

The Application of The C4.5 Algorithm Method Calculation The Number of Sales of Paving Production on CV. SR Lamongan

Agus Setia Budi¹, Muhammad Hasan Wahyudi²

¹ Lamongan Islamic University, Veteran Highway No. 53A Lamongan, Indonesia

E-mail : geniusbudi@unisla.ac.id, hasanwahyudi@unisla.ac.id

*corresponding author : geniusbudi@unisla.ac.id

Article Info

Received 2025-12-15
Revised 2026-03-10
Accepted 2026-03-14

**Keywords : C4.5 Algorithm;
Decision Support System**



Abstract

Paving production is essential for ensuring proper road infrastructure for the community. The amount of paving demand in each month is uncertain, so it is challenging for company owners to maintain optimal inventory levels to maximize profits. The purpose of this research is to design a decision support system (SPK) to help company owners determine sales decisions in the future and make it easier to decide whether or not the paving sales are in demand. The method used in this study is the C4.5 Algorithm method. From the results of this study, it was found that the C4.5 Algorithm method is the most appropriate method in determining sales decisions in the next period. The sample used is data from CV. SR Lamongan has been in sales for the last 2 years. There are 25 paving production data used for the sample. From there, the recommended method in calculating sales in the following year can be obtained, namely using the C4.5 Algorithm method due to its higher accuracy rate, achieving up to 76%.

INTRODUCTION

CV. SR is a private company located at Jl. Raya Laren 01 Laren Lamongan. CV. SR is a company in the field of paving production and selling building tools and materials as well as excavator heavy equipment rental. Paving is a concrete block or brick that is used as an alternative ground surface covering.

Sales of paving at CV.SR often fluctuate, while the production process continues to be carried out every day with the same amount, this can certainly bring losses to the company because when demand increases, there is often only a small supply while when there is a lot of inventory and demand is small. That's the reason why sales predictions need to be done. In making sales predictions this time, the author uses the C4.5 Algorithm method. This method is one of the algorithms used to perform classification or segmentation or grouping and is predictive. The basis of the C4.5 algorithm is the formation of a decision tree. The branches of the decision tree are questions of classification and the leaves are classes or segments. This method was chosen because it has a high level of accuracy in determining decisions.

CV. SR Lamongan often faces difficulties in determining the daily or monthly paving production quantity.

Overstock: If production is too high while demand is low, it leads to an accumulation of goods in the warehouse, increasing the risk of physical damage to the paving (moss/cracks) and storage costs.

Stockout: Conversely, if production is too low when demand surges (for example: during the construction project season), the company loses potential profits because it cannot fulfill customer orders on time.

Daily sales transaction data may have been recorded, but it is merely kept as an archive of paper piles or static Excel files. This data has not been optimally utilized as a basis for decision-making (Data-Driven Decision Making). Without algorithms like C4.5, management finds it difficult to see hidden patterns from thousands of past transactions.

There are external factors that are difficult to predict visibly by humans without computational assistance, such as: Weather Factors: The relationship between rainfall and the intensity of road construction projects.

Product Type: Certain types of paving (e.g., 6cm bricks) may be in high demand in certain months, while others (e.g., hexagons) are only popular for park projects. Budget Timing: Demand surges usually occur at the end of the year when government/private projects need to be completed urgently.

Without clear calculations regarding which products fall into the 'Best-Selling' or 'Not Best-Selling' categories, CV. SR Lamongan finds it difficult to determine promotion strategies or production priorities. This results in inefficient allocation of human resources and raw materials.

Research Flow Visualization

Stages: Brief Description

Problem: Inefficient production (Stock vs Demand).

Weakness: Manual/intuition methods are inaccurate and subjective.

Gap: No rule-based system for paving classification.

Solution: Implementation of the C4.5 Algorithm to form a Decision Tree.

From the above problem arose the idea to create a prediction system with the C4.5 algorithm method in carrying out predictions, a website-based writer system. Therefore, the author took the title "Application of the C4.5 Algorithm Method for Calculating the Sales Amount of Paving Production on CV. SR Lamongan".

