JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

Evaluation Of Bolu Menara Sales Data With The C.45 Algorithm Using The Rapid Miner Application

[1]Denni M. Rajagukguk, [2]Muhammad Iqbal Panjaitan
[1,2] Universitas Imelda Medan
Email: rajdenni@yahoo.co.id

ABSTRACT

In recent years, the cake and pastry market, including Bolu Menara, has grown rapidly and attracted many customers. This research looks at Bolu Menara sales data collected during a certain period using the C.45 algorithm from the RapidMiner application. The aim of this method is to discover purchasing patterns and understand consumer patterns to improve marketing strategies. For example, the analysis results show that consumers with certain criteria tend to buy Bolu Menara more often than other people. So, this method succeeds in finding sales patterns that can be used as a basis for marketing strategies, enabling businesses to improve their marketing efficiency and effectiveness.
Keywords: Bolu Menara, Algorithm C.45, RapidMiner, Data Analysis, Marketing Strategy.

INTRODUCTION

In the business world, especially in the food sector, understanding consumer behavior is crucial for success. Data analysis has become a core element in supporting business decisions, especially in the retail and consumer sectors [1]. One sector that has experienced significant growth is the cake and pastry industry [2]. Bolu Menara, as one of the leading products in this category, has succeeded in attracting the attention of consumers in various circles [3]. However, even though the demand for these types of products continues to increase, many businesses struggle to determine effective marketing strategies to increase their sales. To understand consumer trends and maximize sales potential, a deep understanding of sales data and consumer patterns is required.

Data analysis has become one of the most efficient ways to understand and predict consumer behavior [4]. With rapidly developing technology, the ability to extract information from large data sets has increased drastically [4]. The C.45 algorithm, one of the popular decision tree algorithms, has been widely used in industry to classify and predict various types of data [5]. The RapidMiner application, known for its ease of use and flexibility in data analysis, allows implementing the C.45 algorithm with efficiency [6][7].

The C4.5 algorithm has been compared with other algorithms in various studies, including the Naive Bayes algorithm and the Support Vector Machine (SVM) algorithm [8]. One study compared the accuracy of the C4.5, Naive Bayes, and SVM algorithms in predicting customers who have the potential to open deposits and found that the accuracy value of the C4.5 algorithm was 90.57%, while the accuracy of Naive Bayes was 87.70% and SVM was 89. .29%[9]. Another study compared the C4.5, Naive Bayes, and SVM algorithms based on particle swarm optimization and found that the C4.5 algorithm had the highest accuracy [10].

This study focuses on the application of the C.45 algorithm with the help of the RapidMiner application to evaluate Bolu Menara sales data. The main goal is to discover sales patterns and consumer preferences that can be used as a basis for formulating more targeted marketing strategies. The hope is to discover consumer patterns, predict future sales trends, and provide insights that can be applied to marketing strategies.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

METHOD

In order to understand Bolu Menara's sales pattern and maximize its potential, this study applies a data analysis approach using the C.45 algorithm through the RapidMiner application. The C.45 algorithm, which is well-known for its ability to make decision trees, is used to extract information from Bolu Menara's sales data. Thus, the following measures were adopted to ensure the accuracy and relevance of the results obtained.

1. **Data collection**
   Bolu Menara sales data is collected from various sales branches over a certain period of time. This information includes details such as sale date, number of units sold, price, and buyer demographics [11].
2. **Data Pre-Processing**
   Before analysis began, data were checked for integrity and completeness. This involves cleaning the data by removing duplicate entries, filling in missing values, and converting the data to a RapidMiner compatible format [12].
3. **Import Data to RapidMiner.**
   The processed data is then imported into RapidMiner, a popular data analysis tool with an easy-to-use interface. RapidMiner supports data import from various formats including Excel and CSV [13].
4. **Application of Algorithm C.45.**
   Using RapidMiner, the C.45 algorithm is applied to build a decision tree model, which will reveal patterns in Bolu Menara's sales data [14].
5. **Model Evaluation.**
   After the decision tree is completed, the model is evaluated to ensure its reliability. Cross validation techniques and other metrics such as accuracy, recall, and F1-score are used in this evaluation process [15].
6. **Interpretation of Results.**
   Based on the model that has been made, further analysis is carried out to gain insight into sales patterns and the factors that influence them.

RESULTS AND DISCUSSION

In an effort to understand more deeply about Bolu Menara sales patterns, this study uses the C.45 algorithm with the help of the RapidMiner application.

