



Implementation of Gain Ratio on KNN Method in Predicting Sales of Electronic Sparepart at Panasonic Service Center Lhokseumawe

Samsul Bahri Siagian^{1,*}, Samsudin², Muhammad Dedi Irawan³

^{1,2,3} Program Studi Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Jl. IAIN No. 1, Medan, Indonesia

Email: samsulbahrisiagian@gmail.com¹, samsudin@uinsu.ac.id², muhmaddediirawan@uinsu.ac.id³

* *corresponding author*

ARTICLE INFO

Article history:

Received: Mar 06, 2022

Revised: Mar 24, 2022

Accepted: Mar 28, 2022

Available online Mar 30, 2022

Keywords:

Data Mining;

Gain Ratio;

K-Nearest Neighbor;

Predicting Sales.

ABSTRACT

K-Nearest Neighbor is a good classification technique, but judging by previous studies, the accuracy of the KNN performance obtained is still inferior to other methods. In the classification process, if some characteristics are not good it can cause errors in the new classifier. As for this study, the researcher uses the gain ratio method as a parameter to see the correlation between each attribute in the dataset, and the gain ratio serves as a weighting for each attribute so as to produce a dataset. The correct way of classifying data using the KNN method, this study is very suitable for predicting sales of spare parts at the Panasonic Service Center company, where the company experienced a decline in sales, this research is very useful for predicting sales for the following month. The results of this study produce very precise predictions of distance with an accuracy value of 13%, where the comparison of the highest accuracy value is found in the total attribute with an accuracy distance of 13%, while the lowest accuracy difference is obtained in the month and type of sales dataset with 0.08%. The overall accuracy of all datasets increases by 100% with K=3, and K=5 gets 80% accuracy. So this method can be used to make sales predictions to make it easier for the company.

© 2022 JTI C.I.T. All rights reserved.

1. Introduction

Since the beginning of the emergence of technology, there have been so many various innovations and developments that have been continuously evolving until now. Technology plays a role in providing various conveniences for its users. Things that were previously impossible can become possibilities slowly materialize one by one. Activities that take up a lot of time, energy, and also cost, have now become more practical with the existence of technology [2]. With the development of the use of technology, the data collected in the database is also increasing and causing data stacks. One technique that can be used to calculate large amounts of data is data mining [3] [4].

KNN is one of the good classification techniques, but seen from various previous studies [5] [6], the accuracy of the KNN performance obtained is still less than other methods. The main cause of the low accuracy produced, due to the attribute of the dataset very influential in the classification process, if some characteristics are not good it can cause errors in the classifier new class. As for this research, the researcher uses the gain ratio method as a parameter to see a correlation between each attribute in the dataset, and the gain ratio functions as a weighting each attribute, so as to produce the right dataset for classifying data using the KNN method, this study is very suitable for predicting spare parts sales in Panasonic Service Center companies. The results of this study produce a very precise prediction with an accuracy value of 13%, where the comparison of the highest accuracy value on the total attribute with



an accuracy distance of 13%, while the lowest accuracy difference is obtained in the dataset. month and type of sales by 0.08%. the result of the overall accuracy of all datasets an increase of 100% with K=3, and K=5 getting 80% accuracy. Knn is one of the very good and simple classification methods that works well[5][6]. however, there are several studies with lower performance accuracy of KNN classification compared to other methods, one of which is in study [7] where in this study compares the performance of SVM and KNN. the research results obtained in the quote that the performance of SVM is better than KNN with an accuracy rate of 82.54% while the accuracy value of KNN is 79%. Another study can also be seen in [8] with the KNN and ANN methods, from the results of the research where the accuracy rate is 75% with 5 layers of the ANN method. in study [9] comparing KK with Naive Bayes in diagnosing a disease, with the results obtained an accuracy value of 79.62% while the overall value of KNN is only 64.84% with a distance of K = 10.

From some of the studies above, it can be seen that the accuracy of the KNN method is still lower than the other methods in the research above, one of the reasons is the low attribute value that causes data refraction which affects the classification. as for this study using the gain ratio method as the weighting of each attribute with min-max normalization which is then classified using the KNN method, so that the accuracy results are very good and there is no data bias, this research is also useful for various cases, especially in predicting sales of electronic spare parts at the panasonic lhokseumawe service center, where the problem with the company was the lack of sales of electronic spare parts in the previous year, so the use of data mining is very useful in terms of predicting sales using the gain ratio method on KNN, so that the prediction results of data mining will be better and more accurate in terms of accuracy.

Based on the above problems, the authors help provide solutions by studying the criteria and attribute patterns of selling electronic spare parts at the Panasonic Lhokseumawe Service Center, and processing existing historical data to build a web-based data mining application that can be used as a tool. predict sales of electronic spare parts and can determine the amount of stock of electronic spare parts in the future from spare part sales data. In this study, it will be useful to assist companies in determining the sale of electronic spare parts and can also determine the stock of electronic spare parts based on new information obtained from the results of the data mining process.