The C4.5 algorithm method has been used in predicting sales in various fields of the company, conducting research on the implementation of data mining with the C4.5 algorithm to predict the acceptance of prospective new employees at PT. WISE. From the study, the results of measuring the prediction success rate of prospective new employees were obtained by 71%. (Haryanto Fandi Ferdian and Hansun Seng, 2017).

Hamida et al. 2014 stated that the C4.5 algorithm is one of the algorithms used to perform classification or segmentation or grouping and is predictive. The basis of the C4.5 algorithm is the formation of the decision pohn. The branches of the decision tree are questions of classification and the leaves are classes or segments. This study uses one of the Data Mining methods, namely the C4.5 Algorithm, which is an algorithm that can be used to form a decision tree. Meanwhile, the decision tree is one of the classification methods that is quite easy for humans to interpret. A decision tree is a prediction model using a tree structure or hierarchical structure. The concept of the decision tree is to turn data into a decision tree and decision rules. The main benefit is that it breaks down the complex decision-making process into a simpler so that decision-makers will find it easier to find solutions to existing problems.

In the C4.5 algorithm to build a decision tree, the first thing that is done is to choose the attribute as the root. Then a branch is made for each value in the root. The next step is to divide the cases into branches. Then repeat the process for each branch until all cases on the branch have the same class.

RESEARCH METHOD

At this stage, the author introduces the problems that have been described above in determining steps to determine the most appropriate research method to solve problems in society, including using a descriptive method, which is a method that in the first stage the author does by collecting the necessary data and materials first, in the next stage the author processes and discusses until a conclusion is must have a method used or used in his research to support the search for facts and data collection to build the application that will be designed, so the author conducts several research methods, including the following:

- a. Literature Study, This method is the basis for the preparation of a thesis, namely reading books, journals, articles from the internet, and other readings that support the writing of this thesis.
- b. Interview Method, Interviews are an information gathering technique by asking a number of questions orally to answer orally according to the data needed by the writer.
- c. System Planning, In the process of designing the sales system for the production of paving pads CV.SR Lamongan using the C4.5 algorithm using PHP and mySQL programming languages.
- d. Program creation, programming, Creating a program is creating program code that is in accordance with the results of the system design stage with the aim of optimizing the data to be closer to the machine language and changing the design into a form that is understood by the machine.
- e. Testing, aims to focus on the internal logic of the software, external fingsi and look for any possible errors. And whether the system application is as desired.
- f. Maintenance or maintenance, it is possible that the Software has changed the addition of functions that the officer wants and the possibility of encountering errors when running. The goal is for the program to run optimally.

In designing the system in the application for predicting the first sale of internet package quota, the author makes a design using a usecase, activity diagram, sequence, class diagram.



Figure 1. Research Methods

1. Data Types (Attributes)

In the C4.5 algorithm, data is divided into Independent Attributes (criteria) and Dependent Attributes (target/label). Based on the context of CV. SR Lamongan, the data commonly used includes:

Historical Sales Data: Monthly or weekly sales records.

Attributes (Variables):

Type of Paving: (e.g., Brick Paving, Hexagon, U-Ditch).

Stock: (High, Medium, Low).

Season/Month: (Rainy Season, Dry Season, or the beginning of the project fiscal year).

Price: (Stable, Rising).

Label/Class (Target): The predicted output, categorized as "Bestseller" or "Not Bestseller".

2. Amount of Data (Dataset)

To achieve good accuracy in the C4.5 method: Training Data: Usually covers 70-80% of the total data (for example, sales data from the last 2-3 years). Testing Data: The remaining 20-30% is used to test how well the generated decision tree can correctly predict new data.

- To select attributes with roots, it is based on the highest gain value of the existing attributes. To calculate the gain the following formula is used:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots (1)$$

Description:

S : case set

A : Attribute

n : the number of partitions attribute A

i : number of cases on partition i

S : number of cases in S

Gain : Information Gain

While the formula for calculating entropy can be seen as follows:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \dots\dots\dots (2)$$

Where :

S : Case Set

n : number of partitions S

Pi : The number of cases on the ith partition.

