EduMa
MATHEMATICS EDUCATION LEARNING AND TEACHING
article link: https://syekhnurjati.ac.id/jurnal/index.php/eduma/eduma/article/view/4011

Data Mining Implementation with Clustering Techniques for Drug Inventory Information in Antonius Hospital Pontianak

Antonius Anton*

 Computer Science Department, BINUS Graduate Program - Master of Computer Science Bina Nusantara University
*Corresponding author: mailing addresses Jakarta, 11530 Indonesia, e-mail addresses: sempitak@gmail.com

a r t i c l e   i n f o

a b s t r a c t

DATA MINING IMPLEMENTATION WITH CLUSTERING TECHNIQUES FOR DRUG INVENTORY INFORMATION IN ANTONIUS HOSPITAL PONTIANAK. Competition in the business world, especially in the pharmacy industry, requires developers to find a strategy that can increase sales of special drug sales by maximizing service to consumers. One way is to keep the availability of various types of drugs in the pharmacy warehouse. To find out what medicines are purchased by consumers, it can be done using basket analysis techniques, namely analysis of consumer buying habits. Detection of drugs that are often bought together is called the association rule. Medication is an important factor whose availability must be controlled properly in a hospital. The availability of drugs at the hospital will support hospital performance. One of the data mining methods that can be used from the above analysis is the clustering method, which is grouping data items into small groups so that each group has an essential equation. The evaluation results from this study are grouping drug data with clustering techniques to facilitate the process of grouping drugs based on drugs that are often used in a certain period of time.

K e y w o r d s :

INTRODUCTION

The existence of day-to-day operational activities, the longer the data will increase, the more. If left alone, then the transaction data only becomes meaningless garbage. With the support of technological developments, the ability to collect and process data has also developed (Buulolo, 2013).

Competition in the business world, especially in the pharmacy industry, requires developers to find a strategy that can increase sales of special drug sales by maximizing service to consumers. One way is to keep the availability of various types of drugs in pharmacies warehouse. To find out what medicines are purchased by consumers, it can be done using basket basket analysis techniques, namely analysis of consumer buying habits. Detection of drugs that are often bought together is called the association rule.

Drugs are all single ingredients or mixtures that are used by all creatures for the inside and outside of the body to prevent, alleviate and cure diseases. According to the law, what is meant by medicine is a substance or mixture of materials to be used to determine the diagnosis, prevent, reduce, eliminate, cure diseases or symptoms of illness, injury or physical or spiritual abnormalities in humans or animals to turn the body or parts of the human body (Syamsuni, 2006).

Planning for drug needs is one of the important and decisive aspects in drug management, because planning of drug needs will affect the procurement, distribution and use of drugs in the health service unit, where planning the right drug requirements will make procurement effective and efficient so that and sufficient amount according to health service needs with guaranteed quality and can be obtained at the time needed (Nugraha & Kusumawati, 2014).

RELATED WORKS

The related journals used in the literature review of this study are as follows: Data Mining Implementation Journal for Diagnosis of Hypertension in Pregnancy by Using a Decision Tree. In the Journal written by Ari Muzakir and Rika Anisa Wulandari in 2016, it was explained that the use of a decision tree method for diagnosing hypertension in pregnant women. With this decision tree method, the diagnosis of hypertension in pregnant women can be done more quickly (Muzakir & Wulandari, 2016).

Journal of Medical Record Analysis to Determine Patterns of Disease Groups Using Classification with Decision Tree J48 written by Edy Kurniawan and friends explained the application of the J48 decision tree method for analyzing medical records to determine the pattern of diseases suffered by patients. With the use of the J48 decision tree method, it will greatly help the work of medical personnel to be able to diagnose the disease of the patients being examined (Kurniawan, Purnama & Sumpeno, 2011).

Journal of Implementation of Priori Algorithms in Drug Supply Systems (Case Study: Pharmacy of the Estomihi Hospital in Medan). This journal was written by Efori Buulolo in 2013. In this journal described the implementation of a priori algorithms on the drug inventory system. With the use of a priori algorithm, the dispensary manager can control the available drug supply. In addition, with the implementation of a priori algorithms at this pharmacy, it can help the dispensary manager to determine what drug supplies are widely used by patients(Buulolo, 2013).

