Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

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

Analysis and Visualization of Sales Transaction Patterns using Decision Tree and Tableau Public

Miftahul Akbar[1*], Nining Rahaningsih[2], Irfan Ali[3], Fatihanursari Dikananda[4], Umi Hayati[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], fatiha.dikananda@gmail.com[4], umi@stmik-amikbandung.ac.id[5]*

Abstract

This study aims to analyze sales transaction patterns of rubber waste at PT Mandiri Enviro Technosio by integrating the *Decision Tree* algorithm with interactive visualization using *Tableau Public*. The dataset consists of 405 sales transactions recorded during the 2024–2025 period, comprising attributes such as transaction date, product type, quantity, unit price, total value, delivery region, and buyer category. The research methodology includes data acquisition, preprocessing to ensure data quality and consistency, construction of a *classification* model using the *CART* algorithm, evaluation of model performance through a confusion matrix, and development of interactive dashboards for enhanced interpretability. The Decision Tree model achieved an accuracy of 88.24% in classifying transaction values into low, medium, and high categories. Unit price and transaction period were identified as the most influential attributes in determining transaction value. Visualization using *Tableau Public* effectively presented the distribution of transaction values, sales trends, and geographical patterns, thereby strengthening analytical insights and supporting data-driven decision making. The integration of *classification* techniques and interactive visualization contributes to improving business intelligence capabilities and enables the formulation of more adaptive, *evidence-based* sales strategies.

***Keywords**: Decision Tree; Tableau Public; sales patterns; transaction classification; data visualization.*

1. Introduction

The rapid advancement of information technology has fundamentally transformed the way organizations collect, manage, and analyze operational data. Digital transformation has encouraged companies to treat data as a strategic asset that supports *evidence-based decision making* [1]. In modern business environments, sales transaction data represent a primary source of information that reflects a company's economic activities. These data no longer function solely as administrative records but hold strategic value for deriving analytical insights and supporting *data-driven managerial decisions* [2]. However, the increasing volume, variety, and velocity of sales data have made it difficult for many enterprises to extract meaningful patterns and trends, especially when the data are unstructured or distributed across multiple systems [3]. Consequently, systematic and *technology-driven* analytical approaches are needed to transform raw data into actionable knowledge.

One widely adopted approach in sales data analysis is data mining, which involves discovering hidden patterns within large datasets through statistical algorithms and machine learning techniques [4]. In business contexts, data *mining* assists organizations in understanding customer behavior, predicting sales trends, and optimizing marketing strategies [5]. Common methods include *clustering*, *association rule mining*, and *classification*, with *classification* being frequently used to identify categories or behavioral patterns within transactional data [3]

Among various *classification* algorithms, the Decision Tree remains one of the most popular due to its ability to produce interpretable predictive models. *Decision Trees* partition data based on the most influential attributes related to the target variable, using measures such as information *gain or Gini index* [6]. A key strength of this method lies in its interpretability, enabling companies to systematically identify the dominant factors influencing increases or decreases in transaction values [7].

Nevertheless, classification analysis alone may not provide comprehensive insights for non-technical *decision makers*. Data visualization therefore plays an essential role in bridging analytical outputs and managerial understanding. Visualization involves presenting information through graphical or interactive dashboards to facilitate intuitive interpretation of analytical results [8]. Tools such as *Tableau Public* are widely used to display complex datasets interactively and cohesively. Research by [9] demonstrates that *Tableau* enhances comprehension of sales patterns through dynamic and intuitive visual representations.

The integration of *Decision Tree* algorithms and *Tableau-based* visualization offers an effective approach for analyzing sales transaction data. While the classification model identifies patterns and key determinants within the dataset, *Tableau* visualizes these findings in a format accessible to decision makers. This combined approach enables organizations not only to understand what occurs within their sales activities but also why it occurs, thereby strengthening the foundation for more informed and adaptive business strategies. Based on this rationale, the present study aims to analyze sales transaction patterns using the *Decision Tree* algorithm and visualize the results through *Tableau Public*. This integrated approach is expected to provide a comprehensive understanding of transaction characteristics and trends, ultimately supporting *data-driven decision making* in modern business environments.

