International Journal Software Engineering and Computer Science (IJSECS)
4 (3), 2024, 943-952
Published Online December 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i3.2999.

RESEARCH ARTICLE **Open Access**

Application of Decision Tree Method for Sales Prediction at PT. Cipta Naga Semesta (Mayora Group) North Jakarta for 2023

Richardviki Beay
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: beayrichard@gmail.com.

Frencis Matheos Sarimole *
Informatics Engineering Study Program, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: matheosfrancis.s@gmail.com.

Received: July 29, 2024; Accepted: October 20, 2024; Published: December 1, 2024.

Abstract: The purpose of this study is to forecast sales of PT. Cipta Naga Semesta, one of the companies owned by Mayora Group headquartered in North Jakarta using the Decision Tree method during 2023. Decision Tree was chosen because this model identifies key attributes that greatly affect sales in the data and has the ability to predict outcomes by recognizing patterns in historical data. The database used in this analysis includes monthly records of sales, promotions, prices, and other economic characteristics. The findings of the study indicate that the Decision Tree method is very effective in providing accurate sales predictions with a low margin of error. The forecast provides valuable perspectives for company management, which can help them design tighter sales strategies and make better inventory decisions, thereby maximizing operational efficiency and profitability. In addition, the exploration of sales prediction models is one of the future works proposed in this study, which recommends practitioners to explore alternative methods to improve forecast accuracy and robustness.

Keywords: Decision Tree; Sales Forecasting; Data Analysis Techniques; Predictive Modeling; Operational Efficiency.

1. Introduction

Competition in the business world is getting tighter, and the ability to predict sales with a high degree of accuracy is very important for effective strategic decision making. Accuracy in projecting sales allows companies to better plan production, manage inventory, and design targeted marketing strategies. PT. Cipta Naga Semesta, which is part of the Mayora Group and located in North Jakarta, realizes the importance of implementing technology and data-based analysis to improve operational efficiency and company competitiveness. One method that can be applied for this purpose is the Decision Tree method, which has proven effective in various fields of predictive analysis and data mining. Decision Tree is a technique used to classify and predict by breaking data into several branches based on decisions taken based on certain attributes. The main advantage of this method is its ability to handle complex data and easy-to-understand results. With a clear visual representation, Decision Tree allows decision makers to identify the factors that most influence the results obtained. Therefore, the application of this method at PT. Cipta Naga Semesta can help companies analyze sales patterns and determine more appropriate steps in planning marketing and sales management in the future.

This study aims to apply the Decision Tree method in predicting sales of PT. Cipta Naga Semesta in 2023. This study not only focuses on predicting sales volume, but also aims to identify factors that influence the company's sales performance, such as market trends, consumer preferences, and internal factors related to the company's operations. By utilizing this technique, it is hoped that PT. Cipta Naga Semesta can obtain clearer information regarding the factors that need to be optimized to increase sales.

In previous research, Bahri, Harahap, and Nasution (2024) stated that the application of technology in business analysis, especially for processing big data, is increasingly needed to create efficiency and increase the effectiveness of company operations. They also emphasized the importance of the role of data mining in obtaining information that can be used to support more appropriate business decisions [1]. Rimbarizki and Susilo (2017) added that the use of technology in business not only increases efficiency but also allows companies to be more responsive to changes in the market [2]. The Decision Tree method has many applications in various fields, including sales analysis. According to Mining (2006), this technique has been widely used in various industries to predict trends and patterns that emerge from historical data [3]. One of the main advantages of Decision Tree is its ability to provide transparent and understandable results, allowing company management to make decisions based on easily interpreted information.

In the business realm, the application of the Decision Tree method is also very relevant to sales management strategies. Hartatik *et al*. (2023) in their book discussing data science for business, explains that analysis techniques such as Decision Tree can provide a more structured picture in operational and marketing management [4]. They also explain that by using this method, companies can identify patterns in sales data that may have previously gone undetected, which will be useful for developing more targeted marketing strategies. On the other hand, research by Novareza *et al*. (2024) shows that the application of data mining in the agricultural sector, such as in freshwater fish production, uses the C4.5 method (a type of Decision Tree) to predict production results based on environmental and operational factors. This finding is relevant because the same principle can be applied in predicting product sales involving many interacting variables [5]. In addition, research by Elisa (2022) revealed that the use of the C4.5 algorithm can be adapted in manufacturing companies to predict sales of goods based on historical patterns [6]. This shows the flexibility and effectiveness of this technique in various industrial sectors.