Medicine is one of the important components in terms of health both to prevent, reduce, eliminate or cure a disease or symptom of the disease. For this reason, drugs need to be managed well, effectively and efficiently. Planning for drug needs is an important thing to do to ensure availability and equal distribution of drugs with sufficient types and quantities so that the drugs can be obtained quickly at the right place and time in the agency that is related to health services, be it a hospital, health center, health service and so forth. Planning for drug needs will affect the procurement, distribution and use of drugs in health services Taslim & Fajrizal, 2016).

Journal of Data Mining with the Clustering Method for Drug Inventory Information Processing at Pandanaran Health Center Semarang. This journal written by Joanna Ardhyanti Mita Nugraha in 2013 explained that the use of data mining using the clustering method in processing drug information (Nugraha & Kusumawati, 2014). The application of data mining in information processing is very helpful for the manager. From the results of data analysis, the results are in the form of drugs that are often used for prescription or treatment from each month from the sample for the last 3 years, namely 2011, 2012 and 2013 which can be used as a reference for drug supplies in the following year. In addition there are also drugs that are often used as a drug for prescription or treatment each year, this can be used for recommendations in Pandanaran Health Center in the supply of medicines for the following year or month.

METHODS

The term data mining has several views, such as knowledge discover or pattern recognition. The two terms actually have their respective accuracy, the term knowledge discovery or knowledge discovery is precisely because the main purpose of data mining is used to obtain knowledge that is still hidden in the chunks of data (Bastian, 2018).

Data Mining is the process of extracting information from a data set through the use of algorithms and techniques involving the fields of statistics, machine learning, and database management systems . Data Mining is used for extracting important hidden information from large datasets. With the data mining, it will get a gem in the form of knowledge in a large number of data (Yanto & Khoiriah, 2015)

The development of information is currently growing very rapidly. Likewise data on each organization will increase every day. The use of computers in processing daily data of organizations is very much needed and even computers are as a primary requirement for an organization to be used in processing data. Data that grows a lot every day will certainly make the work of employees increasingly.

In organizations engaged in the health / pharmaceutical sector, drug data is important data. Medicines that enter and leave every day must be recorded properly. The purpose of this data collection of drugs in and out will later be useful for organizations to find out which drugs are used more daily. With the knowledge of drugs that are used daily, the organization can determine which drugs need to be stocked up a lot.

In this study the author tries to make data mining drugs to help group the data needed so that when data is needed, the data can be retrieved quickly so that it can help the work process of an organization. A quick work process will have an impact on the quality of service for an organization,

in this case an organization engaged in the health / pharmaceutical sector.

In the process of collecting data, the thing that will be done by the author is to make direct observations to the place of research. The data source used in this study is drug data from the Pharmacy Hospital of Santo Antonius Pontianak.

After the data collection stage is complete, the next process is to implement the results of the research. In this implementation process, what is implemented is a data mining drug inventory.

The tool that will be used in the data mining process in this research is a tool called RapidMiner. RapidMiner is software that is open (open source). RapidMiner is a solution for analyzing data mining, text mining and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques to provide insight to users so that they can make good decisions (Aprilla, Baskoro, Ambarwati & Wicaksana, 2013).

After the implementation phase is complete, the next process is to analyze the data that has been implemented. The purpose of this analysis is to find out whether the implementation process was successful or failed.

After the data analyst stage, the next step is the evaluation stage. From this evaluation phase the conclusion will be concluded that whether the technique used is successful or not.

Cluster analysis plays an important role in classifying objects. Depending on the application, objects can be signals, customers, patients, news, plants and others. Cluster technique is a nonparametric technique that is very much applied in real cases. Cluster techniques can be grouped into two large classes, namely partitioning cluster and hierarchichal cluster. There are two types of cluster techniques that are quite often used, the first is the k-means cluster (representing partitioning clusters) and the next is hierarchical clustering.

K-Means Algorithm is an iterative grouping algorithm that partition data sets into a number of K clusters that have been set at the beginning. The K-Means algorithm is simple to implement and run, relatively fast, easy to adapt, commonly used in practice. Historically, K-Means became one of the most important algorithms in the field of data Mining (Sulastri & Gufroni, 2017).

