Decision Tree Algorithm Classification for Medan City Laundry Business

Simon Sinaga, Zulham Sitorus

Abstract

This research aims to classify laundry businesses in the city of Medan using the Decision Tree algorithm. As the small and medium-sized enterprise sector continues to grow, the laundry business has become one of the industries that is constantly developing, especially in urban areas like Medan. Although the demand for laundry services is high, many entrepreneurs struggle to determine the right strategy to win the competition. Therefore, this research applies the Decision Tree algorithm to help classify the factors influencing the success of laundry businesses in Medan City. The data used in this study were obtained through surveys of several laundry business owners in Medan City, with variables including location, price, service quality, target market, and promotions conducted. The research results show that the Decision Tree algorithm is capable of identifying patterns and dominant factors that influence the success of laundry businesses. The location and service quality variables are the most important factors determining success in the laundry business, followed by price and promotion strategies. This research is expected to provide insights for laundry entrepreneurs in designing more effective business strategies, by considering the factors that have been proven to influence the performance of laundry businesses in Medan City.

Keywords: *Decision Tree algorithm, Classification, Laundry Business, Medan City, Success Factors*

Simon Sinaga[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: sinaga@bmkg.go.id[1]

Zulham Sitorus[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: zulhamsitorus@dosen.pancabudi.ac.id[2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

In this modern era, the development of small and medium enterprises (SMEs) in Indonesia is increasingly rapid, especially in the service sector. One type of business that continues to grow rapidly in large cities like Medan is the laundry business. Laundry businesses have become an integral part of urban life, especially in large cities with high mobility rates and increasingly busy lifestyles. The existence of laundry businesses offers a practical solution for individuals and groups who need fast and efficient laundry services, thereby increasing the demand for these services.

Medan City, as one of the largest cities in Indonesia with a population of over 2 million people, offers a highly potential market for laundry businesses. The demand for laundry services continues to increase along with the growth in population, shifts in lifestyle, and the rising number of workers with high mobility. However, despite the high demand, laundry entrepreneurs in Medan often face various challenges that hinder the success of their businesses. Some of the challenges faced by laundry entrepreneurs include choosing a business location, setting competitive service prices, managing service quality, and implementing effective promotional strategies to attract customers.

The small and medium-sized enterprise (SME) sector in Indonesia, particularly in the city of Medan, has experienced significant growth in recent years. One of the sectors that has shown rapid development is the laundry business, which is part of the service sector. As a city with a dense population and high mobility, Medan has become one of the potential markets for laundry services. The laundry business not only serves the daily needs of the community but also becomes a practical solution for individuals with busy schedules and business owners who require professional services in washing and maintaining clothes.

It can be inferred that the service quality of Ririn Laundry on G. Obos XII Street, Palangka Raya, is excellent based on the five characteristics of service quality—Tangibility, Reliability, Responsiveness, Assurance, and Empathy. The results of the researcher's interviews with the informants demonstrate this. Additionally, Ririn Palangka Raya's laundry consistently enhances quality to foster connections that are advantageous to both clients and service providers [1]. A web-based information system that administrators, staff, and clients may access has been successfully developed thanks to this research [2].

Although the demand for laundry services continues to increase, laundry entrepreneurs in Medan City often face challenges in determining the factors that can guarantee the success of their business. Among these challenges, entrepreneurs often struggle to identify the key variables that can influence the success level of their business, such as business location, pricing, service quality, and appropriate promotional strategies. Therefore, it is important to conduct a systematic and data-based analysis to help laundry entrepreneurs formulate effective strategies.

One of the approaches that can be used to analyse the factors influencing the success of a laundry business is the Decision Tree algorithm. This algorithm is one of the methods in artificial intelligence used for classification and prediction based on existing data. Decision Tree works by building a decision model that can map the relationship between input and output variables, thereby providing useful information for entrepreneurs to make more accurate decisions. Thus, this research aims to classify the factors influencing the success of laundry businesses in Medan City using the Decision Tree algorithm, in order to provide a more structured guide for laundry entrepreneurs in designing more effective and efficient business strategies.

The research results show that the accuracy of the best-selling product classification model using the Decision Tree C4.5 obtained from this study is 90% with an AUC value of 0.709, which falls into the Good Classification category. Therefore, it can be concluded that the data mining classification model using the Decision Tree C4.5 algorithm is accurate in classifying best-selling products [3]. Because it can model and comprehend complicated relationships among the factors involved in a decision, decision trees are a popular machine learning technique [4]. The results of this study using a decision tree show that the majority of students, amounting to 49 people (49%), are recommended to continue to the vocational high school level (SMK), based on the classification results of 100 students. In the meantime, 26 pupils (26%) and 25 students (25%) are recommended

for the senior high school (SMA) and madrasah aliyah (MA) levels, respectively [5].