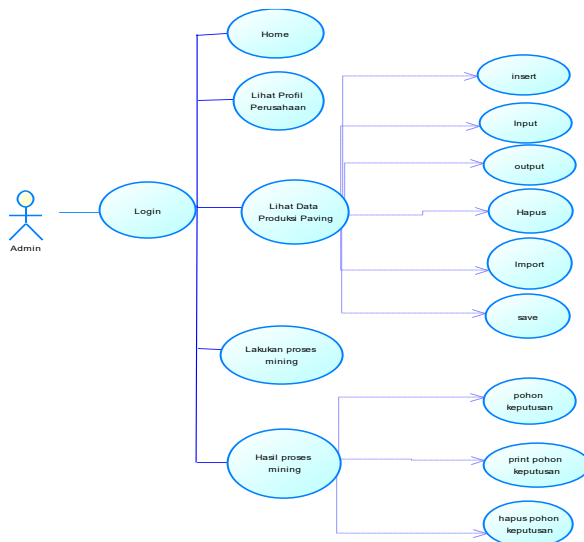


Figure 2. Use Case Diagram

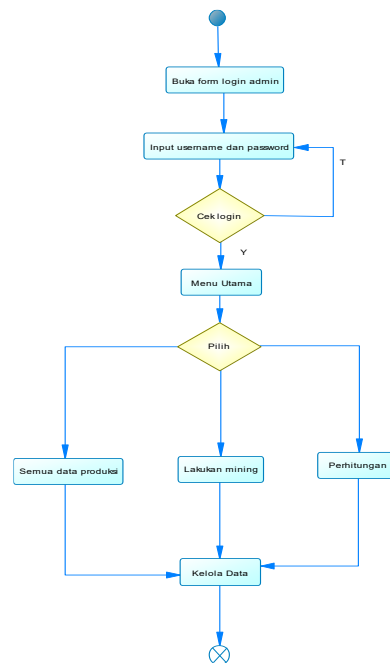


Figure 3. Admin Activity Diagram

4. Data Planning

A use case diagram is a description of the functions of a system from the user's perspective. Use cases work by describing the typical interaction between the user of a system and the system itself through a story of how a system is used. In the Use Case Diagram, the user actor is faced with several menu features that can be used. Use case diagrams are activities that will occur in the system between actors and use cases.

Actors portray the parties who play a role in the system (figure 2). In a system built to predict sales on CV. SR Lamongan has one actor. Actors are also admins who can do anything in the system such as add settings, view mining, add attributes, display processes, view rules show cases.

In the admin activity diagram (figure 3), the overall process carried out by the admin in managing the system will be explained, namely the admin can enter the system by login, then can view and input sales data, paving sales transaction data, carry out the sales prediction process, see the prediction results and see the sales prediction report data and can print it. And the system will display all the data according to what the admin wants.

In the login diagram sequence, the admin process in entering the system will be explained, namely logging in by entering the username and password. If the username entered is correct, it will appear on the main menu page, and if the username and password entered are incorrect, the system will give a warning to re-enter the correct username and password. After logging in, the admin can manage all the data in the system. The process will be described in Figure 4.

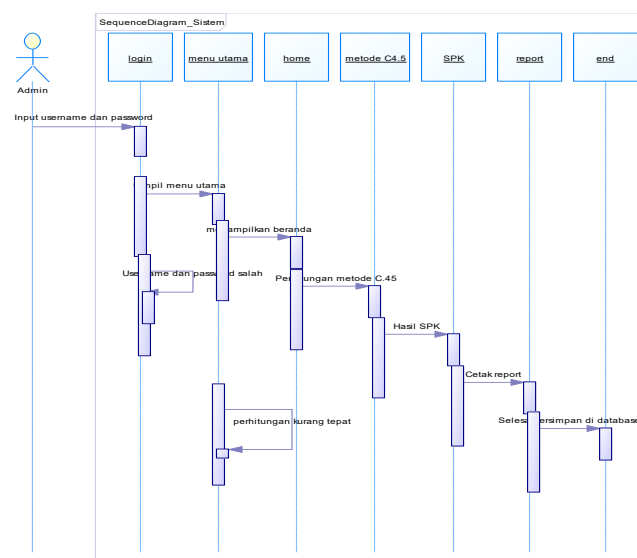


Figure 4. Sequence Diagram System

Class diagram is a type of diagram in UML that is used to display the classes and packages that exist in a system that will be used later, so this diagram can provide an overview of the system and the relationships contained in the system. Diagram class image as figure 5.

