ANALYSIS OF THE C4.5 ALGORITHM IN PREDICTING FERTILIZER AND PESTICIDE SALES AT UD. MR ZEN AGRO

Jijon Raphita Sagala[1], Penda Sudarto Hasugian[2], Melisa Van Breukelen[3], W.S.Usha Nantheni[4]
[1,3,4] *Teknologi Informasi, STMIK Pelita Nusantara, Medan, Indonesia,*
[2]*Teknologi Rekayasa Komputer Jaringan, STMIK Pelita Nusantara, Medan, Indonesia*
E-mail : [1]*sisagala@gmail.com,* [2]*penda.hasugian@gmail.com,* [3]*melisavanbreukelen97@gmail.com,* [4]
ushanandini926@gmail.com

ABSTRACT

Fertilizer and Pesticide sales prediction is a process for finding, assessing, and predicting fertilizer and pesticide sales for the future based on sales results in the past few months. The increasing demand from customers for Fertilizers and Pesticides means UD. Mr Zen Agro, which operates in supplying and selling necessities, needs to record sales transactions so that stock shortages and excess stock of goods do not occur. Using data mining techniques and applying the C4.5 Algorithm can help UD. Mr Zen Agro predicts sales of Fertilizers and Pesticides as a reference or guideline in making decision trees, which are able to predict sales data with a high level of accuracy by calculating the entropy value, information gain, split info, and gain ratio until the matter can no longer be calculated or is zero. An application system exists to predict sales of fertilizers and pesticides, which applies the C4.5 algorithm to UD. Mr Zen Agro, it can be seen how big the sales prediction results obtained can be implemented in a decision tree where from the decision tree, it can be seen which fertilizers and pesticides are in demand and which are not for sale. This system was built using the PHP programming language and MySQL as the database.

Keywords: *Data Mining, Sales Prediction, Fertilizer and Pesticide, Algorithm C4.5.*

INTRODUCTION

UD. Mr Zen Agro is a trading business that trades goods by marketing or selling fertilizers and pesticides, agricultural seeds, and agricultural equipment to farmers. UD. Mr Zen Agro is located in Situnggaling Village, JL. Sipiso-Piso, Brand District, Karo Regency. Achievement strategies in sales make UD. Mr Zen Agro is increasingly improving the quality of services to farmers in Situnggaling Village and its surroundings. Due to the increasing demand for fertilizers and pesticides from farmers, UD. Mr Zen Agro is experiencing more and more problems that arise in providing services to consumers, one of which is giving stock supplies of fertilizers and pesticides, which consumers demand. Then, fertilizers and pesticides are sold to consumers whose sales have many transactions. UD has a vast amount of sales transaction data. Mr. Zen Agro finds it challenging to conduct data analysis, which means looking at patterns in sales data to produce sales predictions for fertilizers and pesticides that are most needed and least needed by farmers,

which will later help supply and distribute fertilizers and pesticides to farmers. Extracting information from this extensive data set can be done using data mining techniques by applying the C4.5 Algorithm. Analysis and application of the C4.5 algorithm in data mining techniques can produce knowledge that can be used as a basis for decision-making in predicting future sales of fertilizers and pesticides (Dewi & Mauladi, 2020).

LITERATUR REVIEW

1. Data Mining Method

The data mining method is a method that is applied but needs to be adapted to the user's goals. There are several divisions of data mining methods, namely as follows (Populix, 2021): Classification, Association, Clustering, Regression, Forecasting, Sequencing, and Descriptive.

2. Prediction

Prediction is translated as a group of activities in making estimates using a specific sequence of processes on something or an item that can be predicted to occur in the present and future based on valid information that has been held in the past (Izyuddin & Wibisono 2020). Prediction systematically estimates something likely to happen in the future based on the available past and present information to minimize errors (the difference between something that happens and the estimated results) (Roza et al., 2020).

3. Sales

Sales is one of the main functions carried out in the marketing activities of a product by a marketing organization, an activity to deliver goods to consumers. This activity is significant, considering the increasingly complex and intense competition in the business world (Suparman, 2018).

4. Algorithm C4.5

The C4.5 algorithm is a method for creating a decision tree based on the training data that has been provided (Azwanti, 2018). Classification is a process in data mining that aims to find valuable patterns from relatively large to extensive data. The C4.5 algorithm is a development of the ID3 algorithm. The problem in data mining classification is determining the unknown dominant factor to decide from an attribute (x) to one of the previously defined label attributes (y). Tree Pruning is carried out to simplify the tree to increase accuracy. There are two approaches to pruning, namely as follows (Wanto et al., 2020)

A. The basic algorithm of C4.5 is as follows (Wanto et al., 2020):

1. Input: Training Samples, Training Labels, and Attributes

2. Output: Decision Tree.

3. The steps for completing the C4.5 Algorithm are as follows:

    a. Creates a root node for the created tree.

    b. If all samples are positive, stop with a tree with one root node and put a (+) sign.

    c. If all samples are negative, stop with a tree with one root node and put a (-) sign.

    d. If the attribute is empty, stop with a tree with a root node with a label corresponding to the highest value in the training labels.

    e. For others, start:

        o A---- attribute that classifies the sample with the best results (based on Gain Ratio)

        o Decision attributes for the root node ----A

        o For every possible value of "vi" for A

            • Add a branch below the root corresponding to A = vi

            • Define the sample "Svi" as a subset of the samples with the value vi for attribute A.