According to Akram Fais' research, 303 data points from the Heart Disease Prediction Dataset were used in this investigation. A Decision Tree was then used to make predictions, yielding an F1 score of 83.33%, recall of 78.12%, accuracy of 83.60%, and precision of 89.28% [6]. The Decision Tree algorithm C4.5 model. The test results obtained in this study are accuracy (IDDRUS 2023). In the context of decision-making, this study investigates the use of decision trees as an analytical technique [4].

This research is expected to make a significant contribution to the understanding of the dynamics of laundry businesses, as well as provide data-driven solutions that can be used to enhance the competitiveness and sustainability of laundry businesses in the city of Medan. Technically, Decision Tree works by dividing data into several categories based on predetermined rules, to produce a clear and easily understandable decision model. This model provides a comprehensive overview of the relationships between factors that influence customer decisions, as well as offering solutions for laundry business owners to optimise their operations.

This research aims to apply the Decision Tree algorithm in analysing and classifying the factors that influence the success of laundry businesses in Medan City. By using data obtained from laundry entrepreneurs, this study will provide a clearer picture of the dominant factors that can enhance the competitiveness and performance of laundry businesses. The results of this study are expected to provide practical insights for laundry entrepreneurs in designing more effective and efficient business strategies, thereby increasing the success and sustainability of their businesses in an increasingly competitive market.

Literature Review

2.1 Data Mining

Data mining is a discipline aimed at digging, mining, or discovering information that has not been directly known from a large dataset. Data mining is a combination of statistical techniques, mathematics, artificial intelligence, and machine learning in extracting hidden information from a large database. Data mining can also be defined as the process of exploring patterns or characteristics of data that were not previously anticipated to assist in decision-making in the future. These patterns will be recognised and studied meticulously by the device, which is expected to provide useful analysis in the future.

Data mining is a computer learning process (machine learning) using one or more learning techniques (algorithms), so it is expected
computers are capable of automatically analysing knowledge. Based on the tasks that can be performed, data mining is divided into several groups, namely: description, estimation, prediction, classification, clustering, and association [7][8]; [9]; [10].

Data mining is one of the stages in the knowledge discovery in database (KDD) process. KDD is the entire process of identifying patterns from data, where these patterns are new and can be beneficial. The KDD process can be broadly explained as follows (Hutabarat, 2018; Wijaya, Hasibuan, and Ramadhani, 2018; Marcania, 2019).

2.2 Decision Tree

According to Kusrini and Luthfi (2017), decision trees are a very powerful and popular classification method. Decision tree can transform very large data into a decision tree that represents rules. Rules can be easily understood in natural language, therefore decision trees are a classification technique that is easy to learn and very popular to use. Decision tree is a technique used to classify a set of objects or data with a tree structure representation. Decision trees are generally used to explore patterns in a dataset and to find relationships between a number of input variables and an output variable. The decision tree diagram consists of three parts, namely: (Hadi, 2017; Itiqomah, Susanti, and Zukhronah, 2019; Marcania, 2019; Manurung, 2020).

1. Root Node: The node that is located at the top of the decision tree.
2. Internal Node: The branching node, where there is only one input and it produces at least two

outputs. 3. Leaf Node: The end node, which has only one input and no output.

Decision tree is a structure that can be used to divide large data sets into smaller sets of records by applying a series of decision rules. With each series of splits, the members of the resulting sets become similar to one another. Data in a decision tree is usually represented in the form of a table with attributes and records. Attributes represent a parameter that is used as a criterion in the formation of the tree. For example, to determine whether to play tennis, the criteria considered are weather, wind, and temperature. One of the attributes is an attribute that represents the solution data per data item, which is called the target attribute. Attributes have values known as instances. Suppose the weather attribute has instances such as sunny, cloudy, and rainy, the data concept in the decision tree can be seen in Table 1.

Tabel 1. Tabel ProsesDecision tree (Mardi, 2017)

Nama	Cuaca	Angin	Temperatur	Main
Faris	Cerah	Keras	Panas	Tidak
Totok	Cerah	Lambat	Panas	Iya
Dinda	Berawan	Keras	Sedang	Tidak
Fendik	Cerah	Lambat	Dingin	Iya

Sample Atribut TargetAtribut

The process in decision trees involves transforming data (tables) into a tree model and simplifying rules (Mardi, 2017). Decision trees are one of the most popular learning methods and are widely used in practice. This method is an approach that seeks to find functions with discrete values. The concept of a decision tree is to transform the data in the decision table into a decision tree and decision rules.

