

Analysis of sales levels of pharmaceutical products by using data mining algorithm C45

Rini Sovia, Abulwafa Muhammad, Syafri Arlis, Guslendra, Sarjon Defit

Department of Computer Science, Universitas Putra Indonesia YPTK Padang, Indonesia

Article Info

Article history:

Received Mar 8, 2020

Revised Dec 5, 2020

Accepted Jan 11, 2021

Keywords:

Algorithm C45

Data mining

Knowledge

Pharmaceutical

ABSTRACT

This research was conducted to analyze the level of sales of pharmaceutical products at a Pharmacy. This is done to find out the types of products that have high and low sales levels. This study uses the C45 Data Mining Algorithm concept that will produce a conclusion on the prediction of sales of pharmaceutical products through data processing obtained from sales transactions at pharmacies. This C45 algorithm will form a decision tree that provides users with knowledge about products that are in great demand by consumers based on sales data and predetermined variables. The final result of the C45 algorithm produces a number of rules that can identify the inheritance of a type of medicinal product. C45 algorithm is able to produce 20 types of categories that will be labeled goals based on the number of pharmaceutical products, since it can be concluded that C45 successfully defines 55% of the existing objective categories.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Abulwafa Muhammad

Department of Computer Science

Universitas Putra Indonesia YPTK Padang, Indonesia

Jl. Raya Lubuk Begalung Padang, Sumatera Barat, Indonesia 25221

Email: abulwafa@upiypk.ac.id

1. INTRODUCTION

The sale of pharmaceutical products to Indonesians can be found in various places, one of which is a pharmacy. Pharmacy is one of the health service facilities in helping to achieve optimal health status for the community. Health services are all efforts made individually or collectively in an organization to maintain and improve their health, prevent and cure diseases, and restore the health of individuals, families, communities [1]. Good health services are supported by the availability of pharmaceutical products, which can meet consumer needs and understand consumer behavior [2]. As for things related to consumers, especially experiences with digital concepts are very necessary for now [3]. To find out the needs and behavior of consumers, we must analyze and process the data we have with various concepts such as data mining or big data [4].

Data mining is the process of mining data to generate new knowledge from very large data [5]. In data mining, there is a classification process, including classifying the promotion of private schools in order to have good branding [6]. Data mining is closely related to data, information and knowledge. A process in data mining that starts with a data extraction process which then produces information [2]. The information that will be generated will then be processed to produce a bias in the form of a pool (pattern). This pattern will be translated into knowledge that can produce decisions [3]. Data mining is part of the knowledge discovery in database (KDD) process, which in the KDD process consists of stages of data cleaning, data integration, data selection, data transformation, data mining, evaluation of patterns and presentation of knowledge, such as Figure 1. There are several algorithms commonly used in data processing and analysis,

including the K-nearest neighbor (KNN) with a global GINI diversity index for subsidized food classification in the city of Semarang, Indonesia in recent years many methods have been used for data classification [7]. For other classification processes, the c45 algorithm can be used [8]. To do predictions, you can also use this c45 algorithm [9].

The C45 algorithm which is one of the algorithms in machine learning can be used to diagnose various types of diseases in the medical world, supported by other algorithms in machine learning [10-12]. Besides the C45 algorithm can also be combined with the naïve Bayes algorithm to analyze a social, academic problem, [5, 13, 14]. Many researchers combine and compare the c45 algorithm with other algorithms such as the comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 methods in weather forecasting that provide decision support [15]. In addition, c45 datamining can also be used to measure the level of customer satisfaction in an institution or organization [16]. The C45 algorithm can also measure the level of service quality in banking companies [17]. In managing product availability, several inventory management techniques can be used, one of which is data mining with the c45 algorithm, k means and others [18]. The c45 method can be used for various prediction processes and is compared with the Support Vector Machine algorithm for identification of pests and diseases in plants [19]. The c45 algorithm can also be used to diagnose covid 19 surveillance classifications which include PDP, ODP, and OTG [20]. The results showed that the COVID-19 surveillance diagnosis using the C4.5 algorithm was successfully modeled into a decision tree with the classification of PDP, ODP, and OTG. The testing process in the form of confusion matrix with 3 (three) classes produces an accuracy rate of 92.86% which is included in the very good classification category.

