Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Application of Decision Tree Algorithms to Classify the Sales Results of Kangen Kripik Sme Products

Adila G Khiqmatiar Muchsin[1*], Nining Rahaningsih[2], Irfan Ali[3], Dadang Sudrajat[4], Saeful Anwar[5]

[1,2,3,4,5]*Teknik Informatika, STMIK IKMI Cirebon, Indonesia*
akbarmiftahul569@gmail.com[1*], niningr157@yahoo.co.id[2], irfanaali0.0@gmail.com[3], d42ngsudrajat2012@gmail.com[4], 2021saeful@gmai.com[5]

Abstract

Micro, Small, and Medium Enterprises (MSMEs) play a vital role in strengthening the national economy; however, many still face challenges in managing and analyzing sales data effectively. This study aims to classify product sales results at UMKM Kangen Kripik Mang Acep by applying the Decision Tree algorithm as a data classification method based on machine learning. A quantitative experimental approach was employed to evaluate the model's performance using one-year sales data, including attributes such as product variants, sales volume, sales channels, and marketing regions. Data processing was conducted using RapidMiner software following the Knowledge Discovery in Databases (KDD) framework, which includes data selection, preprocessing, transformation, data mining, and model evaluation. The results indicate that the Decision Tree algorithm successfully classified sales regions (Garut, Bandung, and Sumedang) with an accuracy rate of **96.48%**, identifying "Units Sold (pcs)" as the most influential attribute for distinguishing marketing areas. These findings demonstrate that the Decision Tree method is not only effective in improving data analysis efficiency but also provides valuable strategic insights for data-driven business decision-making in MSMEs

Keywords: *Decision Tree;data mining, classification; sales prediction; MSME*

1. Introduction

Advances in information and communication technology have resulted in the accumulation of vast amounts of data across various industries. To generate useful knowledge to aid decision-making, this large amount of data requires effective analytical methods. Data-driven decision making has become a strategic necessity for businesses in the modern era to improve competitiveness and operational efficiency [1]Businesses benefit from data mining because it helps them understand market patterns, sales trends, and customer behavior. Through sales data analysis, companies can find the most popular products, peak purchasing times, and customer segmentation based on characteristics and regions [2] By using this process, businesses can create more effective marketing strategies, optimize product inventory, and provide better customer service. Therefore, the ability to convert unused data into usable information is a strength that makes contemporary organizations more competitive [3]

Decision Trees can be used by companies to support strategic decisions based on historical data, predict customer categories, and analyze elements that influence sales [4]. The Decision Tree Method for Sales Classification in Wholesale Stores. The C4.5 Decision Tree algorithm is used in this study to classify sales levels in a large supermarket. Product category, stock quantity, purchase frequency, and transaction time are the input factors. The results show that the "purchase quantity" attribute has the most influence on purchasing decisions and stock rotation, with a model accuracy rate ranging from 88 to 90 percent. According to the results of this study, Decision Tree has the ability to provide interpretations of rules that are easily understood by store managers. Thus, using Decision Tree to improve the restocking process and inventory management more efficiently is something that store managers can do.[5]

2. Page layout

This research was conducted through a series of systematic stages designed to ensure that the entire data analysis process proceeded in a structured, measurable, and replicable manner. Each phase was organized to support methodological rigor and to maintain consistency throughout the study. The sequence of research stages is illustrated in Figure 1.

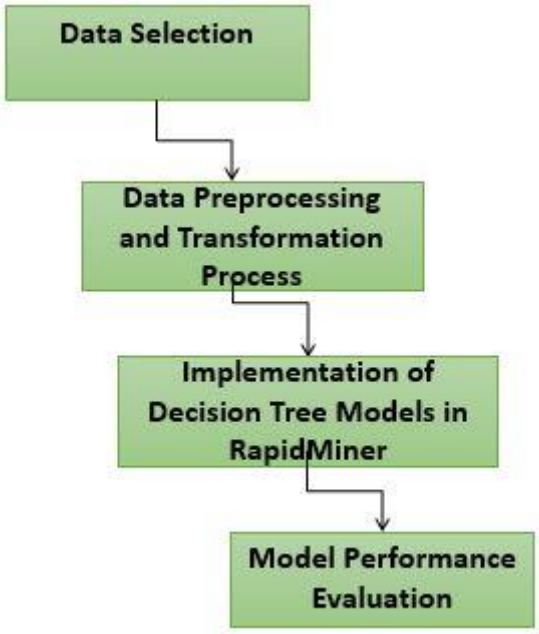


Fig.1 Research Stages

2.1. Data Selection

The first stage of the research involved data Selection, At this stage, researchers conduct an initial selection of relevant data sources and attributes for analysis. The main objective is to select a representative subset of data and variables that have the potential to generate useful insights, so that subsequent analysis is efficient and valid [6] This process includes identifying data entities, defining input and target variables, and selecting data formats and ranges that are appropriate for the research objectives.

