JurnalMantik
Volume 4, Number 4, Feb 2021, pp. 2604-2612 E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

Data Mining in Grouping Indihome Customer Data Using the K–Means Clustering Method at PT.Telkom Akses

Egi Affandi[1], Syahrin Syam Noor Berutu[2], Yohanni Syahra[3], A. Amrullah[4]

[1,2,4]STMIK Triguna Dharma Medan, Jalan Pintu Air I Nomor 73 Medan, 20146, Indonesia
[3]Universitas Pembangunan Panca Budi, Jl. Gatot Subroto No 4, Medan Sunggal,Medan, 20122, Indonesia

Email : egi.afandi46@gmail.com, syahrin.berutu@gmail.com,yohanni.syahra@gmail.com, amrullahmedan1@gmail.com

ARTICLE INFO

Article history:
Received: 01/01/2021
Revised: 10/01/2021
Accepted: 30/01/2021

Keywords:
PT. Telkom, Indihome, Data Mining, Clustering K-means

ABSTRACT

PT. Telkom Akses is part of the PT. Telekomunikasi Indonesia, Tbk (Telkom) company whose shares the stock and fully owned by Telkom.Telkom is a state owned company and as an IndiHome service Provider. So far, PT. Telkom Akses still has problems in classifying customer data. Based on these problems, PTTA requires a computer application that can cluster customer data and use data mining that can help solve problems that occur using the clustering method.. One of the methods offered to solve this problem is the k-means algorithm. The results of this study, the authors design and design a desktop-based information system that can implement the clustering method to produce customer data clusters according to the needs of PT. Telkom Access

Copyright © 2021 Jurnal Mantik.
All rights reserved.

1. Introduction

IndiHome (Indonesia Digital HOME) is a leading digital service using fiber optic technology that offers Triple Play services consisting of Fiber Internet (Fast Internet), Landline Telephone and Interactive TV services. IndiHome also offers Dual Play services consisting of Fiber Internet (Fast Internet) and Landline Telephone. IndiHome is also equipped with a variety of additional services that can be selected according to customer needs and desires such as Telephone Mania, Wifi.id Seamless, Trend Micro Antivirus, IndiHome View and others.

Customers are people or agencies or institutions who buy our services regularly or repeatedly because the goods and services purchased have benefits. And to classify customers, one of the methods used is data mining. Data mining is a method used to classify customer data. To determine the customer group, the K-Means method will be applied in classifying customers. In Research [1], it is explained that data mining is one way to find interesting patterns from large amounts of data, data warehouses and data can be stored in databases or other information storage.

Customers are people or agencies or institutions who buy our services regularly or repeatedly because the goods and services purchased have benefits. And to classify customers, one of the methods used is data mining. Data mining is a method used to classify customer data. To determine the customer group, the K-Means method will be applied in classifying customers. In Research [1], it is explained that data mining is one way to find interesting patterns from large amounts of data, data warehouses and data can be stored in databases or other information storage.

2. Theoritical Basis

2.1 Knowledge Discovery Database (KDD)

KDD is the whole non-trivial process to look for and identify patterns in data, where the patterns found are valid, useful and understandable. Data mining is also often connected by using data storage analysis in a warehouse. Broadly speaking, there are three main techniques in data mining, including classification and cluster regression [2]. As for some KDD processes in general can be explained as follows [3]:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

a. Data Selection
   In selecting a data set from several operational data sets it is very important to do it before the information mining phase in KDD begins. The selected data that will be used for the data mining process is stored in a separate file from the operational data base.
b. Pre-processing (Cleaning)
   In carrying out the data mining process, it is important to carry out the cleaning process on a data which will be the focus of KDD. The cleaning process includes removing duplicate data, checking inconsistent data, and correcting errors in the data, such as printing errors.
c. Trasnformasi (Transformation)
   Coding is a transformation of the selected data, so that the data is suitable for the data mining process. The coding process in KDD is a creative process and really depends on the type or pattern of information to be searched in the database.
d. Data mining
   Data mining is a process for looking for patterns or information. Data mining is a process for finding patterns or interesting information on data that has been selected using a particular technique or method. Choosing the right method or algorithm really depends on the objectives of the KKD process as a whole.
e. Interpretation / Evaluation
   In the pattern of information generated from each data mining process, the importance of being displayed in a form that is easily understood by interested parties. At this stage, it is part of every KDD process called interpretation. This stage includes checking whether the patterns or information found are contrary to the facts.