The implementation of the C45 data mining algorithm can be carried out in all aspects of life including in soil or civil science. With a data mining approach that uses the C4.5 Algorithm decision tree, a classification model will be created where the model functions as a classification of the causes of landslides in Purwakarta district [21]. Twenty-eight goals and thirty-six measures were used, and nine departments were involved to monitor the performance of the goals so that the company achieved the goals set. The research conducted data mining with the C4.5 algorithm [22]. The resulting rules are 11 rules and the level of accuracy achieved is 79.41%. In this study the authors apply the C45 algorithm to analyze customer loyalty. The variables used are based on price, attitude in serving consumers [23]. Based on the analysis of the use of data mining with the C4.5 algorithm, it can be used in customer data sets into strategic management activities so that it can accommodate customers as long as possible properly, this C4.5 algorithm application must be included in the data set. The C45 method is used to predict furniture sales, through this process the product items that are most in demand by customers are found [24, 25].

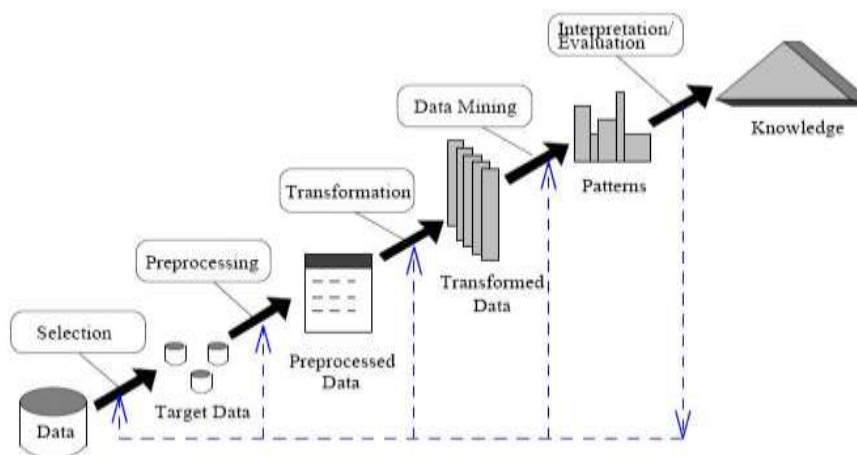


Figure 1. The process of knowledge discovery in database

2. RESEARCH METHOD

One of the decision tree induction algorithms is ID3 (iterative dichotomiser). ID3 was developed by J. Ross Quinlan. In the ID3 algorithm procedure, the input is in the form of training samples, training labels and attributes. C4.5 algorithm is the development of ID3. Algorithm C 4.5: Select attribute as root, create a branch for each value, for cases in the branch, repeat the process for each branch until all cases in the branch have the same class. To select the root attribute, based on the highest GAIN value of the attributes which exists. To get the GAIN value, you must first determine the ENTROPY value.

The research focuses on the process of analyzing the data of drug products entering pharmacies through suppliers with the C4.5 algorithm and using the Weka (tools data mining) program to obtain classification results. There are 6 attributes used in the study, namely: (1) IDs from 1 to 20; (2) names consist of supplier names, namely : A001, A002, A003, A004, A005, A006, A007, A008, A009, A010, A011, A012, A013, A014, A015, A016, A017, A018, A019, dan A020; (3) the number of drugs consisting of 1296, 896, 528, 8000, 180, 200, 132, 132, 144, 400, 340, 456, 200, 3000, 222, 6360, 362, 300, 144, 144; (4) types of drugs consists of generic drugs and patents; (5) Delivery consists of fast and late; (6) Prices consist of 1500, 2008, 3500, 210, 5400, 12400, 10500, 10500, 5700, 7300, 3600, 3200, 3000, 600, 3200, 950, 5200, 3570, 5500, 4700; The C4.5 algorithm starts from the process of selecting the attribute with the highest gain as the root of the tree, then makes a branch for each value, then divides the cases into branches, after that repeats the process for each branch until all cases in the branch have the same class. Flowchart can clearly illustrate the stages and steps in the classification using the C4.5 algorithm. It can be seen in the form of a flowchart in Figure 2. C4.5 Algorithm Classification Technique begins with data processing and transformation so that the raw data used for analysis is data with complete attributes and can produce decision trees. Supplier data that has been obtained is processed to process the C45 data mining process by searching for entropy, after the entropy value is obtained then look for the gain value. The entropy value and the existing gain then look for the highest gain value, because the highest gain value will determine the root node in the decision tree, then get a new node in the decision tree, such us Figure 2.

