Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15 th October 202 5 . Vol. 5 . No. 1; e-ISSN: 2808-4519

Clustering of PLN ULP Binjai Timur Customer Complaints using the K-Means Method

Iyut Sahira [1] *, Akim MH Pardede [2] , Muammar Khadapi [3]

[123] *STMIK Kaputama,Indonesia*

iyuttsahiira@gmail.com[1*],Akimmhp@gmail.com [2],khdafi5@gmail.com [3]

Abstract

The large number of customer complaints received daily by PLN ULP Binjai Timur presents a challenge in providing responsive and accurate service. Irregularities in recording and grouping complaints mean that the available information is less than optimal for supporting decision-making. This study aims to group customer complaints based on similar characteristics for easier analysis. The method used is the K-Means algorithm, a clustering technique in data mining that divides data into several groups based on their proximity to the cluster center (centroid). The analysis was conducted through the Knowledge Discovery in Database (KDD) stages, which include data selection, transformation, and algorithm implementation using MATLAB software. The three main variables used in the grouping process were complaint type, complaint submission medium, and customer address. The implementation results in three main complaint clusters with distinct patterns, providing PLN with insight into the most frequently encountered problems, areas with high complaint rates, and the most frequently used reporting medium. These findings provide an important foundation for PLN in setting treatment priorities, improving service quality, and strengthening customer relationships. The application of the K-Means algorithm has proven effective as a systematic and practical solution for managing complex and large amounts of complaint data.

Keywords *: Clustering, Customer Complaints, K-Means, MATLAB*

1. Introduction

PT. PLN (Persero) is a business entity state -owned enterprises (BUMN) in the field of electricity that serves public throughout archipelago, along with development technology and information that is so rapid , energy electricity become matter basic needs of humans , activities humans in today's era No escape from energy electricity . PT PLN (Persero) is a state-owned company that provides electricity. service to Indonesian society in provision related services with power electricity . Increase need electricity in Indonesia is increasing with high and fast . PT PLN (Persero) can it is said Still often experience obstacles in his efforts . As provider service public , PLN is sued For give optimal service so that it can fulfil needs and satisfaction customers . However , in in reality , often appear various complaint from customer related quality service (Hartawan et al., 2024) .

The State Electricity Company (PLN) of East Binjai is also one of the service units tasked with provide electricity to public around . As provider service PLN public is sued For give optimal service so that it can fulfil needs and satisfaction customers . However , in the reality often appear various complaint from customer related quality service like face problem Serious moment serve customer kWh meter problems , disturbances electricity , and power installation complaints .

the amount of complaint data received every day can become challenge alone in the management and analysis process . If not managed with OK , complaint data This only will become pile difficult information utilized For taking decision . Therefore that , is necessary something capable method grouping complaint customer to in a number of category or cluster that has characteristics similar . Grouping This will make it easier for PLN to understand pattern complaints , so that can plan a more comprehensive handling strategy effective and efficient management complaint effective customer service is essential For increase satisfaction customers and image company . With understand patterns and characteristics frequent complaints appears , PLN can design a better strategy appropriate target in handle problems faced customer .

K-Means Method *Clustering* is one of the technique in *Data Mining* that can used For grouping data based on characteristics certain . In the context of complaint customer , method This can assist PLN East Binjai in identify group main from complaint customers . With existence grouping this , PLN can more focus in designing a resolution strategy problem in accordance priority .

2. Literatur Riview

2.1. Complaint customer

Complaints are a word that is often connotative *negative* for second parties , both for company and for Consumers . *Complaints* in general perceived as mistakes , problems , stress , frustration , anger , conflict , punishment , demands , change loss , and the like . *Complaint* is part from the post -process buy . The process started from consumer feel dissatisfaction after accept service or do transactions.
Every organization or company Certain will face complaint from customers . Along with progress technology and convenience *internet access* , customers now more easy convey experience they , including criticism to product or the services they provide accept . Although complaint can cause impression *negative* , thing the No always means end from connection between customers and companies . Instead , complaints Can become opportunity For do repair as well as increase quality products and services , which ultimately can strengthen loyalty customers . Therefore that , every organization generally have a special strategy in manage as well as maintain connection with Customers . (National & Serving, 2024)

Complaint Customer is form aspirations customers that occur Because dissatisfaction to something goods or services . Various complaint customers basically divided into 2 , namely complaints submitted in a way verbal and complaints submitted in a way written (Handayani et al., 2021) .