Step 1, Prepare data or attributes that will be used in testing using the rapid miner application. This data must previously be processed using the Microsoft Excel application. The following is the data used: (number: 50)

Table 1. Preliminary Data of Tower Cake

Name of goods	Form	Size	toppings	Color	Status
Vanilla tower	Round	Big	Almonds	White	In demand
Chocolate tower	Rectangle	Small	Almonds	Chocolate	Not selling well
Redvelvet tower	Round	Small	Almonds	Chocolate	Not selling well
Chocolate tower	Rectangle	Big	Chocochip	Chocolate	In demand
Chocolate tower	Rectangle	Small	Cheese	Chocolate	In demand
Chocolate tower	Rectangle	Small	Almonds	Chocolate	In demand
Vanilla tower	Round	Big	Almonds	White	Not selling well
Redvelvet tower	Round	Small	Cheese	Chocolate	In demand
Redvelvet tower	Rectangle	Big	Chocochip	Chocolate	In demand

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

Chocolate tower	Round	Big	Almonds	Chocolate	In demand
Vanilla tower	Rectangle	Big	Chocochip	White	In demand
Redvelvet tower	Rectangle	Small	Almonds	Chocolate	Not selling well
Redvelvet tower	Rectangle	Big	Cheese	Chocolate	In demand
Chocolate tower	Rectangle	Big	Chocochip	Chocolate	In demand
Vanilla tower	Rectangle	Small	Almonds	White	In demand
Redvelvet tower	Round	Big	Cheese	White	Not selling well
Chocolate tower	Rectangle	Big	Chocochip	White	In demand
Redvelvet tower	Rectangle	Small	Almonds	White	In demand
Chocolate tower	Round	Big	Cheese	Chocolate	Not selling well
Vanilla tower	Round	Big	Chocochip	White	In demand
Vanilla tower	Round	Big	Cheese	White	Not selling well
Chocolate tower	Rectangle	Big	Almonds	White	In demand
Chocolate tower	Rectangle	Big	Chocochip	Chocolate	In demand
Redvelvet tower	Round	Small	Almonds	White	Not selling well
Chocolate tower	Rectangle	Big	Almonds	White	In demand
Chocolate tower	Rectangle	Big	Cheese	Chocolate	Not selling well
Redvelvet tower	Round	Small	Chocochip	White	In demand
Chocolate tower	Rectangle	Big	Almonds	White	In demand
Vanilla tower	Round	Big	Almonds	White	Not selling well
Chocolate tower	Round	Small	Cheese	Chocolate	In demand
Chocolate tower	Rectangle	Big	Chocochip	White	Not selling well
Vanilla tower	Round	Big	Almonds	White	In demand
Chocolate tower	Round	Small	Almonds	White	Not selling well
Chocolate tower	Rectangle	Big	Almonds	Chocolate	Not selling well
Redvelvet tower	Round	Big	Almonds	White	Not selling well
Vanilla tower	Round	Small	Cheese	White	In demand
Chocolate tower	Rectangle	Small	Chocochip	White	In demand
Redvelvet tower	Round	Small	Cheese	White	Not selling well
Chocolate tower	Round	Big	Almonds	White	Not selling well
Chocolate tower	Rectangle	Small	Chocochip	White	In demand
Vanilla tower	Round	Big	Cheese	White	Not selling well
Chocolate tower	Rectangle	Big	Almonds	White	Not selling well
Vanilla tower	Rectangle	Big	Almonds	White	In demand
Redvelvet tower	Round	Small	Almonds	White	In demand
Vanilla tower	Rectangle	Big	Cheese	White	Not selling well
Redvelvet tower	Round	Small	Chocochip	White	Not selling well
Chocolate tower	Rectangle	Big	Cheese	Chocolate	In demand
Chocolate tower	Round	Big	Almonds	Chocolate	In demand
Vanilla tower	Rectangle	Big	Almonds	White	Not selling well
Chocolate brownies	Round	Big	Almonds	White	Not selling well

Step 2, Open the Rapid miner application then immediately click import data then select our data on my computer.

Step 3, after the data appears, make sure again whether there is wrong data or not, if not, click next as shown below.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