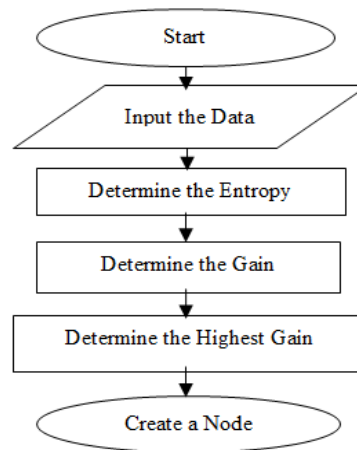


Figure 2. Flowchart proses algoritma C45

3. RESULTS AND ANALYSIS

3.1. Data analysis

Data analysis is the stage for analyzing the data needed for the design of the system to be made, in this case the authors take data through the literature relating to the research theme, to find information compile the theories related to the discussion so that there is a fusion complex between one and the other, such us Table 1.

The classification process for each Sales table field, such us Table 2.

1. Classification of Medicine Amounts
 - a. ≥ 1200 : if the number of medicine is more than or equal to 1200, so the number of medicine is many
 - b. < 1200 : if the amount of medicine is less than 1200 then the number of medicine is few
2. Price Classification
 - a. ≥ 5000 : If the price of the medicine is more than or equal to 5000 per pcs then the drug is expensive
 - b. < 5000 : If the price of medicine is less than 5000 per pcs, the medicine is cheap
3. Classification of types of medicines
 - a. Generic: is a medicine whose patent has expired so that it can be produced by all pharmaceutical companies without the need to pay royalties
 - b. Paten: a new medicine that is produced and marketed by a pharmaceutical company that has a patent

4. Delivery

- a. Fast : Is a supplier that is delivered on time or sooner than the agreed time
- b. Late: Delivery is made by the supplier later than the agreed time.

Table 1. Medicine supplier search

ID	Name	Number of Medicine	Types of Medicine	Delivery	Price	Decision
1	A001	1296	generic	Fast	1500	In Demand
2	A002	1896	generic	Fast	2008	In Demand
3	A003	528	patent	Late	3500	Not In Demand
4	A004	8000	generic	Fast	210	In Demand
5	A005	180	patent	Fast	5400	In Demand
6	A006	200	generic	Fast	12400	Not In Demand
7	A007	132	patent	Late	10500	Not In Demand
8	A008	132	generic	Fast	10500	Not In Demand
9	A009	144	patent	Fast	5700	In Demand
10	A010	400	generic	Fast	7300	Not In Demand
11	A011	340	patent	Fast	3600	In Demand
12	A012	456	generic	Late	3200	Not In Demand
13	A013	200	generic	Late	3000	Not In Demand
14	A014	3000	generic	Late	600	In Demand
15	A015	222	generic	Fast	3200	In Demand
16	A016	6360	patent	Late	950	In Demand
17	A017	362	generic	Fast	5200	Not In Demand
18	A018	300	generic	Fast	3570	In Demand
19	A019	144	patent	Late	5500	Not In Demand
20	A020	144	generic	Fast	4700	In Demand