Impact Complaint customers who do not managed can optimally cause various consequence negative for companies , especially for agency service public such as PT PLN (Persero). One of them the consequences is decline level satisfaction as well as loyalty customers **,** who can have a bad impact image company in the eyes society . If complaint No quick handled , thing This Can spread wide through social media or communication between customers **,** so that make things worse reputation company . In addition , if PLN does not own ability manage and analyze complaint data with okay then will difficult For identify common problems occur and formulate the right solution . Failure in handle complaints can also be slow down the repair process service , causing recurrence the same problem , and in the end hinder PLN's efforts in reach efficiency and professionalism in service .

2.2. Data Mining

Data mining is the process of excavation patterns and information useful from very large data sets . This process consists of from data collection , data extraction , data analysis , and data statistics . In addition , it is also called as invention knowledge , search knowledge , analysis pattern or data, and collection information (Arhami and Nasir, 2020).
Data Mining according to *Gartner Group* is the process of finding connection new ones that have meaning, patterns and habits with sorting out part large amount of data stored in storage media with use technology introduction pattern like mathematics and statistics . *Data Mining* is also a combination from a number of discipline unifying science technique from learning machine , introduction patterns , statistics , databases, and visualizations For handling problem . *Data Mining* present due to condition data overload experienced by various institutions , companies or organization . This abundance of data is accumulation of recorded transaction data months even until years .

2.3. Clustering

Clustering Method is the process of dividing a group data object to in set a part called a cluster. *Cluster* own different objects and have the same characteristics . Through use algorithm *Clustering* , participation done in a way automatically . As a result , *Clustering* is very useful Because can find group or groups that do not known in the data (Annisa et al., 2022) .

Clustering is method data analysis whose purpose is grouping data with the same characteristics to the same area . One of them the approach used in develop clustering method , namely method *K-Means* , where method This is one of the method *nonhierarchical* data grouping (barrier) that tries partitioning data into in form two or more characteristic clusters The same entered to in One the same group (Hasibuan et al., 2022) .

As for the benefits method *Clustering* can seen as following :
1. *Clustering* is method very useful data segmentation For prediction and analysis problem business certain , such as mapping regional zoning , market segmentation , and marketing.
2. Identification object in field various field like *computer vision* and *image processing* .

2.4. Algorithm *K-Means*

Algorithm *K-Means* is algorithm *non- hierarchical* origin from method *Clustering* . This method share data to in different groups so that the data with the same characteristics entered to in the same group and data with different characteristics entered to in different groups (Alhamad & Hasan, 2023) .

Algorithm *K-Means is* also famous Because convenience and capabilities For combine large data sets and *outliers* very quickly . Algorithm *K-Means* share collection of n objects to in k clusters with k input parameters , so that level similarity between member in One high *cluster* whereas level similarity between members in other *clusters are very* low . Proximity object to *cluster* mean value (also known as as *centroid cluster*) shows how much similar member with clusters.

One of objective from this data grouping is For reduce function set objectives in the grouping process , which is usually aim For minimize variation in something groups and more Lots variation between group .

3. Research methodology

Methodology study is a series procedure or steps scientifically structured in a way systematic For obtain appropriate data and support the research process . In the implementation study this , the author follow stages designed methodology in a way sequential and structured to produce recognized research the truth in a way scientific .

