SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of Data Mining for Optimal Drug Inventory in a Hospital

Siringo-ringo, Dewi Sahputri
Universitas Prima Indonesia
Medan, Indonesia
dewisahputri.siringoringo@gmail.com

Tambunan, Razana Baringin Daud
Universitas Prima Indonesia
Medan, Indonesia
razana23tambunan@gmail.com

Yulizar, Dian
Universitas Prima Indonesia
Medan, Indonesia
dianyulizar3@gmail.com

Daulay, Tri Agustina
Universitas Prima Indonesia
Medan, Indonesia
triagustinadaulay683@gmail.com

Amir Mahmud Husein
Universitas Prima Indonesia
Medan, Indonesia
amirmahmud@unprimdn.ac.id

Abstract— The Hospital is a health care institution that conducts complete individual health services that provide inpatient, outpatient and emergency services. Drug inventory management is one thing that is very important for the survival of hospitals, management of the supply of medical equipment that is not optimal including medicines will have an impact on medical services as well as economically, because 70% of hospital revenue comes from drugs. In this study we propose data mining with a focus on contributions to the comparison of the K-Means and K-Nearest Neighbor (KNN) algorithms for disease classification, then the classification results are carried out mapping the correlation of diseases with drugs using Apriori, based on the results of testing the K-Means algorithm more accurately compared KNN in the Apriori method to find the relationship of disease with drugs based on the value of support, trust, support value, trust is expected to be a reference for drug purchase recommendations so that there is no excess or emptiness of the drug.

Keywords—datamining; k-means; knn; apriori

I. INTRODUCTION

Hospital is a health care institution that organizes individual health services in a plenary that provides inpatient, outpatient, and emergency services (Dep Kes RI, 2016).

The application of data mining in the field of health has been proposed many researchers, because it has the ability to extract large amounts of data to obtain information (Harahap, M et al, 2018).

The classifications are one of the many functions of the data mining process to find an almost identical group of objects (Manuel, Ricky, 2017). A clustering method that has an efficient nature and rapid data processing process one of which is the K-means algorithm. The method of clustering the K-Means, grouping data based on the proximity of data to each other corresponds to the Euclidean distance by comparing the majority of the similarities of other objects assigned to each cluster (Arora & Shipra, 2015). The study uses 3 classes, many, medium and little. The data used is the medicine sales dataset sourced from the medical record with the gender variable, age category and diagnosis name of the disease.

After performing the Clustering phase, the classification method is also used to find the most disease groups. Classification is a data analysis process that generates a model to describe the classes that exist in the data. There are many types of classification algorithms, two of which are decision trees and K Nearest Neighbors (KNN). The classification method used is the closest K-Neighbor method, which uses exercise data and data testing (Hermanto, et al, 2019).

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The results of Clustering and classification will be used by the association's rule methods to find links between diseases and medications based on the most 10 diseases. The implementation of the Association rules aims to find information about interconnected items in the form of a rule. Thus the association rules are applied to the most disease data patterns with the medicine using apriori algorithm. Apriori algorithm is a popular algorithm for mining sets of items that often acquire association rules (Rani, 2015).

From the results of the support and confidence obtained from the priori algorithm, this study predicts the medicine needed for the 10 most diseases in optimal inventory.

II. Literature Review

A. Data Mining

Data Mining is the process of converting raw data into data that produces useful and comprehensive information (Arora & Shipra, 2015).

B. Medical Records

In regulation of the Minister of Health No. 269/MenKes/PER/III/2008 concerning medical records stating medical records are files containing records and documents in patients containing identity, examination, treatment, other medical measures on Health care services for outpatient and hospitalization, both government and private (Ramadhana, 2019).

C. Clustering

Clustering is one of the methods of data processing without direction (without supervision) used in the process of data grouping (Khotimah et al, 2106).