In the Class Diagram there are 6 tables, namely admin, item data, C4.5 method data, criteria, and iterations, and the Decision Support System (DSS). From the six tables, it can be explained that the admin will manage all the data in the system, starting with the admin managing the data of goods, then looking at the C4.5 decision tree table and the admin can calculate the sales of paving production. Furthermore, all data will be processed in programming.

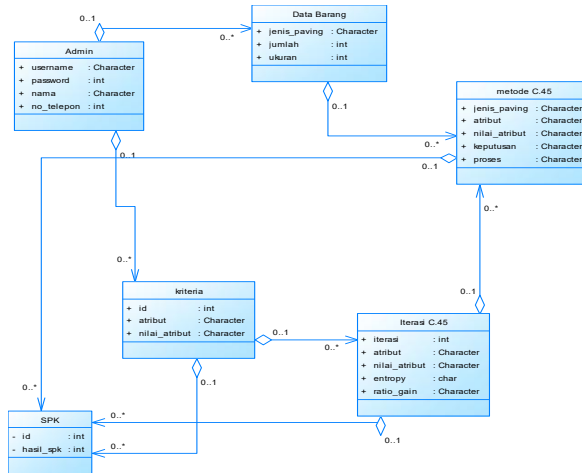


Figure 5. Class Diagram

RESULTS AND DISCUSSION

A. Interface Program

a. Form login

This login form is provided by the user with a username and password, namely username (manager) and password (admin) then click login (figure 6).

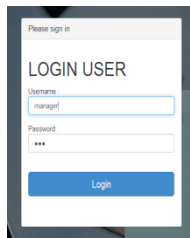


Figure 6. Login Form



Figure 7. Home Page

No	Kode Barang	Jenis Barang	Ukuran/Merek	Tahun	Harga	Status	Action
1	BRG_1021	Paving1	Uskup	2014	Rp 15500000	Tidak Laris	Edit Hapus
2	BRG_6221	Paving2	Segi Empat	2016	Rp 12000000	Laris	Edit Hapus
3	BRG_8005	Paving2	Segi Enam	2017	Rp 13000000	Laris	Edit Hapus
4	BRG_1244	paving1	Segi Empat	2012	Rp 9000000	Laris	Edit Hapus
5	BRG_7334	Paving1	Uskup	2011	Rp 10000000	Laris	Edit Hapus
6	BRG_1337	Paving2	Segi Enam	2017	Rp 13000000	Laris	Edit Hapus
7	BRG_5535	Paving2	Segi Empat	2016	Rp 12000000	Laris	Edit Hapus

Figure 8. Transaction Data

b. Home Page

On this page is the main page that explains about logging in with the user's name (figure 7).

c. Transaction Data

This transaction data contains several transaction items from the sales transactions of paving production.