            • If the Svi sample is empty:

                ✓ Under the branch, add a leaf node with the label = the highest value in the training label.

                ✓ Others add new branches under the current C4.5 (training samples, training labels, and attributes [A].

    f. Stop.

B. Gains

To select an attribute as the root, it must be based on the highest gain value with the existing attributes using the following formula (Fikri & Verina, 2020):

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

Information :

        S = Set of Cases

        A = Attribute

        n = Number of Attribute A Partitions

        $|S_i|$ = Number of Cases in Partition i

        $|S|$ = Number of Cases in S

METHODOLOGY

1. Research Framework

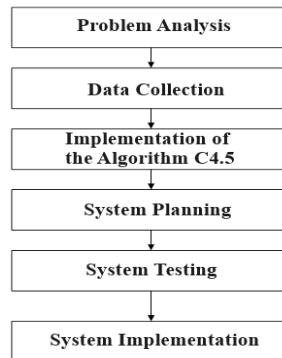The steps taken in this research can be seen in the block diagram below:

Figure 3.1 Research Framework

2. Framework Description

The research steps that will be carried out based on Figure 3.1 will be described as follows:

a. Problem Analysis

Problem analysis is the first step in conducting research. To analyze the problem in this research, what needs to be done is identifying the problem, formulating the problem, and creating research objectives and benefits.

b. Data collection

The stages in data collection in this research are literature study and interviews.

c. Implementation of the C4.5 Algorithm

After data collection has been carried out and the data has been collected, the next stage is to analyze or process the data that has been previously collected using the C4.5 algorithm with the aim that the problem formulation can be resolved so that the data mining application predicts sales of Fertilizers and Pesticides at UD. Mr Zen Agro, implementing the C4.5 Algorithm can be useful.

d. System Planning

The stages in system design are system modeling using UML and system design.

e.   System Testing

System testing is testing the functionality of all systems and whether error handling in the system created is as expected. This system will first be tested on the localhost server, for example, https://localhost/preksi_penjualan_pupuk.

f.   System Implementation

System implementation is the final step in implementing a new system. This stage is the stage of putting the system so that it is ready to operate and can be seen as an effort to realize the system that has been designed.

RESULTS AND DISCUSSION

1.   Data Analysis

Problem analysis is to understand which components are needed and recognize the elements of the situation in the problems UD faces. Mr Zen Agro in predicting fertilizer and pesticide sales. In this research, the data needed to be used as a dataset to predict fertilizer and pesticide sales is fertilizer and pesticide sales data for three months, namely October, November, and December 2023.

2.   Analysis of the Application of the C4.5 Algorithm

Analysis of the application of the C4.5 algorithm, which is a decision tree-type data classification algorithm. The C4.5 Algorithm decision tree is built with several stages, which include selecting an attribute as the root, creating branches for each value, and dividing cases into branches. Based on data on sales of fertilizers and pesticides at UD. Mr Zen Agro, calculations can be carried out using the C4.5 Algorithm, which is as follows:

3.   Root Node Calculation (Root)

To complete the calculation at the root node using the C4.5 Algorithm is as follows:

a.   Entropy Calculation

   1. Entropy Total

$$\text{Entropy(Total)} = \left(-\frac{30}{44} * \text{Log}_2 \left(\frac{30}{44} \right) \right) + \left(-\frac{14}{44} * \text{Log}_2 \left(\frac{14}{44} \right) \right) = 0.902$$

   2. Entropy Type Attribute

$$\text{Entropy(Insektisida)} = \left(-\frac{5}{11} * \text{Log}_2 \left(\frac{5}{11} \right) \right) + \left(-\frac{6}{11} * \text{Log}_2 \left(\frac{6}{11} \right) \right) = 0.994$$

$$\text{Entropy(Herbisida)} = \left(-\frac{1}{4} * \text{Log}_2 \left(\frac{1}{4} \right) \right) + \left(-\frac{3}{4} * \text{Log}_2 \left(\frac{3}{4} \right) \right) = 0.811$$

$$\text{Entropy(Fungisida)}=\left(-\frac{0}{2}*Log_2\left(\frac{0}{2}\right)\right)+\left(-\frac{2}{2}*Log_2\left(\frac{2}{2}\right)\right)=0$$

$$\text{Entropy(Fertilizer)}=\left(-\frac{6}{8}*Log_2\left(\frac{6}{8}\right)\right)+\left(-\frac{2}{8}*Log_2\left(\frac{2}{8}\right)\right)=0.811$$

$$\text{Entropy(Pestisida)}=\left(-\frac{2}{3}*Log_2\left(\frac{2}{3}\right)\right)+\left(-\frac{1}{3}*Log_2\left(\frac{1}{3}\right)\right)=0.918$$