Source : Arafah Pharmacy, Padang Panjang

Table 2. Data supplier pharmacies (Medicine) January 2017-January 2019

ID	Name	Number of Medicine	Types of Medicine	Delivery	Price	Decision
1	A001	Many	Generic	Fast	Cheap	In Demand
2	A002	Few	Generic	Fast	Cheap	In Demand
3	A003	Few	Patent	Late	Cheap	Not In Demand
4	A004	Many	Generic	Fast	Cheap	In Demand
5	A005	Few	Patent	Fast	Expensive	In Demand
6	A006	Few	Generic	Fast	Expensive	Not In Demand
7	A007	Few	Patent	Late	Expensive	Not In Demand
8	A008	Few	Generic	Fast	Expensive	Not In Demand
9	A009	Few	Patent	Fast	Expensive	In Demand
10	A010	Few	Generic	Fast	Expensive	Not In Demand
11	A011	Few	Patent	Fast	Cheap	In Demand
12	A012	Few	Generic	Late	Cheap	Not In Demand
13	A013	Few	Generic	Late	Cheap	Not In Demand
14	A014	Many	Generic	Late	Cheap	In Demand
15	A015	Few	Generic	Fast	Cheap	In Demand
16	A016	Many	Patent	Late	Cheap	In Demand
17	A017	Few	Generic	Fast	Expensive	Not In Demand
18	A018	Few	Generic	Fast	Cheap	In Demand
19	A019	Few	Patent	Late	Expensive	Not In Demand
20	A020	Few	Generic	Fast	Cheap	In Demand

Source : Arafah Pharmacy, Padang Panjang

The entropy search processing formula is performed as follows:

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

Note :

S : case set

A : attribute

n : Number of partitions S

pi : proporsitionfrom Si to S

Gain search processing formula performed as follows:

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

Entropy(S_i)

Note: S : case set

A : attribute

n : number of partitions A

|S_i| : the number of cases in the partition to-i

|S| : number of case in S

3.2. Process of analyze

Decision tree has two types of attribute data consisting of several input attributes of the target attribute and of course supporting the existing problem, its function as a comparison in calculating gain and ratio. In the training data set the sample size and must at least one target attribute whose value is a temporary conclusion of the problem of each instance (record), in this study the value of the target attribute is: bestsellers and less in demand. This data mining design writer uses the C4.5 algorithm. The process in the decision tree is to change the form of data (tables) into a tree model, change the tree model to a rule, and simplify the rule.

3.2.1. Node calculation 1

The process of finding total entropy and gain is done by grouping the data correctly, then calculating the data and using the entropy and gain search formula for each data attribute, such us Table 3. From the calculation results in the Table 3, it can be seen that the largest gain value is the attribute "Number of Medicine" of 0.2646. So that the attribute "Number of Medicine" becomes the root node. On the attribute "Number of Medicine" there are 2 attribute values, namely few and many, then further calculations need to be done. From this process, a temporary tree can be produced, such as Figure 3.

Table 3. The calculation of the highest gain 1

NODE			Number of Case (S)	Less Loyal (S1)	Loyal (S2)	Entropy	Gain
1	Total		20	9	11	0.9710	
	Number of Medicine	Many	5	0	5	0	0.2646
		Few	15	9	6	0.9710	
	Type of Medicine	Generic	13	6	7	0.9957	0.0008
		Patent	7	3	4	0.9852	
	Delivery	Fast	13	4	9	0.8905	0.1119
		Late	7	5	2	0.8631	
	Price	Expensive	8	6	2	0.8113	0.1815
		Cheap	12	3	9	0.8113	

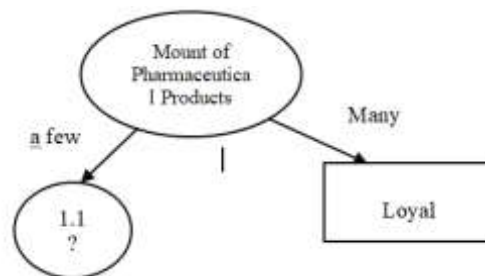


Figure 3. Temporary decision tree node 1

3.2.2. Node calculation 1.1

The process of finding total entropy and gain is done by grouping the data correctly, then calculating the data and using the entropy and gain search formula for each data attribute, such us Table 4. From the results of calculations in Table 4, it can be seen that the largest gain value is the attribute "Shipping" of 0.0637. So the attribute "Delivery" becomes the root of node. At the "Delivery" attribute there are 2 attribute values, namely fast and Late, then further calculations need to be done. From this process, a temporary tree can be produced, such as Figure 4.