Iriadi and Nuraeni (2016) also discussed the application of the C4.5 algorithm in predicting creditworthiness in banks, which can be applied with similar principles to predict the possibility of product sales based on existing data. They highlighted the advantages of Decision Tree in making clearer and more objective decisions based on structured data [7]. This study further strengthens the importance of applying this method in processing large and complex data that is often found in business. In addition, Khormarudin (2016) in his research on data mining techniques, such as the K-Means clustering algorithm, suggests using a combined approach between clustering and Decision Tree to improve accuracy in predicting and grouping sales data into relevant categories. The use of this technique can improve understanding of different market patterns, which is important for product management [8]. Sijabat (2015) in his research on the application of Decision Tree in student data processing also showed that this method can be used to identify important characteristics that influence prediction results. In a business context, this is equivalent to identifying factors that influence sales, such as price, promotion, or market trends [9]. Similar research was also conducted by Muzakir and Wulandari (2016) who used Decision Tree to predict hypertension in pregnant women. This technique has proven effective in providing data-based decisions, which are similar in terms of product sales predictions [10]. Research by Haryati *et al*. (2015) who applied the C4.5 algorithm to predict student study periods, emphasized that this method is useful for projecting possible outcomes within a certain period of time. The same principle can be applied to predict product sales in a certain period [11]. Ahmad *et al*. (2022) also showed the importance of implementing data mining techniques in business analysis to manage information more effectively, which can be adapted to predict sales results at PT. Cipta Naga Semesta [12]. Nasrullah (2021) added that the implementation of Decision Tree to classify best-selling products is also very useful for determining which products need to be prioritized in sales strategies [13]. With a strong theoretical basis and application from various previous studies, this study focuses on the implementation of Decision Tree to predict sales of PT. Cipta Naga Semesta in 2023. It is hoped that by implementing this method, companies can design more effective marketing strategies, manage inventory better, and improve overall operational efficiency.

2. Research Method

The research utilizes the Decision Tree method, focusing on data mining techniques for sales prediction. The process aims to identify key attributes that influence sales and generate accurate predictive models. The methodology consists of the following steps:

1) Data Mining
   Data mining is the process of discovering patterns, relationships, and trends in large datasets using statistical and computational techniques. It extracts valuable information from extensive data by analyzing patterns and correlations. This approach includes classification, clustering, association, and prediction tasks, which are crucial for business analytics. Data mining has been widely applied in various fields such as retail, helping businesses optimize store layouts and enhance customer experience [19]. By analyzing historical sales data, it provides actionable insights that support better decision-making in areas like resource allocation and operational efficiency [18].
2) Classification
   Classification is a core technique in data mining that organizes data into predefined categories based on specific attributes. The classification process consists of two key stages: learning and testing. During the learning phase, a training dataset is analyzed using a classification algorithm to construct a model. In the testing phase, this model is validated with a new dataset to ensure its accuracy. If the model's predictions meet the required level of precision, it is used to classify future data. Decision Tree is a widely used classification method due to its intuitive nature, allowing users to easily interpret the results [15]. The model represents decisions in a flowchart-like structure where internal nodes represent tests on attributes, branches represent outcomes, and leaf nodes show the final classification or predicted value.
3) Decision Tree
   The Decision Tree algorithm is applied to build predictive models for sales classification and forecasting. It employs a "divide-and-conquer" strategy, recursively splitting the dataset based on attribute values to create a tree structure. This approach simplifies complex relationships into easy-to-understand decision rules. Decision Trees are particularly useful for business tasks that require transparent and explainable decision-making, such as sales forecasting and customer segmentation [21]. Each split in the tree is determined by calculating the information gain or entropy of attributes. The attribute with the highest gain is chosen to split the data, continuing until the tree structure is optimized. Decision Trees are beneficial because they can handle large datasets and provide clear visual representations of the decision-making process, which is important for business strategy and operational planning [22].
4) C4.5 Algorithm
   The C4.5 algorithm, a variant of the Decision Tree method, is used to enhance the classification process. It begins by preparing a training dataset, typically consisting of historical data divided into different classes. The algorithm calculates the entropy of each attribute to measure uncertainty, and then selects the attribute with the highest information gain as the root node. The process continues recursively with the algorithm splitting the dataset using the most informative attributes, building a tree structure that can predict future outcomes based on previous data [14]. C4.5 is known for its ability to handle both continuous and discrete attributes, making it suitable for a wide range of prediction tasks. Additionally, it incorporates a pruning technique to avoid overfitting, ensuring that the tree structure remains manageable and does not become excessively complex [14][21].