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.

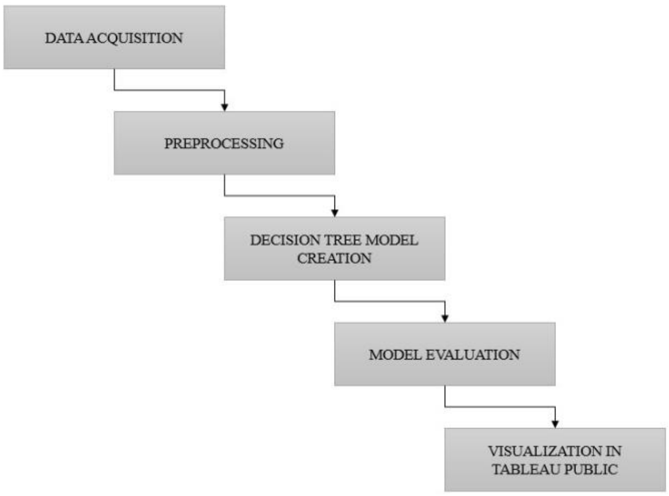

Fig. 1: Research Stages

2.1. Data Acquisition

The first stage of the research involved data acquisition, during which sales data were collected from the company's internal sales management system of PT Mandiri Enviro Technosio. The retrieved data were initially stored in Excel format to facilitate readability and further processing using the Python programming language. As noted by [10], the data acquisition phase serves as a critical foundation in the data analytics pipeline, as the quality of the acquired data directly influences the validity and reliability of subsequent analytical processes. In this study, the sales data were exported from the internal system into Microsoft Excel (.xlsx) format and subsequently converted into *Comma-Separated Values* (*.csv*) to ensure compatibility with Python-based analytical tools and the *Tableau Public* visualization platform.

2.2. Preprocessing

The data preprocessing stage involved cleaning, transforming, and normalizing the dataset prior to model development. This phase is essential for ensuring that the data are properly structured and analytically ready. According to [11], preprocessing plays a critical role in enhancing the performance of machine learning algorithms by ensuring that the dataset is clean, free from duplication, and internally consistent. Through systematic preprocessing, potential sources of bias and noise are minimized, thereby improving the reliability and accuracy of the subsequent analytical and predictive processes.

2.3. Decision Tree Model Creation

According to [6], the Decision Tree algorithm is particularly effective for business-related datasets due to its capability to handle both numerical and categorical attributes while producing models that are highly interpretable. The modeling process in this study began by defining the target variable (*value_class*), which represents the transaction value categories—low, medium, and high—and selecting independent variables such as quantity (QTY), unit price, region, and product name. The model was trained using a designated training subset and subsequently evaluated against a testing subset to assess its classification performance.

The resulting model produced a decision tree structure that illustrates the hierarchical sequence of the most influential attributes affecting transaction value. This structure provides a transparent and easily interpretable *classification* pattern, enabling users to understand the underlying *decision* pathways and variable interactions that shape the final prediction outcomes.

2.4. Model Evaluation

Model evaluation was conducted to assess the accuracy and overall effectiveness of the classification results. The evaluation employed several performance metrics, including the *confusion matrix, precision, recall, F1-score*, and *accuracy*. According to [12], these metrics are essential for understanding the balance between the model's ability to correctly identify actual classes (recall) and the degree of correctness in its predicted outputs (precision).

To ensure the robustness and consistency of the model's performance, this study also utilized the *k-fold cross-validation* technique. This method helps minimize evaluation bias by repeatedly partitioning the dataset into different training and testing subsets, thereby providing a more reliable estimate of the model's generalization capability.

2.5. Visualization in Tableau Public

This visualization stage serves to present the classification results in a clearer and more accessible manner while supporting intuitive, visually driven data exploration. As noted by [9], Tableau Public is an effective visualization tool for communicating analytical outcomes because it enables dynamic data exploration and facilitates evidence-based visual storytelling. Through this visualization process, the analytical results generated by the *Decision Tree* model can be transformed into strategic insights that are immediately useful for corporate management, thereby strengthening *data-driven decision-making* practices.