Table 4. Temporary decision tree node 1.1

NODE			Number of Case (S)	Less Loyal (S1)	Loyal (S2)	Entropy	Gain
1.1	Small Number of Medicine		15	9	6	0.9710	
	Type of Medicine	Generic	9	7	2	0.7642	0.145
		Patent	6	2	4	0.9183	2
	Delivery	Fast	10	5	5	1	0.063
		Late	5	4	1	0.7219	7
	Price	Expensive	9	6	3	0.9183	0.020
		Cheap	6	3	3	1	0

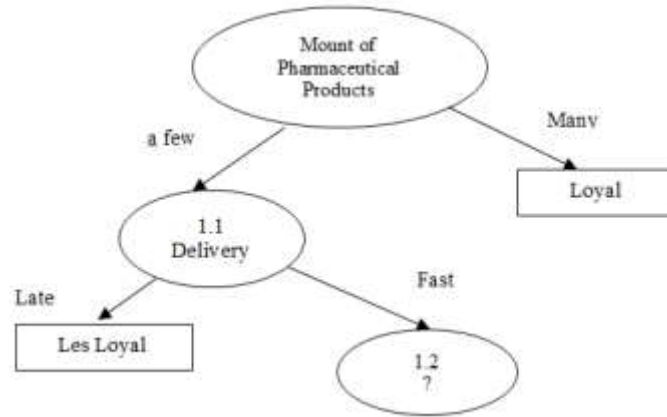


Figure 4. Temporary decision tree node 1.1

3.2.3. Node calculation 1.2

The process of finding total entropy and gain is done by grouping the data correctly, then calculating the data and using the entropy and gain search formula for each data attribute, such as Table 5. From the results of calculations in Table 5, it can be seen that the largest gain value is the attribute "Price" of 0.4200. So the attribute "Price" becomes the root node. At the "Price" attribute there are 2 attribute values, namely Expensive and Cheap, and then further calculations need to be done. From this process, a temporary tree can be produced, such as Figure 5.

Table 5. Calculation of the highest gain node 1.2

NODE			Number of Case (S)	Less Loyal (S1)	Loyal (S2)	Entropy	Gain
1.2	Fast Delivery		10	4	6	0.9710	
	Type of Medicine	Generic	7	4	3	0.9852	0.2814
		Patent	3	0	3	0	
	Price	Expensive	6	4	2	0.9183	0.4200
		Cheap	4	0	4	0	

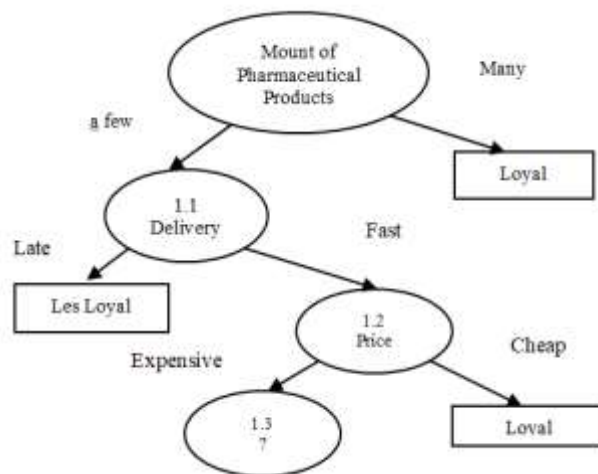


Figure 5. Temporary decision tree node 1.2

3.2.4. Node calculation 1.3

The process of finding total entropy and gain is done by grouping the data correctly, then calculating the data and using the entropy and gain search formula for each data attribute, such as Table 6. The results in the decision tree in Table 6 can be concluded that it can be a decision tree, such as Figure 6.

Table 6. Highest gain calculation node 1.3

NODE		Jumlah Kasus (S)	Kurang Loyal (S1)	Loyal (S2)	Entropy	Gain
1.3	Cheap Price		6	4	2	
	Type of Medicine	Generic	4	4	0	
		Patent	2	0	2	

For more details, the results in the decision tree can produce rules like the following:

Rule 1

IF Number of Medicine = Many
THEN Decision = Loyal

Rule 2 :

IF Number of Medicine = Few
THEN Decision = Less Loyal

Rule 3 :

IF Number of Medicine = Few
Delivery = Late
THEN Decision = Kurang Loyal

Rule 4 :

IF Number of Medicine = Few
Delivery = Fast
THEN Decision = Loyal

Rule 5 :