In this study, the method that will be used by the author in building data mining applications is the classification method by applying the Gain Ratio algorithm with the K-Nearest Neighbor method. Data mining is an analytical activity that has the aim of reviewing a collection of data that is useful in unexpected relationships and can summarize data in different ways, to make it easier to understand and use for data owners, can also be interpreted as a result of solving a problem[10][11]. or Problem solving that can be implemented in the form of a collection of program commands that reduce the attribute bias of each activity, with the use of data mining very useful for companies in solving problems, by determining the sale of electronic spare parts for the next month and so on[12]. Data mining operations according to their nature are divided into 2, namely (1) prediction (prediction driven) to answer what questions and something that is abstract or transparent. Prediction operations are used for hypothesis validation, querying and reporting. (2) discovery (discovery driven) is transparent. Discovery operations are used for exploratory data analysis, predictive modeling, database segmentation, link analysis and deviation detection[13][14].

Gain Ratio is part of the Decision tree which is the stage to determine the more appropriate attribute values, so the resulting attribute values will be more accurate, with the Gain Ratio it can reduce attribute bias that has many branches[13][15]. Gain Ratio has the following properties: 1. Great value if the data is evenly distributed, 2. Small value if all data goes in one branch[1][16]. In the process of finding the gain ratio value, you must first calculate the entropy value of the entire class using the formula (1) and split info, so that the gain ratio value for each attribute will be obtained.

$$\text{Entropy (S)} = \sum_{i=1}^n -P_i * \log_2 P_i \quad (1)$$

Information : S = Case Set, N = Number of Partitions Attribut A, P_i = Proportion of S_i to S.

After calculating the entropy value of the entire class on the attribute with equation (1), the next

step is to calculate the information gain for each attribute by using the following equation.

$$\text{Information Gain } (S, A) = \text{Entropy } (S) - \sum_{i=1}^n \frac{S_i}{S} * \text{Entropy } (S_i) \quad (2)$$

Information : A = Atribut, $|S_i|$ = Number of Cases on Partition I, $|S|$ = Number of Cases in S.

If you have obtained the information gain value in each attribute class, then perform calculations using the splitinfo formula, with the following formula:

$$\text{SplitInfo } (S, A) = \sum_i^c = 1 \frac{S_i}{S} \text{Log2} \frac{S_i}{S} \quad (3)$$

To get the gain ratio value, you must first find the splitinfo value, because the splitinfo value is very influential in terms of weighting the entropy results of the entire class with information gain, using the following formula:

$$\text{Gain Ratio} = \frac{\text{Gain}(S,A)}{\text{Split Info}(S,A)} \quad (4)$$

Gain Ratio as a measuring tool to see the correlation of the attributes in the data set, where the Gain Ratio will be used as the basis for weighting the highest gain attribute [17]. It is hoped that by giving weights to attributes, it can reduce the influence of irrelevant attributes on the results of the classification using the K-Nearest Neighbor method, so as to be able to increase the accuracy of the classification process[18][19]. An overview of the stages of the method proposed in this study is shown in the following figure:

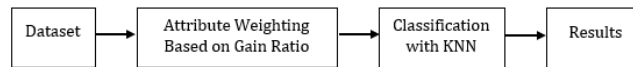


Fig 1. Proposal Method

The K-Nearest Neighbor algorithm is an algorithm method that can classify to find the closest distance between the data to be evaluated with the closest k-neighbors in the data studied. k is the closest number of data, the class that appears the most is the class that is the classification class. In the K-Nearest Neighbor algorithm, there are 5 ways to find the nearest neighbor value, namely Euclidean distance, Manhattan distance, Cosine distance, and Hamming distance[1][20]. In this study the author only uses the Euclidean distance, therefore the formula used in the calculation is:

$$\text{Euclidean} = \sqrt{\sum_{i=1}^k (X_i - Y_i)^2} \quad (5)$$

Information ; X_i = Text Data, Y_i = Sample Test Data, P = Dimension of Data.

The X_i value is the value in the Training data, while the Y_i value is the value in the Testing data. The value of K is the attribute dimension value[21]. The following are the steps to calculate the KNN algorithm: a). Determining the value of K, b). Calculates the square of the Euclidean distance (query instance) of each object to the given training data. c). Then sort the objects into groups that have the smallest Euclidean distance. d). Collects class Y labels (Nearest Neighbor Classification). e). By using the majority category of K-Nearest Neighbors, the calculated query instance value can be predicted.

2. Research Methods

The research method used in this study uses R&D (Research and Development), where this method produces a new product which can be in the form of a software (software), this method is very suitable to be used because it has 10 stages but this study only uses 5 stages, namely, pre-model development, model development, model testing, model implementation, and model revision[22][23]. This method is very suitable because R&D can collect data in the form of observations, interviews, literature studies,

using mixed and numerical datasets.

System development method, the author chose the Rational Unified Process (RUP) development method. RUP is a phased model that identifies four phases in the software process. However, unlike the waterfall model where phases are equated with process activities, the phases in RUP are more closely related to business than technical issues[24]. RUP is a method that uses object-oriented concepts, and has activities that focus on model development using UML (Unified Model Language)[25].