3. Result and Discussion

3.1. Data Acquisition

A snapshot of the sales transaction dataset is presented in tabular format, consisting of several key variables: NO, DATE, PRODUCT NAME, BUYER, QTY, PRICE, TOTAL, and SHIPPING ADDRESS. These variables represent the fundamental information associated with each transaction, covering aspects ranging from the type of product sold to the destination of delivery. An example of the first five rows of the dataset is illustrated in Figure 2.

```
   NO     TANGGAL         NAMA BARANG          PEMBELI       QTY   HARGA  \
0   1  2024-07-11       KARET SIRIHAN  BPK.DENI PATONAH  21800.00   150.0
1   2  2024-07-11  KARET KENYAL HITAM      IBU HERAWATI    191.58  1500.0
2   3  2024-07-11  KARET KENYAL HITAM      IBU HERAWATI   1166.92  1500.0
3   4  2024-07-11         KARET NILEX      IBU HERAWATI     44.32  1000.0
4   5  2024-07-11   PVC HITAM CRUSHER      IBU HERAWATI     69.80  3000.0

       TOTAL ALAMAT PENGIRIMAN
0  3270000.0         TANGERANG
1   287370.0           BANDUNG
2  1750380.0           BANDUNG
3    44320.0           BANDUNG
4   209400.0           BANDUNG
```

Fig. 2 : Data Acquisition

3.2. Preprocessing

```
••• ✅ Kolom tanggal berhasil dikonversi ke datetime
    ✅ Kolom value & weight dikonversi ke numerik dan baris kosong dihapus
    ✅ Nilai kosong pada kolom kategorikal diisi dengan 'Unknown'
    ✅ Kolom 'value_class' berhasil dibuat berdasarkan terciles (Rendah/Sedang/Tinggi)

    📋 Contoh 5 baris data hasil transformasi:
      no        date             product          customer    weight   price  \
    0  1  2024-07-11       KARET SIRIHAN  BPK.DENI PATONAH  21800.00   150.0
    1  2  2024-07-11  KARET KENYAL HITAM      IBU HERAWATI    191.58  1500.0
    2  3  2024-07-11  KARET KENYAL HITAM      IBU HERAWATI   1166.92  1500.0
    3  4  2024-07-11         KARET NILEX      IBU HERAWATI     44.32  1000.0
    4  5  2024-07-11   PVC HITAM CRUSHER      IBU HERAWATI     69.80  3000.0

           value     region value_class
    0  3270000.0  TANGERANG      Tinggi
    1   287370.0    BANDUNG      Rendah
    2  1750380.0    BANDUNG      Sedang
    3    44320.0    BANDUNG      Rendah
    4   209400.0    BANDUNG      Rendah
```

Fig. 3 : Standardization

The data preprocessing stage involved a series of structured operations performed on the sales dataset. Four primary processes were successfully completed.

First, the date column was converted into a standardized datetime format to ensure that temporal information could be analyzed chronologically.

Second, the value and weight columns were transformed into numeric data types, and rows containing invalid numerical entries were removed to reduce potential bias and computational errors.

Third, missing values in categorical fields such as product, customer, and region were handled by replacing empty entries with the label "Unknown", thereby preserving data integrity without discarding transactional information.

Lastly, a new feature, *value_class*, was constructed to categorize transaction values into Low, Medium, and High groups using a *tercile-based approach*, facilitating more effective segmentation analysis of sales performance.

The lower section of the figure presents five sample rows resulting from the transformation process, illustrating that the dataset has been refined to a cleaner and more consistent structure. This enhanced dataset is thus well-prepared for subsequent analytical procedures, including sales pattern exploration, data visualization, and *classification* modeling.