IF Number of Medicine = Few
Delivery = Fast
Price = Cheap
THEN Decision = Loyal

Rule 6 :

IF Number of Medicine = Few
Delivey = Fast
Price = Expensive
THEN Decision = Less Loyal

Rule 7 :

IF Number of Medicine = Few
Delivery = Fast Price = Expensive
Type of Medicine = Patent
THEN Decision = Loyal

Rule 8 :

IF Number of Medicine = Few
Delivey = Fast Price = Expensive
Type of Medicine = Generic
THEN Decision = LessLoyal

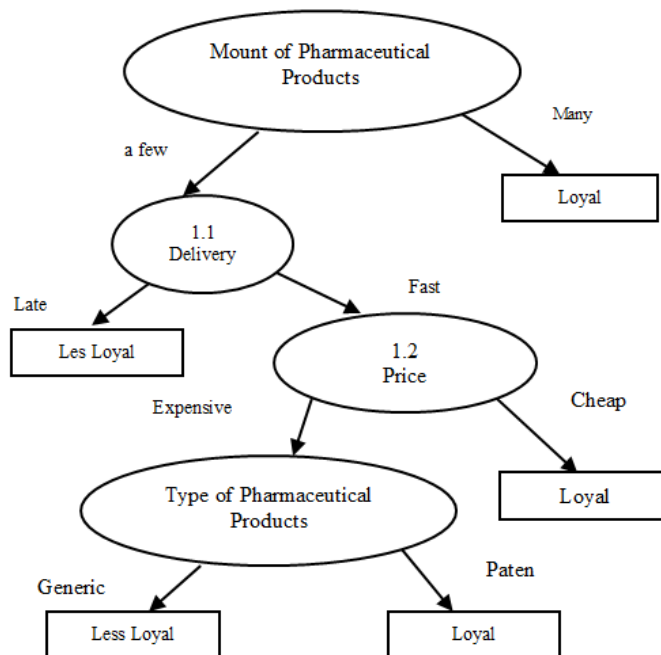


Figure 6. Temporary decision treenode 1.3

3.3. Result of analyze

The results of the analysis showed that the C4.5 algorithm succeeded in 21 types of categories which became the destination label based on the number of medicine, so it can be said that the C4.5 algorithm succeeded in defining 55% of the existing destination. Supplier A001 Loyal and Classified a leaf node(label), A002 Loyal (Classified a leaf node(label), A003 Less Loyal Cases that occur for the name of PT "A003" are very few, only occurred 9 times during January 2017 to January 2019, with a data presentation of 4.5%, A004 Clasification is Loyal and Classified a leaf node(label), A005 Clasification is Loyal and Classified

a leaf node(label), A006 Less Loyal Cases that occur for the name of PT "A006" are very few, only occurred 8 times during January 2017 to January 2019, with data presentation of 4%, A007 Les Loyal Cases that occur for the name of PT "A007" are very few, only occurred 7 times during January 2017 to January 2019, with data presentation of 3.5%, A008 Les Loyal Cases that occur for the name of PT "A008" are very few, only occurred 6 times during January 2017 to January 2019, with data presentation of 3%, A009 Loyal Classified as a leaf node (label), A010 Less Loyal Cases that occur for the name of PT "A010" are very few, only occurred 5 times during January 2017 to January 2019, with data presentation of 2.5%, A011 Loyal Classified as a leaf node (label), A012 Less Loyal Cases that occur for the name of PT "A012" are very few, only occurring 4 times during January 2017 to January 2019, with data presentation of 2%, A013 Less Loyal Cases that occur for the name of PT "A013" are very few, only occurred 3 times during January 2017 to January 2019, with data presentation of 1.5%, A014 Loyal Classified as a leaf node (label), A015 Loyal Classified as a leaf node (label), A016 Loyal Classified as a leaf node (label), A017 Less Loyal Cases that occur for the name of PT "A017" are very few, only occurred 2 times during January 2017 to January 2019, with a data presentation of 1%, A018 Loyal Classified as a leaf node (label), A19 Less Loyal Cases that occur for the name of PT "A019" are very few, only occurred once during January 2017 to January 2019, with data presentation of 0.5%, A20 Loyal Classified as a leaf node (label).

