Journal of Intelligent Decision Support System (IDSS) 5 (1) (2022) 10-19

Published by: IOCSCIENCE

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www. idss.iocspublisher.org

Application of C4.5 Algorithm to Prediction Sales at PT. Sumber Sayur Segar

[1]Fadhila, [2]Penda Sudarto Hasugian

[1,2]Informatics Engineering, STMIK Pelita Nusantara, Sumatera Utara, Indonesia

Article Info

Article history:

Received 01/12/2021
Revised 13/12/2021
Accepted 17/01/2022
Available online 30/01/2022

Keywords:

C4.5 Algorithm
Data Mining
Sales Prediction

ABSTRACT

Fresh vegetables, fruits and fresh meat are one of the basic needs for human life. The need for fresh vegetables, fruits and meat is one of the most important factors for buyers before making a purchase transaction. Likewise with the needs of fresh vegetables, fruit and meat needed by restaurants, cafes, hospitals, hotels and so on. With the increasing number of requests from customers for the needs of fresh vegetables, fruit and meat, companies engaged in the supply and sale of these necessities need to record sales transactions so that there are no stock vacancies and excess stock of goods. Therefore, companies must be more careful in providing fresh vegetables, fruits and meat which are in great demand, so it needs a data processing in the form of data mining using the C4.5 algorithm. In this study, the predicted sales transactions are the last three months of January, February and March 2021. Then for the sales prediction criteria used are in the form of price, type of goods, type of unit and month of sale so that from these criteria can be obtained sales transactions that are selling or not selling. Data mining is a process of mining important information from a very large data. While the C4.5 algorithm is a data classification that has numeric and categorical attributes, where the results of the classification process in the form of rules can be used to predict the value of discrete type attributes from new records. The system was built using the PHP programming language and MySQL as the database. This study obtained predictive results which were implemented in the form of a decision tree, namely the category of types of vegetables belonging to the best-selling sales transactions.

This is an open access article under the CC BY-NC license.

Corresponding Author:

Fadhila,
Informatics Engineering,
STMIK Pelita Nusantara Medan,
Jl. Iskandar Muda No. 1 Medan, 20154, Indonesia.
Email: fadhilla@gmail.com

1. Introduction

Along with the rapid development of information technology in the current era, it is one of the factors that advances a business field, one of which is in the field of marketing and sales. Information technology systems have been widely used in various agencies, organizations and by individuals with the aim that business activities run efficiently and accurately, so that the company's expectations for every job can be done easily and quickly. To support this, we need a system that is designed according to the needs of each company. One of them is to predict sales in the form of vegetables, fish, fruits and meat that are most needed and less needed by consumers.

To predict sales of vegetables, fish, fruits and vegetables using data mining techniques by applying the C4.5 algorithm. Data mining is the extraction of important or interesting information or patterns from existing data in large databases as a set of techniques that are used automatically to thoroughly

explore and bring to the surface complex relationships in very large data sets (Dongoran, 2019). To implement data mining techniques in predicting sales, the c4.5 algorithm is used. The C4.5 algorithm is one method for making a decision tree based on the training data that has been provided and this method is used to classify or group datasets, where the C4.5 algorithm is basically a decision tree formation (Shiddiq, Niswatin, & Farida, 2018).

This study aims to find out how to design and build a data mining application system to predict sales, and how to find out data mining techniques by applying the c4.5 algorithm to predict sales, as well as how to find out how much the result of applying the c4.5 algorithm to predict sales of vegetables, fish , meat and fruit.