3. Entropy Size Attribute

$$\text{Entropy(10 Gram)}=\left(-\frac{1}{1}*Log_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*Log_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(100 ML)}=\left(-\frac{8}{10}*Log_2\left(\frac{8}{10}\right)\right)+\left(-\frac{2}{10}*Log_2\left(\frac{2}{10}\right)\right)=0.722$$

$$\text{Entropy(250 ML)}=\left(-\frac{2}{5}*Log_2\left(\frac{2}{5}\right)\right)+\left(-\frac{3}{5}*Log_2\left(\frac{3}{5}\right)\right)=0.971$$

$$\text{Entropy(50 ML)}=\left(-\frac{1}{3}*Log_2\left(\frac{1}{3}\right)\right)+\left(-\frac{2}{3}*Log_2\left(\frac{2}{3}\right)\right)=0.918$$

$$\text{Entropy(40 Gram)}=\left(-\frac{1}{1}*Log_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*Log_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(250 Gram)}=\left(-\frac{1}{1}*Log_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*Log_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(500 Gram)}=\left(-\frac{2}{2}*Log_2\left(\frac{2}{2}\right)\right)+\left(-\frac{0}{2}*Log_2\left(\frac{0}{2}\right)\right)=0$$

$$\text{Entropy(400 ML)}=\left(-\frac{1}{2}*Log_2\left(\frac{1}{2}\right)\right)+\left(-\frac{1}{2}*Log_2\left(\frac{1}{2}\right)\right)=1$$

$$\text{Entropy(500 ML)}=\left(-\frac{6}{8}*Log_2\left(\frac{6}{8}\right)\right)+\left(-\frac{2}{8}*Log_2\left(\frac{2}{8}\right)\right)=0.811$$

$$\text{Entropy(1 Liter)}=\left(-\frac{4}{9}*Log_2\left(\frac{4}{9}\right)\right)+\left(-\frac{5}{9}*Log_2\left(\frac{5}{9}\right)\right)=0.991$$

$$\text{Entropy(200 ML)}=\left(-\frac{1}{1}*Log_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*Log_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(1000 Gram)}=\left(-\frac{0}{1}*Log_2\left(\frac{0}{1}\right)\right)+\left(-\frac{1}{1}*Log_2\left(\frac{1}{1}\right)\right)=0$$

4. Sold Attribute Entropy

$$\text{Entropy(1 s/d 50)}=\left(-\frac{6}{18}*Log_2\left(\frac{6}{18}\right)\right)+\left(-\frac{12}{18}*Log_2\left(\frac{12}{18}\right)\right)=0.918$$

$$\text{Entropy(51 s/d 100)}=\left(-\frac{16}{17}*Log_2\left(\frac{16}{17}\right)\right)+\left(-\frac{1}{17}*Log_2\left(\frac{1}{17}\right)\right)=0.323$$

5. Entropy of Price Attributes

$$\text{Entropy(Rp.1.000-Rp.25.000)}=\left(-\frac{3}{4}*\text{Log}_2\left(\frac{3}{4}\right)\right)+\left(-\frac{1}{4}*\text{Log}_2\left(\frac{1}{4}\right)\right)=0.811$$

$$\text{Entropy(Rp.26.000-Rp.50.000)}=\left(-\frac{6}{8}*\text{Log}_2\left(\frac{6}{8}\right)\right)+\left(-\frac{2}{8}*\text{Log}_2\left(\frac{2}{8}\right)\right)=0.811$$

b. Split Info Calculation

1. Split Type Attribute Info

$$\text{Split Info(Type)}=\left(-\frac{18}{44}*\text{Log2}\left(\frac{18}{44}\right)\right)+\left(-\frac{9}{44}*\text{Log2}\left(\frac{9}{44}\right)\right)+\left(-\frac{6}{44}*\text{Log2}\left(\frac{6}{44}\right)\right)$$

$$\left(-\frac{8}{44}*\text{Log2}\left(\frac{8}{44}\right)\right)+\left(-\frac{3}{44}*\text{Log2}\left(\frac{3}{44}\right)\right)=2.099$$