From the report we have 11 Classification of Supplier Loyal and 9 Supplier Less Loyal, Loyal Supplier is A001, A002, A004, A005, A009, A011, A014, A015, A016, A018 and A020 and A003, A006, A007, A008, A010, A012, A013, A017, A019 is Less Loyal Supplier.

4. CONCLUSION

C45 algorithm helps the pharmacy to determine which medicine are in demand and those that are not in demand based on the number of medicine, types of medicine, delivery and price. The final result of the C45 algorithm produces a number of rules that can identify the inheritance of a type of medicinal product. This research is able to produce a classification of pharmaceutical product sales. The data that is processed is the data of 20 pharmacy suppliers which are used as input, so that the classification of 11 suppliers is categorized as loyal, and there are 9 suppliers that are categorized as less loyal suppliers. This of course still has some drawbacks, it is expected that this data mining program can run optimally, of course it needs further development. It is expected that further research will develop this application by adding a number of other calculation variables so that the results of decisions can be more accurate and determination of supplier loyalty with other data mining methods by adding new features in accordance with changing pharmacy needs.

ACKNOWLEDGMENT

This work is supported by Universitas Putra Indonesia YPTK Padang, West Sumatera, Indonesia.

REFERENCES

- [1] H. H. Chang, K. H. Wong, dan P. W. Fang, "The effects of customer relationship management relational information processes on customer-based performance," *Decis. Support Syst.*, vol. 66, pp. 146-159, 2014, doi: 10.1016/j.dss.2014.06.010.
- [2] G. Alfian, M. F. Ijaz, M. Syafrudin, M. A. Syaekhoni, N. L. Fitriyani, dan J. Rhee, "Customer behavior analysis using real-time data processing: A case study of digital signage-based online stores," *Asia Pacific J. Mark. Logist.*, vol. 31, no. 1, pp. 265-290, 2019, doi: 10.1108/APJML-03-2018-0088.
- [3] R. N. Bolton *et al.*, "Customer experience challenges: bringing together digital, physical and social realms," *J. Serv. Manag.*, vol. 29, no. 5, pp. 776-808, 2018, doi: 10.1108/JOSM-04-2018-0113.
- [4] S. J. Miah, H. Q. Vu, J. Gammack, dan M. McGrath, "A Big Data Analytics Method for Tourist Behaviour Analysis," *Inf. Manag.*, vol. 54, no. 6, pp. 771-785, 2017, doi: 10.1016/j.im.2016.11.011.
- [5] S. Dwiasnati dan Y. Devianto, "Utilization of Prediction Data for Prospective Decision Customers Insurance Using the Classification Method of C.45 and Naive Bayes Algorithms," *J. Phys. Conf. Ser.*, vol. 1179, no. 1, 2019, doi: 10.1088/1742-6596/1179/1/012023.
- [6] G. Garaika, S. Sumatera, E. Ogan, K. Ulu, dan S. Sumatera, "The Classification Of Promotion Media Of A College To Improve The Management Of Promotion For A Private," vol. 11, 2020.
- [7] D. Ispriyanti, A. Prahutama, dan Mustafid, "K-nearest neighbor (KNN) with global GINI diversity index for classification subsidy food in Semarang city, Indonesia," *J. Phys. Conf. Ser.*, vol. 1524, no. 1, 2020, doi: 10.1088/1742-6596/1524/1/012034.
- [8] T. Sousa, A. Silva, dan A. Neves, "Particle Swarm based Data Mining Algorithms for classification tasks," *Parallel*