3. Result and Discussion

3.1 Results

3.1.1 Determining Attribute Values

The first step in constructing a decision tree is selecting the attribute that will serve as the root node. This is done by calculating the information gain for each attribute to determine which one will be chosen as the root. Information gain measures how well an attribute splits the data based on the results of testing the attribute. The higher the information gain, the better the attribute is at dividing the data into more homogeneous (pure) classes. The formula for calculating information gain is based on entropy, which measures the uncertainty of a dataset. The formula for entropy for a dataset S with two classes (positive and negative) is as follows:

$$Entropy(S) = -P(+)Log_2P(+) - P(-)Log_2P(-)$$

Where:
P(+) is the proportion of the positive class in the dataset,
P(−)is the proportion of the negative class in the dataset.

After calculating entropy for each attribute, information gain is computed by subtracting the entropy after splitting the dataset based on the attribute from the original entropy. The attribute with the highest information gain is selected as the root node of the decision tree. For this study, the RapidMiner 8.1 software is used to implement the decision tree method. The data is processed using Microsoft Excel 2010 and imported into RapidMiner using the read excel operator, which allows the file to be processed. Once the data is imported, the Decision Tree operator is used to build the tree model, which will then generate predictions based on the processed dataset.

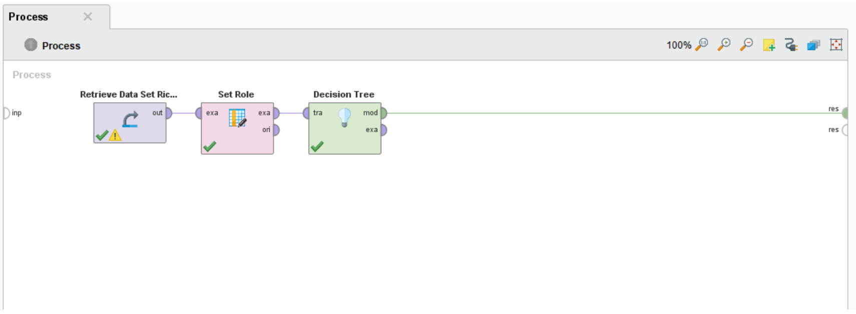


Figure 1. Initial Process in RapidMiner

3.1.2 Setting Parameters

After selecting attributes based on information gain, the next step is to set the parameters for the decision tree model. One key parameter used in this research is the gain_ratio. The gain ratio is used to adjust for bias towards attributes with many values (multi-value attributes). Bias occurs when attributes with many possible values tend to be chosen more often because they divide the data into many branches, even though they may not provide better separation. The gain ratio helps mitigate this by considering both the number of branches and the size of each branch when selecting attributes. Another important parameter is maximum depth, which defines the maximum length of the tree from the root to the leaf nodes. Limiting the maximum depth helps prevent the tree from growing too deep and complex, which can lead to overfitting—where the model fits the training data too closely, losing its generalization ability. Properly setting the maximum depth ensures the tree remains manageable and the model does not become overly complicated.

