Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Classification of Payment Patterns for Toyota Car Sales Using the Decision Tree Algorithm

Martua Hami Siregar[1], Hafdiarsya Saiyar[2], Desmulyati[3], Mohammad Noviansyah[4]
[1,2,3,4]Bina Sarana Informatika University. Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450

Article Info
Keywords:
Decision Tree,
Classification,
Car Sales,
Payment Patterns,
Sales Strategy

ABSTRACT
Car sales represent a business sector highly dependent on the implementation of appropriate payment strategies to enhance customer satisfaction and operational efficiency. This study aims to classify payment patterns in Toyota car sales using the Decision Tree algorithm. Historical sales data were utilized to identify various attributes influencing payment methods, such as cash, credit, or leasing. Through processes of preprocessing, feature selection, and model training, the Decision Tree algorithm successfully established clear classification patterns based on variables such as payment type, gender, car type, and car category. The research findings indicate that the Decision Tree method not only provides a high level of accuracy in classifying payment patterns but also produces models that are easily interpretable by business decision-makers. Thus, the implementation of this classification technique is expected to assist companies in designing more effective and targeted sales and promotional strategies.

Corresponding Author:
Martua Hami Siregar
Bina Sarana Informatika University. Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450
martua.mhe@bsi.ac.id

INTRODUCTION

The increasing complexity of business activities has amplified the need for effective and efficient data management. In the retail sector, sales transaction data not only reflects current business operations but also serves as a valuable foundation for future strategic decision-making. One of the key technologies supporting this need is the data warehouse, a structured and multidimensional data storage system that enables the rapid, accurate, and informative processing of large datasets (Handika, 2020).

A data warehouse serves not only to store historical data but also to transform it into relevant and actionable information, thereby supporting strategic decision-making. When combined with analytical tools such as the decision tree algorithm, a data warehouse can assist businesses in identifying consumer payment patterns, whether through cash or credit transactions. Understanding these payment trends is essential for developing more effective and targeted sales strategies (Atsila Imanda et al., 2024).

In this context, the integration of business intelligence systems becomes equally important. Business intelligence encompasses databases, architectures, applications,

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

analytical tools, and methodologies that facilitate interactive data access, enabling business analysts and managers to conduct efficient analyses (Hidayat & Putri, 2024; Pratama et al., 2024). Through business intelligence, organizations can visualize sales data comprehensively, revealing customer preferences and market trends, thereby enhancing the quality of business decisions (Handika, 2020). However, many companies have not yet optimized the use of these technologies. Sales transaction data is often archived without further analysis, missing opportunities for deeper insights that could influence business strategies. This gap highlights the importance of adopting advanced analytical platforms such as RapidMiner, an open-source data analysis software that enables predictive modeling without requiring extensive programming skills (Maulida et al., 2024; Pradhana et al., 2024).

RapidMiner employs both descriptive and predictive techniques, facilitating user-friendly data processing and model development. Designed with an open-core approach, it combines open-source elements with additional features available in its paid versions, thus offering flexibility for various analytical needs. Its capability to process data warehouses efficiently makes it highly relevant for applications in sales management. Specifically, in the field of sales, a data warehouse is designed to store, manage, and analyze large volumes of structured transactional data, including aspects such as sales volume, payment methods, product categories, sales locations, and transaction timing. By organizing this data in a multidimensional manner, businesses are better equipped to perform strategic analyses and make informed decisions.

Given these considerations, this study aims to utilize a data warehouse and the decision tree algorithm through the RapidMiner platform to classify Toyota car sales payment patterns based on cash and credit methods. This approach is expected to assist companies in optimizing their marketing strategies and improving decision-making processes based on more accurate and comprehensive data analysis.

METHODS

The data collection process in this study was conducted through several systematic stages. The research dataset was obtained from an official source, namely the website https://www.dataindustri.com/produk/laporan-data-penjualan-mobil-secara-detail-di-indonesia/, which provides detailed reports on car sales in Indonesia. The dataset consists of 900 sales records from various brands and types of vehicles. From the entire dataset, this study specifically selected sales data for vehicles under the Toyota brand. The selection was carried out based on the brand attribute to ensure that the data used was highly relevant to the focus of the study. The selected dataset includes several key attributes, namely payment method (cash or credit), car brand, car type, car category, buyer's gender, and buyer's age.