3.3. Decision Tree Model Creation

The dataset was partitioned into 80% training data and 20% testing data using the stratified split method to ensure proportional representation of each class within both subsets. The *DecisionTreeClassifier* model was then trained with a maximum depth parameter set to four (max_depth = 4) to prevent overfitting and enhance the model's generalization capability. During the training process, the model generated decision rules derived from a combination of key attributes, including purchase quantity (QTY), product type, and delivery region. The results of the model evaluation and the corresponding decision tree structure are presented in Figure 4.

```
Kolom dataset: ['no', 'date', 'product', 'customer', 'weight', 'price', 'value', 'region', 'value_class', 'Bulan']

Akurasi pada data uji: 0.8824 (88.24%)

Laporan Klasifikasi:
              precision    recall  f1-score   support

      Rendah       1.00      0.91      0.95        34
      Sedang       0.78      0.88      0.83        33
      Tinggi       0.88      0.86      0.87        35

    accuracy                           0.88       102
   macro avg       0.89      0.88      0.88       102
weighted avg       0.89      0.88      0.88       102


Model dan encoder telah disimpan ke file .pkl

✅ Dataset siap digunakan di Tableau: data_penjualan_untuk_tableau.csv
```

Fig. 4 : Data classification results using Decision Tree

The trained model, together with the corresponding label encoder, was saved in .pkl format to enable future reuse and integration within subsequent analytical workflows. Additionally, the finalized dataset was exported as a CSV file, making it ready for further visualization and exploratory analysis in Tableau.

The resulting *decision tree* model produced through *the Decision Tree classification* algorithm is illustrated in the accompanying figure. This visualization presents the hierarchical decision structure constructed from the attributes used during the model training process. Each node in the tree represents a splitting condition based on the value of a particular variable, while the branches indicate the outcomes of those decision rules.

The visualization provides a comprehensive depiction of how the Decision Tree model systematically partitions the dataset into three target classes—Low, Medium, and High—using specific threshold values. Through this graphical representation, the classification patterns become clearly observable, and the leaf nodes indicate the final classification outcomes derived from the trained model. This structure highlights the logical and interpretable nature of the Decision Tree, demonstrating how decisions are formed based on the underlying characteristics of the sales transaction data.

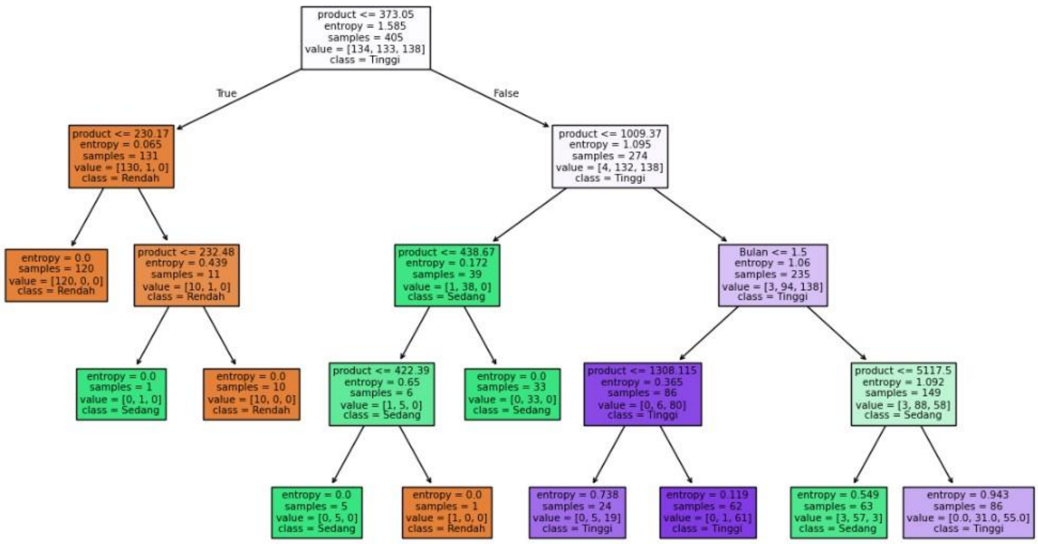

Fig. 5 : Decision tree visualization

The figure illustrates the visualization of an *entropy-based classification* model (information gain) employed to predict categorical classes Low, Medium, and High based on the primary variables product and Month. Each node presents the corresponding *splitting* condition derived from variable values, along with information on entropy, the number of samples, class distribution (value), and the dominant class. The entropy value reflects the level of uncertainty within a node; the lower the entropy, the more homogeneous the data contained in that node.