Of the most simple and commonly known clustering techniques is the k-means clustering. In the k-means clustering the value of k must be determined first. Usually the user or user already has initial information about the object being studied; including how many clusters are most appropriate. In detail the user can use a measure of dissimilarity. This dissimilarity can be translated into the concept of distance. If the distance of two objects is close enough, then the two objects are similar. Getting closer means the higher the resemblance. The higher the distance value, the higher the dissimilarity. The k-means algorithm can be summarized as follows:

a. Select the number of cluster K
b. The initialization of the cluster center can be done in
c. various ways. The most often done is by random means. Cluster centers are given initial values with random numbers. Place each data / object into the closest cluster. The proximity of the two objects is determined based on the distance of the two objects. Likewise the proximity of a data to a particular cluster is determined by the

distance between the data and the cluster center.

d. Recalculate the cluster center with the current cluster membership. The cluster center is the average of all data / objects in a particular cluster. If desired, you can also use the median of the cluster. So the average (mean) is not the only size that can be used.

e. Assign each object again using the new cluster center. If the cluster center is no longer changing, the clustering process is complete. Or go back to step c until the center of the cluster doesn't change again.

RESULT AND DISCUSSION

Data preparation is done manually which is using Ms. Excel. The 1st column is the drug id, the second column is the antibiotic, the third column is antiulkus, the fourth column is the antihistamine, the fifth column is cardiovascular. The results of the data presented in the initial process (pre-processing) are taken 5 data samples of each drug.

Research Result Analyst

1. K-Means Algorithm

K-Means is a method of grouping non-hierarchical data that attempts to partition existing data into two or more groups. This method partitioned the data into groups so that data with different characteristics were grouped into other groups.

The auxiliary program used in the process of data mining here is using the RapidMiner studio Trial 9.2 application. To make it easier to use RapidMiner to make clustering.

After the data is created in the form of an xls format table, then Importing Data into the Repository, look for the Microsoft Excel table that was created and input it into the Local Repository.

To make k-means clustering using RapidMiner, we need a K-Means operator. Import data that will be clustered into the RapidMiner program. Click import data, then select the data to be clustered

Row No.	id	cluster	ANTIBIOTIK	ANTIULKUS	ANTIHISTAM...	KARDIOVAS...
1	1	cluster_0	8	6	5	8
2	2	cluster_0	4	5	6	10
3	3	cluster_2	2	10	2	5
4	4	cluster_1	5	5	10	10
5	5	cluster_1	3	10	9	8

Figure 1 Display of data that has been clustered

Figure 2 Display statistics of data that has been clustered

The picture above is explained by statistics from the results of cluster data processed by RapidMiner. In the statistics it is explained that the minimum value for biotic drugs is 2 (two), the maximum value is 5 (five) and the average value is 4 (four). On anti-infectious drugs, the minimum value is 5 (five), the maximum value is 10 (ten) and the average value is 7.2 (seven point two). For antihistamines, the minimum value is 2 (two), the maximum value is 10 (ten), and the average value is 6.4 (six point four). On cardiovascular drugs, the minimum value is 5 (five), the maximum value is 10 (ten) and the average value is 8.2 (eight point two).
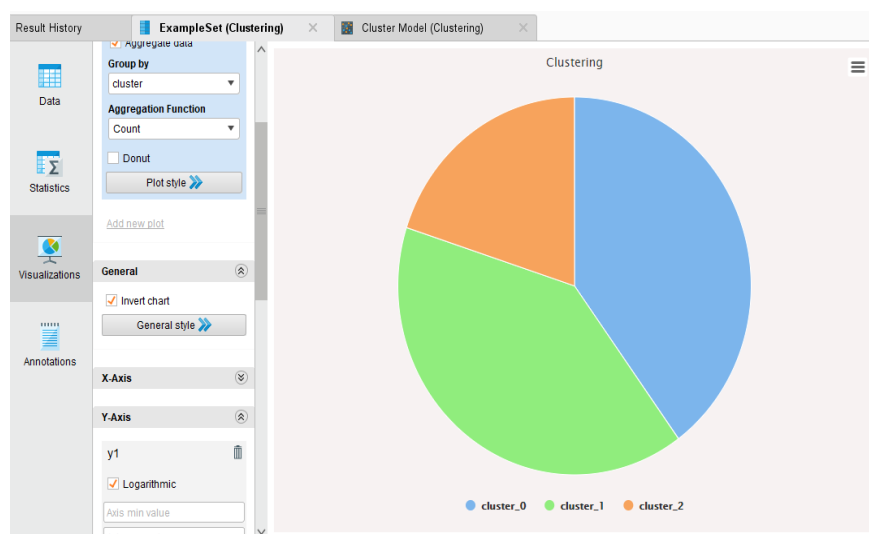