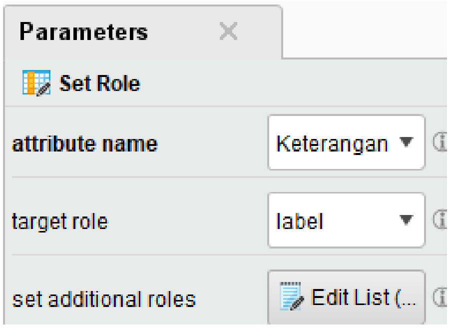
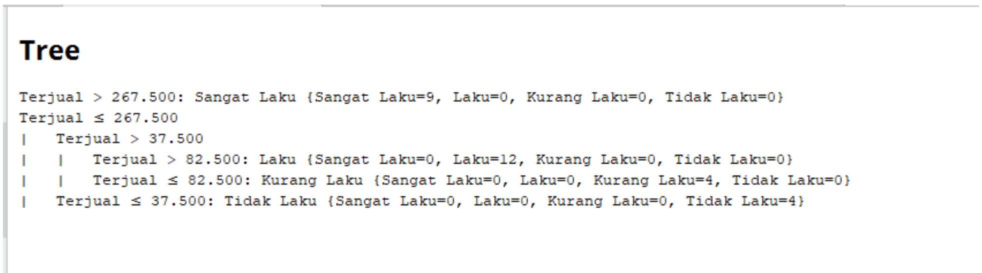


Figure 2. Parameter Settings in RapidMiner

3.1.3 Describing the Results of Classification

Once the decision tree is built, the next step is to generate a description of the classification results. The description provides insights into how the dataset was classified and how the decision tree made its decisions at each node. It explains how the data was split according to the selected attributes, and how the final classification was determined at the leaf nodes. This description also helps in understanding how well the model performs and whether it aligns with expected outcomes. It illustrates the decisions made at each split in the tree and shows the final class predicted for the input data. The description provides a clear breakdown of how each data point is processed and classified based on the constructed tree.



Tree

```
Terjual > 267.500: Sangat Laku {Sangat Laku=9, Laku=0, Kurang Laku=0, Tidak Laku=0}
Terjual ≤ 267.500
|   Terjual > 37.500
|   |   Terjual > 82.500: Laku {Sangat Laku=0, Laku=12, Kurang Laku=0, Tidak Laku=0}
|   |   Terjual ≤ 82.500: Kurang Laku {Sangat Laku=0, Laku=0, Kurang Laku=4, Tidak Laku=0}
|   Terjual ≤ 37.500: Tidak Laku {Sangat Laku=0, Laku=0, Kurang Laku=0, Tidak Laku=4}
```

Figure 3. Description of the Classification with Example Set

3.1.4 Graph View

The Graph View is a visual representation of the decision tree. It displays the structure of the tree, with each node showing the attribute used to split the data, and each branch representing the outcome of that split. The leaf nodes, at the end of the tree, show the final classification or prediction made by the model. The Graph View is important because it allows users to visually track how the model arrives at a decision. By following the path from the root node to the leaf node, users can see how attributes are tested and how the data is split at each stage. This visual representation helps in understanding the decision-making process, making it easier to interpret and communicate the results of the model.

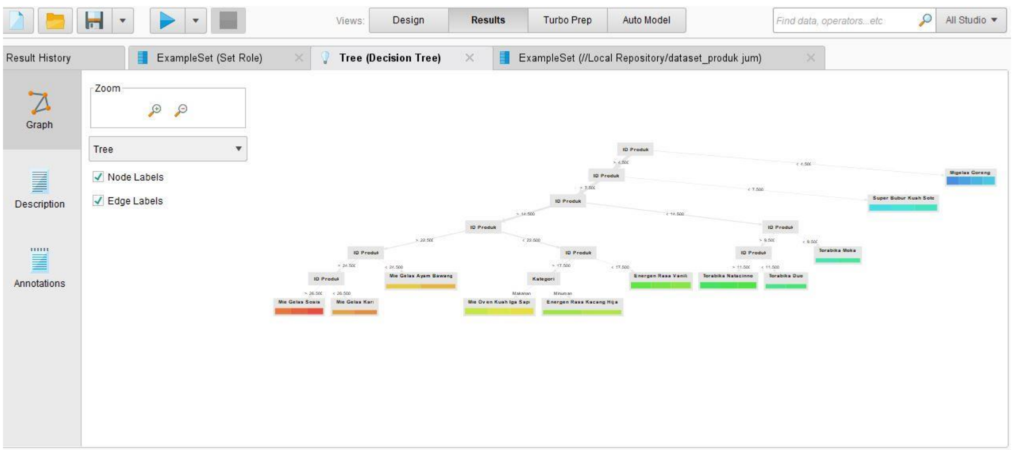


Figure 4. Resulting Decision Tree

This study applies the Decision Tree method using RapidMiner 8.1 to predict the sales of PT. Cipta Naga Semesta in 2023. The process begins by preparing the data, importing the Excel file containing the sales data, and constructing the decision tree using the Decision Tree operator in RapidMiner. The attributes are selected based on information gain, and key parameters such as gain_ratio and maximum depth are set to avoid overfitting and improve the model's accuracy. After the decision tree is built, the classification results are analyzed through the description and graph view to ensure that the model provides accurate predictions. By using this decision tree model, the company can plan its sales strategies, optimize inventory management, and enhance operational efficiency for 2023.