3. Research Methodology

The research method is a scientific stage that is used to collect theories and matters related to the research being carried out and the process undertaken to complete the research. The following are some of the research methods used, namely:

a. Data Collecting
   a. Observation
   b. Interview
b. Library Research
c. Data Analyze
d. System Design
e. System Implementation
f. Testing

3.1 Clustering

Clustering is a part of Data Mining which has no direction. Clustering is a process for dividing data into clusters based on the level of similarity. Clustering is a job that can separate data or vectors into a number of groups or clusters according to their respective characteristics. Data that has a similarity for each characteristic will gather in the same group or cluster. Each data has different characteristics, which will gather in a different group or cluster. The main purpose of the clustering method is to group a number of data or objects into a cluster or group so that a cluster will contain data as closely as possible [4].

3.2 Steps of the K-Means Methods

Therefore the steps of k-means methods as follows adalah[5]:

a. Determine the number of clusters
b. Initialize k as center cluster (give random values)
c. Allocating each data or object to the nearest cluster and the proximity of the two objects is determined based on the distance between the two objects. The closest distance between one data and one particular cluster will determine which data will enter into which cluster
d. Recalculate cluster centers with new cluster members. The center of the cluster is the average of all data in the cluster.
e. Continue again with each object using the new cluster center. If the cluster center has not changed anymore, then each clustering process is complete..
f. Go back to step three until the cluster center won't change anymore

3.3 Data Analyze

The primary data from this company are as follows:

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Table 1
Company's Primary Data

Customer Name	Customer Loyalty	Jenis Package			Timely Payment
		Package Name	Mbps	Price/Month	
Sandy Satriya Putra	18 December 2012	Package Phonix	10 Mbps	Rp280,000	15 May 2019
Dede Irfan	28 January 2013	Package Gamers	50 Mbps	Rp780,000	20 May 2019
Nova Hutabarat	04 February 2014	Package Phonix	10 Mbps	Rp280,000	19 May 2019
Harry Putra	03 March 2014	Package Phonix	20 Mbps	Rp345,000	23 May 2019
Tri Wahyudi	06 May 2014	Package Streamix	50 Mbps	Rp615,000	20 May 2019
Fahmi	08 July 2014	Package Gamers	10 Mbps	Rp380,000	16 May 2019
Yuliyanti	12 January 2015	Package Phonix	50 Mbps	Rp575,000	20 May 2019
Apriansyah	19 February 2015	Package Prestige	20 Mbps	Rp515,000	21 May 2019
Denny Dharma	27 March 2015	Package Phonix	20 Mbps	Rp345,000	28 May 2019
Mariana Sitepu	15 April 2015	Package Phonix	10 Mbps	Rp280,000	20 May 2019
Denny Naibaho	18 May 2015	Package Gamers	20 Mbps	Rp480,000	19 May 2019
Samuel Sitorus	27 January 2016	Package New Netizen 2	10 Mbps	Rp320,000	17 May 2019
Ferry Andrian	08 February 2016	Package Phonix	20 Mbps	Rp345,000	20 May 2019
Joel Sinaga	01 March 2016	Package Phonix	10 Mbps	Rp280,000	19 May 2019
Muhammad Faisal	08 March 2016	Package Streamix	20 Mbps	Rp385,000	20 May 2019
Freddy Nasution	27 May 2016	Package Phonix	50 Mbps	Rp575,000	19 May 2019
Hiskia	07 July 2016	Package Phonix	10 Mbps	Rp280,000	21 May 2019
Bernad Ardiansyah	05 April 2017	Package Prestige	20 Mbps	Rp515,000	20 May 2019
Bery Darmawan	10 May 2017	Package Streamix	20 Mbps	Rp385,000	20 May 2019
Hendro Tri Cahyono	13 March 2017	Package Gamers	100 Mbps	Rp1380000	19 May 2019
Amrida	09 May 2017	PackagePhonix	10 Mbps	Rp280,000	22 May 2019
Abbas Wahyudi	12 June 2017	Package Netizen 1 Plus	10 Mbps	Rp315,000	23 May 2019
Kumala Sari	15 June 2017	Package Phonix	10 Mbps	Rp280,000	20 May 2019
Devi Sari Hiskia	02 July 2018	Package Phonix	10 Mbps	Rp280,000	18 May 2019
Eva Heriyanti	17 August 2018	Package Phonix	10 Mbps	Rp280,000	17 May 2019
Yogi Somantri	12 October 2018	Package Gamers	10 Mbps	Rp380,000	20 May 2019
Nicholas Adji	19 February 2019	Package Gamers	40 Mbps	Rp780,000	19 May 2019