2. Split Size Attribute Info

$$\text{Split Info(Size)}=\left(-\frac{1}{44}*\text{Log2}\left(\frac{1}{44}\right)\right)+\left(-\frac{10}{44}*\text{Log2}\left(\frac{10}{44}\right)\right)+\left(-\frac{5}{44}*\text{Log2}\left(\frac{5}{44}\right)\right)$$

$$\left(-\frac{3}{44}*\text{Log2}\left(\frac{3}{44}\right)\right)+\left(-\frac{1}{44}*\text{Log2}\left(\frac{1}{44}\right)\right)+\left(-\frac{1}{44}*\text{Log2}\left(\frac{1}{44}\right)\right)$$

$$\left(-\frac{2}{44}*\text{Log2}\left(\frac{2}{44}\right)\right)+\left(-\frac{2}{44}*\text{Log2}\left(\frac{2}{44}\right)\right)+\left(-\frac{8}{44}*\text{Log2}\left(\frac{8}{44}\right)\right)$$

$$\left(-\frac{9}{44}*\text{Log2}\left(\frac{9}{44}\right)\right)+\left(-\frac{1}{44}*\text{Log2}\left(\frac{1}{44}\right)\right)+\left(-\frac{1}{44}*\text{Log2}\left(\frac{1}{44}\right)\right)$$

$$=3.048$$

c. Gain Ratio Calculation

1. Gain Ratio Attribute Type

$$\text{Gain Ratio (Type)}=\frac{\text{Gain(Total,Type)}}{\text{Split Info (Type)}}=\frac{0.004}{2.099}=0.002$$

2. Gain Ratio Size Attribute

$$\text{Gain Ratio (Size)}=\frac{\text{Gain(Total,size)}}{\text{Split Info (size)}}=\frac{0.17}{3.048}=0.056$$

3. Gain Ratio Attributes Sold

$$\text{Gain Ratio (Sold)}=\frac{\text{Gain(Total,Sold)}}{\text{SplitInfo(Sold)}}=\frac{0.328}{1.811}=0.181$$

4. Gain Ratio Price Attribute

$$\text{Gain Ratio(Price)}=\frac{\text{Gain(Total,Price)}}{\text{SplitInfo(Price)}}=\frac{0.205}{2.161}=0.095$$

Based on the entropy, gain, split info, and gain ratio calculations above, the calculation results for the root node can be obtained as follows:

Table 1 Root Node Calculation Results (Root)

NODE	Attribute	Number of Data (S)	Number of Selling (S1)	Unsold Amount (S2)	Entropy	Gain	Split Info	Gain Ratio
1	Total	44	30	14	0.902			
Type	Insektisida	18	12	6	0.918	0.004	2.099	0.002
	Herbisida	9	6	3	0.918			
	Fungisida	6	4	2	0.918			
	Fertilizer	8	6	2	0.811			
	Pestisida	3	2	1	0.918			
Attribut	10 Gram	1	1	0	0	0.17	3.048	0.056
	100 ML	10	8	2	0.722			
	250 ML	5	2	3	0.971			
	50 ML	3	1	2	0.918			
	40 Gram	1	1	0	0			
	250 Gram	1	1	0	0			
	500 Gram	2	2	0	0			
	400 ML	2	1	1	1			
	500 ML	8	6	2	0.811			
	1 Liter	9	4	5	0.991			
	200 ML	1	1	0	0			
	1000 Gram	1	0	1	0			
Sold	1 s/d 50	18	6	12	0.918	0.328	1.811	0.181
	51 s/d 100	17	16	1	0.323			
	101 s/d 150	4	3	1	0.811			
	151 s/d 200	1	1	0	0			
	>200	4	4	0	0			
Price	Rp. 1.000 - Rp.	4	3	1	0.811	0.205	2.161	0.095
	Rp. 26.000 - Rp.	8	6	2	0.811			
	Rp. 51.000 - Rp.	7	7	0	0			
	Rp. 76.000 - Rp.	8	7	1	0.544			
	>Rp. 100.000	17	7	10	0.977			

Based on the table above, it can be obtained that the highest gain ratio value is 0.181, which is in the "Sold" attribute so that the sold attribute is not calculated for the next node but is used as the root, where the attribute used as the root is the "1 to 50" attribute. At this stage, the decision tree formed for the root node is as follows:
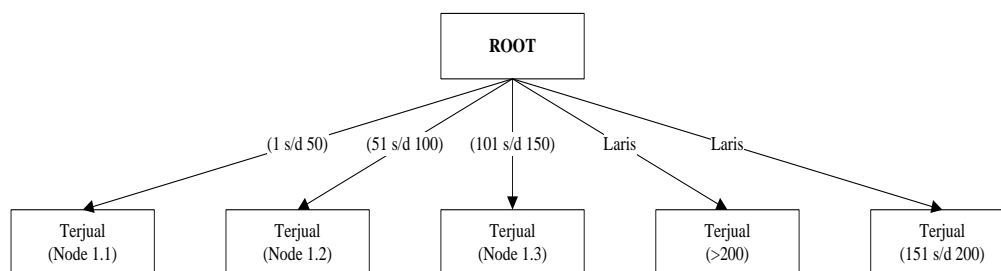