2.2 Data Preprocessing and Transformation Processes

This stage consists of two interrelated parts: pre-processing and data transformation. Pre-processing involves cleaning the data of duplicates, handling missing values, and standardizing the data format so that its quality is adequate for further analysis. After that, data transformation is carried out by changing the data structure—such as encoding categorical variables, normalizing numerical scales, and feature engineering—so that the classification algorithm can work optimally [7] Both of these parts are crucial because incorrect decisions at this stage can hinder the quality of the model and the final interpretation.

2.3 Implementation of Decision Tree Models in RapidMiner

The technical implementation was carried out using RapidMiner Studio software. The implementation procedure included: data import, target and input attribute determination (set role), data division, modeling with the Decision Tree operator, and model testing. The Decision Tree algorithm was chosen for its ability to produce classification rules that are easy to understand and interpret [8]

1. Import Data
   Use the Read Excel operator to import the sales dataset.
2. Set Role & Select Attributes
   Determine the target attribute (label) and independent variables.
3. Decision Tree Operator
   Build a classification model using the parameters criterion = gain ratio, minimum gain = 0.01, and max depth = 20.
4. Apply Model
   Use the model to classify the test data.
5. Performance (Classification)
   Calculate performance metrics to measure the accuracy of the model.

The resulting Decision Tree model shows the relationship between variables and generates rules that explain sales patterns based on marketing regions.

2.4 Model Performance Evaluation

The evaluation and interpretation stage is the final stage to measure the extent to which the model built can be used practically and provide significant results. Evaluation uses metrics such as accuracy, precision, recall, and F1-score to provide a quantitative picture of the model's performance [9] In addition, interpreting the results through a confusion matrix and analysis of the rules generated by the decision tree model helps decision makers understand the patterns found in the data. Comprehensive evaluation ensures that the model is not only statistically "good" but also relevant and applicable in a business context.

3. Result and Discussion

3.1. Data Selection

Data selection is the first step in the data processing process, which aims to select relevant and necessary attributes for the analysis process. The data selection process in this study was conducted using RapidMiner Studio software. The data used included 199 records of goods sales in several locations, such as Garut, Bandung, and Sumedang.

From the data, the attributes to be used are Month, Flavor Variant, and Sales Volume, while Sales Location is used as the label. A complete overview can be seen in the following image.

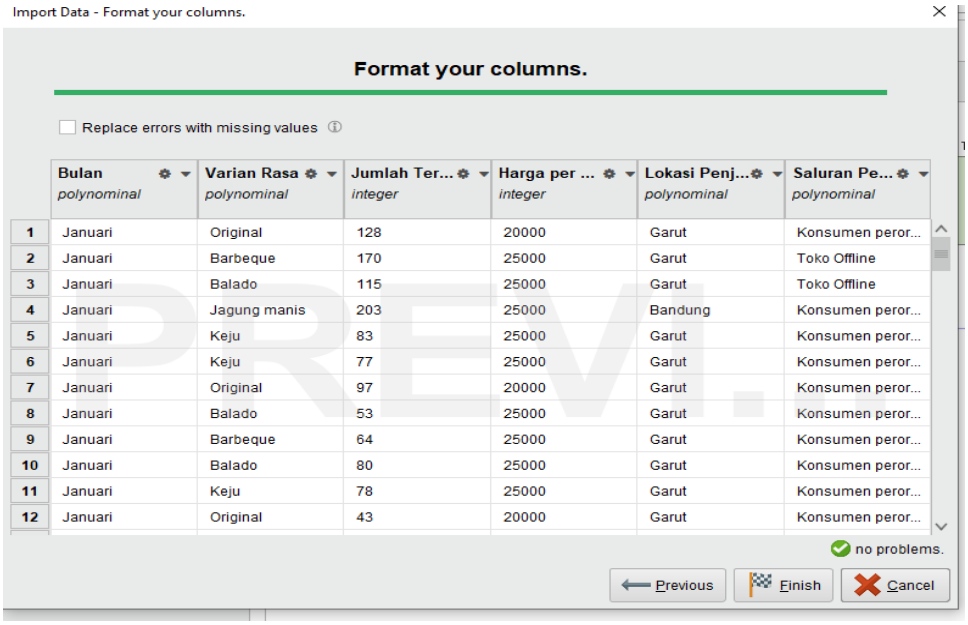
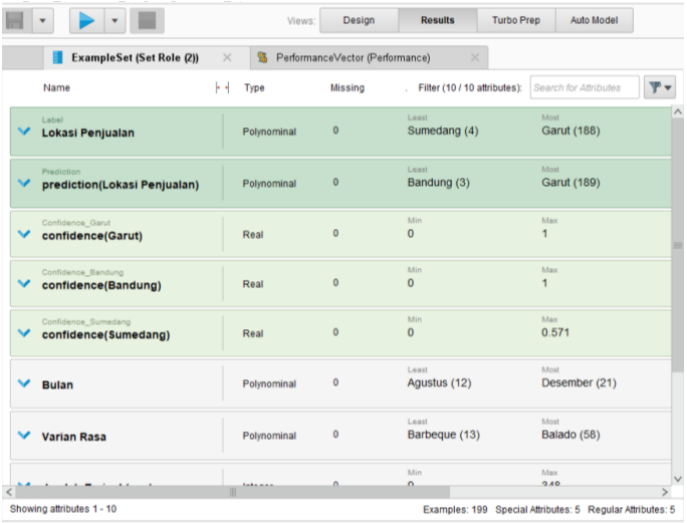