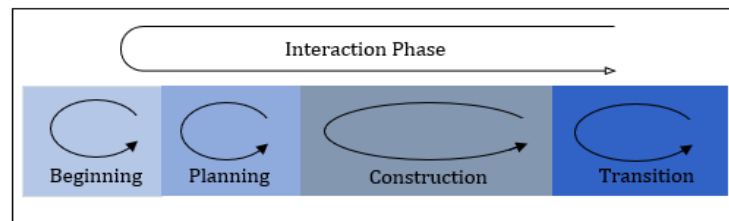


Fig 2. RUP Stage

Using the RUP method so that in this study to improve company management in good quality software that meets the desired criteria. The existence of feedback between users with interests in the system, allows the process of determining the risks that are handled at the beginning of the process, the implementation of which can affect the previous iteration [25]. The stages of the Gain Ratio and K-Nearest Neighbor Algorithm can be seen in the picture below:

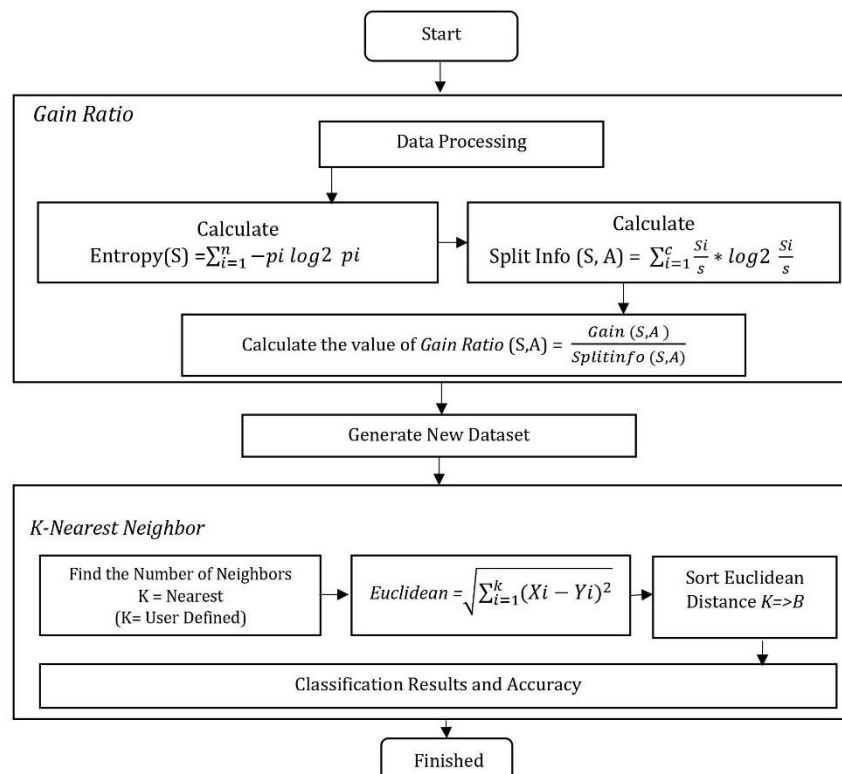


Fig 3. Gain Ratio Algorithm and K-Nearest Neighbor

In this algorithm the author starts the Gain Ratio algorithm, from data processing performs the process of recognizing data in raw data collection, then in the next process, calculating the Entropy value, to separate class objects, then performing SplitInfo calculations to select a number of attribute values from the categories used, that way will get the value of information Gain Ratio in accordance with the

provisions of the desired attribute. In the next Knn process, the calculation process determines the K value to be processed based on the specified User K value, then performs the Euclidean Data process to calculate the distance of each neighbor, then sorts the results by distance, starting from the smallest to the largest, then performs the process of determining or voting.

3. Results and Discussion

This research was conducted based on sales data obtained at the company, the dataset examined at the company consisted of 6 types of products, 34 kinds of electronic spare parts, namely as follows:

Table 1.
Electronic Sparepart Products

No	SPARE PART TYPE		
1	Dryer Motor	Resistor	Propeller
	Washing Motor	Inductor	Fan Boss
	Dryer Timer	Capacitor	Ear Boss
	Engine Gearbox	Dioda	Dynamo Box
	Machine Capacitor	Transistor	Front Caver
	Fan Belt	PCB LED	Gear Box
	Drain House	Led Light	Motor Timing
	Overload	Freon	
	Self-Timer	Double	
	Filter	Isolation	
2	Compressor	Blower	
	Condenser	Strainer	
	Evaporator	Capillary Pipe	
		Capillary PCB	
		Body Cover	

In this study, there is a dataset with a total of 1277 records after data processing, where the data has 5 attributes and 1 class label, then the numeric data transformation process is carried out on the specified criteria. The details of the data can be seen in table 2.

Table 2.
Sales data transformation

No	Spare Part Type	Month	Year	Sale Type	Total	Status
1	Body Caver	12	1	2	1	Decrease
2	Gear Box	12	1	2	1	Decrease
3	Fan Belt	12	1	2	5	Decrease
4	Blower	12	1	2	7	Increase
5	Boss Ears	12	1	2	1	Decrease
6	Dynamo Box	12	1	2	1	Decrease
7	Dryer Timer	12	1	2	1	Decrease
8	Engine Gearbox	12	1	2	2	Decrease
9	Filter	12	1	2	1	Decrease
10	Motor Timing	12	1	2	1	Decrease
11	Fan Boss	12	1	2	9	Increase
12	Transistor	12	1	1	12	Increase
13	Inductor	12	1	1	12	Increase
14	Led Lights	12	1	2	5	Decrease
15	Capillary Pipe	12	1	2	2	Decrease
16	Drain Hose	12	1	2	8	Increase
17	Machine Capacitor	12	1	2	6	Decrease
18	Control PCB	12	1	2	9	Icrease

No	Spare Part Type	Month	Year	Sale Type	Total	Status
19	Caver Body	11	1	2	2	Decrease
20	Fan Belt	12	1	2	1	Decrease
.....
.....
1276	Inductor	10	6	2	19	Increase
1277	Dryer Timer	10	6	1	8	Increase

From these data, there are 5 criteria used in the research process, where the 3 criteria were carried out by data transformation, namely on the attributes of month, year, type of sales. The following is a description of the data that has been transformed:

Table 3.