Figure 2 Root Node Decision Tree (Root

4. Node Calculation 1.3 in Sold Branches (101 to 150)

The next step is to complete the calculation at node 1.3 for branches sold "101 to 150," which are as follows:

1. Entropy Calculation

    a. Entropy Total

    $$\text{Entropy(Total)}=\left(-\frac{3}{4}*\text{Log}_2\left(\frac{3}{4}\right)\right)+\left(-\frac{1}{4}*\text{Log}_2\left(\frac{1}{4}\right)\right)=0.811\text{zrrrry}$$

    b. Entropy Type Attribute

    $$\text{Entropy(Insektisida)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    $$\text{Entropy(Fertilizer)}=\left(-\frac{2}{3}*\text{Log}_2\left(\frac{2}{3}\right)\right)+\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0.918$$

    c. Entropy Size Attribute

    $$\text{Entropy(500 ML)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    $$\text{Entropy(1 Liter)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

    $$\text{Entropy(100 ML)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    $$\text{Entropy(50 ML)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    d. Entropy Price Attribute

    $$\text{Entropy(Rp.1.000-Rp.25.000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    $$\text{Entropy(Rp.51.000-Rp.75.000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

    $$\text{Entropy(>Rp.100.000)}=\left(-\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)+\left(-\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)=1$$

2. Gain Calculation

    a. Gain Attribute Type

    $$\text{Gain(Total,Type)}=0.811-\left(\frac{1}{4}*0\right)-\left(\frac{1}{4}*0.918\right)=0.123$$

    b. Size Attribute Gain

    $$\text{Gain (Total, Size)}=0.811-\left(\frac{1}{4}*0\right)-\left(\frac{1}{4}*0\right)-\left(\frac{1}{4}*0\right)-\left(\frac{1}{4}*0\right)=0.811$$

c. Price Attribute Gain

$$\text{Gain (Total, Price)} = 0.811 - \left(\frac{1}{4}*0\right) - \left(\frac{1}{4}*0\right) - \left(\frac{2}{4}*0\right) = 0.311$$

3. Split Info Calculation

a. Split Type Attribute Info

$$\text{Split Info" ("Type")} = \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4}*Log2\left(\frac{3}{4}\right)\right) = 0.811$$

b. Split Size Attribute Info

$$\text{Split Info (Size)} = \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) + \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) + \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right)$$

$$\left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) = 2$$

c. Split Price Info

$$\text{Split Info (Price)} = \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) + \left(-\frac{1}{4}*Log2\left(\frac{1}{4}\right)\right) + \left(-\frac{2}{4}*Log2\left(\frac{2}{4}\right)\right) = 1.5$$

4. Gain Ratio Calculation

a. Gain Ratio Attribute Type

$$\text{Gain Ratio (Type)} = \frac{\text{Gain(Total,Type)}}{\text{Split Info (Type)}} = \frac{0.123}{0.811} = 0.151$$

b. Gain Ratio Size Attribute

$$\text{Gain Ratio (Size)} = \frac{\text{Gain(Total,size)}}{\text{Split Info (size)}} = \frac{0.811}{2} = 0.406$$

c. Gain Ratio Price Attribute

$$\text{Gain Ratio (Sold)} = \frac{\text{Gain(Total,Price)}}{\text{SplitInfo(Price)}} = \frac{0.311}{1.5} = 0.208$$

Based on the entropy, gain, split info, and gain ratio calculations above, the calculation results for node 1.3 for branches sold "101 to 150" can be obtained as follows:

Table 2 Node Calculation Results 1.3

NODE	Attribute	Number of Data (S)	Number of Selling (S1)	Unsold Amount (S2)	Entropy	Gain	Split Info	Gain Ratio
1.3	Total	4	3	1	0.811			
Type	Insektisida	1	1	0	0	0.12	0.81	0.15
	Fertilizer	3	2	1	0.918	2	1	1

Attribut	500 ML	1	1	0	0	0.81	2	0.40
	1 Liter	1	0	1	0	1		6
	100 ML	1	1	0	0			
	50 ML	1	1	0	0			
Price	Rp. 1.000 - Rp.	1	1	0	0	0.31	1.5	0.20
	Rp. 51.000 - Rp.	1	1	0	0	1		8
	>Rp. 100.000	2	1	1	1			