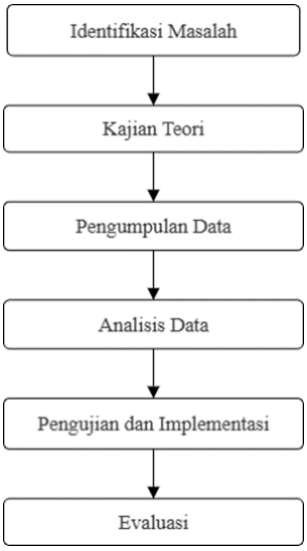

Fig. 1: Research Methodology Flow

3.1. Research Design

This research apply method quantitative with approach descriptive analytical . The goal is For grouping complaint data PLN ULP East Binjai customers based on location address , origin complaints , as well as type complaints . For the process of grouping data based on similarities characteristics , used K-Means Clustering algorithm identify pattern significant relationship .

3.2. Input Data

Input data used in system This originate from complaint PLN ULP East Binjai customers . The data saved using Microsoft Office Excel as storage media . Each data entry later changed to in numeric format so that it can processed use method clustering . Transformation process done with set mark specific to each category in variables used , such as explained following This :

1. Input Data

File Name : data_keluhan.xlsx
Amount Data: 519 data
Variable : X = Customer Address
Y = Source Complaint
Z = Type of Complaint

2. Number of Grouping Clusters
   Determined as many as 3 clusters
3. Data Transformation Values on Variables

Table 1: Data Transformation Values on Variables

No	Variables	Transformation	Transformation Values
1	Customer Address	Like to Progress	1
		Sei Semayang	2
		Lalang Village	3
		Right Sunggal	4
		Krio Field	5
		Geli Bay	6
		Purwodadi	7
		Sei Mencirim	8
		Cape Gusta	9
		Muliorejo	10
2	Source Complaint	Call PLN 123	1
		PLN Mobile	2
		Facebook	3
		PLN Mobile Live Chat	4
		Twitter	5
		Instagram	6
		E-mail	7
3	Type of Complaint	New Installation	1
		Power Change	2

	Connection Temporary	3
	Condition Network	4
	Electricity bills	5
	Token	6
	Prepaid kWh Meter Problems	7
	Postpaid kWh Meter Problems	8
	kWh Meter Shows "CHECK"	9
	Disturbance Installation Customer	10
	Power outage	11

4. Implementation

Implementation is stage implementation and testing system new , and become the moment when the application executed in condition real. At the stage this , implementation done as step end from implementation Clustering method for grouping complaint PLN ULP East Binjai customers based on the design that has been arranged previously.Process implementation done with utilise device MATLAB software , in which the Clustering method is applied use K-Means algorithm in program code . With method this , system capable do complaint data grouping customer in accordance with stages that have been designed. Based on results the design that has been made , stage implementation succeed apply Clustering method with K-Means algorithm using MATLAB software for grouping complaint PLN ULP East Binjai customers in general effective and optimal.

Results

a. Results with 3 clusters

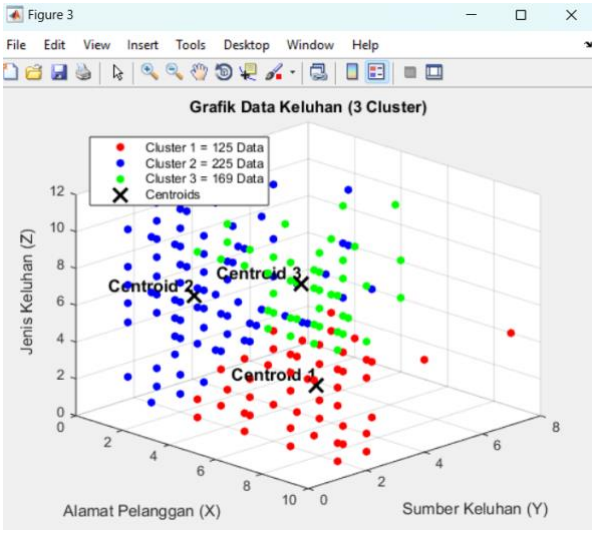

Fig. 2: Chart results of the grouping process

Output from testing system This is the results of the data obtained through the clustering process with using complaint data PLN ULP Binjai Timur customers who have processed in device MATLAB 2014b software . The graph displayed show results grouping based on level similarities complaints , with objective For help PLN in identify patterns frequent complaints appear so that can increase quality appropriate response and strategy target in service to customer .