Some of the literature sources used in this study are based on previous research related to the current research topic in several journals entitled Analysis of the C4.5 Algorithm for prediction of agricultural drug sales at Dewi Sri Stores where in this study, the c4.5 algorithm is very feasible to use. in predicting sales of agricultural drugs with an accuracy rate of 75% (Dewi & Mauladi, 2020). Then the research entitled C4.5 Algorithm Analysis To Predict Motorcycle Sales At PT. Capella Dinamik Nusantara Muka Kuning Branch, this study analyzes the pattern of sales data so that it can produce motorcycle sales predictions which will later be useful for motorcycle distribution in several regions (Azwanti, 2018). The following research is the Application of Data Mining for Wallpaper Sales Prediction Using the C4.5 Algorithm, this study predicts wallpaper sales transactions by applying the c4.5 algorithm in the form of a decision tree type data classification that classifies sales in terms of brand, color, motif, material quality, size and price (Eska, 2018). Then the research entitled AC Sales Prediction Application Using Decision Tree With C4.5 Algorithm, where the application of the c4.5 algorithm in this study as an alternative to predict the best-selling product that can be seen in the decision tree by calculating the value of gain information until the value cannot be calculated (Putri & Waspada, 2018). Subsequent research in the journal entitled Application of Data Mining for Readymix Sales Prediction Using the C4.5 Algorithm Method at PT. Remicom Widyaprima, this research analyzes Readymix sales transactions using product quality, delivery and product warranty variables. From the results of the application of the c4.5 algorithm, it shows that the c4.5 algorithm is very suitable for use in predicting sales transactions (Hendra, 2020). Then research on the Application of Data Mining for Prediction of Medical Device Sales Using the C4.5 Algorithm PT. Murni Indah Sentosa. Predicting the sale of medical devices with the variables used are category of goods, price and amount sold (Fikri & Verina, 2021). Another study entitled Prediction of Sales of Bread Products Using the C4.5 Algorithm at PT. Prima Top Catering. This study predicts sales of bakery products so that the stock is always stable with the variables used in the form of types of goods, prices, and quantities sold (Effendi & Rahmawati, 2019). Then another study entitled Data Mining Analysis of Tire Sales Using the C4.5 Algorithm. This study predicts tire sales transactions with the aim of optimizing tire stock in the warehouse (Anggraini, Defit, & Nurcahyo, 2018). The last research study used in this research is Product Sales Prediction Analysis Using the C4.5 Method (Case Study: PT. Kawan Lama Ace Hardware). This study predicts product sales using the c4.5 algorithm which results can be seen through a decision tree (Lubis, 2019).

2. Method

2.1 Research Framework

Research stages are steps or techniques that are arranged regularly to solve the problems to be discussed. The purpose of the research stage is to collect information or data relating to the problem to be studied.

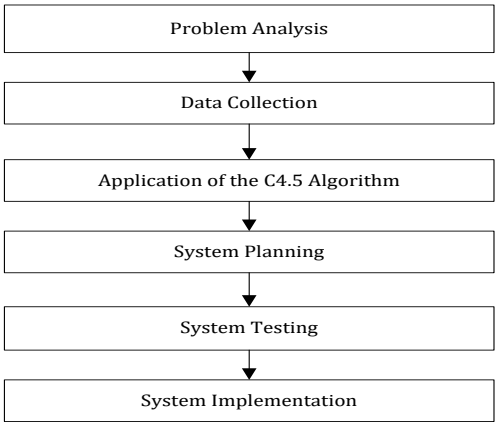

Figure 1. Research Framework

2.2 Description of the Framework

The description of the research framework is a further explanation of the steps in the research carried out:

a. Problem Analysis
Problem analysis is the first step carried out in this study which is divided into three types of analysis, namely:
1). Identifying Problems
To search, find, collect, research, register, record data and information needed in building sales prediction data mining applications
2). Formulating the Problem
Formulate the current problem in predicting the sales of vegetables, fish, meat and fruits with the aim of stabilizing the supply of goods every day.
3). Creating Research Objectives and Benefits
Determine what is the purpose of building a data mining application to predict sales and their benefits for the company.

b. Data Collection
Data collection is done by using R&D (Research And Development) technique.
1). Study of Literature
The literature study used in this research is in the form of searching for library sources or references from several sources such as books, national journals, research report articles, and internet site searches.
2). Interview
Interviews were conducted to obtain information directly by asking the respondents, the respondents referred to in this study were Mr. David, B.B.A as the President Director of PT. Sumber Sayur Segar.