Criteria attribute data transformation

No	Month	Numeric Format	No	Month	Numeric Format
1	January	1	7	July	7
2	February	2	8	August	8
3	March	3	9	September	9
4	April	4	10	October	10
5	May	5	11	November	11
6	June	6	12	Desember	12

Table 4.

Transformation on criteria

No	Year	Numeric Format
1	2021	1
2	2020	2
3	2019	3
4	2018	4
5	2017	5
6	2016	6

Table 5.

Transformation On Criteria

No	Sales Type	Numeric Format
1	Dealer	1
2	Company	2

The first process in this study is the determination of the attribute weights, by calculating the entropy value of the Training data, then calculating the information gain value and then calculating the Gain Ratio value for each attribute in the dataset. Then calculate the weight based on Gain Ratio max. The process for calculating attribute weights is as follows: No vertical lines in table.

$$\text{Entropy (Spare Part Type 1)} = (-16/31) \times \text{Log}_2 (16/31) + (-15/31) \times \text{Log}_2 (15/31) = 0.99924925$$

$$\text{Entropy (Spare Part Type 2)} = (-10/29) \times \text{Log}_2 (10/29) + (-19/29) \times \text{Log}_2 (19/29) = 0.92936363$$

$$\text{Entropy (Spare Part Type 3)} = (-34/61) \times \text{Log}_2 (34/61) + (-27/61) \times \text{Log}_2 (27/61) = 0.99047997$$

$$\text{Entropy (Spare Part Type 4)} = (-18/59) \times \text{Log}_2 (18/59) + (-41/59) \times \text{Log}_2 (41/59) = 0.88741793$$

$$\text{Entropy (Spare Part Type 5)} = (-28/55) \times \text{Log}_2 (28/55) + (-27/55) \times \text{Log}_2 (27/55) = 0.99976152$$

$$\text{Entropy (Spare Part Type 6)} = (-45/67) \times \text{Log}_2 (45/67) + (-22/67) \times \text{Log}_2 (22/67) = 0.91324016$$

$$\text{Entropy (Spare Part Type 7)} = (-23/48) \times \text{Log}_2 (23/48) + (-25/48) \times \text{Log}_2 (25/48) = 0.9987473$$