Gain (S, A) = Entropy (s) $\sum_{i=1}^{n} * Entripy\ (Si)$

Explanation:

S : set
A : attribute
N : number of partitions of attribute A
| Si | : number of cases in the i-th partition
| S | : number of cases in S

Calculating the Value of Entropy: Entropy (s) $\sum_{i=1}^{n} -pi * log2\ pi$

Explanation:

S: set of cases
A: feature
n: number of partitions S
pi: the proportion of Si to S

2.3 Understanding Laundry

Small businesses within a country's economy play a very important role. Selling with the increasing economic growth rate of Indonesia, the growth of the laundry business in Indonesia currently shows quite significant development. Not only in big cities, but even in remote areas, the laundry business is now flourishing with various offers of facilities and services. Laundry is

the process of washing clothes that usually uses water and detergent. Of course, this is done by laundry services, but it can also be done at home with a washing machine and regular detergent. There is a common understanding in society that the term "Laundry" is only used to refer to the service of washing clothes by laundry services, whereas it actually also applies to the activity of washing at home.

According to several experts, the term laundry has various meanings such as:
Rumekso (2001): Laundry is the division responsible for the laundry sent to it (Rumekso:2001). Agustinus Darsono (1995): Laundry is a section in the hotel responsible for washing clothes, whether they belong to guests, employee uniforms, or all hotel linens. Bagyono (2006): Laundry is part of the housekeeping division tasked with and responsible for washing guest clothes, employee uniforms, and linens from both inside and outside the hotel.

Research Methodology

The methodology of this research aims to classify the factors influencing the success of laundry businesses in Medan City using the Decision Tree algorithm. The steps taken in this research are as follows:

1. Research Design

This study uses a quantitative approach with a descriptive-explanatory research type. This research aims to describe and analyse the factors influencing the success of laundry businesses in Medan City and to identify patterns that can be explained through the Decision Tree algorithm. The data used in this study are primary data obtained through surveys distributed to laundry business owners in Medan City.

2. Population and Sample

The population in this study is all laundry businesses operating in the city of Medan. The sample taken consists of several laundry businesses randomly selected from the list of laundry businesses registered with the Medan City Trade Office. The criteria for sample selection are laundry businesses that have been operating for at least one year and have sufficient data related to the variables to be analysed, such as business location, price, service quality, promotion, and number of customers.

The sampling method used is random sampling, with the aim that the selected sample can represent the characteristics of all laundry businesses in Medan City. The number of samples used in this study is 50 laundry businesses spread across various districts in Medan City.

3. Research Variables

This research analyses several variables that are considered to influence the success of laundry businesses in the city of Medan. The variables used include:

a. Business Location: Strategies for selecting a business location that can affect customer accessibility.
b. Service Pricing: Setting competitive prices that match the quality of the services provided.
c. Service Quality: The level of customer satisfaction with the results of the laundry service.
d. Promotion: Promotional strategies implemented to attract customers, such as discounts, loyalty programs, or social media.
e. Number of Customers: The number of customers who come regularly and the frequency of laundry service usage.

4. Data Collection

The data used in this research is primary data obtained through survey techniques using questionnaires. The questionnaire is designed to gather information related to the variables mentioned earlier. Each respondent was asked to provide information regarding the business location, service prices, quality of services provided, and the promotional strategies they use.

In addition, secondary data was also collected from various relevant sources, such as annual reports from laundry entrepreneurs, publications related to the laundry industry, and statistical data from the Medan City Trade Office. This data will be used to support the analysis

and provide a more comprehensive picture of the laundry industry conditions in Medan City.

Results

Based on data from UMKM Kota Medan, the development of laundry businesses in the Medan City Government area amounts to approximately 1000 businesses. However, the reality after re-surveying the field is that not many internet cafes have gone bankrupt or closed down because they cannot cope with such strong competition and cover the operational costs of the laundry. The data that will be used in the form of a decision tree for analysis is only a portion of the total number of laundry businesses, approximately 464 businesses, as a sample data.