Fig. 2. Data Selection

3.2. Data Preprocessing and Transformation Processes

3.2.1 Prepocessing



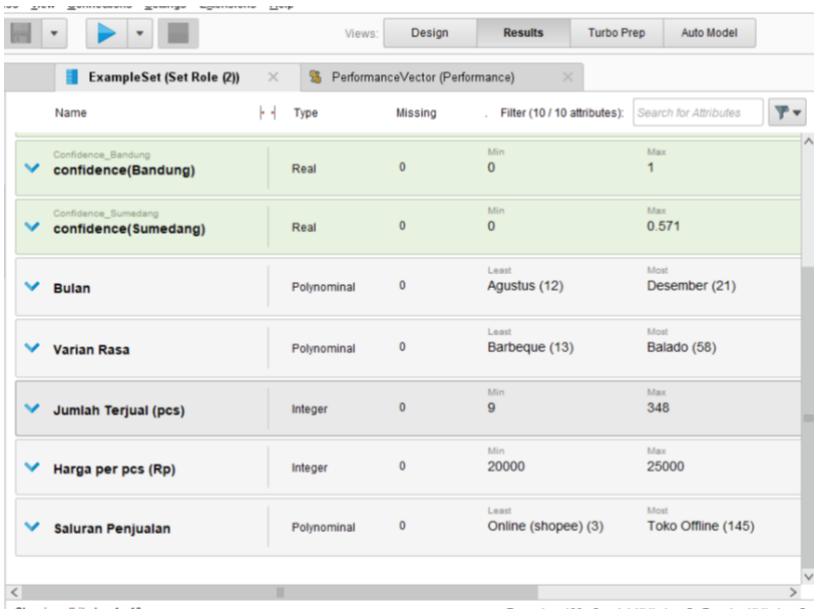


Fig. 3. Statistical result

Cleaning or data cleansing in the preprocessing step that is missing or has inconsistent values Before starting this process, an analysis is carried out to ensure that the selected dataset has consistent data and that the attributes have no missing values. There are no attributes with missing values, as shown by the dataset statistics, as shown in the image above. The dataset used is checked directly for each record to ensure that the data is consistent with its values.

3.2.1 Transformation

The purpose of the transformation is to prepare the data so that it is ready to be used to build a Decision Tree model. In this study, the data was imported from an Excel file using the Read Excel operator, then certain attributes, such as "Month," "Flavor Variant," and "Sales Amount," were selected using Select Attributes. Next, the Set Role operator was used to assign roles to attributes in the analysis, for example, to determine the target attribute.