Based on the calculation results for node 1.3 in the table above, the decision tree formed for node 1.3 can be obtained as follows:
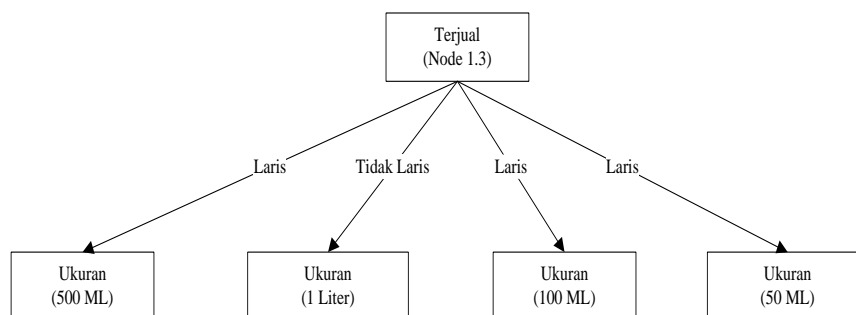

Figure 3 Node Decision Tree 1.3

After analyzing the application of the C4.5 Algorithm in the fertilizer and pesticide sales prediction application at UD. Mr Zen Agro can conclude that:

1. If the number of sales is 1 to 50, the price is Rp. 51,000-Rp. 75,000 is included in the "Best Selling" prediction while the price is Rp. 1,000-Rp. 25,000 is included in the "Not Selling" prediction

2. If the number of sales is 51 to 100, the types of Herbicides, Fungicides, Insecticides, and Pesticides are classified as "Best Selling." At the same time, the Fertilizer types with sizes of 1 liter and 500 ML are declared Best Selling, and the 100 ML size is declared "Not Selling."

3. If the number of sales is 101 to 150, the 500 ML, 100 ML, and 50 ML sizes are declared "Selling Selling" while the 1 Liter size is declared "Not Selling".

5. Results

The results of activities carried out through planning and referring to specific rules for implementing or implementing the results of the system built to achieve the objectives of the activity or system design.

a. Main page

The following is the implementation of the main page of the website when the user first accesses it via https://localhost/preksi_penjualan_pupuk

Figure 4 Main page

b. Admin Main Page

This page will appear if an admin has logged in as an admin using the correct and valid email and password.
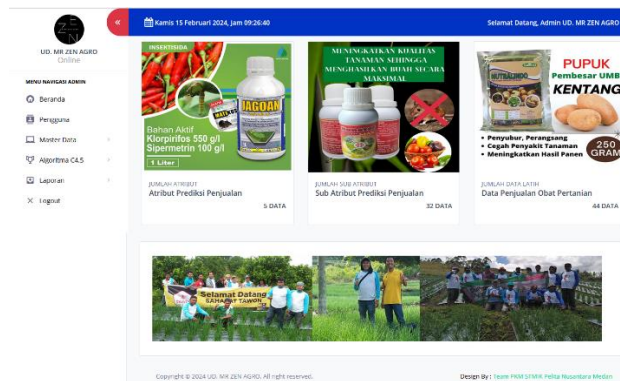

Figure 5 Admin Main Page

c. Manage Attribute Data Form

Display the output form of attribute data information on sales predictions for Fertilizers and Pesticides. This form will be displayed if the fertilizer and pesticide sales prediction attribute data has been added and stored in the database in the "tb_attribute" table.
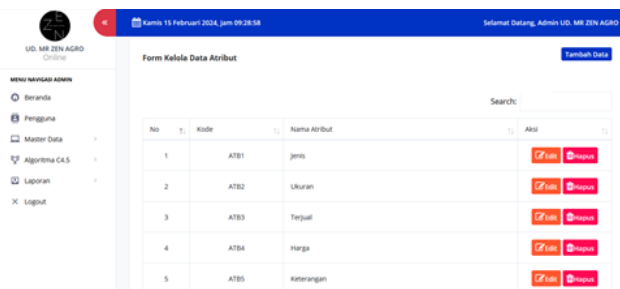

Figure 6 Manage Attribute Data Form

d. Manage Sales Data Form

Display the output form for information on fertilizer and pesticide sales data. This form display can only be accessed on the main admin page.
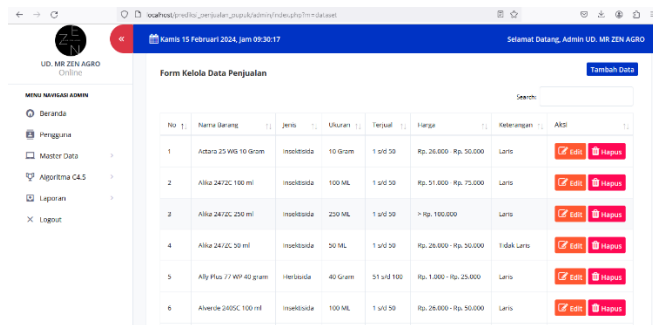
Figure 7 Form for Managing Sales Data