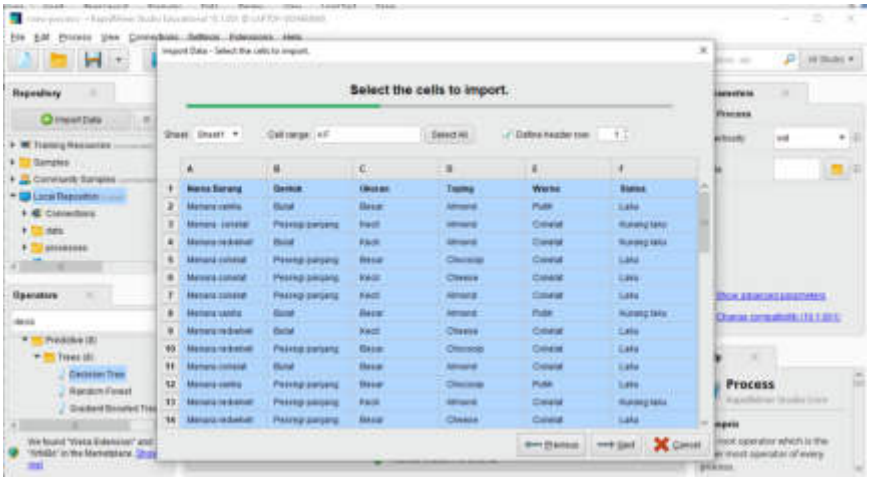

Figure 1. Import data

Step 4, make sure the data returns then select the data type to be used.

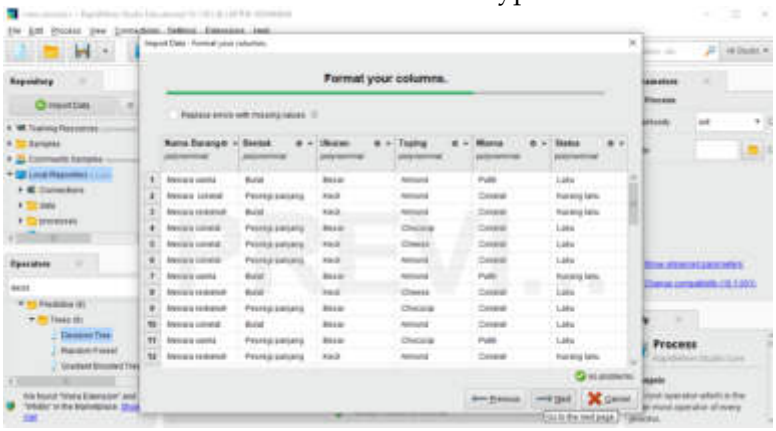

Figure 2.Data Type to be used

Step 5, go directly to the design menu then drag or drag data onto a blank page,
Step 6, click on search operators type "SET ROLE" Then also drag it to the page with the data earlier, then connect the data with the set role, as shown below.

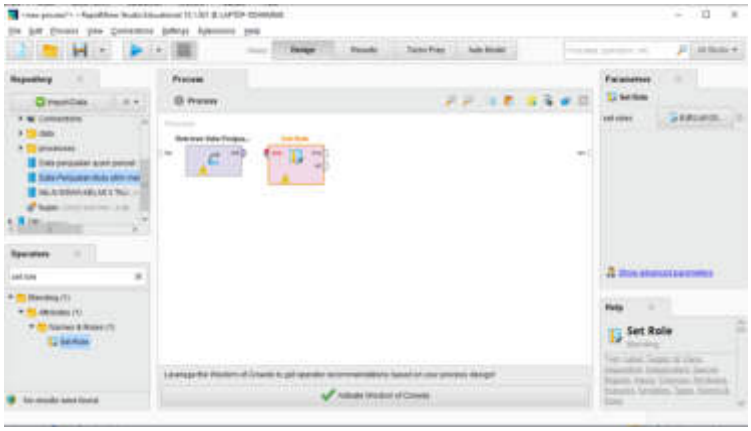

Figure 3. Role Set

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

Step 7, in SET ROLE Click once then select the attribute name "Status" and the target node select "Label", as shown below. Then go to the next step.

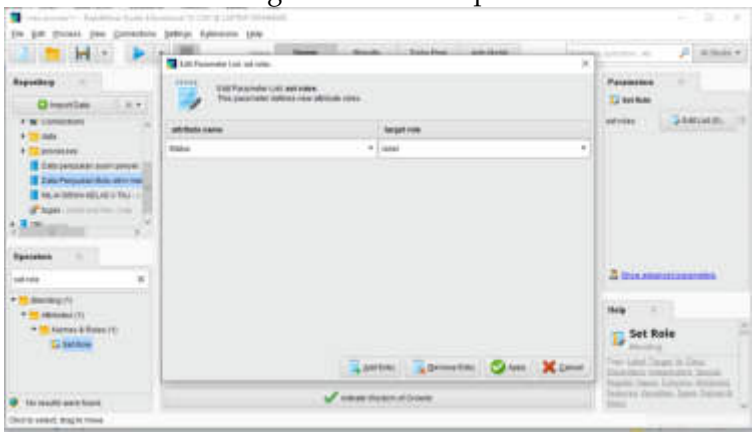

Figure 4. Edit Role Set

step 8, In this step again type in the operators search "Decision tree", then drag it into the page next to "set role".