c. Application of the C4.5 Algorithm
This stage is an analysis or process of data that has been previously collected using the C4.5 algorithm, with the aim that the problem formulation can be resolved so that the sales prediction data mining application at PT. Sumber Sayur Segar can be useful.

d. System Planning
System design is a stage to describe the system to be built.
1). System Modeling with UML
The process of building or forming a model of a real system in a particular formal language. To design the system using use case diagrams, activity diagrams, and class diagrams.
2). System Design

The stages of this system design are describing the system to be built in the form of main page design, input menu design and output menu design.

e. System Testing
This stage is testing the functionality of all systems and error handling on the system made whether it is as expected. Testing this system will be first carried out on the localhost server, for example: http://localhost/predik-penjualan-algoritma-c45/.

f. System Implementation
This stage is the stage of putting the system so that it is ready for operation and can be seen as an effort to realize the system that has been designed. The steps in this implementation phase are the sequence of activities from start to finish that must be carried out in realizing the system that has been designed.

3. Results and Discussion

3.1 Analysis C4.5 Algorithm

The following is an analysis of sales of vegetables, fish, meat and fruit at PT. Sumber Sayur Segar period January-March 2021.

a. Select Root Attribute
To facilitate the explanation of the C4.5 Algorithm, it is necessary to determine the root attribute based on the highest gain value for each attribute.

TABLE 1.
SALES DATA ANALYSIS WITH C4.5 ALGORITHM

Code	Name of Goods	Unit Type	Types of Goods	Month/Year	Price (IDR)	Conclusion
KTG	Potato	Kg	Fruit	Mar-21	1.470.000	Selling
SWJ	Mustard Greens	Kg	Vegetable	Mar-21	480.000	Selling
SAP	Chinese	Kg	Vegetable	Jan-21	320.000	Selling
KPH	White cabbage	Kg	Vegetable	Feb-21	518.000	Selling
TMT	Tomatoes	Kg	Fruit	Jan-21	630.000	Selling
TMN	Cucumber	Kg	Fruit	Jan-21	292.000	Selling
DSP	Celery	Kg	Vegetable	Jan-21	396.000	Selling
DPE	Prey Leaves	Kg	Vegetable	Mar-21	560.000	Not Selling
SLD	Lettuce	Kg	Vegetable	Mar-21	666.000	Selling
BMB	Onion Bombay	Kg	Vegetable	Feb-21	600.000	Not Selling
CMP	Red Chili	Kg	Vegetable	Feb-21	1.575.000	Not Selling
CHP	Green Chili	Kg	Vegetable	Mar-21	1.064.000	Selling
JMT	Old Ginger	Kg	Vegetable	Jan-21	224.000	Selling
WTL	Carrot	Kg	Vegetable	Mar-21	210.000	Not Selling
JGK	Peeled Corn	Kg	Fruit	Jan-21	200.000	Selling
PSK	Banana Kepok	SSR	Fruit	Feb-21	270.000	Not Selling
PSB	Banana	SSR	Fruit	Jan-21	700.000	Selling
JKN	Lime	Kg	Fruit	Jan-21	84.000	Not Selling
JKM	Honey Orange	Kg	Fruit	Jan-21	750.000	Selling
MLM	Melon	Kg	Fruit	Mar-21	375.000	Not Selling
SMK	Watermelon	Kg	Fruit	Mar-21	390.000	Not Selling
ALP	Avocado	Kg	Fruit	Mar-21	448.000	Not Selling
BKG	Jicama	Kg	Fruit	Mar-21	110.000	Selling
AMF	Red Apple	Kg	Fruit	Feb-21	320.000	Not Selling
BRS	Rice	Kg	Basic Food	Jan-21	5.600.000	Selling
MGCH	Cooking Oil	Kg	Basic Food	Jan-21	1.800.000	Selling
BKH	White Rice	Kg	Basic Food	Jan-21	200.000	Selling
BKP	Black Stiky Rice	Kg	Basic Food	Mar-21	340.000	Selling
GUP	Sugar	Kg	Basic Food	Jan-21	1.120.000	Selling
TAE	Chicken Eggs	PPN	Basic Food	Jan-21	3.510.000	Selling