e. C4.5 Algorithm Analysis

The output form displays information on data from fertilizer and pesticide sales prediction calculations using the C4.5 algorithm. This form display can be accessed on the main admin page.
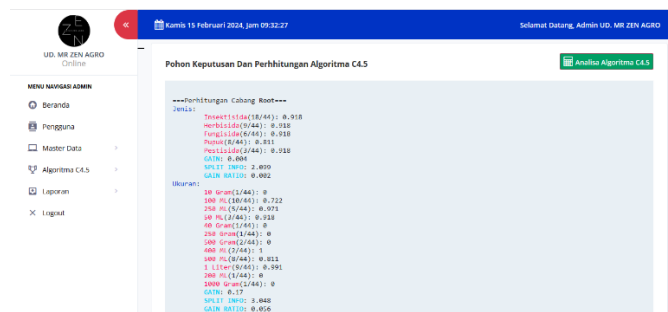

Figure 8 C4.5 Algorithm Analysis

f. Decision Trees And C4.5 Algorithm Calculations

Display of the output form for decision tree data information on fertilizer and pesticide sales predictions. This form display can be accessed on the main admin page.

Figure 9 Decision Tree and C4.5 Algorithm Calculations

g. Manage Data Input Rule Accuracy Test Form

Display output form information on data from accuracy test results of fertilizer and pesticide sales prediction rules. This form display can be accessed on the main admin page.
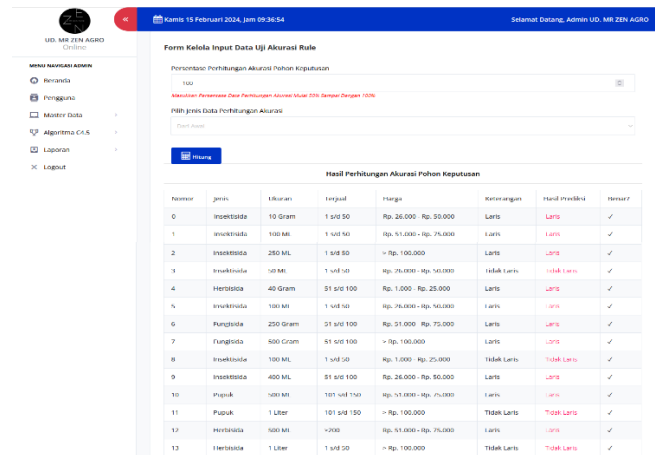
Figure 10 Manage Data Input Rule Accuracy Test Form

h. Form to Manage Data on Fertilizer and Pesticide Sales Prediction Results

Display the output form information on sales prediction data (test data). This form display can be accessed on the main admin page.
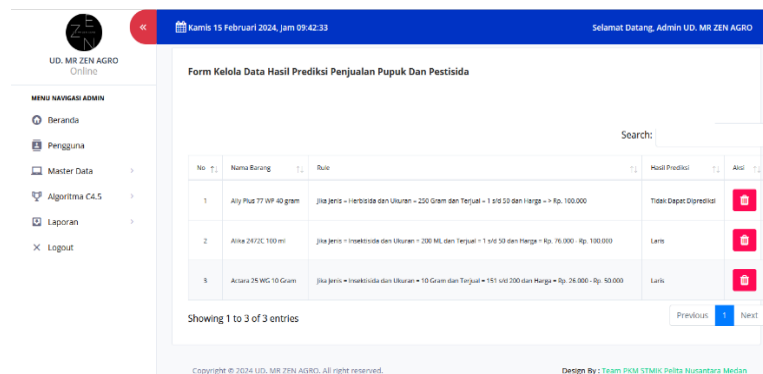

Figure 11 Form for Managing Data on Fertilizer and Pesticide Sales Prediction Results

CONCLUSION

After completing the application design, the C4.5 Algorithm was implemented to predict fertilizer and pesticide sales at UD. Mr Zen Agro, several things can be concluded as follows:

1. Using data mining techniques and applying the C4.5 Algorithm can help UD. Mr. Zen Agro predicts sales of Fertilizers and Pesticides as a reference or guideline in making decision trees, which can predict sales data with a high level of accuracy by calculating the entropy value, information gain, split info, and gain ratio until the value can no longer be calculated or is zero. An application system exists to predict sales of fertilizers and pesticides, which applies the C4.5 algorithm to UD. Mr Zen Agro, it can be seen how big the sales prediction results obtained can be implemented in a decision tree where from the decision tree, it can be seen which fertilizers and pesticides are in demand and which are not in demand for sale.

2. Data mining application system to predict sales of fertilizers and pesticides using the C4.5 algorithm at UD. Mr Zen Agro was designed using UML design or modeling, using use case diagrams, activity diagrams, and class diagrams, and was built using the PHP programming language and MySQL database.