3.2 Discussion

The Decision Tree method used in RapidMiner 8.1 was effective in classifying sales data based on key attributes such as pricing, promotional activities, and seasonal trends. By selecting the attribute with the highest information gain, the model identified the variables that contributed most to the sales prediction. Information gain measures how well an attribute splits the data into homogenous subsets, and the attribute with the highest value of gain is chosen for further splits. This approach led to an accurate model with minimal error. In this study, attributes such as promotional campaigns, pricing, and seasonal factors were identified as significant in predicting sales. This result is consistent with findings from previous research, which also showed that factors like promotions and pricing directly influence consumer behavior. The ability to visualize these relationships in the form of a Decision Tree allows for an intuitive understanding of how each factor impacts the sales prediction. Using the gain_ratio parameter to adjust for bias towards attributes with many possible values helped improve the model’s accuracy. Bias tends to occur when attributes with numerous categories are chosen more often, simply because they divide the data into many branches, even if the splits do not result in more accurate predictions. The gain ratio addresses this by considering both the number of branches and their size when selecting the best attributes. This makes the model more reliable and less prone to overfitting. Additionally, the maximum depth parameter helped control the growth of the tree, preventing it from becoming too complex and overfitting the data. Limiting the depth of the tree ensures the model remains general and does not become overly tailored to the training data, which can result in poor performance on unseen data.

The accuracy of the Decision Tree model was validated through its ability to predict sales patterns based on historical data. The Graph View in RapidMiner provided a visual representation of how the tree made decisions at each node and how those decisions ultimately led to the final classification. The model's predictions were found to be accurate, with minimal error, making it a reliable tool for sales forecasting. The model’s predictive performance could be enhanced by incorporating additional data variables. While the model was effective in predicting sales based on the available data, additional factors such as customer demographics, competitor actions, and broader economic conditions could improve its accuracy. For instance, including customer preferences or feedback could provide more granular insights, leading to more precise predictions of sales volume. Several studies have shown that including such additional data can improve the predictive power of Decision Trees [15][18].

One of the major advantages of the Decision Tree model is its interpretability. Unlike other machine learning algorithms that may be considered "black-box" models, Decision Trees provide a transparent structure that can be easily understood. Each node represents a decision based on an attribute, and each branch represents the result of that decision. The final classification or prediction is shown at the leaf nodes. This structure makes the model particularly useful for businesses like PT. Cipta Naga Semesta, where decision-makers need to understand not only the predictions but also the reasoning behind those predictions. The Graph View feature in RapidMiner was especially valuable in presenting the tree structure in a clear, visual format. This allowed business managers and stakeholders to trace how the model arrived at its decisions. For example, if the Decision Tree model showed that sales tend to increase during specific promotional campaigns, the company could plan to run more targeted promotions during these high-impact periods. The ability to visualize and interpret these results is a critical advantage in real-world business applications, as it facilitates more informed and actionable decision-making.

While the Decision Tree model performed well, there are several ways it could be enhanced. One possible improvement would be to incorporate additional data sources, such as customer feedback, competitor analysis, and macroeconomic indicators, which could provide a fuller picture of the factors influencing sales. This could increase the model’s predictive power and make it even more relevant for decision-making. Another suggestion is to explore the use of ensemble methods such as Random Forests or Gradient Boosting. These techniques combine multiple decision trees to reduce variance and improve the accuracy of predictions. Ensemble methods are known to reduce overfitting and capture more complex patterns in the data, making them particularly effective when dealing with large datasets or datasets with a high level of noise [22]. Additionally, while the maximum depth parameter was used to control the complexity of the tree, other pruning techniques could be considered to refine the model further. For example, cost-complexity pruning is a method that can help remove unnecessary branches from the tree, improving its simplicity without sacrificing predictive accuracy.