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Customer Name	Customer Loyalty	Jenis Package			Timely Payment
		Package Name	Mbps	Price/Month	
Feby Anatasya	04 March 2019	Package Phonix	10 Mbps	Rp280,000	15 May 2019
Felix	07 March 2019	Package Netizen 1 Plus	10 Mbps	Rp315,000	20 May 2019
Anto Darmawan	06 May 2019	Package Streamix	20 Mbps	Rp385,000	21 May 2019

4. Result and Implementation

4.1 K-Means Method Clustering Process

Table2
Result of Primary Data

No	Customer Name	C1	C2	C3
1	Sandy Satriya Putra	100	50	100
2	Dede Irfan	100	80	100
3	Nova Hutabarat	80	50	100
4	Harry Putra	80	50	10
5	Tri Wahyudi	70	70	100
6	Fahmi	70	50	100
7	Yuliyanti	70	70	100
8	Apriansyah	70	70	10
9	Denny Dharma	70	50	10
10	Mariana Sitepu	70	50	100
11	Denny Naibaho	70	60	100
12	Samuel Sitorus	70	50	100
13	Ferry Andrian	70	50	100
14	Joel Sinaga	70	50	100
15	Muhammad Faisal	70	50	100
16	Freddy Nasution	50	70	100
17	Hiskia	50	50	10
18	Bernad Ardiansyah	50	70	100
19	Bery Darmawan	50	50	100
20	Hendro Tri Cahyono	50	100	100
21	Amrida	50	50	10
22	Abbas Wahyudi	50	50	10
23	Kumala Sari	50	50	100
24	Devi Sari Hiskia	50	50	100
25	Eva Heriyanti	50	50	100
26	Yogi Somantri	50	50	100
27	Nicholas Adji	50	80	100
28	Feby Anatasya	50	50	100
29	Felix	50	50	100
30	Anto Darmawan	50	50	10

After the data is collected, manual calculations are performed as follows:

a. Determine the initial Cluster Center Table

Table 3
Initialize the cluster start center point

NO	Centroid	Centroid Name

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

1	M1	Centroid 1
2	M2	Centroid 2
3	M3	Centroid 3

Determination of the number of clusters (K), namely 3 clusters. After setting the amount cluster, determine the starting center point of the cluster (centroid). Here is the selected centroid:

Table 4
The starting point of the cluster

NO	Centroid	Customer Loyalty	Package Type	Timely Payment
1	1st Data (M1)	100	80	100
2	2nd Data (M2)	70	50	100
3	3rd Data (M3)	50	50	10

b. Calculation of the distance from the cluster center with the formula:
D(X_(2,) X_1)=|(|X_2-X_1 |)|= √(∑_(i=1)^p▒〖|X_2j-X_1j |^2 〗)

c. Calculating the distance between the variables from each data sample and the centroid. Here is the iteration :

a) Distance between 1st data to m1 point

$$= \sqrt{\sum_{i=1}^{p} |X_{2j} - X_{1j}|^2}$$

$$= \sqrt{(100-100)^2 + (80-50)^2 + (100-100)^2}$$

= 30,000

b) Distance between 2nd data to m2 point

$$= \sqrt{\sum_{i=1}^{p} |X_{2j} - X_{1j}|^2}$$

$$= \sqrt{(70-100)^2 + (50-50)^2 + (100-100)^2}$$

= 30,000

d. Calculation of the distance until the 30th data and find the closest distance for each cluster

e. Calculate the value of WCV (Within Cluster Variation) by ranking the closest distance to the cluster and adding up each value of WCV