$$\begin{aligned}
\text{Entropy (Spare Part Type 8)} &= (-5/20) \times \text{Log}_2 (5/20) + (-15/20) \times \text{Log}_2 (15/20) = 0.81127812 \\
\text{Entropy (Spare Part Type 9)} &= (-27/58) \times \text{Log}_2 (27/58) + (-31/58) \times \text{Log}_2 (31/58) = 0.99656637 \\
\text{Entropy (Spare Part Type 10)} &= (-7/28) \times \text{Log}_2 (7/28) + (-21/28) \times \text{Log}_2 (21/28) = 0.81127812 \\
\text{Entropy (Spare Part Type 11)} &= (-13/25) \times \text{Log}_2 (13/25) + (-12/25) \times \text{Log}_2 (12/25) = 0.99884554 \\
\text{Entropy (Spare Part Type 12)} &= (-20/48) \times \text{Log}_2 (20/48) + (-28/48) \times \text{Log}_2 (28/48) = 0.97986876 \\
\text{Entropy (Spare Part Type 13)} &= (-10/17) \times \text{Log}_2 (10/17) + (-7/17) \times \text{Log}_2 (7/17) = 0.97741782 \\
\text{Entropy (Spare Part Type 14)} &= (-25/48) \times \text{Log}_2 (25/48) + (-23/48) \times \text{Log}_2 (23/48) = 0.9987473 \\
\text{Entropy (Spare Part Type 15)} &= (-10/24) \times \text{Log}_2 (10/24) + (-14/24) \times \text{Log}_2 (14/24) = 0.97986876 \\
\text{Entropy (Spare Part Type 16)} &= (-2/16) \times \text{Log}_2 (2/16) + (-14/16) \times \text{Log}_2 (14/16) = 0.54356444 \\
\text{Entropy (Spare Part Type 17)} &= (-16/34) \times \text{Log}_2 (16/34) + (-18/34) \times \text{Log}_2 (18/34) = 0.99750255 \\
\text{Entropy (Spare Part Type 18)} &= (-10/36) \times \text{Log}_2 (10/36) + (26/36) \times \text{Log}_2 (26/36) = 0.85240518 \\
\text{Entropy (Spare Part Type 19)} &= (-12/33) \times \text{Log}_2 (12/33) + (-21/33) \times \text{Log}_2 (21/33) = 0.9456603 \\
\text{Entropy (Spare Part Type 20)} &= (-15/42) \times \text{Log}_2 (15/42) + (-27/42) \times \text{Log}_2 (27/42) = 0.94028596 \\
\text{Entropy (Spare Part Type 21)} &= (-14/33) \times \text{Log}_2 (14/33) + (-19/33) \times \text{Log}_2 (19/33) = 0.98337619 \\
\text{Entropy (Spare Part Type 22)} &= (-3/19) \times \text{Log}_2 (3/19) + (-16/19) \times \text{Log}_2 (16/19) = 0.62924922 \\
\text{Entropy (Spare Part Type 23)} &= (-22/48) \times \text{Log}_2 (22/48) + (-26/48) \times \text{Log}_2 (26/48) = 0.99498483 \\
\text{Entropy (Spare Part Type 24)} &= (-4/19) \times \text{Log}_2 (4/19) + (-15/19) \times \text{Log}_2 (15/19) = 0.74248757 \\
\text{Entropy (Spare Part Type 25)} &= (-34/62) \times \text{Log}_2 (34/62) + (-28/62) \times \text{Log}_2 (28/62) = 0.99323382 \\
\text{Entropy (Spare Part Type 26)} &= (-42/68) \times \text{Log}_2 (42/68) + (-26/68) \times \text{Log}_2 (26/68) = 0.95968689 \\
\text{Entropy (Spare Part Type 27)} &= (-17/43) \times \text{Log}_2 (17/43) + (-26/43) \times \text{Log}_2 (26/43) = 0.96816473 \\
\text{Entropy (Spare Part Type 28)} &= (-16/37) \times \text{Log}_2 (16/37) + (-21/37) \times \text{Log}_2 (21/37) = 0.98678672 \\
\text{Entropy (Spare Part Type 29)} &= (-6/10) \times \text{Log}_2 (6/10) + (-4/10) \times \text{Log}_2 (4/10) = 0.97095059 \\
\text{Entropy (Spare Part Type 30)} &= (-4/18) \times \text{Log}_2 (4/18) + (-14/18) \times \text{Log}_2 (14/18) = 0.76420451 \\
\text{Entropy (Spare Part Type 31)} &= (-21/50) \times \text{Log}_2 (21/50) + (-29/50) \times \text{Log}_2 (29/50) = 0.9814539 \\
\text{Entropy (Spare Part Type 32)} &= (-1/10) \times \text{Log}_2 (1/10) + (-9/10) \times \text{Log}_2 (9/10) = 0.46899559 \\
\text{Entropy (Spare Part Type 33)} &= (-7/25) \times \text{Log}_2 (7/25) + (-18/25) \times \text{Log}_2 (18/25) = 0.85545081 \\
\text{Entropy (Spare Part Type 34)} &= (-30/56) \times \text{Log}_2 (30/56) + (-26/56) \times \text{Log}_2 (26/56) = 0.99631652
\end{aligned}$$

Next, calculate the Entropy value of the entire class on all existing electronic spare parts sales datasets using the Entropy formula on the Gain Ratio algorithm.

$$\begin{aligned}
\text{Entropy (Whole Class)} &= (-567/1277) \times \text{Log}_2 (567/1277) + (-710/1277) \times \text{Log}_2 (710/1277) \\
&= 0.990935453 \\
\text{Gain Info (Type of Spare Part)} &= \text{Entropy (Whole Class)} - \sum_{i=1}^n \frac{S_i}{S} \times \text{Entropy}(S_i) \\
&= 0.990935453 - \sum_{i=1}^n \frac{S_i}{S} \times \text{Entropy}(S_i) \\
&= 0.050215 \\
\text{Gain Ratio (Type of Spare Part)} &= (\text{Information Gain}) / (\text{Split Info}) \\
&= 0.050215 / 4.9351736 = 0.010175
\end{aligned}$$

Then calculate the entropy, information gain, and gain ratio for the attributes of month, year, type of sales, and total sales, resulting in the max gain ratio value as follows:

Table 6.

The results of the weight calculation for each attribute

Calculation	Spare Part Type	Month	Year	Sales Type	Total
Information Gain	0.05022	0.01829	0.05288	0.28608	0.70352
Gain Ratio	0.0101750	0.005121	0.021713	0.340798	0.732772
Bobot	0.1	0.1	0.1	0.1	1

As for the weighting, if the Gain Ratio is min, it is 0.1, if the Gain Ratio is max, then it is 1, after getting the max gain ratio value, then the process of multiplying the highest Gain Ratio weight with the initial Training dataset is carried out, resulting in a new dataset that will be classified using the k-nearest neighbor method. , the following is a new dataset that will be classified with K-Nearest Neighbor:

Table 7.

The new dataset is the result of the gain ratio calculation process

No	Spare Part Type	Month	Year	Sale Type	Total	Status
1	Body Caver	12	1	2	0.73	Decrease
2	Gear Box	12	1	2	0.73	Decrease
3	Fan Belt	12	1	2	3.66	Decrease
4	Blower	12	1	2	5.13	Increase
5	Boss Ears	12	1	2	0.73	Decrease
6	Dynamo Box	12	1	2	0.73	Decrease
7	Dryer Timer	12	1	2	0.73	Decrease
8	Engine Gearbox	12	1	2	1.47	Decrease
9	Filter	12	1	2	0.73	Decrease
10	Motor Timing	12	1	2	0.73	Decrease
11	Fan Boss	12	1	2	6.59	Increase
12	Transistor	12	1	1	8.79	Increase
13	Inductor	12	1	1	8.79	Increase
14	Led Lights	12	1	2	3.66	Decrease
15	Capillary Pipe	12	1	2	1.47	Decrease
16	Drain Hose	12	1	2	1.47	Increase
17	Machine Capacitor	12	1	2	4.40	Decrease
18	Control PCB	12	1	2	6.59	Icrease
19	Caver Body	11	1	2	1.47	Decrease
20	Fan Belt	12	1	2	2.20	Decrease
.....
.....
1276	Inductor	10	6	2	13.92	Increase
1277	Dryer Timer	10	6	1	5.86	Increase