Code	Name of Goods	Unit Type	Types of Goods	Month/Year	Price (IDR)	Conclusion
TET	Wheat	Kg	Basic Food	Mar-21	50.000	Not Selling
MTG	Butter	Kg	Basic Food	Mar-21	255.000	Selling
KPS	Cheese Prochis	PACK	Basic Food	Feb-21	160.000	Selling
TPY	Quail eggs	BTR	Basic Food	Mar-21	9.000	Not Selling
AYT	Whole Chicken	Kg	Meat	Mar-21	40.160.000	Selling
ATG	Broilers	Kg	Meat	Jan-21	3.132.500	Selling
AKS	Kalasan Chicken	Kg	Meat	Jan-21	576.000	Not Selling
AYK	Kampong	Kg	Meat	Jan-21	1.156.000	Not Selling
TLA	Chicken Bone	Kg	Meat	Feb-21	36.000	Not Selling
CKA	Chicken Feet	Kg	Meat	Feb-21	360.000	Selling
LMKA	Chicken fat	Kg	Meat	Jan-21	14.000	Not Selling
HAM	Gizzar Heart	Kg	Meat	Feb-21	56.000	Not Selling
FDD	Breast Filler	Kg	Meat	Mar-21	5.225.500	Selling
FPH	Thigh Fillers	Kg	Meat	Jan-21	1.182.000	Not Selling
DKG	Lamb	Kg	Meat	Jan-21	665.500	Not Selling
DSA	Beef	Kg	Meat	Feb-21	13.585.000	Selling
DHD	Special Meat in	Kg	Meat	Feb-21	10.125.000	Selling
OXT	Oxtail	Kg	Meat	Feb-21	5.500.000	Selling
IGAS	Beef Ribs	Kg	Meat	Jan-21	660.000	Not Selling
SAF	Fiesta Chicken	BKS	Meat	Feb-21	3.984.000	Selling
BFFLS	Buffalo	Kg	Meat	Feb-21	1.658.000	Selling
PBBI	Pork Ribs	Kg	Meat	Feb-21	3.835.000	Selling
CKTBI	Pork Chop	Kg	Meat	Mar-21	1.625.000	Selling
DKPP	Crab Meat	PACK	Fish	Feb-21	870.000	Selling

b. Create a Branch for Each Case

TABLE 2.
NODE 1 CALCULATION FOR PRICE

NODE	Attribute	Attribute Value	Amount Case (S)	Selling (S1)	Not Selling (S2)	Entropy	Gain
1	Total	Total	55	35	20	0.9457	
	Types of Goods	Vegetables	11	7	4	0.9457	0.0353
		Fruit	14	8	6	0.9852	
		Meat	19	11	8	0.9819	
		Fish	1	1	0	0	
		Basic of Foods	10	8	2	0.7219	
		Bks	1	1	0	0	
		Pack	2	2	0	0	
		Ppn	1	1	0	0	0.0764
		Btr	1	0	1	0	
		SSR	2	1	1	1	
	Month/Year	Jan/2021	22	15	7	0.9024	
		Feb/2021	17	11	6	0.9367	0.0076
		Mar/2021	16	9	7	0.9887	
	Price (IDR)	9000	1	0	1	0	0.873
		14000	1	0	1	0	
		36000	1	0	1	0	
		50000	1	0	1	0	
		56000	1	0	1	0	
		84000	1	0	1	0	
		110000	1	1	0	0	
		160000	1	1	0	0	
		200000	2	2	0	0	
		210000	1	0	1	0	
		224000	1	1	0	0	
		255000	1	1	0	0	
		480000	1	1	0	0	
		270000	1	0	1	0	