Prior to further analysis, the data underwent a preprocessing stage, which involved cleaning duplicate and irrelevant entries, verifying the completeness and consistency of attributes, and filtering the dataset to focus solely on Toyota vehicles. Once the data selection and cleaning processes were completed, the data were stored in a data warehouse and processed using the RapidMiner platform. This step aimed to create a dedicated Toyota dataset to facilitate a more focused classification analysis of payment patterns. Through this

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

method of data collection and processing, the research ensures that the dataset is relevant, well-structured, and ready for analysis using the decision tree algorithm.

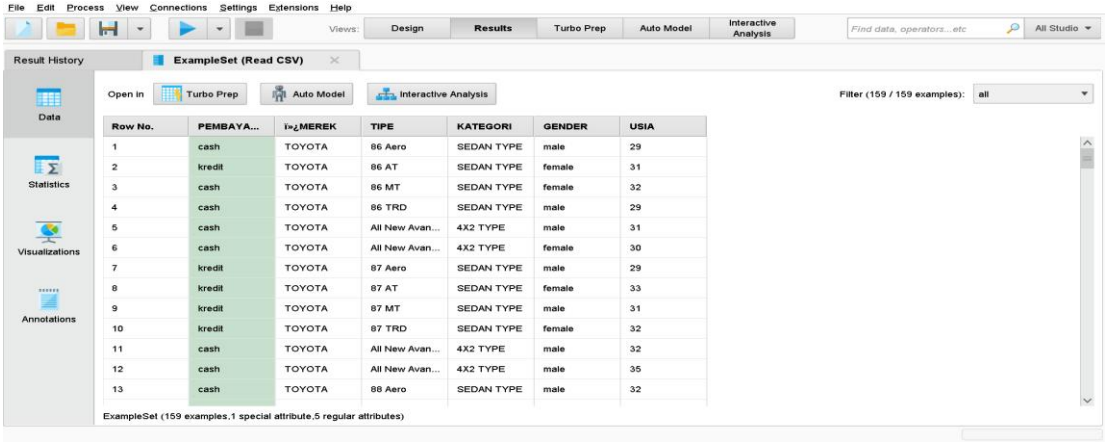

Figure 1. Toyota Car Dataset

Based on the sales dataset used in this study, the attributes include brand, type, category, gender, age, and payment method as the label. The descriptive analysis conducted in this study focused on several key aspects of the dataset. First, the payment distribution was examined, which consisted of two primary categories: cash and credit payments. This analysis aimed to understand the dominant methods used by customers when purchasing Toyota vehicles. Second, the customer gender distribution was analyzed, revealing the proportion of male and female buyers. This aspect provided insights into the demographic composition of Toyota car buyers and helped in identifying potential patterns based on gender. Further analysis was conducted on the types of vehicles sold, which included various models such as the 86 Aero, 86 AT, 86 MT, 86 TRD, and the All New Avanza. By examining different car types, the study aimed to capture the diversity of products preferred by customers.

The data processing in this research was carried out systematically to ensure that the dataset was ready to be analyzed using the decision tree algorithm. The steps applied in data processing are as follows:

1. Initial Data Exploration
   a. Loading the dataset obtained from the official source into the data processing software, RapidMiner.
   b. Conducting an initial review to ensure that the data format was appropriate, such as data types (numerical, categorical), and verifying the presence of all required attributes.
2. Data Cleaning
   a. Removing duplicate records to prevent negative impacts on the analysis.
   b. Handling missing values through imputation techniques, such as replacing with mean, mode, or median values, depending on the data type.
   c. Ensuring data consistency, such as standardizing date formats or capitalizing product names uniformly.

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

3. Data Selection
   a. Filtering the dataset to include only vehicles of the Toyota brand.
   b. Ensuring the availability of key attributes, such as vehicle type, payment method, sales quantity, transaction time, and sales location, in every data entry.
4. Data Transformation
   a. Converting categorical data (e.g., payment method: cash or credit) into numerical format using label encoding or one-hot encoding techniques to enable processing by the decision tree algorithm.
   b. Normalizing numerical attributes, such as sales quantity, to ensure uniform data scaling and prevent any attribute from dominating the analysis.
5. Dataset Splitting
   The dataset was divided into two main parts:
   a. Training Data: 80% of the dataset used to train the decision tree model.
   b. Testing Data: 20% of the dataset used to test the model's performance and ensure its accuracy.
6. Decision Tree Model Building
   a. The processed dataset was input into RapidMiner.
   b. The decision tree algorithm was applied by setting the target attribute (payment method) and the other predictor attributes.
7. Model Evaluation
   a. The resulting model was evaluated using performance metrics such as accuracy, precision, recall, and F1-score.
   b. nterpretation was conducted to understand the resulting patterns, particularly the attributes that most significantly influenced the payment method.