WCV = 30.0002 + 0.0002 + 10.0002 + 30.0002 + 20.0002 + 0.0002 + 20.0002 + 28.2842 + 20.0002 + 0.0002 + 10.0002 + 0.0002 + 0.0002 + 0.0002 + 0.0002 + 28.2842 + 0.0002 + 28.2842 + 20.0002 + 53.8522 + 0.0002 + 0.0002 + 20.0002 + 20.0002 + 20.0002 + 20.0002 + 36.0562 + 20.0002 + 20.0002 + 0.0002

WCV = 12600

f. Calculate the value of BCV (Between Cluster Variation) by adding the results of the distance between each centroid.

a) d(m1,m2) = √((m1-m2)^2)
=√((100-70)^2+(80-50)^2+(100-100)^2)
= 42.426

b) d(m1,m3) = √((m1-m3)^2)
= √((100-50)^2+(80-50)^2+(100-10)^2)
= 107.238

c) d(m2,m3) = √((m2-m3)^2)
=√((70-50)^2+(50-50)^2+(100-10)^2)
= 92.195

Nilai BCV = d(m1,m2) + d(m1,m3) + d(m2,m3)
= 42.426 + 107.238 + 92.195
= 241.860

g. Calculating the value of the ratio by comparing the values of BCV and WCV.

h. Rasio = BCV/WCV = 241.860/12600 = 0.019

Because this step is iteration 1, continue to the next step until the iteration results do not change. From the results of the clustering above, it can be concluded that:

a) The characteristics of the customers in cluster 1 are the average on time payment, the types of packages are expensive and include loyal customers who have been subscribed for more than 6 years with cluster members 5.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

b) Characteristics of customers in cluster 2 are the average payment on time, the types of packages are expensive and include loyal customers who have been subscribed for 5 to 6 years with cluster members 18.
c) Characteristics of customers in cluster 3 on average are not punctual in wifi payments, the types of packages are cheap and include customers who are not loyal to the number of clusters 7.

4.2 Implementasi

In order to realize a system that has been designed in an information system, other supporting facilities are needed in principle, commonly referred to as technical aspects, namely:

a. Hardware
b. Software
c. Brainware

1) Login Form

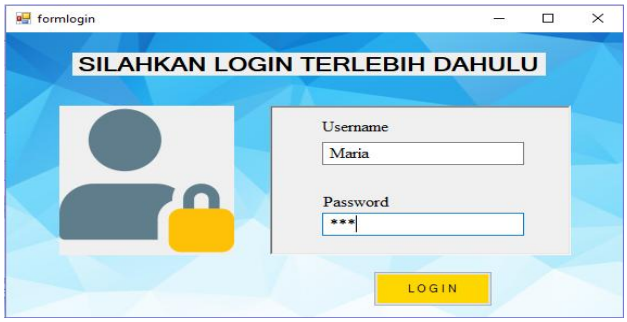

Fig 1. Login Form

2) Main Form

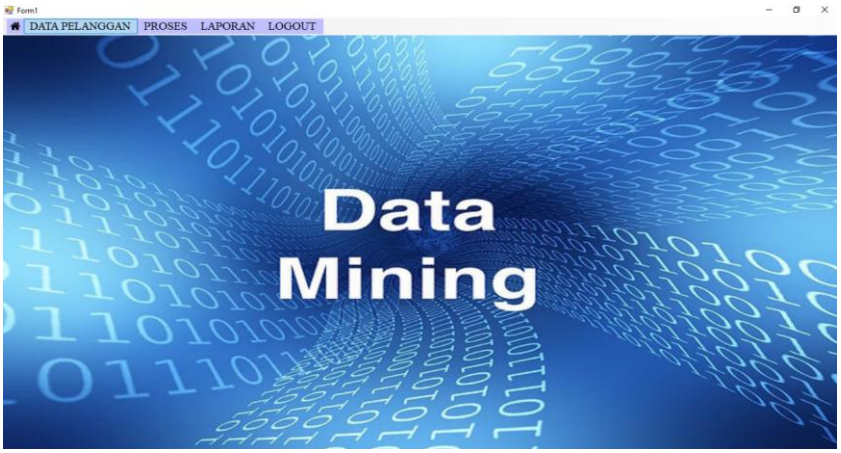

Fig 2. Main Form

3) Customer Input Form

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

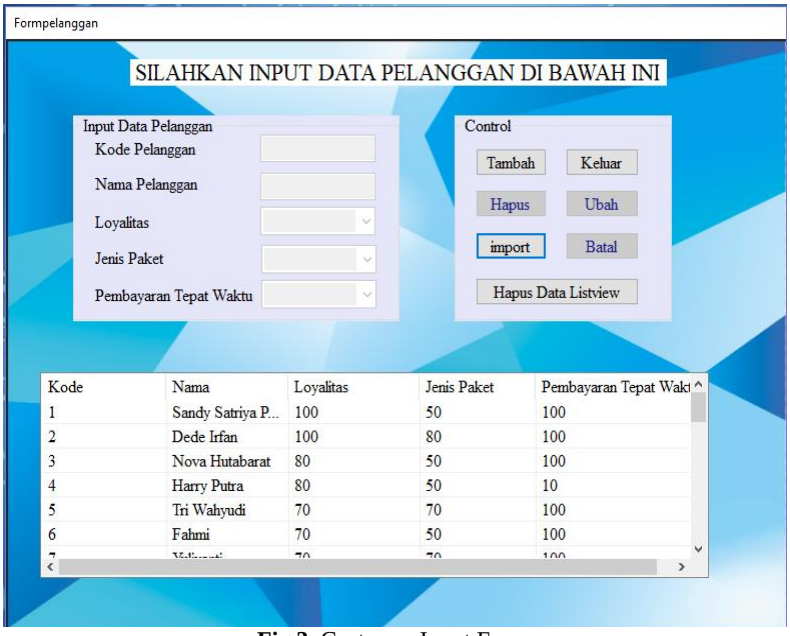

Fig 3. Customer Input Form

4) Centroid Determination Form

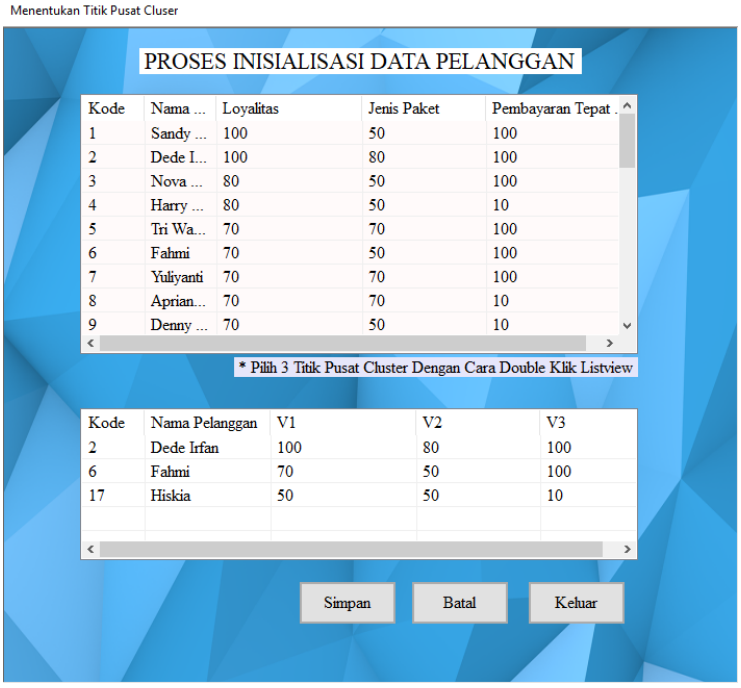

Fig 4. Centroid Determination Form

5) Clustering Process Form

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

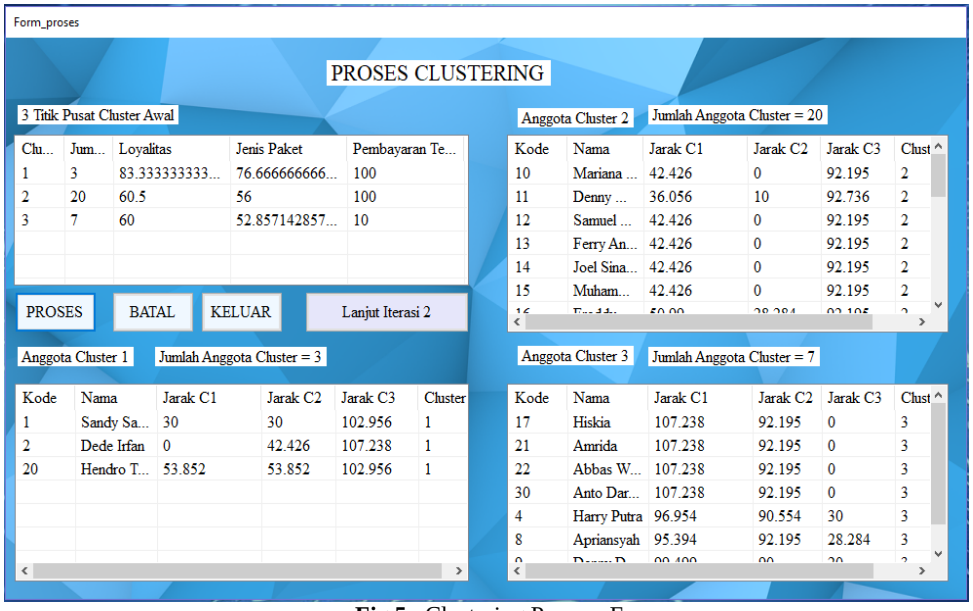

Fig 5. Clustering Process Form

6) Report of K-Means Clustering Results Form

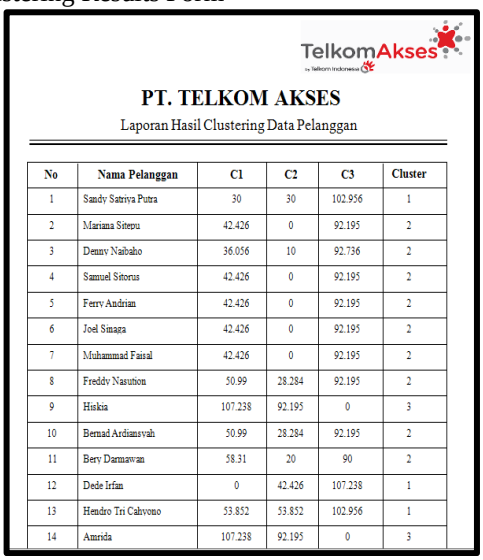

Fig 6. Report of K-Means Clustering Results Form

5. Conclusion

Based on the results of this study it can be concluded:

a. The system that has been built can classify customer data by applying the K-Means method.
b. Can apply the calculation results with the K-Means method in grouping customer data at PT.Telkom Akses.
c. Can implement the K-Means method in a system which will later be used to classify customer data.

Thank you for :

a. STMIK Triguna Dharma Medan
b. DRPM and the Ministry of Education and Culture of the Republic of Indonesia who have supported the implementation of this research.
c. PT.Telkom Medan City.
d. Researcher's family and friends

6. References

[1] G. Abdillah *et al.*, "PENERAPAN DATA MINING PEMAKAIAN AIR PELANGGAN UNTUK MENENTUKAN

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

KLASIFIKASI POTENSI PEMAKAIAN AIR PELANGGAN BARU DI PDAM TIRTA RAHARJA MENGGUNAKAN ALGORITMA K-MEANS," *SENTIKA 2016*, vol. 2016, no. Sentika, pp. 18–19, 2016.

[2] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.

[3] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.

[4] Y. Siyamto, "Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor Di PT Winstar Batam," *Media Inform. Budidarma*, vol. 1, no. 2, pp. 28–31, 2017.

[5] S. Rony, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta)," *J. Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2016.

Jurnal Mantik is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).