NODE	Attribute	Attribute Value	Amount Case (S)	Selling (S1)	Not Selling (S2)	Entropy	Gain
		292000	1	1	0	0	
		320000	2	1	1	1	
		340000	1	1	0	0	
		375000	1	0	1	0	
		390000	1	0	1	0	
		396000	1	1	0	0	
		448000	2	1	1	1	
		518000	1	1	0	0	
		560000	1	0	1	0	
		576000	1	0	1	0	
		600000	1	0	1	0	
		630000	1	1	0	0	
		660000	1	0	1	0	
		665000	0	0	0	0	
		666000	1	1	0	0	
		700000	1	1	0	0	
		750000	1	1	0	0	
		870000	1	1	0	0	
		1064000	1	1	0	0	
		1120000	1	1	0	0	
		1182500	1	0	1	0	
		1575000	1	0	1	0	
		1625000	1	1	0	0	
		1800000	1	1	0	0	
		3152500	0	0	0	0	
		3510500	0	0	0	0	
		3835000	1	1	0	0	
		5500000	1	1	0	0	
		5600000	1	1	0	0	
		10125000	1	1	0	0	
		13595000	0	0	0	0	
		40160000	1	1	0	0	

c. Calculating Entrophy and Gain Values for Node 1

a. Node 1 Calculation for Total

$$\text{Entropy(Total)}=\left(-\frac{S_1}{S}*\text{Log}_2\left(\frac{S_1}{S}\right)\right)+\left(-\frac{S_2}{S}*\text{Log}_2\left(\frac{S_2}{S}\right)\right)=\left(-\frac{35}{55}*\text{Log}_2\left(\frac{35}{55}\right)\right)+\left(-\frac{20}{55}*\text{Log}_2\left(\frac{20}{55}\right)\right)=0.4150$$

$$+0.5307 = 0.9457$$

b. Entropy Calculation

- *Entropy* Types of Goods

$$\text{Entropy(Total,Vegetables)}=\left(-\frac{7}{11}*\text{Log}_2\left(\frac{7}{11}\right)\right)+\left(\frac{4}{11}*\text{Log}_2\left(\frac{4}{11}\right)\right)=0.9457$$

$$\text{Entropy(Total,Fruits)}=\left(-\frac{8}{14}*\text{Log}_2\left(\frac{8}{14}\right)\right)+\left(\frac{6}{14}*\text{Log}_2\left(\frac{6}{14}\right)\right)=0.9852$$

$$\text{Entropy(Total,Meat)}=\left(-\frac{11}{19}*\text{Log}_2\left(\frac{11}{19}\right)\right)+\left(\frac{8}{19}*\text{Log}_2\left(\frac{8}{19}\right)\right)=0.9819$$

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

$$\text{Entropy(Total,Basic of Food)}=\left(-\frac{8}{10}*\text{Log}_2\left(\frac{8}{10}\right)\right)+\left(\frac{2}{10}*\text{Log}_2\left(\frac{2}{10}\right)\right)=0.7219$$

- *Entropy* Unit

$$\text{Entropy(Total,Kg)}=\left(-\frac{30}{48}*\text{Log}_2\left(\frac{30}{48}\right)\right)+\left(\frac{18}{48}*\text{Log}_2\left(\frac{18}{48}\right)\right)=0.9544$$

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

$$\text{Entropy(Total,Pack)}=\left(-\frac{2}{2}*\text{Log}_2\left(\frac{2}{2}\right)\right)+\left(\frac{0}{2}*\text{Log}_2\left(\frac{0}{2}\right)\right)=0$$

$$\text{Entropy(Total,Ppn)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{2}*\text{Log}_2\left(\frac{0}{2}\right)\right)=0$$

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

$$\text{Entropy(Total,SSR)}=\left(-\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)+\left(\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)=1$$