Figure 3 Clustered graphical display / data diagram

After the statistical data, the rapidminer can also be seen visualizing in the form of a diagram. Several diagram choices. This visualization uses the Pie diagram. Blue shows two clusters of 0 (zero), green shows cluster 1 (one) as much as two data, and orange color shows cluster 2 (two) as much as 1 data.
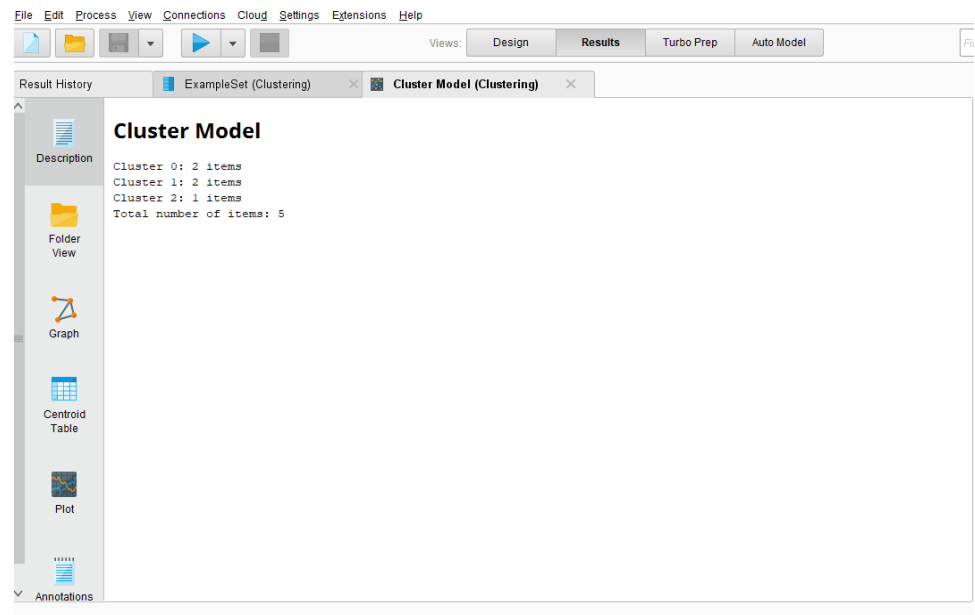

Figure 4 Display description of data that has been clustered

The picture above explains the part of data that is clustered. In the sample data used, there are five data available. In the description section of the rapidminer a cluster is displayed and the amount of data in each cluster is displayed. Cluster 0 (zero) as many as two data, cluster 1 (one) as many as two data, and cluster 2 (two) as many as one data.

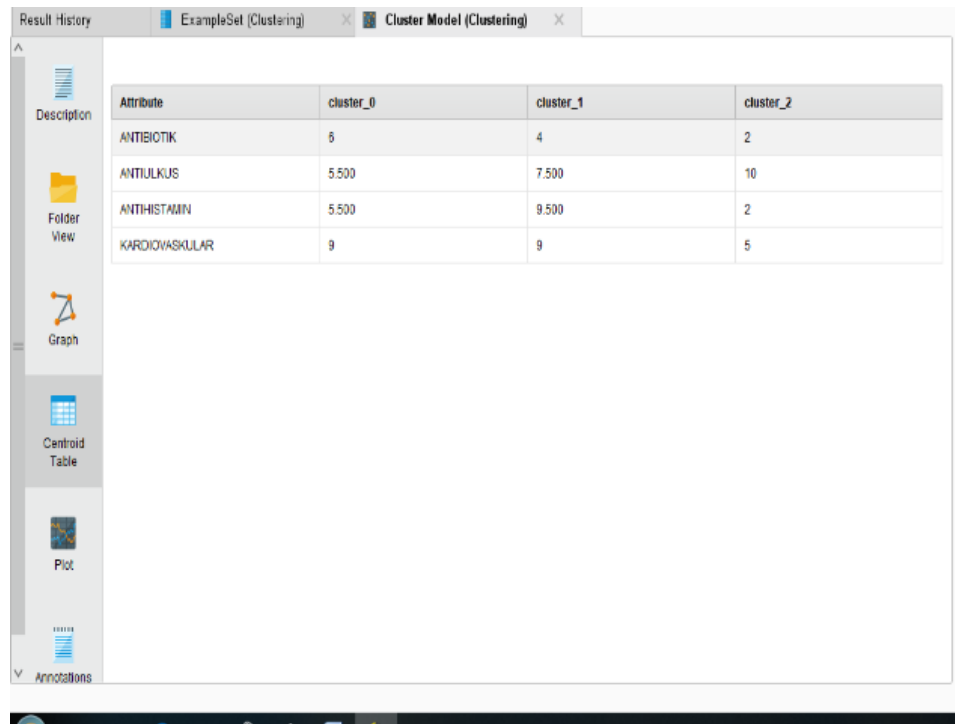| Attribute	cluster_0	cluster_1	cluster_2
ANTIBIOTIK	6	4	2
ANTIULKUS	5.500	7.500	10
ANTIHISTAMIN	5.500	9.500	2
KARDIOVASKULAR	9	9	5