The K-means algorithm is one of the well-known partitioning methods for Clustering. The method of grouping K-means is the grouping of data based on the proximity of data to each other according to the Euclidean distance by comparing the majority of the similarities of other objects assigned to each cluster (Arora & Shipra, 2015).

D. Classification

Classification is the process of data analysis that produces a model to describe the existing classes in the data, this model is called classification (Hermanto et al, 2109).

The K-NN method is a supervised method. At the KNN, the nearby K neighbors were found for unknown samples and determined the class that had a nearby neighbor. That is, this principle for neighbouring X closest among the samples is known. KNN can be easily integrated with other machine learning algorithms except lazy learning defects and dependence on class K (Alimjan et al, 2018).

E. Association

The method of association is one of the data mining techniques in finding combinations or patterns of a set of items. This method is often used in a convenience store sales transaction.

F. Weka Tool

Weka Tool is the most powerful data mining tool. It is also an open-source tool. It has features such as preprocessing filters, selection, classification and regression, grouping, Discovery Associations, visualizations (Kodati et al, 2019).

III. Proposed Method

The methods used in this study are as follows:

A. Clustering Method

The procedure used to do the grouping using K-means, is as follows:

Select the desired number of k clusters

1. Initialize the center of the cluster K (centroid) randomly
2. Place each data or object to the nearest cluster. The proximity of two objects is determined by distance. The distance used in the K-means algorithm is Euclidean Distance.
3. Recalculate the cluster center with the current cluster membership. The new cluster Center is the average (mean) of all the data or objects in a specific cluster.
4. Assign again each object by using the new cluster center. If the cluster center does not change anymore, then the processing process is complete. Or, go back to step 3 until the center of the cluster does not change anymore/stabilized or there is no significant decrease of the SSE value (number of squared errors).

B. Classification Method

The classification method used is K Nearest Neighbor method which uses training data and testing data.

The KNN algorithm stage is:

1. Specify the K parameter (number of closest neighbors)

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

2. Calculate the squared distance from the Euclidean (instance Queri) of each object against the given sample data
3. Then sort the distance into a group that has the smallest Euclid (order result No. 2 ascending)
4. Collecting Y (the classification of nearby neighbors) categories based on the K value or retrieving neighboring data nearby
5. By using the nearest neighbor category the most majority will result in classification.

C. Association Method

The method of association used is a priori algorithm, in which the algorithm seeks support, trust and lifting value. Lifts are a measure of a priori to know the strength of the association rules that have formed.

Stages of the apriori algorithm are:

1. Looking for a support value
2. Analyze in order to determine the data related to each other
3. After all high frequencies are found, the researchers proceed to find an association rule that meets the minimum requirements for confidence by calculating the rules of associations A and B
4. After getting the result of the support and confidance value, the next step is to look for the lift value. Lifts are a useful measure to know the strength of the association rules that have formed

IV. RESULT AND DISCUSSION

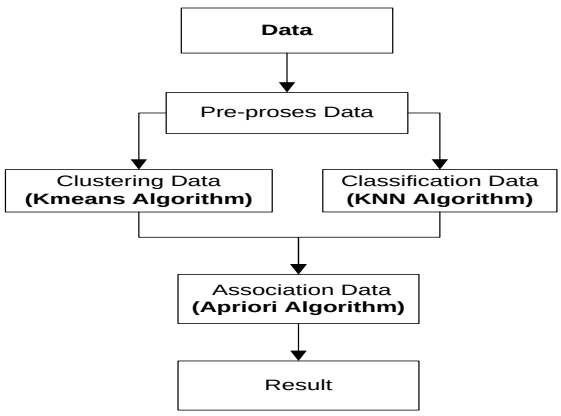

Figure 1. Stages of work process

A. Data

The Data used is the medicine sales dataset sourced from the medical record at RSU X Medan in 2017. The sales Data of the medicine obtained consists of No. Prescription, patient name, date of sale, patient ID, gender, date of birth, diagnosis code of the patient's disease (ICD-10), diagnosis name of the patient's disease, name of the case type of disease, medicine ID, amount of medicine purchase, medicine price per item and total price of medicine sales With a total of 1041 data in the form of Excel files. The later processed attributes are the diagnose code (ICD-10), age and gender categories that have been in pre-process data.