– *Entropy* Month/Year

$$\text{Entropy(Total,Jan/2021)}=\left(-\frac{15}{22}*\text{Log}_2\left(\frac{15}{22}\right)\right)+\left(\frac{7}{22}*\text{Log}_2\left(\frac{7}{22}\right)\right)=0.9024$$

$$\text{Entropy(Total,Feb/2021)}=\left(-\frac{11}{17}*\text{Log}_2\left(\frac{11}{17}\right)\right)+\left(\frac{6}{17}*\text{Log}_2\left(\frac{6}{17}\right)\right)=0.9367$$

$$\text{Entropy(Total,Mar/2021)}=\left(-\frac{9}{16}*\text{Log}_2\left(\frac{9}{16}\right)\right)+\left(\frac{7}{16}*\text{Log}_2\left(\frac{7}{16}\right)\right)=0.9887$$

– *Entropy* Price (IDR)

$$\text{Entropy(Total,9000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,14000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,36000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,56000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,480000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,320000)}=\left(-\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)+\left(\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)=1$$

$$\text{Entropy(Total,375000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,390000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,448000)}=\left(-\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)+\left(\frac{1}{2}*\text{Log}_2\left(\frac{1}{2}\right)\right)=1$$

$$\text{Entropy(Total,518000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,560000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,666000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,700000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,750000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,1470000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,1575000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,1625000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,1658000)}=\left(-\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)+\left(\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)=0$$

$$\text{Entropy(Total,3152500)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,3510500)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,2984000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

$$\text{Entropy(Total,40160000)}=\left(-\frac{0}{1}*\text{Log}_2\left(\frac{0}{1}\right)\right)+\left(\frac{1}{1}*\text{Log}_2\left(\frac{1}{1}\right)\right)=0$$

c. Gain Calculation

- Calculation of Information Gain Value for Items

$$\text{Gain(Total,Types of Food)}=0.9457-\left(\frac{11}{55}*0.9457\right)-\left(\frac{14}{55}*0.9852\right)\left(\frac{19}{55}*0.9819\right)-\left(\frac{1}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{10}{55}*0.7219\right)=\mathbf{0.0353}$$

- Calculation of Information Gain Value for Unit Type

$$\text{Gain(Total,Unit Type)}=0.9457-\left(\frac{48}{55}*0.9544\right)-\left(\frac{1}{55}*0\right)-\left(\frac{2}{55}*0\right)-\left(\frac{0}{55}*1\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{2}{55}*1\right)=\mathbf{0.0764}$$

- Calculation of Information Gain Value for Month/Year

$$\text{Gain(Total,Month/Year)}=0.9457-\left(\frac{22}{55}*0.9024\right)-\left(\frac{17}{55}*0.9367\right)-\left(\frac{16}{55}*0.9887\right)=\mathbf{0.0076}$$

- Calculation of Information Gain Value for Price (IDR)

$$\text{Gain(Total,Price(IDR))} = 0.9457-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{2}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{2}{55}*1\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{2}{55}*1\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{1}{55}*0\right)-\left(\frac{0}{55}*0\right)-\left(\frac{1}{55}*0\right)=\mathbf{0.873}$$

3.2 System Design

a. Use Case Diagram

Use Case Diagram aims to find out what are the functions that exist in the data mining system to predict sales at PT. Sumber Sayur Segar uses the c4.5 algorithm which is shown below:

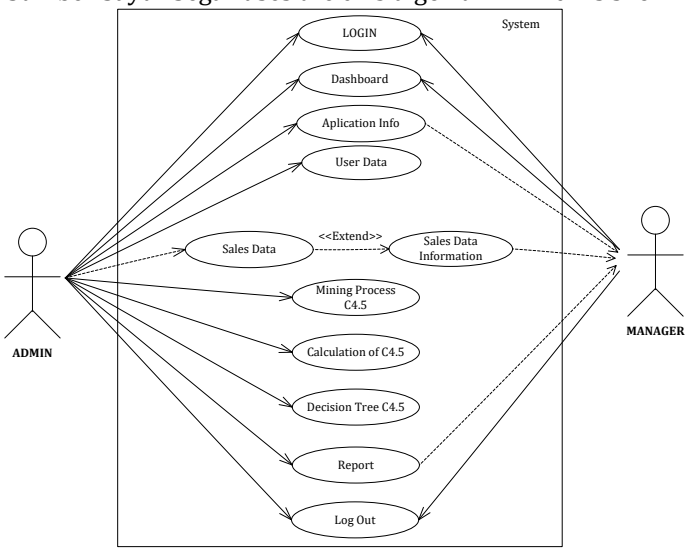

Figure 2. Use Case Diagram

b. Output Result

After designing the system, the system can be implemented from the output predictions of sales at PT. Sumber Sayur Segar uses Algorithm c4.5 as shown in the following image:

LAPORAN PERHITUNGAN PREDIKSI PENJUALAN DENGAN ALGORITMA C4.5

NO	ATTRIBUTE GAIN RATIO MAX	ATRIBUT	NILAI ATRIBUT	TOTAL KASUS	JUMLAH LARIS	JUMLAH TIDAK LARIS	ENTROPY	GAIN
1	harga	Total	Total	55	35	20	0.9457	
2	harga	jenis_barang	Sayur-Sayuran	11	7	4	0.9457	0.0353
3	harga	jenis_barang	Buah-Buahan	14	8	6	0.9852	0.0353
4	harga	jenis_barang	Daging	19	11	8	0.9819	0.0353
5	harga	jenis_barang	Ikan	1	1	0	0	0.0353
6	harga	jenis_barang	Bahan kering	0	0	0	0	0.0353
7	harga	jenis_barang	Sembako	10	8	2	0.7219	0.0353
8	harga	satuan	Kg	48	30	18	0.9544	0.0764
9	harga	satuan	Bks	1	1	0	0	0.0764
10	harga	satuan	Pack	2	2	0	0	0.0764
11	harga	satuan	Ekor	0	0	0	0	0.0764
12	harga	satuan	Ppn	1	1	0	0	0.0764
13	harga	satuan	Btr	1	0	1	0	0.0764
14	harga	satuan	SSR	2	1	1	1	0.0764
15	harga	bulan_tahun	Januari/2021	22	15	7	0.9024	0.0076
16	harga	bulan_tahun	Februari/2021	17	11	6	0.9367	0.0076
17	harga	bulan_tahun	Maret/2021	16	9	7	0.9887	0.0076
18	harga	bulan_tahun	April/2021	0	0	0	0	0.0076
19	harga	bulan_tahun	Mei/2021	0	0	0	0	0.0076
20	harga	bulan_tahun	Juni/2021	0	0	0	0	0.0076
21	harga	bulan_tahun	Juli/2021	0	0	0	0	0.0076
22	harga	bulan_tahun	Agustus/2021	0	0	0	0	0.0076
23	harga	bulan_tahun	September/2021	0	0	0	0	0.0076
24	harga	bulan_tahun	Oktober/2021	0	0	0	0	0.0076
25	harga	bulan_tahun	November/2021	0	0	0	0	0.0076
26	harga	bulan_tahun	Desember/2021	0	0	0	0	0.0076
27	harga	harga	9000	1	0	1	0	0.873
28	harga	harga	14000	1	0	1	0	0.873
29	harga	harga	36000	1	0	1	0	0.873
30	harga	harga	50000	1	0	1	0	0.873
31	harga	harga	56000	1	0	1	0	0.873
32	harga	harga	84000	1	0	1	0	0.873
33	harga	harga	110000	1	1	0	0	0.873
34	harga	harga	160000	1	1	0	0	0.873
35	harga	harga	200000	2	2	0	0	0.873
36	harga	harga	210000	1	0	1	0	0.873
37	harga	harga	224000	1	1	0	0	0.873
38	harga	harga	255000	1	1	0	0	0.873
39	harga	harga	480000	1	1	0	0	0.873
40	harga	harga	270000	1	0	1	0	0.873
41	harga	harga	292000	1	1	0	0	0.873
110	bulan_tahun	jenis_barang	Sembako	0	0	0	0	0
111	bulan_tahun	satuan	Kg	2	1	1	1	0
112	bulan_tahun	satuan	Bks	0	0	0	0	0
113	bulan_tahun	satuan	Pack	0	0	0	0	0
114	bulan_tahun	satuan	Ekor	0	0	0	0	0
115	bulan_tahun	satuan	Ppn	0	0	0	0	0
116	bulan_tahun	satuan	Btr	0	0	0	0	0
117	bulan_tahun	satuan	SSR	0	0	0	0	0
118	bulan_tahun	bulan_tahun	Januari/2021	1	1	0	0	1
119	bulan_tahun	bulan_tahun	Februari/2021	0	0	0	0	1
120	bulan_tahun	bulan_tahun	Maret/2021	1	0	1	0	1
121	bulan_tahun	bulan_tahun	April/2021	0	0	0	0	1
122	bulan_tahun	bulan_tahun	Mei/2021	0	0	0	0	1
123	bulan_tahun	bulan_tahun	Juni/2021	0	0	0	0	1
124	bulan_tahun	bulan_tahun	Juli/2021	0	0	0	0	1
125	bulan_tahun	bulan_tahun	Agustus/2021	0	0	0	0	1
126	bulan_tahun	bulan_tahun	September/2021	0	0	0	0	1
127	bulan_tahun	bulan_tahun	Oktober/2021	0	0	0	0	1
128	bulan_tahun	bulan_tahun	November/2021	0	0	0	0	1
129	bulan_tahun	bulan_tahun	Desember/2021	0	0	0	0	1