Figure 5 Display of clustered centroid data tables

The picture above describes the centoid part of the data cluster on Rapidminer. Centorid data is taken from the middle value of data that has been clustered. On antibiotics, zero is the middle value is 6, clustered one middle value is 4 and clustered two middle values are 2. On the antiulcus drug, the middle value in the zero cluster is 5.50, in the cluster one the middle value is 7.50, and in the cluster the two middle values are 10. On antihistamine drugs, the middle value in the zero cluster is 5.50, in the cluster one is 9.50, and in the cluster of two middle values 2. On the cardiovascular drug the middle value in the zero cluster is 9, in the cluster one is the middle value of 9, and in the cluster of two middle values 5.

2. Decision Tree Method

Classification is the process of finding a model based on classes that are used as a differentiator between one class and another class. Decision tree is the application of the most popular classification method, with this method an item can be grouped and modeled on a decision tree, so that it can be easily understood. The following is the Application of the decision tree method on drug data using the RapidMiner Application

Row No.	Sering digunakan	ANTIBIOTIK	ANTIULKUS	ANTIHISTAM...	KARDIOVASKULAR
1	Ya	8	6	5	8
2	tidak	4	5	6	10
3	ya	2	10	2	5
4	ya	5	5	10	10
5	Tidak	3	10	9	8

Figure 6 Cluster Data Display with Decision Tree Method

The picture above is the result of the decision tree method. The following is a description of the graphics produced by Rapidminer using the Decision Tree method.
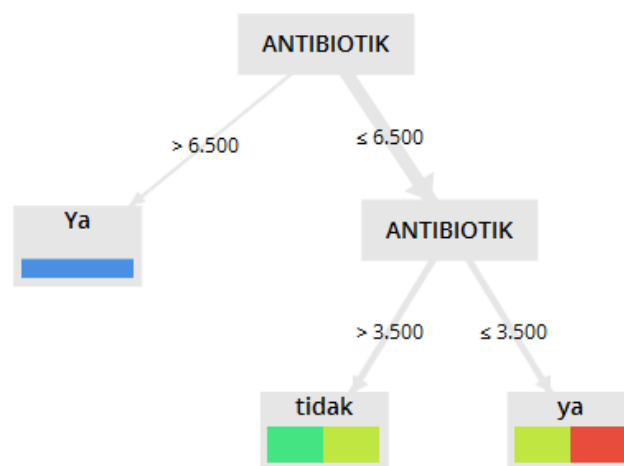

Figure 7 Image of Decision Tree from data of drug processed

Of the five rows of data available on antibiotic drugs there are two drug data that are often used with a value index above 6.5, while three data indexes below 6.5 drugs fall into the rarely used category.

3. Linear Regression

Regression analysis is a statistical technique for modeling and investigating the relationship of two or more variables. What is often used is simple linear regression. In regression analysis one or more independent / predictor variables which are usually represented by x

notation and one response variable are usually represented by y notation

Row No.	ANTIBIOTIK	ANTIULKUS	ANTIHISTAM...	KARDIOVAS...
1	8	6	5	8
2	4	5	6	10
3	2	10	2	5
4	5	5	10	10
5	3	10	9	8

Figure 8 Linear Regression Data

Figure 8 is the data that will be used for the liner registration process. This data will be processed and divided into two columns, namely attribute columns and weight columns.

attribute	weight
ANTIBIOTIK	1
ANTIULKUS	0.261
ANTIHISTAMIN	0.905
KARDIOVASKULAR	0

Figure 9 Regressed data

In figure 9 is the regression data. In the column attribute there are names of drugs used in the linear regression process and the weight column is the value generated from the linear regression process.