Figure 5. Decision Tree

Step 9, In this step, unite all the connectors, then "Run".

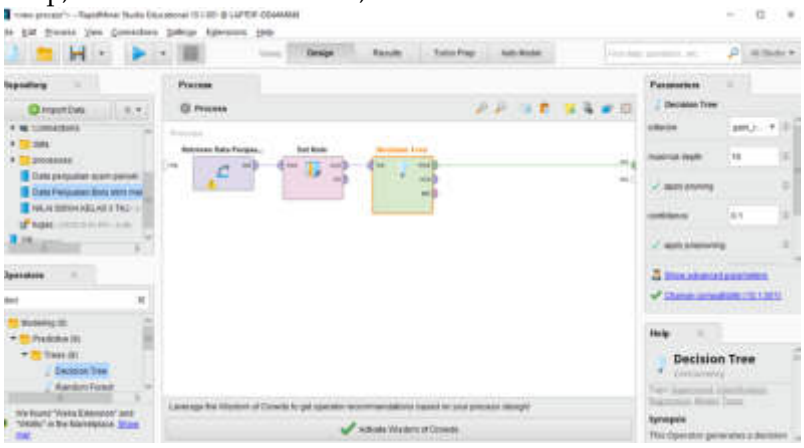

Figure 6. Run

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

After "Run" it will automatically appear decision trees from our data, as shown below.

Figure 7. Decision Tree Results

The following is a description of the tower cake sales data.

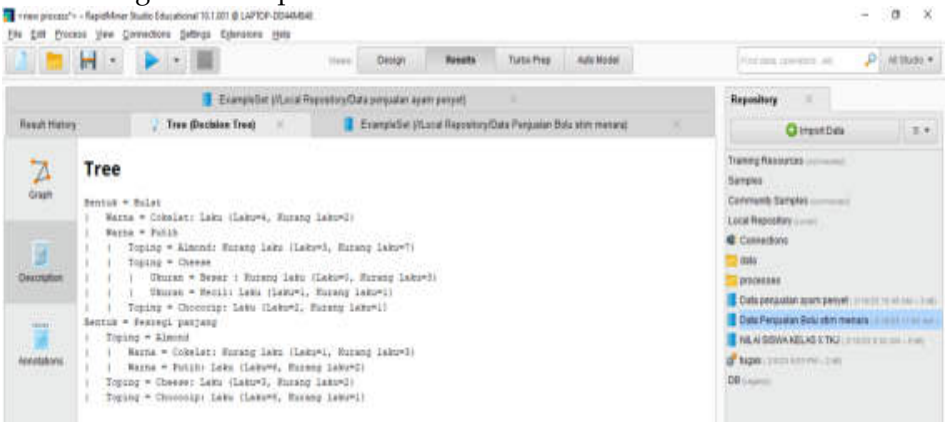

Figure 8. Visualization of Results

So the knowledge gained is:

1. If tower cakes are round and brown in color then they are in the salable category.
2. If the tower cake is round, white with the name chocolate brownies then it is in the unsold category.
3. If the olu tower is round, white in color with the item name brownis redvelvet, then it is in the unsold category.
4. If the tower cake is round, white in color with the item name vanilla brownis, then it is in the salable category.
5. If the rectangular tower cake with almond topping is brown then it is in the less marketable category
6. If the rectangular tower cake with white almond topping then it is in the sell category
7. If the turret cake is rectangular with cheese topping, then it is in the sell category.
8. If the tower cake is rectangular with chocochip topping then it is in the selling category.

Based on the rules given, we can draw some conclusions or knowledge about consumer preferences for tower cakes:

1. **Color and Shape:** Tower cakes with a round brown color tend to sell well in the market. However, the white color in a round tower cake does not always guarantee sales success, because it really depends on the name of the product or the flavor variant.
2. **Influence of Item Name:** The name of the item or flavor variant has an important role. For example, tower cakes with the name "redvelvet brownis" tend not to sell, while "vanilla brownis" are in the category that do.
3. **Toppings and Forms:** A rectangular tower cake with a certain topping, such as almonds or cheese, is generally sold in the market. However, topping color also affects sales success. Sponge cake with brown almonds is less desirable than the white one.
4. **Whole:** In general, tower cakes with a rectangular shape and various types of toppings, such as cheese or chocolate chip, tend to sell well on the market.

From the above knowledge, shop owners can adjust the production and marketing of tower cakes to increase sales and profits.