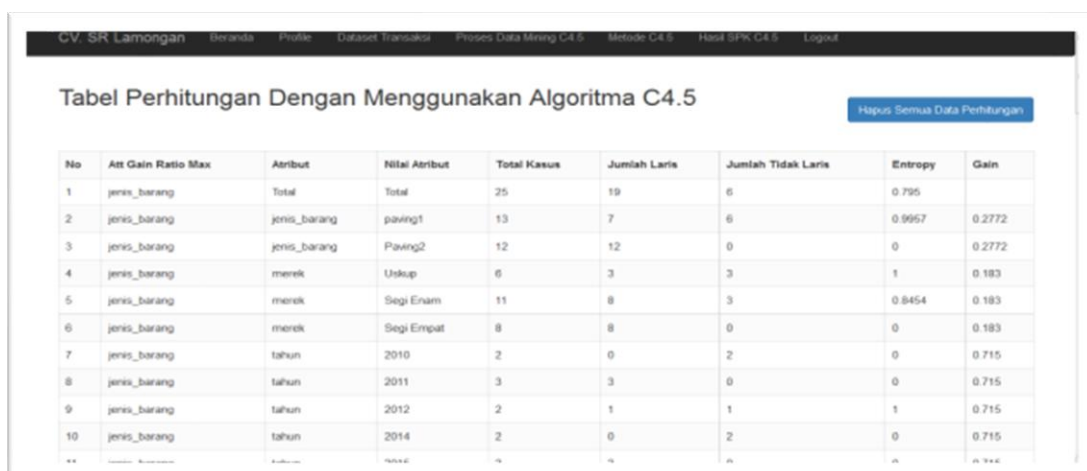
B. Evaluate

From the results in this research tool, the results of the number of cases for best-selling decisions, the number of cases for non-best-selling decisions, the entropy of all cases and cases divided based on the attributes of the number of goods, goods sold, and the rest of goods. After that, do a gain calculation for each attribute. The results of the calculation are shown by Table 1.

Table 1. Node 1 Calculation

Variabel	Atribut	Jumlah Kasus	Tidak Terjual (S1)	Terlaris (S2)	Entropy	Gain
Total		25	6	19	0,790	
Harga						0,203
	A	7	2	5	0,860	
	B	6	2	4	0,540	
Merek	C	2	2	0	0,000	
						0,180
	Rectangle	8	0	8	0,000	
Jenis	Uskup	6	3	3	1,000	
						0,277
	Paving 1	12	0	12	0,000	
Tahun	Paving 2	13	6	7	0,990	
						0,143
	A	8	3	5	0,950	
	B	17	3	14	0,671	

The application of the C4.5 method in the calculation of the number of sales of paving production on CV. SR Lamongan uses a website programming language. In operating this program, the admin must first do the login process to enter the main menu, after the admin successfully logs in, then the admin inputs the type of item, brand of goods, number of goods and price of goods. After all the data is completed, then the admin can carry out the sales calculation process using the C4.5 algorithm method which gets the results as shown in Figure 9.



Tabel Perhitungan Dengan Menggunakan Algoritma C4.5

No	Att Gain Ratio Max	Atribut	Nilai Atribut	Total Kasus	Jumlah Laris	Jumlah Tidak Laris	Entropy	Gain
1	jenis_barang	Total	Total	25	19	6	0.795	
2	jenis_barang	jenis_barang	paving1	13	7	6	0.9957	0.2772
3	jenis_barang	jenis_barang	Paving2	12	12	0	0	0.2772
4	jenis_barang	merek	Uskup	6	3	3	1	0.183
5	jenis_barang	merek	Segi Enam	11	8	3	0.8454	0.183
6	jenis_barang	merek	Segi Empat	8	8	0	0	0.183
7	jenis_barang	tahun	2010	2	0	2	0	0.715
8	jenis_barang	tahun	2011	3	3	0	0	0.715
9	jenis_barang	tahun	2012	2	1	1	1	0.715
10	jenis_barang	tahun	2014	2	0	2	0	0.715

Figure 9. Calculation Table of Method C4.5

CONCLUSION

Based on the discussion and evaluation of the previous chapter, the following conclusions can be drawn:

1. This system was built using the C4.5 Algorithm Method so that it can be used to solve the problem in this study, namely a decision support system for sales for paving production using the C4.5 algorithm method with an accuracy of up to 76%.
2. This system is presented with a comparison between items and the rules of association that exist in the C4.5 algorithm method, classifying goods that are in demand and not in demand in the C4.5 algorithm method, so that it is able to determine consumer purchasing patterns for paving purchases based on the C4.5 algorithm method and to know the goods that are in demand and not in demand based on the C4.5 algorithm method.

SUGGESTIONS

Based on the results of the research, there are several suggestions that should be made to develop this system better, including the following:

1. Combining the C4.5 algorithm method with other mathematical methods can make the resulting decision-supporting values more accurate and detailed.
2. Development by adding a larger volume of data so that more data associations between goods are obtained.