The tree begins at the root node, where the initial split is determined by the condition product ≤ 373.05. The left branch leads to the Low category, while the right branch proceeds to additional splits based on both product and Month to classify observations into either the Medium or High categories.

Overall, the model depicts a hierarchical *decision-making* process in supervised learning *classification*. Its objective is to derive structured *decision* rules capable of predicting performance categories or specific levels based on numerical input variables. This visualization thus provides a clear representation of how the model partitions the data and arrives at interpretable *classification* outcomes.

3.4. Model Evaluation

The *confusion matrix* is used to evaluate the performance of the *classification* model by presenting a comparison between the predicted outcomes and the actual class labels—Low, Medium, and High. Through this analysis, the model's ability to correctly identify each category can be assessed, as well as the extent to which misclassifications occur. The *confusion matrix* thus provides a comprehensive overview of classification accuracy at the class level and serves as an essential tool for understanding the strengths and weaknesses of the predictive model. The corresponding *confusion matrix* results are presented in Figure 6.

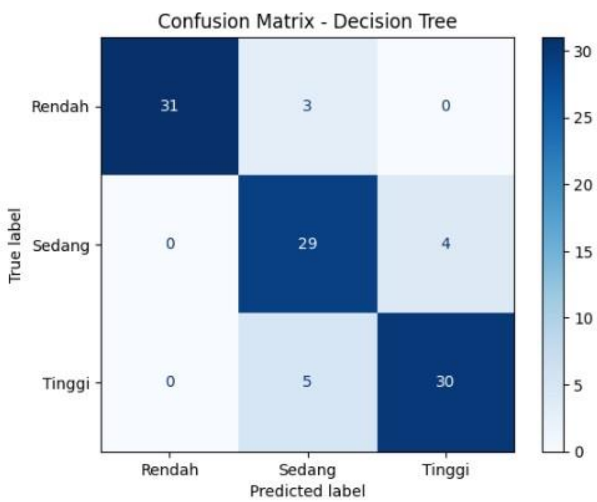

Fig. 6 : Confusion Matrix – Decision Tree

The figure presents the *confusion matrix* visualization for the *Decision Tree classification* model used to evaluate predictive performance across the three categorical classes: Low, Medium, and High. The matrix displays the distribution of the model's predictions against the actual test data, offering detailed insights into class-level accuracy.

For the Low class, the model demonstrates strong predictive capability, correctly classifying 31 samples with only 3 misclassified as Medium. The Medium class shows 29 correct predictions, while 4 samples were incorrectly classified as High. For the High class, the model accurately predicted 30 samples, although 5 were misclassified as Medium.

Overall, the dominant diagonal distribution indicates that the *Decision Tree* model performs well, achieving a high level of *classification accuracy*. Most misclassifications occur between the Medium and High classes, suggesting similarities in the underlying feature patterns of these categories. This observation highlights the model's ability to recognize the primary structure of the data, while also indicating the need for further refinement to more effectively distinguish between classes with closely overlapping value ranges.