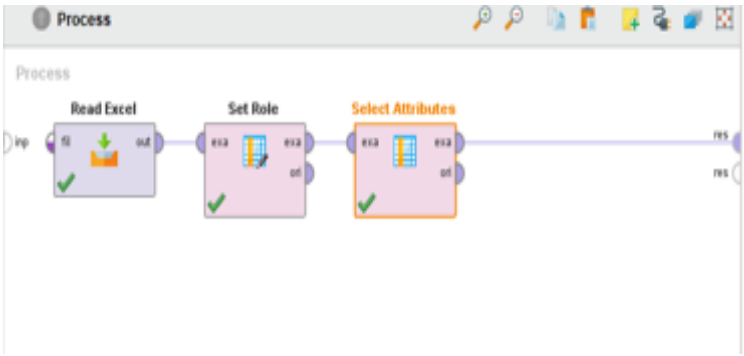


Fig. 4. 1. Process model up to the transformation step

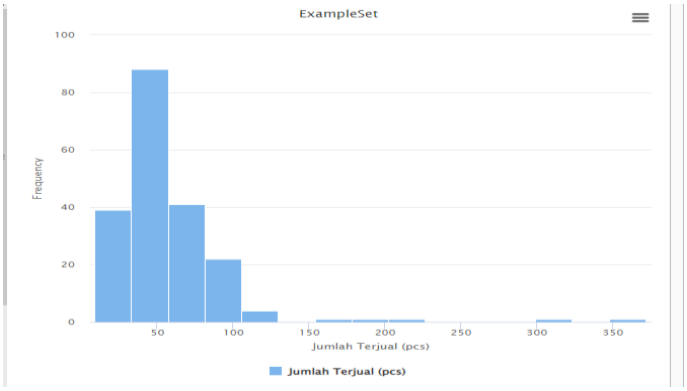


Fig. 4. 2. Histogram chart

This graph shows that the data has a fairly normal distribution, there are no missing values, and it can be further processed using the Decision Tree algorithm without the need for additional complex transformations. This distribution also supports the finding that "Number Sold (pcs)" is the most dominant attribute in determining the marketing area classification results.

3.3. Implementation of Decision Tree Models in RapidMiner

Fig. 5. Operator decision tree

This algorithm is used to train models with learning data. After the attributes are selected, this process builds a decision tree. Each branch of the tree shows the rules created to separate data according to specific features. After the Decision Tree algorithm is trained to create a decision tree based on patterns in the data, the Decision Tree operator creates rules to predict sales results based on selected attributes. Once the model is trained, the Apply Model operator is applied to the test data to measure the model's progress.

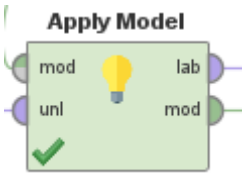


Fig. 6. Apply model operator in RapidMiner

To predict sales results, the trained model is applied to the test data. The result of this stage is a decision tree model, which can be used for prediction or classification.

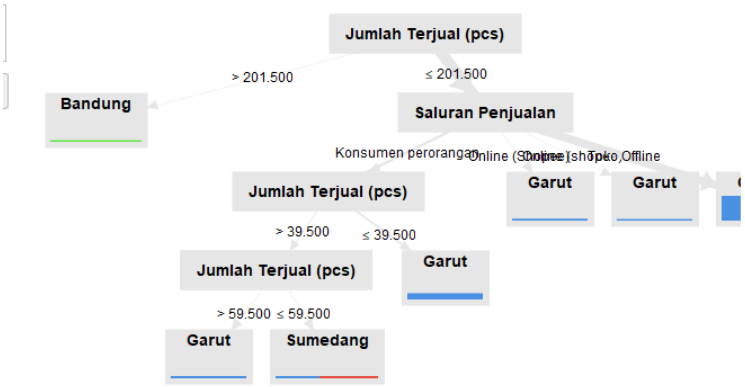


Fig. 7. Rule results on predicted data

Graphs and other numbered figures should appear throughout the text as close to their mention as possible. Figures shouldn't infringe upon the page borders. To measure the performance of classification activities on the dataset, the Performance (Classification) operator is used. Operator Performance (Classification), as shown in the image below.

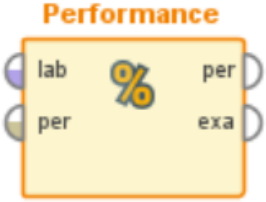


Fig. 8. Operator Performance (Classification) in Rapidminer

The performance of the classification model that has been built is evaluated using the Performance Operator (Classification) in RapidMiner. The process model in RapidMiner up to the Data Mining step is shown in the image below.