REFERENCES

- [1] Adani, M. R. (2021, June 22). *Data Flow Diagram(DFD): Pengertian, Jenis, Fungsi & Contoh*. <https://www.sekawanmedia.co.id/blog/dfd-adalah/>.
- [2] Fenia, R. W. (2024, January 12). *Bootstrap: Pengertian, Fungsi, dan Kelebihannya*. <https://www.kmtech.id/post/bootstrap-pengertian-fungsi-dan-kelebihannya>.
- [3] Ferdian Harryanto, F., & Hansun, S. (2017). Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. In *March* (Vol. 3, Issue 2).
- [4] Haryanto Fandi, Ferdian and Hansun, Seng. 2017 . “*The application of the c4.5 algorithm to predict the acceptance of prospective new employees at PT. "Wise"*”, <http://www.mdp.ac.id/jatisi/vol-3-no-2/2.%20Jurnal%20Fandy%20Ferdian%20Harryanto.pdf>, Retrieved 02 June 2018.
- [5] Hamida et al. 2014. "*Analysis and Application of C4.5 Algorithm in Data Mining to Predict Student Study Period Based on Academic Score Data*", <http://repository.akprind.ac.id/sites/files/A263-272%20silvia%20lorena.pdf> , accessed 04 June 2018.
- [6] Hamzah Kadar, & Agus Budiyantra. (2024). Penentuan Kelayakan Karyawan Baru Menggunakan Data Mining Dengan Algoritma Decision Tree (C4.5). *Merkurius : Jurnal Riset Sistem Informasi Dan Teknik Informatika*, 2(6), 29–41. <https://doi.org/10.61132/mercurius.v2i6.389>
- [7] Mushthofa. (2022, June 22). *Bagan Alir*. https://id.wikipedia.org/wiki/Bagan_alir.
- [8] Nusantara, P. B. G. (2022, September 5). *Apa itu XAMPP?* <https://www.biznetgio.com/news/apa-itu-xampp>.

- [9] Penelitian, L., Pembelajaran, P., Kepada Masyarakat, P., Caesar Kurniawan, T., Zainal Abidin Pagar Alam No, J., & Lampung, B. (2017). *SEMINAR IBI DARMAJAYA PENERAPAN ALGORITMA C4.5 DALAM PENERIMAAN CALON KARYAWAN PT. TELKOM AKSES AREA LAMPUNG BERBASIS WEBSITE*.
- [10] Raharja, U. (2023, November 13). *Apa Itu PHP? Pengertian, Sejarah, Keunggulan dan Fungsinya*. <https://Raharja.Ac.Id/2023/11/13/Apa-Itu-Php-PengertianSejarahKeunggula-Dan-Fungsinya/>.
- [11] Ramadhan, H. (2023, February 12). *APA ITU HTML Pengertian dan Informasi*. <https://Sko.Dev/Wiki/Html/>.
- [12] Rivera, S. (2023, April 18). *3 examples of conceptual data models for data and analytics*. <https://www.thoughtspot.com/data-trends/data-modeling/conceptual-data-model-examples>.
- [13] Tasya khairally, E. (2024, July 31). *Apa itu studi literatur? Ini Ciri-ciri dan Teknik Pengumpulam Data Lainnya Baca artikel detikedu, "Apa Itu Studi Literatur? Ini Ciri-ciri dan teknik pengumpulan Data Lainnya" selengkapnya* <https://www.detik.com/edu/detikpedia/d-7465287/apa-itu-studi-literatur> ini ciri-ciri-dan-teknik-pengumpulan-data lainnya <https://www.detik.com/edu/detikpedia/d-7465287/apa-itu-studi-literatur>
- [14] Ticong, L.(2024, January 5). *What is a physical Data Model? Definition and Example*. <https://www.datamation.com/big-data/physical-data-models/>.
- [15] Wicaksana, Y. A. (2024, Februari 20). *Mengenal Apa itu Database serta Fungsi dan Jenisnya*. <https://www.bhinneka.com/blog/database-adalah/>.