3.5. Visualization in Tableau Public

1. Distribution of Transaction Value Classes

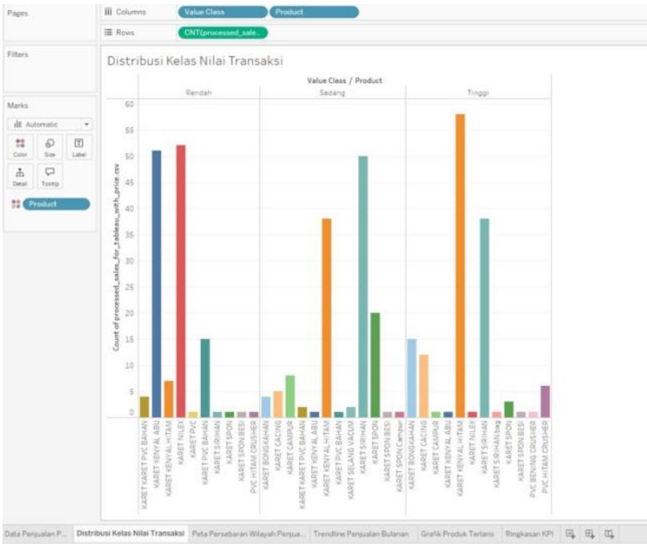

Fig. 7 : Transaction value distribution graph

The figure illustrates the distribution of transaction value classes across different product types. *Rubber Kenyal Hitam* and *Rubber Kenyal Abu* dominate the Low and High categories, whereas *Rubber PVC Abu* appears most frequently in the Medium class. Meanwhile, several products such as *Rubber Sirihan* and *PVC Hitam Crusher* show consistently low sales frequencies across all categories. This pattern indicates that product type significantly influences the level of transaction value, demonstrating its role as a key determinant in sales performance.

2. Sales Distribution Map

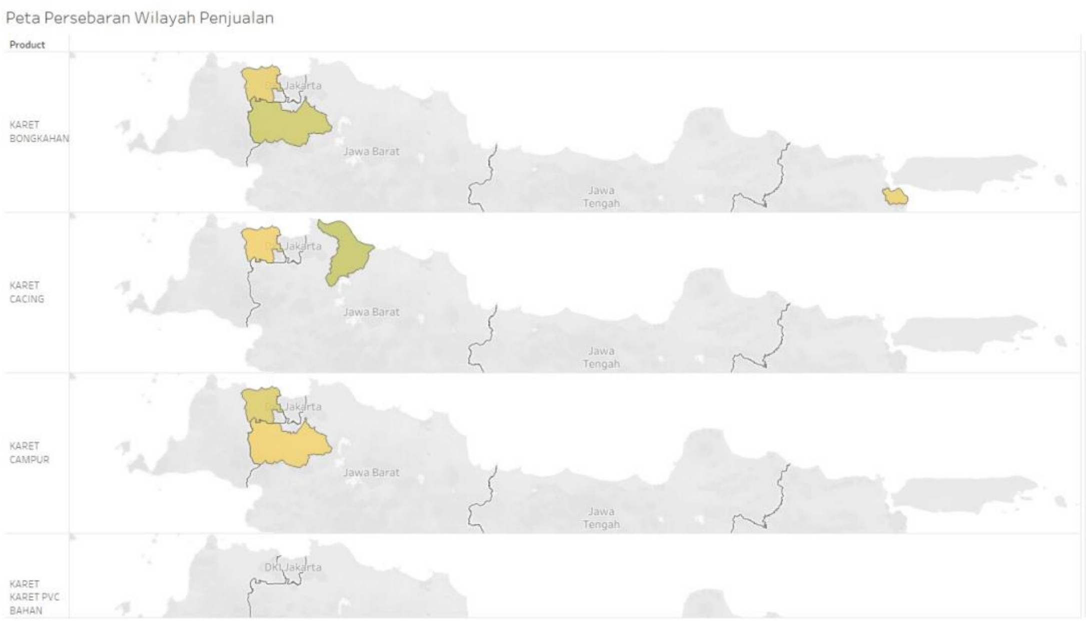

Fig. 8 : Distribution of sales areas

The figure presents a geographic distribution map of sales regions for various types of rubber waste across Java Island. The visualization indicates that sales of products such as *Rubber Bongkahan*, *Rubber Cacing*, *Rubber Campur*, and *Rubber PVC Bahan* are predominantly concentrated in DKI Jakarta, West Java, and parts of East Java. This pattern suggests that the company's primary distribution activities are centered in the western region of Java, with limited expansion toward the eastern areas. Such spatial concentration highlights the strategic importance of these regions in supporting the company's operational and market penetration efforts.