The results of this data processing are expected to provide deep insights into consumer payment preferences based on Toyota vehicle sales data.

RESULTS AND DISCUSSION

The objective of this study is to analyze the percentage distribution of customer payment methods using the Decision Tree method. Prior to processing with RapidMiner, the dataset used consisted of 159 records, which were classified based on the payment attribute.

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Row No.	ï»¿MER...	TIPE	KATEGORI	GENDER	PEMBAYA...	USIA
1	TOYOTA	86 Aero	SEDAN TYPE	male	cash	29
2	TOYOTA	86 AT	SEDAN TYPE	female	kredit	31
3	TOYOTA	86 MT	SEDAN TYPE	female	cash	32
4	TOYOTA	86 TRD	SEDAN TYPE	male	cash	29
5	TOYOTA	All New Avan...	4X2 TYPE	male	cash	31
6	TOYOTA	All New Avan...	4X2 TYPE	female	cash	30
7	TOYOTA	87 Aero	SEDAN TYPE	male	kredit	29
8	TOYOTA	87 AT	SEDAN TYPE	female	kredit	33
9	TOYOTA	87 MT	SEDAN TYPE	male	kredit	31
10	TOYOTA	87 TRD	SEDAN TYPE	female	kredit	32
11	TOYOTA	All New Avan...	4X2 TYPE	male	cash	32
12	TOYOTA	All New Avan...	4X2 TYPE	male	cash	35
13	TOYOTA	88 Aero	SEDAN TYPE	male	cash	32

Figure 2. Data Before Classification

The data were grouped into various categories, where the classification function itself was used to categorize or group the data into different classes or labels based on the features present in the dataset. The use of labels in the payment attribute was intended to identify the distribution of customer payments.

Row No.	PEMBAYA...	ï»¿MEREK	TIPE	KATEGORI	GENDER	USIA
1	cash	TOYOTA	86 Aero	SEDAN TYPE	male	29
2	kredit	TOYOTA	86 AT	SEDAN TYPE	female	31
3	cash	TOYOTA	86 MT	SEDAN TYPE	female	32
4	cash	TOYOTA	86 TRD	SEDAN TYPE	male	29
5	cash	TOYOTA	All New Avan...	4X2 TYPE	male	31
6	cash	TOYOTA	All New Avan...	4X2 TYPE	female	30
7	kredit	TOYOTA	87 Aero	SEDAN TYPE	male	29
8	kredit	TOYOTA	87 AT	SEDAN TYPE	female	33
9	kredit	TOYOTA	87 MT	SEDAN TYPE	male	31
10	kredit	TOYOTA	87 TRD	SEDAN TYPE	female	32
11	cash	TOYOTA	All New Avan...	4X2 TYPE	male	32
12	cash	TOYOTA	All New Avan...	4X2 TYPE	male	35
13	cash	TOYOTA	88 Aero	SEDAN TYPE	male	32

Figure 3. Dataset Classified Based on Payment Attribute

Subsequently, an analysis of the Decision Tree illustrated below will be conducted. The purpose of this analysis is to gain a more detailed understanding of the decision tree structure and the relationships between variables reflected in the figure.