Next, classify predictions using the K-Nearest Neighbor method, using the Eucludien Distance formula and class distances totaling 3, to calculate the distance between training data and test data, researchers use 20 data to be calculated, as follows:

$$d_1 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093$$

$$d_2 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093$$

$$d_3 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (3,66-9)^2} = 12.22765718$$

$$d_4 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (5.13-9)^2} = 11.66091334$$

$$d_5 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093$$

$$d_6 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093$$

$$d_7 = \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093$$

$$\begin{aligned}
d_8 &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (1.47-9)^2} = 13.3304501 \\
d_9 &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093 \\
d_{10} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (0.73-9)^2} = 13.7620093 \\
d_{11} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (6.59-9)^2} = 11.26091027 \\
d_{12} &= \sqrt{(12-1)^2 + (1-1)^2 + (1-2)^2 + (8.79-9)^2} = 11.04735715 \\
d_{13} &= \sqrt{(12-1)^2 + (1-1)^2 + (1-2)^2 + (8.79-9)^2} = 11.04735715 \\
d_{14} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (3.66-9)^2} = 12.22765718 \\
d_{15} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (1.47-9)^2} = 13.3304501 \\
d_{16} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (1.47-9)^2} = 13.3304501 \\
d_{17} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (4.40-9)^2} = 11.92308685 \\
d_{18} &= \sqrt{(12-1)^2 + (1-1)^2 + (2-2)^2 + (6.59-9)^2} = 11.26091027 \\
d_{19} &= \sqrt{(11-1)^2 + (1-1)^2 + (2-2)^2 + (1.47-9)^2} = 12.51802301 \\
d_{20} &= \sqrt{(11-1)^2 + (1-1)^2 + (2-2)^2 + (2.20-9)^2} = 12.09297317
\end{aligned}$$

Furthermore, the sorting of the resulting data is calculated, the distance that has been obtained is then sorted from the closest data to the most distant (*ascending*). Then determine the test data group based on the majority label from the k closest neighbors, because the value of k = 5, then a data is taken from the closest ascending distance from the smallest 3, namely on data and . By using the K-Nearest Neighbor category from the majority of the data, it can predict the number of sales in the next period, and can determine the stock of electronic spare parts in the future.

In this study, the writer also made a comparison between the predictions of the coming month with the previous month, which can be seen from the results of the graphic analysis that in December sales predictions were made, then the writer made a comparison of predictions in the previous month so that the classification results can be seen in the following figure. Statements that serve as captions for the entire table do not need footnote letters.

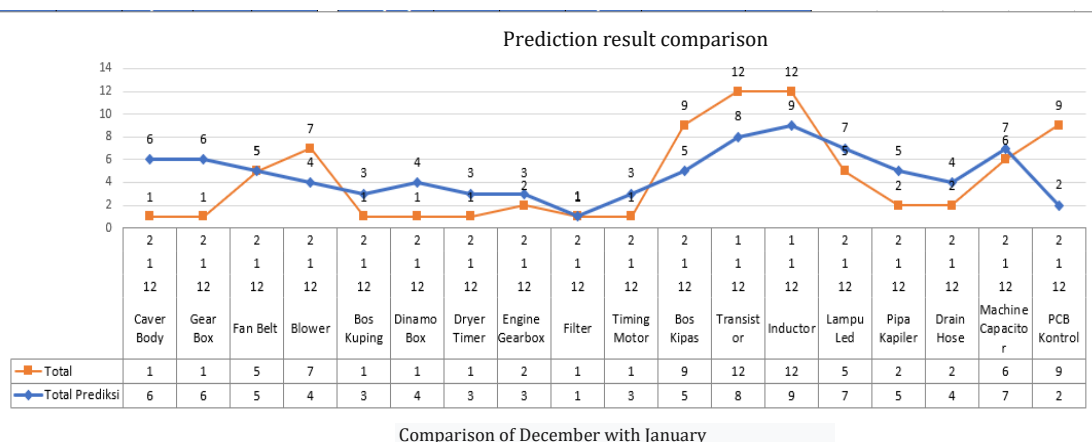


Fig 4. Prediction Comparison Chart

From the results of the graph above, it can be seen that the prediction results in December have a slight difference from the predictions made, using the K-Nearest Neighbor algorithm, the Gain Ratio

algorithm on the K-Nearest Neighbor is very influential in the prediction classification of electronic spare parts sales.

