 Key Benefits and Practical Implications of the Research Results

The implementation of the C4.5 algorithm provides significant benefits for CV. Harapan Karya Mandiri, not only in terms of the accuracy of predictions but also in terms of practical applications in business operations:

1. Improved Stock and Inventory Planning: By identifying which products (such as Advance and Helix) have the highest sales potential, the decision tree helps CV. HKM predict future demand. This insight allows the company to optimize inventory levels, reducing the risk of both overstocking and understocking. Products with consistent demand can be stocked in higher quantities, while those with lower demand can be managed more carefully, leading to better cost efficiency.
2. Optimized Sales Strategy: The decision tree also provides clear segmentation of products based on attributes like brand and area, which allows CV. HKM to tailor its sales strategies. For example, products like Helix may require targeted marketing efforts in regions where they perform well, such as Bengkulu Selatan, while others, like Rimula, may need to be promoted more in Bengkulu Utara. This segmentation ensures that the right product reaches the right customer at the right time, enhancing the effectiveness of marketing campaigns.
3. Data-Driven Decision Making: The research emphasizes the power of data mining techniques to empower decision-making. With an accuracy of 79.29%, the algorithm's ability to predict future sales trends gives management actionable insights, ensuring that decisions on stock orders, marketing efforts, and even promotional campaigns are informed and effective. This leads to more strategic resource allocation and better overall business performance.
4. Increased Profitability: The ability to predict the best-selling products, forecast demand more accurately, and allocate resources more effectively ultimately results in higher sales, improved customer satisfaction, and maximized profits. By aligning stock with predicted demand and avoiding overstock, CV. HKM can reduce

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

wasted costs and improve its overall profit margin.

4. CONCLUSIONS

This research, titled "Data Mining Analysis of Shell Oil Sales Using the C4.5 Algorithm at CV. Harapan Karya Mandiri," provides the following conclusions based on the findings:

1. Analysis of Sales Patterns: Based on the results from applying the C4.5 algorithm, it was found that Shell oil with the brand Advance is sold more than other oil brands. This finding aligns with the research objective to identify sales patterns and trends using the data mining technique.
2. Accuracy of the Predictive Model: The C4.5 algorithm was successfully applied to analyze the Shell oil sales data, achieving an accuracy rate of 79.29%. This indicates that the model effectively classified sales patterns, fulfilling the aim of building a predictive model that can assist in forecasting customer demand and optimizing sales strategies.
3. Effectiveness of the C4.5 Algorithm: The C4.5 algorithm proved to be an effective method for analyzing and classifying Shell oil sales data. It demonstrated that decision tree models can accurately predict future sales trends based on historical data, supporting better decision-making for inventory planning and sales strategy development.
4. Cross-Validation Method: The use of the cross-validation operator in the testing phase enhanced the performance of the C4.5 algorithm, ensuring that the model's predictions were more reliable and generalizable. This method helped improve the accuracy of the model, aligning with the research objective to optimize sales forecasting and stock planning efficiency.
5. Root Node Identification: The manual calculation identified the SANDAL attribute as the root node of the decision tree with a gain ratio of 0.75072863, indicating its importance in determining the sales patterns. This finding contributes to a deeper understanding of the factors driving Shell oil sales
6. Best-Selling Products and Areas: The analysis also revealed that Advance is the best-selling brand, with PCMO being the most popular oil type. The region of Bengkulu City emerged as the area with the highest sales, while the Premium Advance variant was identified as the top-selling product. These insights are crucial for guiding sales strategies and resource allocation.

REFERENCES

[1] J. B. Ginting, "Kemajuan Teknologi Informasi dalam Perkembangan Bisnis Global Advances in Information Technology in Global Business Development," Student Scientific Creativity Journal (SSCJ), vol. 2, no. 4, p. 72, 2024.

[2] R. Purba, "Peran Teknologi Informasi Dalam Meningkatkan Efisiensi Operasional Bisnis Internasional," Digital Bisnis: Jurnal Publikasi Ilmu Manajemen dan E-Commerce, vol. 2, no. 4, p. 455, 2023.

[3] D. Septari, "Implementasi Machine Learning Untuk Prediksi Penjualan Oli Shell Pada CV. Harapan Karya Mandiri Bengkulu," Jurnal Media Infotama, vol. 20, no. 2, p. 425, 2024.

[4] M. P. U. O. P. P. S. H. C. V. B. K. M. M. B. K. S. D. K. I. & J. J. Learning, "Implementasi Machine Learning Untuk Prediksi," Jurnal Media Infotama, vol. 20, no. 2, p. 341139, 2024.