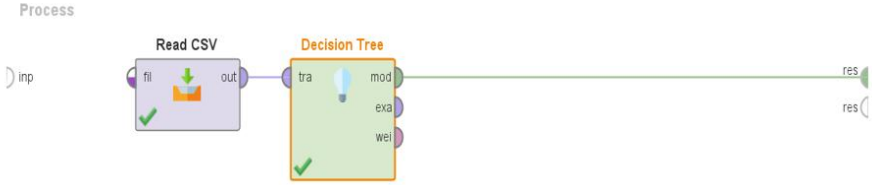

Figure 4. Decision Tree Design Process

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

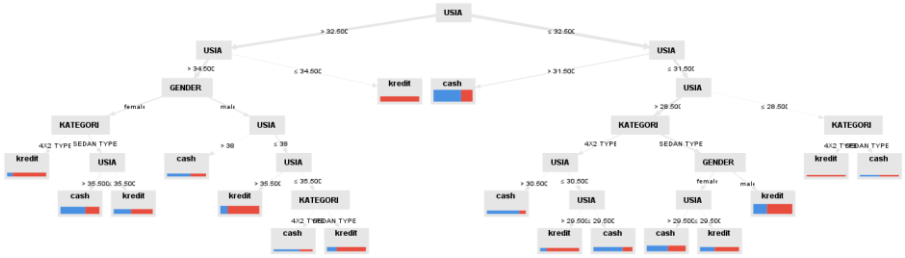

Figure 5. Output of the Process

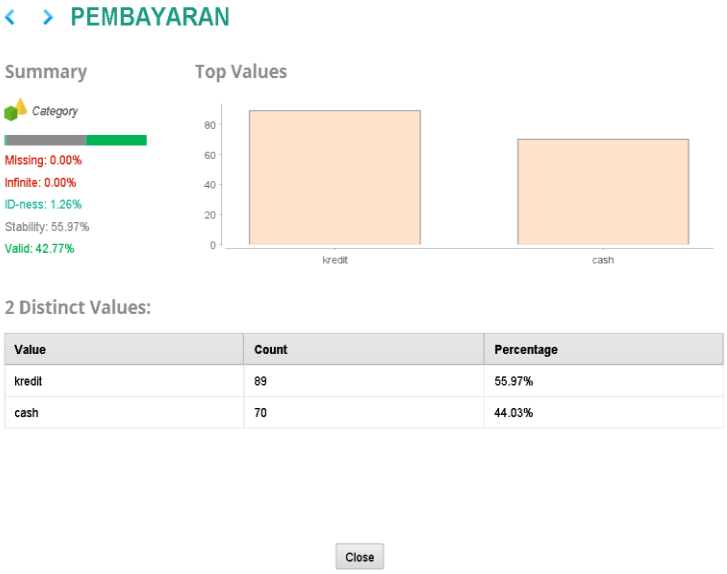

Figure 6. Percentage of Payment Distribution

The process of using the Apply Model and Performance (Classification) operators is carried out to determine the prediction accuracy of the payment distribution, as shown below:

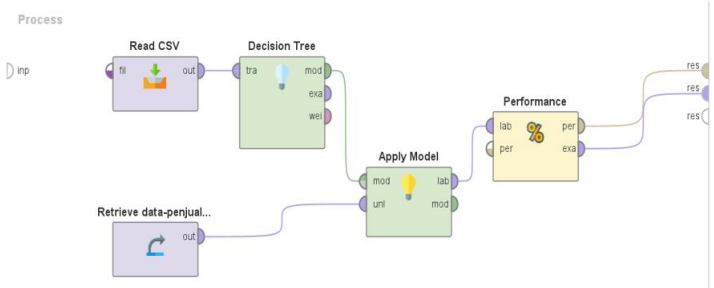

Figure 7. Apply Model – Performance Design Process

From the above process, the predicted accuracy of the data can be obtained, as shown in the figure below:

Jurnal Multidisiplin Sahombu
Volume 5, Number 04, 2025, DOI 10.58471/jms.v5i04
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