The Decision Tree model has several practical applications for PT. Cipta Naga Semesta. First, the model can help the company plan its sales strategies by identifying which factors most influence sales. For example, if the model shows that promotions during the holiday season are highly effective in driving sales, the company can allocate more resources to such campaigns in the future. Similarly, the insights gained from the model can help optimize inventory management, ensuring that popular products are stocked at the right levels to meet customer demand without overstocking. The interpretability of the model also makes it accessible to business managers who may not have a deep technical background. This allows non-technical decision-makers to understand the logic behind the model's predictions and to use the insights for operational decision-making. For example, managers can identify high-impact periods for sales and adjust marketing campaigns accordingly, improving both the efficiency and profitability of operations.

4. Related Work

In recent years, the use of Decision Tree algorithms in predicting business outcomes, such as sales forecasting, has been widely explored. This section reviews some relevant studies that have applied similar methodologies, focusing on their findings and how they relate to the current research. Decision Trees have been widely applied in various domains to predict sales and improve decision-making processes. A study by Nasrullah (2021) explored the use of the Decision Tree algorithm for classifying popular products in retail businesses, emphasizing its effectiveness in predicting future demand. Nasrullah’s research demonstrates how attributes such as product features, customer demographics, and purchase history can be used to build accurate predictive models. Similarly, in this study, PT. Cipta Naga Semesta used a Decision Tree to classify sales based on attributes like promotions, pricing, and seasonal trends, showing a direct connection in terms of predicting consumer behavior and optimizing product offerings. Sijabat (2015) also implemented a Decision Tree for data classification in a case study of educational data, where the goal was to predict student performance based on various input variables [9]. Although the application was in an educational setting, the study’s methodology highlights the versatility of the Decision Tree algorithm in handling diverse datasets and producing interpretable results. This is similar to the approach used in the current research, where the Decision Tree helped predict sales by considering a variety of influencing factors like promotional activities, which directly affect sales performance.

While Decision Trees are effective for many tasks, they are also prone to overfitting when they become too complex. This issue has been addressed by various studies that utilize ensemble methods to combine multiple decision trees and improve model accuracy. In the research conducted by Khoeri & Mulyana (2021), the C4.5 algorithm was used in conjunction with machine learning techniques for employee recruitment [21]. The study highlighted the importance of using ensemble methods, such as Random Forests, to combine predictions from multiple decision trees to achieve more stable and accurate results. The Random Forest algorithm, which combines multiple decision trees, was proven to mitigate overfitting and increase prediction accuracy, especially in large, complex datasets. Similar to this approach, ensemble methods could potentially be applied in future work with PT. Cipta Naga Semesta to refine the predictive model by reducing overfitting and improving sales forecasts. In the case of Fauziningrum & Sulistyaningsih (2021), the authors explored the application of the Decision Tree for measuring English proficiency in maritime education [22]. Although the focus was on educational data, the research stressed the importance of enhancing the model's performance by combining multiple decision trees. The use of ensemble methods like Boosting could offer a similar advantage in the sales prediction model for PT. Cipta Naga Semesta, allowing the model to adapt to complex relationships within the sales data and provide more accurate predictions.

A significant body of research has focused on integrating promotional and seasonal data to improve sales predictions. Elisa (2022) conducted a study on predicting product sales at PT. Batam Bangun Prathama using the C4.5 algorithm, emphasizing the role of marketing activities and price changes in sales forecasting [6]. This study demonstrated that Decision Trees could successfully capture the influence of marketing campaigns, promotions, and price fluctuations on sales, providing actionable insights for future strategies. Similarly, in the present research, promotional activities and seasonal trends were key attributes in predicting sales at PT. Cipta Naga Semesta. Rachmawati (2011) explored the impact of marketing mix elements, such as product, price, promotion, and place (the 4Ps), on the increase in restaurant sales [16]. The study found that sales were significantly impacted by promotional efforts and pricing strategies. This research aligns with the findings of the current study, where promotional activities and pricing were identified as crucial factors influencing sales. By incorporating such factors into the Decision Tree model, businesses can more accurately predict future sales and optimize their marketing efforts.