In December 2021, it can be seen from the graph results that the type of spare part cover body has a predicted distance of 5, while the Machine Capacitor has a prediction distance of 1, in the rules of the K-Nearest Neighbor algorithm the most calcification results will be the results of predictions made, as in the Machine Capacitor it is predicted experienced a decrease where with a value of $K = 5$ the results decreased by 2 and increased by 1, this will be the determination of the results of the predictions made on this type of spare part. The next is the display of the implementation of data mining applications that have been designed according to the results of calculations in this study, here are some views of the application system used, using the Web-based PHP programming language:



Fig 5. Home View System Implementation

In Figure 5 is a display of the implementation of the home system. Figure 6 shows the system display in the attribute weighting process using the gain ratio method

No	Jenis Sparepart	Bulan	Tahun	Jenis Penjualan	Total	Klasifikasi
1	6	12	1	2	1	Menurun
2	19	12	1	2	1	Menurun
3	15	12	1	2	5	Menurun
4	2	12	1	2	7	Meningkat
5	4	12	1	2	1	Menurun

Fig 6. Attribute Weighting System View

No	X1 - X	X2 - X	X3 - X	X4 - X	Total Perhitungan	Klasifikasi
6555	16	1	1	589.0329	24.638037665266	Menurun
6556	16	1	1	589.0329	24.638037665266	Menurun
6557	16	1	1	456.2496	21.777272556498	Menurun
6558	16	1	1	396.01	20.347235684485	Meningkat
6559	16	1	1	589.0329	24.638037665266	Menurun
6570	16	1	1	589.0329	24.638037665266	Menurun
6571	16	1	1	589.0329	24.638037665266	Menurun
6572	16	1	1	554.1316	23.91927256419	Menurun
6573	16	1	1	589.0329	24.638037665266	Menurun
6574	16	1	1	589.0329	24.638037665266	Menurun
6575	16	1	1	340.0336	18.921775815182	Meningkat
6576	16	1	0	264.0625	16.764918729239	Meningkat
6577	16	1	0	554.1316	18.764918729239	Meningkat

Fig 7. K-Nearest Neighbor Calculation Display

The following is a system display in the classification process, which will produce predetermined predictions based on the class desired by the user, in this study the selected class is $K = 5$, so that the prediction results obtained will be compared, if the results are majority more, then it will become the result of the prediction classification which is processed by the K-Nearest Neighbor method

No	Jenis Sparepart	Bulan	Tahun	Jenis Penjualan	Total	Klasifikasi
1	23	6	2	1	25	-----

No	Total Perhitungan	Klasifikasi
7055	2.796553770261	Meningkat
7063	2.9900938914366	Meningkat
7646	3.1468237955119	Meningkat

Akurasi	Nilai
Akurasi K=3	100 %
Akurasi K=5	80 %

Fig 8. Display Of Prediction Results Of Electronic Spare Parts Sales

The results of the predictions obtained, it turns out that most of all jakar class $K = 5$, get prediction results with an accuracy value of 80%, and produce predictions of 25 total spare parts sales for one type of spare part in the following month.

4. Conclusion

Based on the research that has been carried out in the discussion described previously, it can be concluded that, the prediction results carried out in research using the gain ratio method as attribute weighting get an accuracy value with a distance of 13%, where the level of accuracy in the classification is in the total attribute sales, while using the KNN method in classifying sales datasets to get a distance accuracy value of 0.008% where the results of the distance $K=3$ with 100% while at the KNN class distance $k=5$ get a value of 85%, from here it can be seen that the gain ratio is very influential in attribute class baling so that the classifier process becomes better, and is good without biasing the dataset. from the test results for each attribute, it can be concluded that the difference in the accuracy values increases with respect to the weighting of the gain ratio. KNN is a good classifier but when applied by this method to real or nominal data, all performance parameters vary according to the size of the dataset collected. So that the gain ratio method greatly influences the accuracy value of the KNN method. The gain ratio method can also be combined with other algorithmic methods such as Naive Bayes, SVM, Fuzzy, and K-Means so that the classification becomes more accurate with attribute weighting.

References

- [1] A. A. Nababan, O. S. Sitompul, And Tulus, "Attribute Weighting Based K-Nearest Neighbor Using Gain Ratio," J. Phys. Conf. Ser., Vol. 1007, No. 1, Pp. 1–6, 2018, Doi: 10.1088/1742-6596/1007/1/012007.
- [2] Aryati, Samsudin, And M.Fakhriza, "Sistem Seleksi Penerimaan Tenaga Kerja Outsourcing Menggunakan Algoritma C5.0 Berbasis Android (Studi Kasus : Pt. Sinergi Indo Prima Medan) .," J. Teknol. Dan Sist. Inf. Univrab, Vol. 7, No. 1, Pp. 52–63, 2022.
- [3] G. Rizka, U. Sinaga, And Samsudin, "Implementasi Framework Laravel Dalam Sistem Reservasi Pada Restoran Cindelar Kota Medan," J. Janitra Inform. Dan Sist. Inf., Vol. 1, No. 2, Pp. 73–84, 2021, Doi: 10.25008/Janitra.V1i2.131.
- [4] S. Siagian And M. Adilla, "Analisis Sistem Penjualan Pemasaran Produk Pada Pt. Panasonic Gobel Indonesia Cabang Medan," J. Fasikom, Vol. 11, No. 1, Pp. 26–31, 2021, Doi: 10.37859/Jf.V11i1.2429.