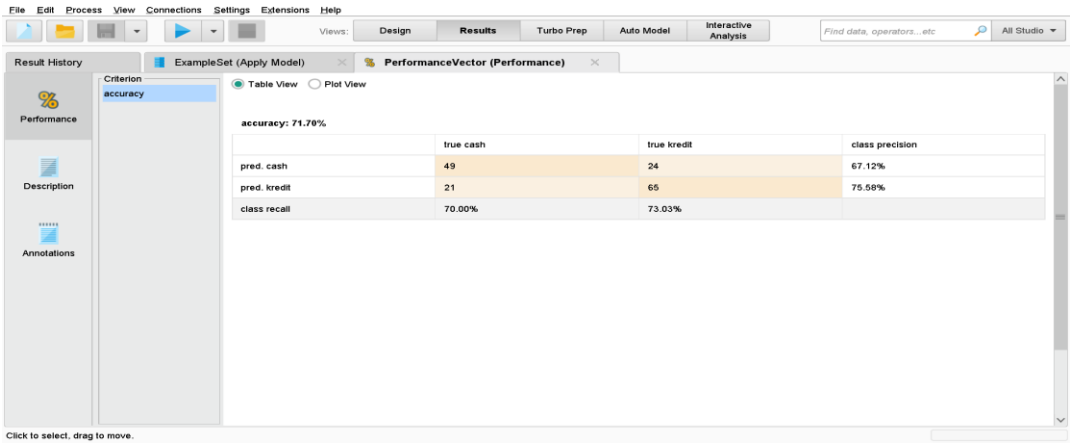

Figure 8. Data Accuracy Results

CONCLUSION

The decision tree algorithm successfully identified consumer payment patterns, whether through cash or credit. The analysis results show that attributes such as vehicle type, sales volume, location, and transaction time have a significant influence on the preference for payment methods. From the model analysis, the main factors affecting the choice of payment method are the vehicle type and the number of units sold. Consumers tend to prefer cash payments for more affordable vehicles, while credit payments are more commonly used for higher-priced vehicles. The resulting decision tree model demonstrated a good level of accuracy in predicting payment methods. This indicates that the approach used in this study is effective in exploring transaction patterns in vehicle sales. The insights gained from the analysis can be utilized by companies to develop more effective sales strategies, such as offering special promotions based on the payment methods most aligned with consumer preferences.

REFERENCE

Atsila Imanda, R., Suroso, S., Fauzi, A., Simanjuntak, H. F., Azizah, Z., Destianty, A., Mutmainnah, A. Z., & Zarka Zahira Shaffa, G. (2024). Pengaruh Data Warehouse Terhadap Pengambilan Keputusan. Jurnal Portofolio : Jurnal Manajemen Dan Bisnis, 3(1), 31–39. https://www.jurnalprisanicendekia.com/index.php/portofolio/article/view/282

Bhaskara, M. R. S., & Wasesa, M. (2023). Enhancing Car Showroom Profitability through K-Means Clustering for Customer Segmentation. *International Journal of Current Science Research and Review*, *6*.

Handika, I. P. S. (2020). Perancangan datawarehouse dan teknologi business intelligence untuk analisa penjualan pada perusahaan retail PT. ABC. Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab, 5(2), 76–85. https://doi.org/10.36341/rabit.v5i2.1309

Hidayat, M. K., & Putri, D. (2024). Business Intelligence Untuk Memantau Perkembangan Harga Pangan Provinsi DKI Jakarta. Jurnal Infortech, 6(1), 30–36. https://doi.org/10.31294/infortech.v6i1.21832

Jaras, V. (2024). *A model for determining the market price of a car* (Doctoral dissertation, Vilniaus universitetas).

Jayadeera, T. P., & Jayamanne, D. J. (2019, November). Fair price prediction system for used cars in Sri Lanka using machine learning and robotic process automation. In *2019 International Conference On Business Innovation (ICOBI), Colombo*.

Maulida, R., Haryadi, D., Ratu Counsela, S., & Rahmania Az Zahra, N. (2024). Implementasi Algoritma K-Means dan Rapidminer untuk Clustering Data Kemiskinan Provinsi Banten. Journal of Informatics and Communications Technology, 6(1), 17–32.

Msiza, Z. A., & Owolawi, P. A. (2023, December). Advancing Used Car Price Prediction in South Africa: An Empirical Examination of Machine Learning Techniques. In *International Conference on Artificial Intelligence and its Applications* (pp. 189-196).

Pradhana, A. H., Irfa, M., Ali, A., Ristyawan, A., & Daniati, E. (2024). Penerapan Regresi Linear Menggunakan RapidMiner Untuk Memprediksi Penjualan dan Persediaan. 8, 291–297.

Pratama, A. A., Pitoyo, B. S., Fauzi, A., Rivaldi, A., & Andini, D. A. (2024). Peran Implementasi Intelijen Bisnis Dalam Menentukan Peramalan Bisnis. Jurnal Portofolio: Jurnal Manajemen Dan Bisnis, 3(1), 60–70.

Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. *Int. J. Inf. Comput. Technol*, *4*(7), 753-764.

Rusmee, K., & Chumuang, N. (2019, October). Predicting system for the behavior of consumer buying personal car decision by using SMO. In *2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (pp. 1-6). IEEE.

Tucker, C. S., Hoyle, C., Kim, H. M., & Chen, W. (2009, January). A comparative study of data-intensive demand modeling techniques in relation to product design and development. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (Vol. 49026, pp. 371-383).

Varol, S. (2016). *Analyzing brand loyalty in automotive sector using the hidden Markov model and support vector machine*. Lamar University-Beaumont.

Wahono, S., & Ali, H. (2021). Peranan Data Warehouse, Software Dan Brainware Terhadap Pengambilan Keputusan (Literature Review Executive Support Sistem for Business). Jurnal Ekonomi Manajemen Sistem Informasi, 3(2), 225–239. https://doi.org/10.31933/jemsi.v3i2.781