After the regression process is complete, matrix results are generated in the linear regression data.

Attributes	ANTIBIOTIK	ANTIULKUS	ANTIHISTAM...	KARDIOVASKULAR
ANTIBIOTIK	1	-0.646	0.142	0.403
ANTIULKUS	-0.646	1	-0.313	-0.811
ANTIHISTAMIN	0.142	-0.313	1	0.745
KARDIOVASKULAR	0.403	-0.811	0.745	1

Figure 10 Matrix Correlation

In Figure 10 the correlation matrix of the drug is divided into two attribute columns and produces each value according to its linear regression. The first attribute column correlates with the second attribute column which produces a value from the linear regression results.

CONCLUSION AND IMPLICATION

a. Conclusion

The conclusion in this writing is that in every business activity in an organization utilizing the system certainly helps the organization in making decisions. Drug clustering techniques in this study provide an overview of drug groupings that are often used in certain periods. From this grouping, it is hoped that it can be useful for organizations in making decisions to supply drugs. Medicines with a high intensity of use will be backed up more so that when needed medicine is always available. The Clustering method used in this study uses the K-meas and Decision method, when viewed from the description of the results, the decision tree is easier to read the results, but when viewed from the technique of ease of use, the K-means method is easier to use.

b. Implicasion

In this study there are still many shortcomings both from the writer and the data provider side. The data obtained is not maximal so that the data processing becomes less maximal. It is expected that if there is further research, the data used as an example of testing variables so that more drug data can be used so that when the clustering method is used then the results obtained from the study are expected to be better.

REFERENCES

Buulolo, E. (2013). Implementasi Algoritma Apriori Pada Sistem Persediaan Obat (Studi Kasus: Apotik Rumah Sakit Estomihi Medan). *Pelita Inform. Budi Dharma*, *4*, 71-83. Retrieved from https://www.researchgate.net/publication/312495968_IMPLEMENTASI_ALGORITMA_APRIORI_PADA_SISTEM_PERSEDIAAN_OBAT_STUDI_KASUS_APOTIK_RUMAH_SAKIT_ESTOMIHI_MEDAN

Syamsuni, H. (2006). *Farmasetika Dasar dan Hitungan Farmasi*. Jakarta: Buku Kedokteran EGC

Nugraha, J. A. M., & Kusumawati, Y. (2014). Data Mining dengan Metode Clustering untuk Pengolahan Informasi Persediaan Obat pada Puskesmas Pandanaran Semarang. *Universitas Dian Nuswantoro*. Retrieved from http://eprints.dinus.ac.id/13153/

Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, *3*(1), 19-26. https://doi.org/10.15294/sji.v3i1.4610

Kurniawan, E., Purnama, I. K. E., & Sumpeno, S. (2011). Analisa Rekam Medis untuk Menentukan Pola Kelompok Penyakit Menggunakan Klasifikasi dengan Decision Tree J48. *Institut Sepuluh Nopember, Surabaya*.1-8. Retreived from https://mmt.its.ac.id/publikasi/analisa-rekam-medis-untuk-menentukan-pola-kelompok-penyakit-menggunakan-metode-klasifikasi-dengan-algoritma-decision-tree-j48/

Taslim, T., & Fajrizal, F. (2016). Penerapan Algorithma K-Mean untuk Clustering Data Obat pada Puskesmas Rumbai. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, *7*(2).108-114. https://doi.org/10.31849/digitalzone.v7i2.602

Bastian, A. (2018). Penerapan Algoritma K-Means Clustering Analysis pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi*, *14*(1), 28-34. https://doi.org/10.21609/jsi.v14i1.566

Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Creative Information Technology Journal*, *2*(2), 102-113. https://doi.org/10.24076/citec.2015v2i2.41

Aprilla, D., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining dengan Rapid Miner*. Jakarta: Gramedia Pustaka Utama.

Sulastri, H., & Gufroni, A. I. (2017). Penerapan data mining dalam pengelompokan penderita thalassaemia. *Jurnal Nasional Teknologi dan Sistem Informasi*, *3*(2), 299-305. https://doi.org/10.25077/TEKNOSI.v3i2.2017.29