B. Pre-proses Data

1. Cleaning Data

Cleaning data that has incomplete information because the field is empty, or populate the data with a mean value from all data in that column. Because the data that researchers do not have incomplete data then the data cleanup process is not done.

2. Selection Data

Grouping attributes after the cleaning data stage.

Table 1. Dataset Grouping

j_k	tgl_lhr	diagnosa_nama
male	2/24/1940	Diabetes mellitus is not dependent on insulin
male	2/24/1940	Fluid, electrolyte, and acid-base disorders
. . .	. . .	. . .
male	7/19/1975	**Septicemia, unspecified**

3. Transformation Data

The process of initializing data from nominal to numerical can be processed by the KMeans algorithm and the KNN algorithm.

a. Gender

Table 2. Gender Initials

Gender	Initials
Female	0
Male	1

b. Birth dates will be grouped into 5 age categories namely infants, toddlers, children, adults and the elderly.

Table 3. Age Category Initials

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Age Category	Initials
Babies	0
Toddler	1
Children	2
Adult	3
Elderly	4

c. The name of the diagnosis is obtained from diagnostic data (disease data).

Table 4. Data Initialization Results

No	gender	Age_c	diagnosa_nama
1	1	4	2012
2	1	4	2336
3	0	2	4288
.	.	.	.
.	.	.	.
1041	1	3	226

4. Pre-process Data

Save this data as a .csv file. Preprocess data is performed on the Weka 3.8 application by entering the selected data, then the data is stored in a .arrf file which can later be read by Weka.

C. Clustering Method in Weka

1. KMeans Algorithm

The grouping results obtained from the K-Means Algorithm is:

Full data: 1041 data with 3 attributes

Dengan nilai SSE = 51,17 %

The grouping process produces 2 iterations with 417 test data, as follows:

a. cluster 0 (many) there were 223 people or 53% of 417 records with the number of diseases 77.

b. cluster 1 (medium) there are 87 people or 21% of 417 records with 37 diseases.

c. cluster 2 (little) there are 107 people or 26% of 417 records with the number of diseases 34.

Table 5. Weka K-Means Clustering Results

gender	Age_c	diagnose_code	Cluster
0	2	6	cluster1
1	4	3929	cluster0
1	4	5526	cluster0
.	.	.	.
.	.	.	.
1	1	4340	cluster0

Table 6. Clustering Analysis Results

Results *Cluster* 1	Results *Cluster* 2
Consisting of 233 people with a diagnosis of the diseases: • Typhoid fever = 11 people • Diarrhea and gastroenteritis thought to be caused by infection = 8 people . . • Chf + Pjk = 5 people By age category: • Babies = 10 people • Toddler = 10 people • Children = 16 people • Adult = 99 people • Elderly = 88 people And gender: Male = 223 people	Consisting of 87 people with disease diagnoses: • Typhoid fever = 12 people • Diarrhea and gastroenteritis suspected to be caused by infection = 2 people . . • Tonsillopharyngitis = 1 people By age category: • Babies = 5 people • Toddler = 3 people • Children = 11 people • Adult = 68 people And gender: Female = 87 people

Table 6. Continued

Results *Cluster* 3
Consisting of 107 people with diagnoses of the disease: • Typhoid fever = 1 people • Anemia, not specified = 3 people . . • Chf + Pjk = 4 people By age category: Elderly = 107 people And gender: Female = 107 people

D. Classification Method in Weka

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

1. KNN Algorithm

The classification results obtained from the KNN method are:

Complete data: 1041 divided into 2 data, namely training data and testing data.

Class classification is divided into 3 parts, namely many, medium and little.