3.1 Calculating The Information Value (I) From All Training Data

To calculate the information value from all the training data of the Warnet Data, the formula below is used:

$$- p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}$$

=-(254/464) * LOG ((254/464),2) -(210/464) * LOG ((210/464),2)
I = 0.99 -◊ Information Value of All Laundry Data
Explanation:
A score of 254 represents a healthy Laundry.
A score of 210 represents an unhealthy Laundry.
A score of 464 represents the total data of all Laundry.

Table 2. Total Data Information of Laundry

Jml Laundry	P(+)	P(-)	N.Inf All
464	**254**	**210**	**0.9935**

3.2 Calculating The Value (i) Of Each Attribute In The Training Data.

a. Value of Location Attribute Information
The Location Attribute has 2 instances, namely good and bad. The value of the Location Attribute information determines the strategic places as explained in the table below:

Table 3. Location Attribute Information Value

Location	laundry Condition	Amount	Total
Good	Healthy	236	**297**
Good	Unhealthy	61	
Good	Healthy	20	**167**
Good	Unhealthy	147	
TOTAL		**464**	**464**

q1 = - (236/297) * LOG ((236/297),2) - (61/297) * LOG ((61/297),2)
= **0.732572531**
q2 = -(20/167) * LOG ((20/167),2) - (147/167) * LOG ((147/167),2)

= 0.528671978

q1 = It is a good criterion.
q2 = It is a bad criterion.

Location Entropy Value = (297/464) * 0.73+(167/464) * 0.53
= **0,66** → Location Entropy Results
Information Value of All Laundry Location Training Data - Location Entropy Location Information Gain = 0.99-0.66 = 0.33 → Location Inf Gain Result

b. Value of Time Attribute Information
The value of the time attribute information determines the time that is good healthy, good unhealthy, bad healthy, and bad unhealthy as explained in the table below:

Table 4. Information Value of Time Attribute

Time	Loundry	Amount	Total
Good	Healthy	234	**257**
Good	Unhealthy	23	
Good	Healthy	20	**207**
Good	Unhealthy	187	
TOTAL		**464**	**464**

q1 = -(234/257) * LOG ((234/257),2) - (23/257) * LOG ((23/257),2)
= **0.434779143**

q2 = -(20/207) * LOG ((20/207),2) - (187/207) * LOG ((187/207),2)
= 0.458183451
q1 = It is a good criterion.
q2 = It is a bad criterion.

Time Entropy Value = (257/464) * 0.43 + (207/464) * 0.45
= **0.445220289** → Time Entropy Result.
Value of Information for All Laundry Training Data - Time Entropy Information Gain Time = 0.99 - 0.44 = 0.55 → Result of Time Inf Gain

c. Value of Administrative
Attribute Information The value of this administrative attribute information determines complete healthy, complete unhealthy, incomplete healthy, and incomplete unhealthy, as explained in the table below:

Table 5. Administrative Attribute Information Value

Administration	Laundry	Amount	Total
Complete	Healthy	224	**224**
Complete	Unhealthy	0	
Incomplete	Healthy	30	**240**
Incomplete	Unhealthy	210	
TOTAL		**464**	**464**

q1 = -(224/224) * LOG ((224/224),2) - (0/224) * LOG ((0/224),2)
= **0**
q2 = -(30/240) * LOG ((30/240),2) - (210/240) * LOG ((210/240),2)
= 0.543564443
q1 = It is a complete criterion.

q2 = It is an incomplete criterion.

Entropy Value Admin = (224/464) * 0+(240/464) * 0.54
= **0.279310345** → Entropy Admin Results.

The Information Value of All Laundry Training Data - Entropy Admin Information Gain Admin = 0.99 - 0.27 = 0.71 ◊ Admin Inf Gain Result

d. Value of Parent Control Attribute Information

The value of the Parent Control Attribute Information determines whether there is something healthy, something unhealthy, and whether there is something healthy or something unhealthy, as explained in the table below:

Table 6. Information Value of the Parent Control Attribute

P. Control	Laundry	Amount	Total
there is	Sehat	236	**263**
there is	Unhealthy	27	
No	Healthy	18	**201**
No	Unhealthy	183	
TOTAL		**464**	**464**

q1 = -(236/263) * LOG ((236/263),2) - (27/263) * LOG ((27/263),2)
= **0.477376319**

q2 = -(18/201) * LOG ((18/201),2) - (183/201) * LOG ((183/201),2)
= 0.43497348

q1 = It is one of the existing criteria.
q2 = It is a criterion that does not exist.

Entropy Value P. Control = (263/464) * 0.47+(201/464) * 0.43
= **0.452672414** → Result P. Control.