- [5] R. Febrilia, T. Wulandari, And D. Anubhakti, "Implementasi Algoritma Support Vector Machine (Svm) Dalam Memprediksi Harga Saham Pt . Garuda Indonesia Tbk," Indones. J. Inf. Syst., Vol. 4, No. 2, Pp. 250–256, 2021.
- [6] M. Nanja And Purwanto, "Metode K-Nearest Neighbor Berbasis Forward Selection Untuk Memprediksi Harga Komoditi Lada," J. Pseudocode, Vol. 2, No. 1, Pp. 53–64, 2017.
- [7] J. Sistem, I. Bisnis, And S. U. Kupang, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa Ppa Dan Bbm," J. Sist. Infromasi Bisni, Vol. 1, No. 2, Pp. 52–62, 2017.
- [8] "Klasifikasi Sentimen Masyarakat Terhadap Harga Tiket Pada Twittter Knn Dengan Aan," J. Instek, Vol. 4, No. 2, Pp. 236–245, 2019.
- [9] M. J. Islam, Q. M. Jonathan, And M. Ahmadi, "Investigating The Performace Of Naive Bayes Clasifiers And K-Nearest Neighbor Classifiers," J. Conver. Inf. Technol., Vol. 5, No. 2, Pp. 133–137, 2016.
- [10] M. D. Irawan, F. Nurhidayahti, R. Hsb, And A. Widarma, "Decision Support System Determining Computer Virus Protection Applications Using Simple Additive Weighting (Saw) Method," J. Comput. Networks , Archit. High Perform. Comput., Vol. 3, No. 1, Pp. 68–79, 2021.
- [11] S. Samsudin, Diktat Pengantar Ilmu Kompuer. Medan: Uinsu, 2018.
- [12] A. Almira, A. Ikhwan, And Suendri, "Implementasi Data Mining Menggunakan Algoritma Fp-Growth Pada Analisis Pola Pencurian Daya Listrik," J. Inform. Univ. Pamulang, Vol. 6, No. 2, Pp. 442–448, 2021.
- [13] A. Ikhwan And N. Aslami, "Implementasi Data Mining Untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means," Jurti (Jurnal Teknol. Informasi), Vol. 4, No. 2, 2020.
- [14] M. D. Irawan And Herviana, "Implementasi Logika Fuzzy Dalam Menentukan Jurusan Bagi Siswa Baru Sekolah Menengah Kejuruan (Smk) Negeri 1 Air Putih," (Jurnal Teknol. Inf., Vol. 2, No. 2, Pp. 129–137, 2018.
- [15] R. Hutami And E. Z. Astuti, "Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Furniture Pada Cv.Octo Agung Jepara," Univ. Dian Nuswantoro Semarang, Vol. 3, No. 2, Pp. 40–51, 2016.
- [16] N. H. Harani And F. S. Damayanti, "Implementasi Algoritma C5.0 Untuk Menentukan Pelanggan Potensial Di Kantor Pos Cimahi," J. Sitech, Vol. 4, No. 1, Pp. 70–76, 2021.
- [17] B. Aisyah Farahdiba And Y. S. Nugroho, "Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio," J. Tek. Elektro, Vol. 8, No. 2, Pp. 43–46, 2017.
- [18] I. M. Sudarma, "Implementasi Algoritma C5.0 Pada Penilaian Kinerja Pegawai Negeri Sipil," Maj. Ilm. Teknol. Elektro, Vol. 17, No. 3, Pp. 1–6, 2018.
- [19] Y. Adani And W. Nurjaya, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Keputusan Calon Nasabah Dan Nasabah Tetap Bank Bri Syariah Menerima Penawaran Program Deposito Berjangka," J. Tek., Vol. 5, No. 2, Pp. 13–24, 2ad.
- [20] T. Setiyorini And R. T. Asmono, "Implementation Of Gain Ratio And K-Nearest Neighbor For Classification Of Student Performance," J. Pilar Nusa Mandiri, Vol. 16, No. 1, Pp. 19–24, 2020, Doi: 10.33480/Pilar.V16i1.813.
- [21] F. G. Dewanto, J. J. M. R. Londok, R. A. V Tuturoong, And W. B. Kaunang, "Pengaruh Pemupukan Anorganik Dan Organik Terhadap Produksi Tanaman Jagung Sebagai Sumber Pakan,," Zootec, Vol. 32, No. 5, Pp. 1–8, 2017, Doi: 10.35792/Zot.32.5.2013.982.
- [22] Triase And Samsudin, "Implementasi Data Mining Dalam Mengklasifikasikan Ukt (Uang Kuliah Tunggal) Pada Uin Sumatera Utara Medan," J. Teknol. Inf., Vol. 4, No. 2, Pp. 370–376, 2020.
- [23] K. Imtihan, Ernawati, And L. Mutawalli, "Penerapan Research And Development (R&D) Dalam Membangun Alat Penyiraman Tanaman Otomatis Berbasis Arduino," J. Manaj. Inform. Sist. Inf., Vol. 5, 2022.
- [24] M. J. Christianto, D. Tjahjadi, And C. Juliane, "Designing The Architecture Of Population Administration Information Systems Using Methods Rational Unified Process (Rup)," J. Sos. Dan Teknol., Vol. 2, No. 2, Pp. 107–124, 2022.
- [25] Suendri, "Implementasi Diagram Uml (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: Uin Sumatera Utara Medan)," J. Ilmu Komput. Dan Inform., Vol. 3, No. 1, Pp. 1–9, 2018.