The classification accuracy level is 68% and the remaining 31% is the wrong classification.

284 data are classified correctly, while 133 data are not classified correctly.

a. "Many" there are 174 people

b. "Medium" there are 116 people

c. "Little" there are 127 people

Table 7. Results of KNN Classification

jk	k_umur	d_nama	p_klasifikasi	classification
0	3	3683	Little	Many
0	2	6	Many	Many
0	3	4267	Medium	Medium
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1	3	226	Little	Little

Table 8. Results of Prediction Analysis of KNN Classification

Many	Medium
Consists of 174 people with a diagnosis of the disease: • Typhoid fever = 18 people • Diarrhea and gastroenteritis suspected to be caused by infection = 3 people . . • Respiratory failure, unspecification = 1 people	Consisting of 116 people with a diagnosis of the disease: • Typhoid fever = 1 people • Diarrhea and gastroenteritis are thought to be caused by infection = 1 people . . • Chf + Pjk = 5 people
By age Category: • Toddler = 1 people • Children = 10 people • Adult = 66 people • Elderly = 97 people And gender: Female = 86 people Male = 88 people	By age category: • Babies = 3 people • Children = 3 people • Adults = 52 people • Elderly = 58 people And gender: • Female = 42 people • Male = 74 people

Table 8. Continued

Little
Consists of 127 people with a diagnosis of the disease: • Diarrhea and suspected gastroenteritis caused by infection = 4 people • Pulmonary tuberculosis, confirmed in an unspecified manner = 3 people . . • Tonsillopharyngitis = 2 people By age category: • Babies = 9 people • Toddler = 6 people • Children = 18 people • Adult = 66 people • Elderly = 28 people And gender: • Female = 48 people • Male = 79 people

E. Association Method in Rapid Miner

1. Apriori Algorithm

The result of the best accuracy of the disease used by the apriori algorithm is the result of the Kmeans algorithm. The disease generated data is found in Table 9, as follows:

Table 9. Most diseases of KMeans algorithm

ICD-10	Name of the disease
J18.0	Bronkopneumonia, not specified
N18.9	Chronic kidney failure, not specified
I50.0	Congestive heart failure
E87.8	Electrolyte abnormalities and fluid balance, not elsewhere classified
A01.0	Typhoid fever

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

E11.9	Diabetes mellitus is not insulin dependent without complications
I63.9	Infark Serebri, not specified
A09	Diarrhea and Gastroenteritis suspected to be caused by infection
I10	Essential Hypertension (Primary)
A91	Dengue hemorrhagic fever

From the above table the researchers describe 10 diseases with medicine and medical devices that are interrelated. Data related to medicine and medical devices are shown in table 10:

Table 10. Medicine and medical devices related to the 10 Most Diseases

ICD-10	Diagnose name	Medicine name
J 18.0	Bronkopneumonia, not specified	CEFOTAXIME 1 GR INJEKSI
		VELUTINE INJEKSI
		CORTIDEX INJEKSI
		SPUIT 10 CC
		ALCOHOL SWAB
		SPUIT 3 CC
		MASKER EARLOOP (KARET)
		SARUNG TANGAN NON STERIL (S)
		CAIRAN 4:1 WIDA (WIDA D5-1/4 NS)
		3-WAY BD CONNECTA PLUS 3 WHITE
		ABBOCATH NO 24 TERUMO
N 18.9	Chronic kidney failure, not specified	INTRAFIX SAFESET (DEWASA)
		PHENYTOIN INJEKSI
		STERIL WATER 1000 ML OTSU
		RINGER LACTAT 500 ML WIDA
		ONDANSETRON 4 MG INJEKSI
		SPUIT 5 CC
		NACL 500 ML WIDA
		CEFTRIAXONE 1 GR INJEKSI
		ALLUPURINOL 100 MG TABLET
		KETOSTERIL TABLET
		SPUIT 10 CC