The Information Value of All Laundry Training Data - Entropy P. Control Information Gain P. Control = 0.99 - 0.45 = 0.54 ◊ Result of Inf Gain P. Control

3.3 Calculating the Entropy Value of Each Attribute

1. Calculating Location Attribute Entropy
   The informational value of the location attribute is as follows:
   (297/464) * 0.73+(167/464) * 0.53 = **0.66**
2. *Calculating Time Attribute Entropy*
   The informational value of the Time attribute is as follows:
   (257/464) * 0.43+(207/464) * 0.45 = **0.44**
3. Calculating the Entropy of Administrative Attributes
   The value of information from administrative attributes is as follows:
   (224/464) * 0+(240/464) * 0.54 = **0.28**
4. Calculating the Entropy of the Parent Control Attribute is as follows
   (263/464) * 0.47+(201/464) * 0.43 = **0.45**

From the description above, the Entropy values for each attribute can be ordered as follows: 1. Entropy Administration (0.28) 2. Entropy Time (0.44) 3. Entropy Parent Control (0.45) 4. Entropy Location (0.66).

Conclusion

This research identifies that there are several factors that significantly influence the success of laundry businesses in the city of Medan. Among them, business location, service quality, price, and promotional strategy are the most dominant variables. A strategic location provides easy access for customers, while high-quality service creates a high level of customer satisfaction, which in

turn contributes to loyalty and repeat usage of services. Competitive prices and effective promotional strategies also play a role in attracting more customers.

The Decision Tree algorithm has proven effective in classifying the factors that influence the success of a laundry business. The model built provides a clear picture of the relationships between the existing variables, and helps entrepreneurs make decisions based on structured and easily understandable data. By using this model, laundry entrepreneurs can prioritise aspects that need improvement or attention to enhance their business performance. Based on the findings of this study, laundry entrepreneurs in Medan City are advised to pay more attention to selecting strategic locations, improving service quality, and setting competitive prices with other laundry businesses. In addition, entrepreneurs are also advised to use effective promotions, whether through social media, customer loyalty programs, or seasonal discounts to attract new customers and retain old ones. This research provides significant contributions to laundry entrepreneurs in designing more effective and data-driven business strategies. Additionally, this research also contributes to the development of knowledge in the field of data analysis and the application of Decision Tree algorithms in the small and medium-sized enterprise sector in Indonesia, particularly in the laundry industry. With a deeper understanding of the factors influencing business success, it is hoped that laundry entrepreneurs will be better prepared to face increasingly fierce competition in the market.

References

[1] J. Pendidikan and I. Pengetahuan, “Desember 2019 244,” vol. 2019, no. 11, pp. 244–251, 2019.

[2] D. Putra and D. Sakethi, “Jurnal Pepadun Pengembangan Sistem Pengelolaan Laundry Berbasis Web (Studi Kasus Arin Laundry) Jurnal Pepadun,” vol. 2, no. 3, pp. 375–384, 2021.

[3] P. Issn, “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK,” vol. 7, no. 2, pp. 45–51, 2021.

[4] R. N. Ramadhon, A. Ogi, and A. P. Agung, “Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank,” vol. 3, pp. 1860–1874, 2024.

[5] P. Sekolah, L. Di, M. A. Medan, A. B. Putra, and M. Azhari, “Implementasi Algoritma Decision Tree Untuk Klasifikasi,” vol. 9675, pp. 128–137, 2025.

[6] A. Fais, M. R. A. Lubis, A. Aulia, and I. Syafitri, “Implementasi Algoritma Decision Tree untuk Klasifikasi Serangan Jantung,” vol. 1, no. 4, 2023.

[7] M. A. K-means, S. M. Hutabarat, and A. Sindar, “Data Mining Penjualan Suku Cadang Sepeda Motor,” vol. 2, no. 2, pp. 126–132, 2019.

[8] I. M. Sudarma, “Implementasi Algoritma C5 . 0 pada Penilaian,” vol. 17, no. 3, pp. 1–6, 2018.

[9] A. F. Cahyaningrum, Y. H. Chrisnanto, and A. K. Ningsih, “Enrichment : Journal of Multidisciplinary Research and Development Classification of Sentiment Towards BPJS Services Using the C50 Algorithm,” vol. 1, no. 8, pp. 484–491, 2023.

[10] R. N. Amalda, N. Millah, and I. Fitria, “Implementasi algoritma c5.0 dalam menganalisa kelayakan penerima keringanan ukt mahasiswa itk,” vol. 7, no. 1, pp. 101–116, 2022.