[5] A. N. A. d. H. Nurdiyanto, "Data Mining Clustering Dalam Pengelompokan Buku Perpustakaan Menggunakan Algoritma K-Means," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 4, no. 2, p. 156, 2023.

[6] S. A. Rahmah, "Review Terbaru Tentang Klasterisasi Data Mining Menggunakan Metode K-Means: Tantangan dan Aplikasi," Djtechno: Jurnal Teknologi Informasi, vol. 5, no. 4, p. 468, 2024.

[7] P. S. S. M. & R. F. Azura, "Analisa Penjualan Produk Oli Dengan Metode Data Mining Asosiasi Algoritma Apriori Pada PT. Mitra Petra Sejahtera," Jurnal …, vol. 2, no. 1, pp. 8–15, 2021.

[8] R. R., K. K., I. Zulfa, "Implementasi Data Mining untuk Menentukan Strategi Penjualan Buku Bekas dengan Pola Pembelian Konsumen Menggunakan

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56

Metode Apriori (Studi Kasus: Kota Medan)," TEKNIKA: Jurnal Sains dan Teknologi, vol. 16, no. 1, pp. 69–82, 2020.
[9] S. M., L. B. Alwendi, "Aplikasi Data Mining untuk Menentukan Lulusan Mahasiswa Tepat Waktu," Jurnal Advance Research Informatika, vol. 3, no. 1, p. 3025, 2024.
[10] R. C. N. S. Syafiatun Ihsani Luthfiyah, "Sistem Pendukung Keputusan (SPK) Penentuan Algoritma dan Metode Penelitian dengan Metode Simple Additive Weighting (SAW)," JIRE (Jurnal Informatika & Rekayasa Elektronika), vol. 5, no. 2, p. 6754, 2022.
[11] P. A. B. H. F. S. Kiki Ananda Mustari, "Implementasi Data Mining Pada Instansi Pemerintahan (Systematic Literature Review)," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 8, no. 3, p. 3354, 2024.
[12] N. Nuris, "Analisis Prediksi Harga Rumah Pada Machine Learning Menggunakan Metode Regresi Linear," EXPLORE, vol. 14, no. 2, pp. 108–112, 2024.
[13] R. Sari, "Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali," IJNS – Indonesian Journal on Networking and Security, vol. 5, no. 1, p. 768, 2016.
[14] N. Syaiful Zuhri Harahap, "Teknik Data Mining Untuk Penentuan Paket Hemat Sembako dan Kebutuhan Harian dengan Menggunakan Algoritma FP-Growth (Studi Kasus di Ulfamart Lubuk Alung)," Informatika: Jurnal Ilmiah Fakultas Sains dan Teknologi Universitas Labuhanbatu, vol. 7, no. 3, p. 2516, 2019.
[15] S. T. Sri Astuti, "Penerapan Data Mining Dalam Menentukan Penerima Beasiswa UPZ (Unit Pengumpulan Zakat) Menggunakan Algoritma K-Means," JSI; Jurnal Sistem Informasi (E-Journal), vol. 13, no. 2, p. 2289, 2021.
[16] M. & A. Hijrah, "Analisis RapidMiner dan Weka dalam Memprediksi Kualitas Kinerja Karyawan Menggunakan Metode Algoritma C4.5," Multi Data Palembang Journal, vol. 9, no. 2, p. 1965, 2022.
[17] H. A. Evicienna, "Algoritma C4.5 untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta," Techno Nusa Mandiri, vol. 9, no. 1, p. 675, 2013.
[18] H. & S. H. L. Syafputra, "Klasifikasi Penjualan Perhiasan Menggunakan Metode Decision Tree Algoritma C4.5 (Studi Kasus: Toko Emas Berkat Famili)," JISI, vol. 20, no. 2, p. 563, 2024.
[19] Syafi, M. Q., & Alamsyah. (2022). Increasing accuracy of heart disease classification on C4.5 algorithm based on Information Gain Ratio and Particle Swarm Optimization using Adaboost ensemble. Journal of Advances in Information Systems and Technology, 4(1), 100–112.
[20] M. T. F. Rusito, "Implementasi Metode Decision Tree dan Algoritma C4.5 untuk Klasifikasi Data Nasabah Bank," INFOKAM, vol. 13, no. 1, p. 5, 2015.

DOI : https://doi.org/10.21063/jtif.2025.V13.2.48-56