In the broader scope of data mining applications in business, Hartatik *et al*. (2023) explored various sectors where data science methods, including Decision Trees, are applied for strategic decision-making. They highlighted that data mining techniques are increasingly utilized in business settings to improve performance, identify market trends, and understand consumer behavior [4]. This aligns with the current research, where the Decision Tree model helps PT. Cipta Naga Semesta make data-driven decisions to optimize sales and resource allocation. Moreover, Rimbarizki & Susilo (2017) focused on data mining in the context of educational institutions to enhance student learning motivation [2]. Their research emphasized the utility of classification algorithms like Decision Trees to analyze and predict behavior. Although their focus was on education, the underlying principle—that classification algorithms can reveal actionable patterns from large datasets—translates directly to sales prediction. By applying similar classification techniques to sales data, PT. Cipta Naga Semesta can enhance its operational decisions and strategies.

Despite the success of Decision Trees in business forecasting, there are limitations that should be addressed in future research. As Zhang *et al*. (2016) noted in their study on medical applications of data mining, Decision Trees are sensitive to noisy data, which can lead to overfitting or underfitting if not properly pruned or tuned [17]. This challenge was addressed in the current study through the use of maximum depth and gain_ratio parameters to prevent overfitting. Future studies could further enhance the model by experimenting with other pruning methods or integrating ensemble learning approaches like Random Forests or Gradient Boosting, which are known to improve predictive accuracy in noisy or complex datasets.

5. Conclusion and Recommendations

The application of the Decision Tree method in predicting sales at PT. Cipta Naga Semesta (a subsidiary of Mayora Group) in North Jakarta for the year 2023 has proven to be highly beneficial. The model successfully identified key factors influencing sales, such as seasonality, promotional activities, and pricing strategies, through the analysis of historical data. With the ability to provide accurate sales predictions, the Decision Tree model enables the company to devise more effective and efficient sales strategies, ultimately improving profitability and optimizing resource allocation. As a result, the use of Decision Trees has proven to be a powerful tool in supporting better business decision-making at PT. Cipta Naga Semesta, helping the company make informed decisions that can drive growth and operational efficiency.

To enhance the accuracy of the Decision Tree model, it is recommended that PT. Cipta Naga Semesta gather more detailed and comprehensive sales data. Additional information, such as customer preferences, customer feedback, and competitor data, could offer deeper insights into the factors that influence sales. Expanding the dataset would allow the model to capture a broader range of variables, thereby improving its predictive power. Furthermore, integrating ensemble methods such as Random Forest or Gradient Boosting could further boost the model's performance. These techniques, by combining multiple decision trees, would help reduce overfitting and better capture the complexities in the data, ultimately leading to more accurate and stable predictions.

References

[1] Bahri, S., Harahap, R. R., & Nasution, N. (2024). Perencanaan, organisasi, dan kontrol operasi pemasaran global. *Jurnal Masharif Al-Syariah: Jurnal Ekonomi dan Perbankan Syariah, 9*(1). https://doi.org/10.30651/jms.v9i1.21361

[2] Rimbarizki, R., & Susilo, H. (2017). Penerapan pembelajaran daring kombinasi dalam meningkatkan motivasi belajar peserta didik paket C vokasi di pusat kegiatan belajar masyarakat (PKBM) Pioneer Karanganyar. *J+ Plus Unesa, 6*(2), 1-12.

[3] Mining, W. I. D. (2006). *Introduction to data mining* (pp. 2-12). Pearson Education.

[4] Hartatik, H., Kwintiana, B., Nengsih, T. A., Baradja, A., Harto, B., Sudipa, I. G. I., & Gugat, R. M. D. (2023). *Data science for business: Pengantar & penerapan berbagai sektor*. PT. Sonpedia Publishing Indonesia.

[5] Novareza, A., Herdiana, R., & Iin, I. (2024). Analisis data mining pada produksi ikan air tawar di Kecamatan Panjalu menggunakan metode C4.5. *JATI (Jurnal Mahasiswa Teknik Informatika, 8*(1), 252-258. https://doi.org/10.36040/jati.v8i1.8209

[6] Elisa, E. (2022). Penerapan algoritma C4.5 untuk memprediksi penjualan barang pada PT Batam Bangun Prathama. *Computer and Science Industrial Engineering (COMASIE), 7*(1), 127-133.