After obtaining medicine that are associated with most diseases, researchers look for a minimum support value of 5% for 1 itemset candidates, self-confidence values, and lift values. Here are the results of medicine and medical devices related to most diseases, namely:

Table 11. Medicine results with most diseases

Diagnosa_N	N_Obat	Sup	Co	Lift
Bronkopneumonia, not specified	CEFOTAXIME 1 GR INJEKSI	0.2	0.5	0.625
	SPUIT 10 CC	0.2	0.5	0.72
	ALCOHOL SWAB	0.2	0.5	0.84
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
	SARUNG TANGAN NON STERIL (S)	0.3	0.5	1.25
	ABBOCATH NO 24 TERUMO	0.2	0.5	2.5
Chronic kidney failure, not specified	RINGER LACTAT 500 ML WIDA	0.2	0.5	0.72
	ONDANSETRON 4 MG INJEKSI	0.2	0.5	0.72
	SPUIT 5 CC	0.4	0.5	1
	CEFTRIAXONE 1 GR INJEKSI	0.2	0.5	16.67
	SPUIT 10 CC	0.2	0.5	0.72
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
	SARUNG TANGAN NON STERIL (S)	0.3	0.5	1.25
Congestive heart failure	SPUIT 10 CC	0.2	0.5	0.72
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
	SARUNG TANGAN	0.3	0.5	1.25

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236
e-ISSN : 2541-2019
p-ISSN : 2541-044X

	NON STERIL (S)			
	SPUIT 5 CC	0.4	0.5	1
	RANITIDINE 25 MG INJEKSI	0.2	0.5	0.84
Electrolyte abnormalities and fluid balance, not elsewhere classified	SARUNG TANGAN NON STERIL (M)	0.2	0.5	2.5
	SPUIT 5 CC	0.4	0.5	1
	ALCOHOL SWAB	0.2	0.5	0.84
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
Typhoid fever	SPUIT 10 CC	0.2	0.5	0.72
	SARUNG TANGAN NON STERIL (S)	0.3	0.5	1.25
	RINGER LACTAT 500 ML WIDA	0.2	0.5	0.72
	SPUIT 5 CC	0.4	0.5	1
	CEFTRIAXONE 1 GR INJEKSI	0.2	0.5	16.67
	RANITIDINE 25 MG INJEKSI	0.2	0.5	0.84
Diabetes mellitus is not insulin dependent without complications	SPUIT 10 CC	0.2	0.5	0.72
	SPUIT 5 CC	0.4	0.5	1
	RANITIDINE 25 MG INJEKSI	0.2	0.5	0.84
Infark Serebri, not specified	SPUIT 10 CC	0.2	0.5	0.72
	SPUIT 3 CC	0.4	0.5	1.25
	SPUIT 5 CC	0.4	0.5	1
Diarrhea and Gastroenteritis suspected to be caused by infection	RANITIDINE 25 MG INJEKSI	0.2	0.5	0.84
	CEFTRIAXONE 1 GR INJEKSI	0.2	0.5	16.67
	RINGER LACTAT 500 ML WIDA	0.2	0.5	0.72
	ONDANSETRON 4 MG INJEKSI	0.2	0.5	0.72
	ALCOHOL SWAB	0.2	0.5	0.84
	SPUIT 3 CC	0.4	0.5	1.25
Essential Hypertension (Primary)	SPUIT 10 CC	0.2	0.5	0.72
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
	SARUNG TANGAN NON STERIL (S)	0.3	0.5	1.25
	ONDANSETRON 4 MG INJEKSI	0.2	0.5	0.72
	SPUIT 5 CC	0.4	0.5	1
Dengue hemorrhagic fever	SARUNG TANGAN NON STERIL (M)	0.2	0.5	2.5
	RINGER LACTAT 500 ML WIDA	0.2	0.5	0.72
	ONDANSETRON 4 MG INJEKSI	0.2	0.5	0.72
	SPUIT 5 CC	0.4	0.5	1
	ALCOHOL SWAB	0.2	0.5	0.84
	SPUIT 3 CC	0.4	0.5	1.25
	MASKER EARLOOP (KARET)	0.3	0.5	1.25
	SARUNG TANGAN NON STERIL (S)	0.3	0.5	1.25

Medicines and medical devices that will be recommended for future supplies, that is CEFTRIAXONE 1 GR INJEKSI, MASKER EARLOOP (KARET) dan ABBOCATH NO 24 TERUMO.