3. Sales Trends by Time

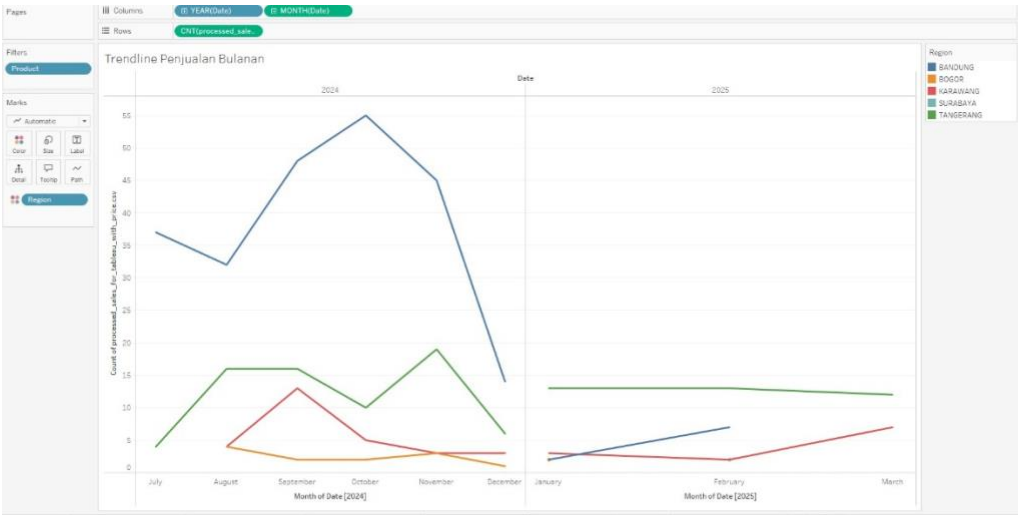

Fig. 9 : Sales trendline

The figure illustrates the monthly sales trends of rubber waste across several regions from 2024 to 2025. The highest sales volume occurred in Bandung in October 2024, followed by a significant decline toward the end of the year. Tangerang demonstrates relatively stable sales activity throughout the observed period, whereas Karawang and Surabaya exhibit minimal fluctuations. At the beginning of 2025, a slight increase is observed in both Bandung and Surabaya, suggesting a potential recovery in overall sales trends.