Figure 3. Output Result

4. Conclusion

Data mining application system to predict sales using the C4.5 Algorithm at PT. XXX is designed by using UML design or modeling using use case diagrams, activity diagrams and class diagrams, and is built using the PHP programming language and MySQL database. 4. With an application system to predict sales that applies the C4.5 Algorithm at PT. XXX, it can be seen how much sales prediction results are obtained which can be implemented in a decision tree, where the decision tree can be seen that the category of vegetable goods belongs to the sales transaction of the best-selling types of goods.

References

Anggraini, S., Defit, S., & Nurcahyo, G. W. (2018). Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4. 5. *Jurnal Ilmu Teknik Elektro Komputer Dan Informatika (JITEKI)*, *4*(2), 136.

Azwanti, N. (2018). Analisa Algoritma C4. 5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Inform. Mulawarman J. Ilm. Ilmu Komput*, *13*(1), 33.

Dewi, K. R., & Mauladi, K. F. (2020). Analisa Algoritma C4. 5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, *4*(3), 109–114.

Dongoran, N. S. (2019). *PENERAPAN ALGORITMA APRIORI DAN ECONOMIC ORDER QUANTITY UNTUK PENGENDALIAN PERSEDIAN PRODUK DI ALMIRA KIDS*. Universitas Islam Negeri Sultan Syarif Kasim Riau.

Effendi, M. M., & Rahmawati, D. (2019). Prediksi Penjualan Produk Roti Menggunakan Algoritma C4. 5 Pada PT. Prima Top Boga. *Jurnal SIGMA*, *9*(1), 51–66.

Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5*.

Fikri, A., & Verina, W. (2021). PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN ALAT MEDIS MENGGUNAKAN ALGORITMA C4. 5 PT. MURNI INDAH SENTOSA. *INFOSYS (INFORMATION SYSTEM) JOURNAL*, *5*(1), 70–83.

Hendra, H. (2020). *Penerapan Data Mining untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4. 5 pada PT Remicon Widyaprima*. Prodi Sistem Informasi.

Lubis, M. R. (2019). Analisa Prediksi Penjualan Produk Dengan Menggunakan Metode C4. 5 (Studi Kasus: PT. Kawan Lama Ace Hardware). *JURIKOM (Jurnal Riset Komputer)*, *6*(5), 545–549.

Putri, R. P. S., & Waspada, I. (2018). Penerapan Algoritma C4. 5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, *4*(1), 1–7.

Shiddiq, A., Niswatin, R. K., & Farida, I. N. (2018). Ahmad Shiddiq Analisa Kepuasan Konsumen Menggunakan Klasifikasi Decision Tree Di Restoran Dapur Solo (Cabang Kediri). *Generation Journal*, *2*(1), 9–18.