V. Conclusion and Suggestion

A. Conclusions

1. K-Means algorithm and KNN algorithm can be used to find the 10 most diseases. The results of the algorithm are used by the a priori algorithm to find the relationship of the disease with the medicine, so that it can recommend medicines that will be provided in the future.
2. The best results from the 10 most diseases are from the Clustering KMeans algorithm.
3. As for medicines and medical equipment that will be recommended for future supplies, that is CEFTRIAXONE 1 GR INJEKSI,

Journal Publications & Informatics Engineering Research
Volume 4, Number 1, October 2019 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v4i1.10236 p-ISSN : 2541-044X

MASKER EARLOOP (KARET) dan ABBOCATH NO 24 TERUMO.

B. Suggestions

1. For further research can do a combination comparison with Clustering methods, classifications and different associations to the medicine inventory.
2. For further research can add attributes in the clustering and classification of patient disease data.

VI. REFERENCES

Alimjan, Gulnaz, et al. 2018. "A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN." *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 7, pp. 1–23.

Arora, Preeti, and Shipra Varshney. *"Analysis of K-Means and K-Medoids Algorithm For Big Data." Procedia - Procedia Computer Science*, vol. 78, no. December 2015, Elsevier Masson SAS, 2016, pp. 507–12, doi:10.1016/j.procs.2016.02.095.

Dep Kes RI, 2016. Keputusan Menteri Kesehatan Republik Indonesia Tentang Standar Pelayanan Kefarmasian Di Rumah Sakit, Jakarta

Harahap, M., Husein, A. M., Aisyah, S., Lubis, F. R., & Wijaya, B. A. (2018, April). Mining association rule based on the diseases population for recommendation of medicine need. *In Journal of Physics: Conference Series* (Vol. 1007, No. 1, p. 012017). IOP Publishing.

Hermanto, et al. 2019. Comparison of Naïve Bayes Algorithm, C4.5 and Random Forest for Service Classification Ojek Online. Vol. 3, no. 2.

Husein, A. M., Harahap, M., Aisyah, S., Purba, W., & Muhazir, A. (2018, March). The implementation of two stages clustering (k-means clustering and adaptive neuro fuzzy inference system) for prediction of medicine need based on medical data. *In Journal of Physics: Conference Series* (Vol. 978, No. 1, p. 012019). IOP Publishing.

Khotimah, Bain Khusul, et al. 2016. A Genetic Algorithm for Optimized Initial Centers K-Means Clustering in SMEs. *Journal of Theoretical and Applied Information Technology*, vol. 90, no. 1, pp. 23–30.

Kodati, Sarangam, et al. 2019. Soft Computing and Signal Processing. Vol. 898, Springer Singapore.

Manuel, Ricky. 2017. Analisa Penentuan Skala Prioritas Obat Berdasarkan Klaster Penyakit Menggunakan Fuzzy C-Means *(Studi Kasus : Kecamatan Sirimau Kota Ambon)*.

Ramadhana, Fanny. 2019. Klasifikasi Data Rekam Medis Berdasarkan International Statistical Classification of Diseases and Related Health Problem (ICD-10) Menggunakan Algoritma K-Nearest Neighbor (K-NN).

Rani, Nisha. 2015. Improving the Performance of Apriori Algorithm by Combining with Clustering Techniques. Vol. 3, no. 2, pp. 13–15.