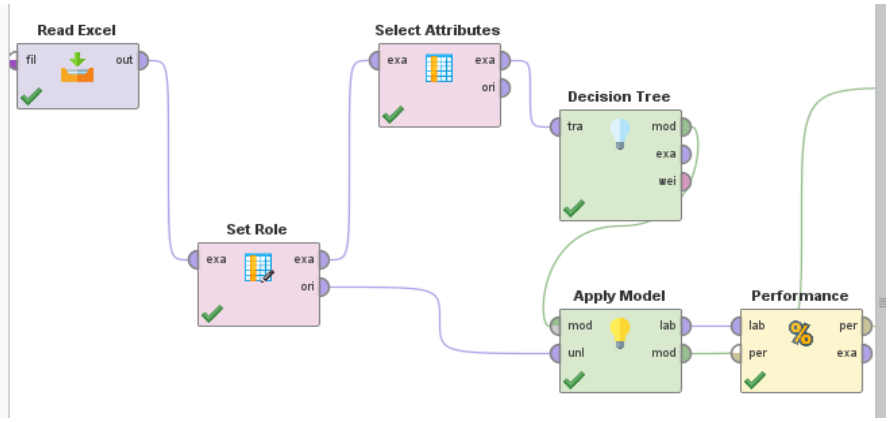


Fig. 9. Process model in RapidMiner up to the Data Mining step

3.4. Model Performance Evaluation

The evaluation conducted on the results of the experiment on the dataset obtained the following results.

a. Model accuracy
With an accuracy of 96.48%, this model demonstrates remarkable ability to classify data.

b. Confusion Matrix

Table View Plot View

accuracy: 96.48%

	true Garut	true Bandung	true Sumedang	class pre
pred. Garut	185	4	0	97.88%
pred. Bandung	0	3	0	100.00%
pred. Sumedang	3	0	4	57.14%
class recall	98.40%	42.86%	100.00%	

Fig. 10. confusion matrix

The performance of the Decision Tree classification model when using RapidMiner to analyze sales data. The evaluation results are displayed in the form of a confusion matrix that shows the comparison between the true class and the predicted class for three sales regions: Garut, Bandung, and Sumedang.

The classification results are described as follows: For the Garut class, there were 185 correctly predicted data out of 188 actual data, with a recall class value of 98.40% and precision of 97.88%. For the Bandung class, all data identified as Bandung data were correctly predicted by the model, with a precision value of 100% and a recall class of 42.86%, indicating that most of the Bandung data are still classified.

c. Precission and Recall

Tabel 2 . Precission and Recall

Kelas	Precision	Recall
Garut	**97,88%**	98,40%
Bandung	**100,00%**	42,86%
Sumedang	**57,14%**	100,00%

Each area class has a different precision and recall value, according to the Decision Tree model evaluation results. The Garut class's accuracy is 97.88%, while Bandung's is 100%. This suggests that the model can lower erroneous positive predictions, particularly for Bandung, where every forecast is correct. Nonetheless, the model can effectively identify nearly all Garut data, as evidenced by the Garut class recall of 98.40%. However, Bandung's recall of only 42.86% indicates that a large portion of the data in Bandung is incorrectly identified, leading to its classification as a different class.

4. Conclusion

The Decision Tree algorithm effectively classifies sales outcomes with an accuracy of 96.48%. The number of units sold is identified as the most influential factor in determining sales regions. The model can assist MSMEs in making informed decisions regarding product distribution and marketing strategies.

Future studies may incorporate algorithms such as Random Forest or SVM to improve classification performance.

References

[1] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Cambridge, MA: Morgan Kaufmann, 2022.

[2] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge, MA: Morgan Kaufmann, 2021.

[3] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA: MIT Press, 2020.

[4] Y. Zhang, L. Zhang, and W. Wang, "Application of Decision Tree Algorithms in Sales Forecasting and Strategic Business Analysis," *Journal of Business Analytics*, vol. 6, no. 2, pp. 101–117, 2023, doi: 10.1080/2573234X.2023.114567.

[5] A. I. Shafarindu, A. W. Wardhana, and E. Patimah, "Sales Data Classification in Supermarkets Using the Decision Tree Method," in *Seminar Nasional Mahasiswa Informatika dan Komputer (SENAMIKA)*, Depok, Indonesia: Universitas Pembangunan Nasional Veteran Jakarta, 2021. [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1389

[6] S. Kaknjo and L. Turulja, "Data Selection Strategies for Machine Learning Efficiency," *(Nama jurnal tidak disebutkan)*, 2025.

[7] M. Hochkamp, "Feature Engineering for Optimal Machine Learning Performance," *(Nama jurnal tidak disebutkan)*, 2025.

[8] A. Riyadi and M. Fahmi, "Implementasi Algoritma Decision Tree di RapidMiner," 2024.

[9] et al. Mukhlisin, "Evaluasi Kinerja Model Decision Tree," 2025.