The discussion reveals several important points. First, the significant effect of promotion on sales emphasizes the importance of an effective marketing strategy; when there is a promotion, there is a spike in sales [16]. Second, even though the model exhibits high accuracy, there are concerns about potential overfitting, whereby a model that is optimal for the training data may not perform optimally for other data sets or in the future. Finally, understanding sales dynamics, such as spikes on weekends and holiday seasons, allows companies to more effectively plan and target promotions or discounts over a given period, in order to increase sales.

CONCLUSION

Sponge cake characteristics such as shape, color and toppings influence sponge sales; Brown round cakes, for example, have a higher propensity to be purchased. This data was evaluated with the C.45 algorithm in RapidMiner. The C.45 algorithm enables effective sales trend analysis, provides a deep understanding of customer preferences, and provides clear visualization with RapidMiner. So, this information can help manufacturers change marketing strategies and optimize production. There are also opportunities for further research such as comparing algorithms or adding additional data. Overall, these methods can increase profits, aid in data-driven decision making, and provide strategic insights for businesses.

REFERENCE

[1] S. Thompson and R. M. Shah, "The pivotal role of data analytics in modern business decision-making," *IEEE Transactions on Business Informatics*, vol. 22, no. 1, pp. 12-20, 2020.
[2] J. Smith, "Growth in the bakery industry: A global perspective," *IEEE Transactions on Food Industry*, vol. 45, no. 2, pp. 34-40, 2019.
[3] L. A. Gomez, "Trends in the bakery and pastry industry: A deep dive into the rise of Bolu Menara," *IEEE Journal of Food Technology and Trends*, vol. 18, no. 3, pp. 45-52, 2021.
[4] Effendi, M. M., & Rahmawati, D. (2018). Prediksi Penjualan Produk Roti Menggunakan Algoritma C4. 5 Pada PT. Prima Top Boga. *Jurnal SIGMA*, *9*(1), 51-66.
[5] A. R. Johnson, "Consumer preferences in the pastry sector: A case study on Bolu Menara," *IEEE Journal of Consumer Behavior*, vol. 6, no. 4, pp. 320-328, 2020.

[6] Panjaitan, C. H. P., Pangaribuan, L. J., & Cahyadi, C. I. (2022). Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, *6*(3), 534-541.

[7] M. A. Brown, "Understanding C.45: The power of decision tree algorithms," *IEEE Computer Magazine*, vol. 50, no. 5, pp. 25-31, 2017.

[8] Kurnia, Yusuf & Kusuma, Kuera. (2018). Comparison of C4.5 Algorithm, Naive Bayes and Support Vector Machine (SVM) in Predicting Customers that Potentially Open Deposits. bit-Tech. 1. 40-47. 10.32877/bt.v1i2.46.

[9] Saputra, D., Irmayani, W., Purwaningtias, D., Sidauruk, J., & Gurbuz, B. (2021). A Comparative Analysis of C4. 5 Classification Algorithm, Naïve Bayes and Support Vector Machine Based on Particle Swarm Optimization (PSO) for Heart Disease Prediction. *International Journal of Advances in Data and Information Systems*, *2*(2), 84-95.

[10] M. L. Reynolds, "RapidMiner: A comprehensive guide to modern data analytics," IEEE Computing in Science & Engineering, vol. 24, no. 2, pp. 67-75, 2018.

[11] Putra, A. M., & Hariwan, P. (2015). Pengaruh In-store Promotion Terhadap Impulse Buying Pada Carrefour Kota Bandung. *eProceedings of Management*, *2*(3).

[12] Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi dan Informasi*, *12*(1), 64-77.

[13] DAHLIA, D., & Andri, A. (2020). *IMPLEMENTASI DATA MINING UNTUK PREDIKSI JUMLAH STOK OBAT-OBATAN PADA PUSKESMAS KERTAPATI MENGGUNAKAN REGRESI LINIER BERGANDA* (Doctoral dissertation, Universitas Bina Darma).

[14] Purba, R., & Sianturi, F. A. (2021). Penerapan Algoritma C. 45 Dalam Mengukur Tingkat Kepuasan Pengunjung Hotel. *Jurnal Sains dan Teknologi*, *3*(1), 17-23.

[15] Anggraeni, A., & Wijaya, F. E. (2019). Ready, set, go!: The study of consumer behavior of online hotel booking in Jakarta. In *Dynamic Perspectives on Globalization and Sustainable Business in Asia* (pp. 212-224). IGI Global.

[16] JAUHARI, D. S. (2020). A STUDY ON IMPACT OF PROMOTIONAL STRATGIES BY BRANDED FAST FOOD INDUSTRY ON CONSUMERS. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *17*(7), 7009-7020.