REFERENCES

Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer, 13(1), 33. https://doi.org/10.30872/jim.v13i1.629

Dewi, K. R., & Mauladi, K. F. (2020). Analisa Algoritma C4.5 untuk Prediksi Penjualan Pupuk Dan Pestisida di Toko Dewi Sri. 6.

Fikri, A., & Verina, W. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 Pt. Murni Indah Sentosa. Infosys (Information System) JOURNAL, 5(1), 70. https://doi.org/10.22303/infosys.5.1.2020.70-83

Helmud, E. (2021). Optimasi Basis Data Oracle Menggunakan Complex View Studi Kasus : PT. Berkat Optimis Sejahtera (Pt.Bos) Pangkalpinang.7(1), 7.

Hendra, & Harman, R. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4.5 Pada PT Remicon Widyaprima. 3(3), 10.

Hermiati, R., Asnawati, A., & Kanedi, I. (2021). Pembuatan E-Commerce Pada Raja Komputer Menggunakan Bahasa Pemrograman Php Dan Database Mysql. Jurnal Media Infotama, 17(1). https://doi.org/10.37676/jmi.v17i1.1317

Indriyani, F., Yunita, Y., & Muthia, D. A. (2019). Analisa Perancangan Sistem Informasi. Graha Ilmu.

Izyuddin, A., & Wibisono, S. (2020). Aplikasi Prediksi Penjualan AC Menggunakan Decission Tree Dengan Algoritma C4.5. MISI (Jurnal Manajemen Informatika & Sistem Informasi), 3(2).

Lalo, A. K., Batarius, P., & Siki, Y. C. H. (2021). Implementasi Algoritma C4.5 Untuk Klasifikasi Penjualan Barang di Swalayan Dutalia. Jurnal Teknik Informatika UNIKA Santo Thomas, 1–12. https://doi.org/10.54367/jtiust.v6i1.1089

Lubis, M. R. (2019). Analisa Prediksi Penjualan Produk Dengan Menggunakan Metode C4.5 (Studi Kasus: PT. Kawan Lama Ace Hardware). 6(5), 5.

Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. Indonesian Journal of Information Systems, 1(1), 47–56. https://doi.org/10.24002/ijis.v1i1.1727

Muflikhah, L., Ratnawati, D. E., & MP, R. R. (2018). Data Mining. Universitas Brawijaya Press.

Nst, C. J., Putri, N., & Lawita, N. F. (2021). Perancangan Basis Data (Emos Marketplace) Sebagai Transaksi PT. Primarintis Sejahtera. Jurnal Pendidikan Tambusai, 5(2), 11.

Populix. (2021, Juni 22). Apa itu Data Mining? Pengertian, Metode, Tahapan, dan Contoh Terbaru. Populix. https://www.info.populix.co/post/data-mining-adalah

Putra, W. I., Siregar, B., & Suhatsyah, M. (2021). Sistem Informasi Geografis Sumber Daya Air Berbasis Webgis Di Badan Perencanaan Penelitian Dan Pengembangan Kabupaten Karimun. 2(1), 13.

Roza, R., Fauzan, M. N., & Rahayu, W. I. (2020). Tutorial Sistem Informasi Prediksi Jumlah Pelanggan Menggunakan Metode Regresi Linier Berganda Berbasis Web Menggunakan Framework Codeigniter. Kreatif.

Saretta, I. R. (2021, Nopember 16). Data Mining: Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya - Cermati.com. https://www.cermati.com/artikel/data-mining-pengertian-fungsi-penerapan-dan-metode-pengambilannya

Setiawan, S. (2022, Januari 14). Pengertian Penjualan—Jenis, Tujuan, Faktor, Pasar, Para Ahli. https://www.gurupendidikan.co.id/pengertian-penjualan/

Setiawan, & Shiddieqy, M. H. A. (2021). Rancang Bangun Sistem Informasi Presensi Karyawan Pada PT. Infiniti Business Synergy di Jakarta. JURNAL LENTERA ICT, 7.

Setyo, W. N., & Wardhana, S. (2019). Implementasi Data Mining Pada Penjualan Produk Di CV Cahaya Setya Menggunakan Algoritma Fp-Growth. Petir, 12(1). https://doi.org/10.33322/petir.v12i1.416

Suparman, D. (2018). Pengaruh Harga Dan Kualitas Pelayanan Terhadap Penjualan Spare Part Motor di PT. Slm (Selamat Lestari Mandiri). Jurnal Ekonomedia, 07(2), 16.

Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori. J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD), 3(1), 118. https://doi.org/10.53513/jsk.v3i1.202

Wanto, A., Siregar, M. N. H., Windarto, A. P., Hartama, D., Ginantra, N. L. W. S. R., Napitupulu, D., Negara, E. S., Lubis, M. R., Dewi, S. V., & Prianto, C. (2020). Data Mining: Algoritma dan Implementasi. Yayasan Kita Menulis.