[7] Iriadi, N., & Nuraeni, N. (2016). Kajian penerapan metode klasifikasi data mining algoritma C4.5 untuk prediksi kelayakan kredit pada Bank Mayapada Jakarta. *Jurnal Teknik Komputer AMIK BSI, 2*(1), 132-137.

[8] Khormarudin, A. N. (2016). Teknik data mining: Algoritma K-Means clustering. *J. Ilmu Komput*, 1-12.

[9] Sijabat, A. (2015). Penerapan data mining untuk pengolahan data siswa dengan menggunakan metode decision tree (studi kasus: Yayasan Perguruan). *Vol. V*, 7-12.

[10] Muzakir, A., & Wulandari, R. A. (2016). Model data mining sebagai prediksi penyakit hipertensi kehamilan dengan teknik decision tree. *Scientific Journal of Informatics, 3*(1), 19-26. https://doi.org/10.15294/sji.v3i1.4610

[11] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama, 11*(2). https://doi.org/10.37676/jmi.v11i2.260

[12] Ahmad, I., Samsugi, S., & Irawan, Y. (2022). Implementasi data mining sebagai pengolahan data. *J. Teknoinfo, 16*(1), 46.

[13] Nasrullah, A. H. (2021). Implementasi algoritma Decision Tree untuk klasifikasi produk laris. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar, 7*(2), 45-51. https://doi.org/10.35329/jiik.v7i2.203

[14] Ardiansyah, D. (2019). Algoritma C4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa SMP dengan menggunakan aplikasi RapidMiner. *Jurnal Inkofar, 1*(2). https://doi.org/10.46846/jurnalinkofar.v1i2.29

[15] Sopiyanti, I. A., Pudjiantoro, T. H., & Santikarama, I. (2021, October). Klasifikasi penjualan jus dengan menggunakan algoritma K-Nearest Neighbor (K-NN) untuk penerapan konsep up-selling. In *SNIA (Seminar Nasional Informatika dan Aplikasinya)* (Vol. 5, pp. E7-13).

[16] Rachmawati, R. (2011). Peranan bauran pemasaran (marketing mix) terhadap peningkatan penjualan: Sebuah kajian terhadap bisnis restoran. *Jurnal Kompetensi Teknik, 2*(2), 143-150.

[17] Zhang, J., Cai, M. X., Lu, T., Sun, X. H., & Jia, J. L. (2016). Lumbar spinal stenosis treated with polyetheretherketone pedicle screw fixation combined with interbody fusion: A follow-up assessment focusing on bone fusion rate. *Chinese Journal of Tissue Engineering Research, 20*(12), 1684. https://doi.org/10.3969/j.issn.2095-4344.2016.12.002

[18] Amalia, R. (2020). Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode Naïve Bayes. *Jurnal Informatika dan Sistem Informasi, 6*(1), 33-42.

[19] Rizki, M., Devrika, D., & Umam, I. H. (2020). Aplikasi data mining dalam penentuan layout swalayan dengan menggunakan metode MBA. *Jurnal Teknik Industri: Jurnal Hasil Penelitian dan Karya Ilmiah dalam Bidang Teknik Industri, 5*(2), 130-138. https://doi.org/10.24014/jti.v5i2.8958

[20] Gymnastiar, S., & Bahtiar, A. (2024). Penerapan algoritma K-Means clustering untuk mengelompokkan data kejadian kekeringan di Kabupaten Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika), 8*(2), 2325-2331. https://doi.org/10.36040/jati.v8i2.8948

[21] Khoeri, I., & Mulyana, D. I. (2021). Implementasi machine learning dengan decision tree algoritma C4.5 dalam penerimaan karyawan baru pada PT. Gitareksa Dinamika Jakarta. *Jurnal Sosial Teknologi, 1*(7), 615-623. https://doi.org/10.59188/jurnalsostech.v1i7.126

[22] Fauziningrum, E., & Sulistyaningsih, E. I. (2021). Penerapan data mining metode decision tree untuk mengukur penguasaan bahasa Inggris maritim (studi kasus di Universitas Maritim Amni). *Jurnal Sains dan Teknologi Maritim, 22*(1), 41-50.