4. Interactive Dashboard

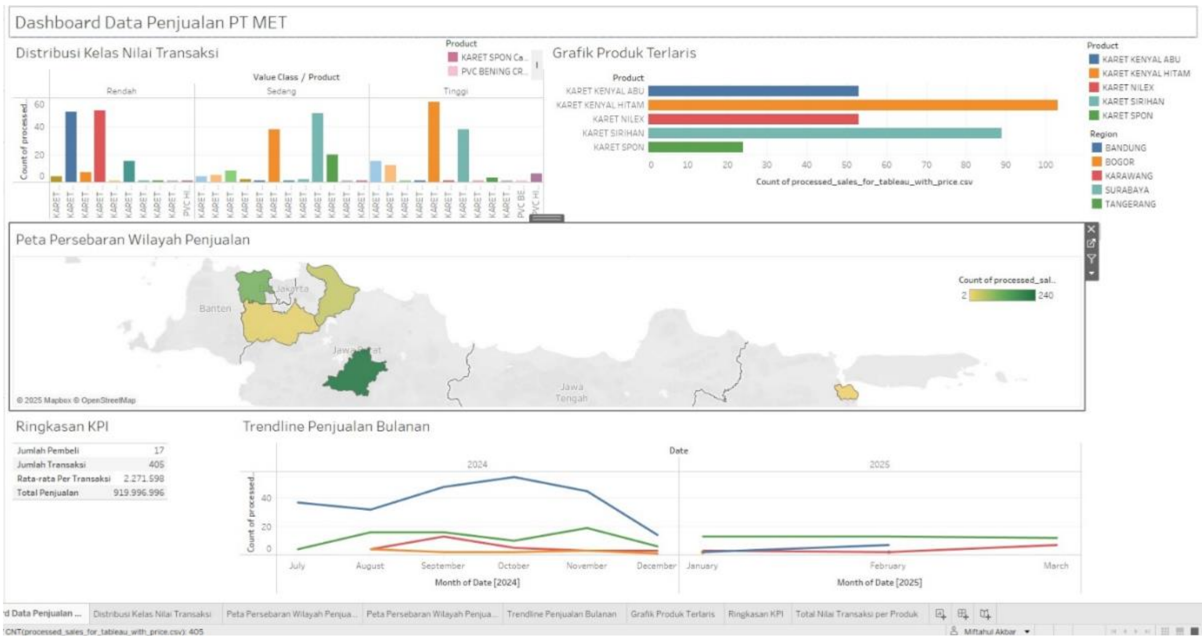

Fig. 10 : Dashboard

The distribution of transaction values indicates that most transactions fall within the low to medium categories. The best-selling products are *rubber Kenyal Hitam* and *rubber Kenyal Abu*, with dominant sales occurring in Bandung, Tangerang, and other regions within West Java. The geographic distribution map further demonstrates that sales activity is most concentrated in the western part of Java Island. The KPI summary records a total of 405 transactions from 17 buyers, amounting to a total value of Rp 919,996,996. The trendline shows fluctuating sales patterns, with a notable increase in October 2024 and relative stabilization at the beginning of 2025.

4. Conclusion

This study demonstrates that the *Decision Tree (CART)* algorithm is effective for analyzing sales transaction patterns of *rubber waste* at PT Mandiri Enviro Technosio. Using a dataset of 405 transactions recorded between 2024 and 2025, the model successfully classified transaction values into low, medium, and high categories with an accuracy of 88.24%. Unit price and transaction period were identified as the most influential attributes in determining transaction value. The integration of *Tableau Public* enabled the generation of interactive visualizations, including transaction distribution, sales trends, and geographic dispersion, thereby enhancing interpretability for *decision makers*. The combination of data *mining* techniques and visualization tools proves highly effective in supporting *data-driven* business *decision-making*.

References

[1] R. Henna *et al.*, "Analisis Pola Transaksi Penjualan Menggunakan Algoritma Decision Tree dan Visualisasi Tableau Public," 2025.
[2] A. Utomo *et al.*, "Analisis Pola Transaksi Penjualan Menggunakan Pendekatan Data Mining dan Algoritma Decision Tree," 2025.
[3] Ritonga, "Implementasi Metode Klasifikasi pada Data Transaksi Penjualan," *Jurnal Sistem Informasi dan Komputer*, p., 2024.
[4] Mawadati, "Penerapan Data Mining dalam Pengolahan Data Penjualan," *Jurnal Teknologi Informasi dan Komputer*, p., 2021.
[5] Marselina, "Pemanfaatan Data Mining dalam Analisis Tren Penjualan dan Pemasaran," *Jurnal Bisnis dan Teknologi Informasi*, p., 2023.
[6] Solmaz, "Struktur dan Interpretasi Model Decision Tree dalam Data Mining," *Jurnal Informatika dan Pembelajaran Mesin*, 2025.
[7] G. Komarasamy and K. Ravishankar, "Penerapan Algoritma Decision Tree dalam Analisis Data Penjualan," *Jurnal Sains Komputer dan Sistem Informasi*, p., 2023.
[8] Darajuanti, "Peran Visualisasi Data dalam Peningkatan Pemahaman Hasil Analisis," *Jurnal Teknologi Informasi dan Visualisasi*, p., 2025.
[9] Dewi and et al., "Pemanfaatan Tableau Public untuk Analisis Tren dan Pola Penjualan," *Jurnal Sistem Informasi Bisnis*, p., 2024.
[10] Beay and Sarimole, "Penerapan Decision Tree untuk Prediksi Penjualan pada Perusahaan Distribusi," *Jurnal Teknologi dan Sistem Informasi*, 2024.
[11] A. I. Zalukhu and M. Iqbal, "Analysis of Product Demand Prediction Using Decision Tree on Sales Data of Ceria Toys Store-Anzas Ibezato Zalukhu et.al Analysis of Product Demand Prediction Using Decision Tree on Sales Data of Ceria Toys Store," *Journal Of Data Science*, vol. 3, no. 01, 2025, doi: 10.58471/Science.v3i01.
[12] Sharma and et al., "Evaluating Decision Tree Models in Modern Business Analytics," *Journal of Business Data Science*, 2023.