- Comput.*, vol. 30, no. 5-6, pp. 767-783, 2004, doi: 10.1016/j.parc.2003.12.015.
- [9] C. Wirawan, E. Khudzaeva, T. H. Hasibuan, Karjono, dan Y. H. K. Lubis, "Application of Data mining to Prediction of Timeliness Graduation of Students (A Case Study)," *2019 7th Int. Conf. Cyber IT Serv. Manag. CITSM 2019*, pp. 18-21, 2019, doi: 10.1109/CITSM47753.2019.8965425.
- [10] K. K. Joshi, K. K. Gupta, dan J. Agrawal, "A review on application of machine learning in medical diagnosis," *2nd Int. Conf. Data, Eng. Appl. IDEA 2020*, 2020, doi: 10.1109/IDEA49133.2020.9170733.
- [11] D. Ayu *et al.*, "an Educational Data Mining for Student Academic Prediction Using K-Means Clustering and Naïve Bayes," pp. 155-160, 2020.
- [12] A. Byna, "Comparative Analysis of Machine Learning Algorithms for classification about Stunting Genesis," no. August, 2020, doi: 10.4108/eai.23-11-2019.2298349.
- [13] Y. Alkhalifi, A. Zumarniansyah, R. Ardianto, N. Hardi, dan A. E. Augustia, "Comparison of Naive Bayes Algorithm and C.45 Algorithm in Classification of Poor Communities Receiving Non Cash Food Assistance in Wanasari Village Karawang Regency," *J. Techno Nusa Mandiri*, vol. 17, no. 1, pp. 37-42, 2020, doi: 10.33480/techno.v17i1.1191.
- [14] A. Wibowo, D. Manongga, dan H. D. Purnomo, "The Utilization of Naive Bayes and C.45 in Predicting The Timeliness of Students' Graduation," *Sci. J. Informatics*, vol. 7, no. 1, pp. 99-112, 2020, doi: 10.15294/sji.v7i1.24241.
- [15] Y. Findawati, I. R. I. Astutik, A. S. Fitriani, I. Indrawati, dan N. Yuniasih, "Comparative analysis of Naïve Bayes, K Nearest Neighbor and C.45 method in weather forecast," *J. Phys. Conf. Ser.*, vol. 1402, no. 6, 2019, doi: 10.1088/1742-6596/1402/6/066046.
- [16] B. G. Sudarsono, A. B. Abd Rahman, S. P. Lestari, dan A. U. Bani, "Systematic review of the builder method and product of the information system strategic planning at higher education institutions in indonesia," *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. 6, pp. 939-947, 2020, doi: 10.5373/JARDCS/V12I6/S20201113.
- [17] M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, dan A. Wanto, "Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012002.
- [18] H. Li, Y. Xi, H. Lu, dan X. Fu, "un co rre ct ed oo f v er si un co rre ct ed pr oo f v," vol. 1, pp. 1-13, 2019, doi: 10.3233/JCM-193794.
- [19] M. Irfan, N. Lukman, A. A. Alfauzi, dan J. Jumadi, "Comparison of algorithm Support Vector Machine and C4.5 for identification of pests and diseases in chili plants," *J. Phys. Conf. Ser.*, vol. 1402, no. 6, 2019, doi: 10.1088/1742-6596/1402/6/066104.
- [20] W. Wiguna dan D. Riana, "Diagnosis of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 71-80, 2020, doi: 10.33480/pilar.v16i1.1293.
- [21] Y. Handrianto dan M. Farhan, "C.45 Algorithm for Classification of Causes of Landslides," *Sinkron*, vol. 4, no. 1, pp. 120, 2019, doi: 10.33395/sinkron.v4i1.10154.
- [22] S. A. Safdar, T. Yue, S. Ali, dan H. Lu, *Using multi-objective search and machine learning to infer rules constraining product configurations*, vol. 27, no. 1-2. Springer US, 2020.
- [23] K. Anagnostopoulos, "Strategic Plan in a Greek Manufacturing Company: A Balanced Scorecard and Strategy Map Implementation," *Int. J. Bus. Manag.*, vol. 5, no. 2, pp. 12-25, 2010, doi: 10.5539/ijbm.v5n2p12.
- [24] I. N. Budi, I. Ranggadara, I. Prihandi, N. R. Kurnianda, dan Suhendra, "Prediction using C4.5 method and RFM method for selling furniture," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 535-541, 2019, doi: 10.35940/ijeat.A9665.109119.
- [25] B. Mrzygłód, M. Hawryluk, M. Janik, dan I. Olejarczyk-Woźńska, "Sensitivity analysis of the artificial neural networks in a system for durability prediction of forging tools for forgings made of C45 steel," *Int. J. Adv. Manuf. Technol.*, vol. 109, no. 5-6, pp. 1385-1395, 2020, doi: 10.1007/s00170-020-05641-y.