Table 2: Data Transformation Values on Variables

No	Cluster Center	Variables			Amount of Data
		X	Y	Z	
1	Centroid 1	7.9840	1.8960	3.8960	125
2	Centroid 2	2.2844	2.2311	6.3156	225
3	Centroid 3	7.4201	1.8284	9.2189	169

Based on the table results classification that has been served above , information the obtained through the testing process with using complaint data available customers . The data has analyzed and grouped use device MATLAB software , so that the results displayed indicates that :

1. In cluster 1 with the number of data 125 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (7.9840) namely “Sei Mencirim ”, the most frequent customer report complaint they from source complaint as variable (Y) with value (1.8960) namely "PLN Mobile", and Type of complaint as variable (Z) with value (3.8960) is the most frequent complained about " is Condition Network ".
2. In cluster 2 with the number of data 225 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (2.2844) namely “Sei Semayang ”, the most frequent customer report complaint they from source complaint as variable (Y) with value (2.2311) which is “PLN Mobile”. And the type complaint as variable (Z) with value (6.3156) is the most frequent complained about is “ Token Problem ”.

3. In cluster 3 with the number of data 169 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (7.4201) namely " Purwodadi ", the most frequent customer report complaint they from source complaint as variable (Y) with value (1.8284) namely PLN Mobile, and type complaint as variable (Z) with value (9.2189) is the most frequent complained about is kWh Meter problem shows “CHECK”.

b. Results with 4 clusters

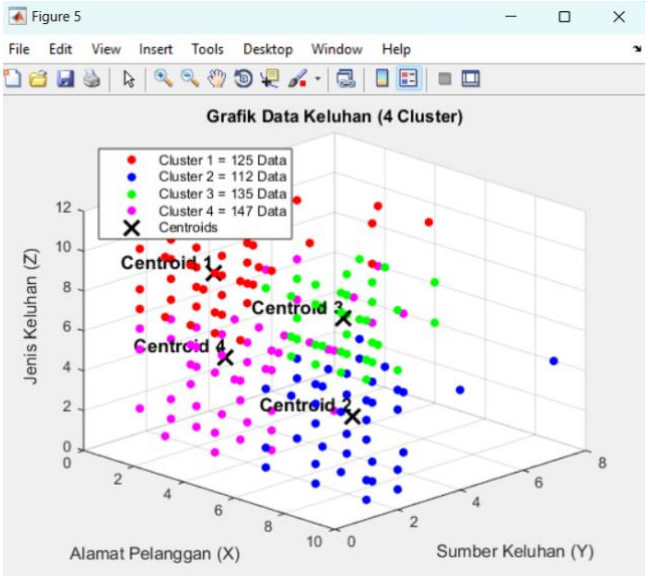

Fig. 3: Clustering results with 4 clusters

Output from testing system This is the results of the data obtained through the clustering process with using complaint data PLN ULP Binjai Timur customers who have processed in device MATLAB 2014b software . The graph displayed show results grouping based on level similarities complaints , with objective For help PLN in identify patterns frequent complaints appear so that can increase quality appropriate response and strategy target in service to customer .

Table 3: Data Cluster Center with 4 Centroids

No	Cluster Center	Variables			Amount of Data
		X	Y	Z	
1	Centroid 1	2,8000	1,9040	9.0480	125
2	Centroid 2	8,3571	1.8839	4.1071	112
3	Centroid 3	8.1704	1,7407	9,0074	135
4	Centroid 4	2.5646	2.4769	4.4830	147

Based on the table results classification that has been served above , information the obtained through the testing process with using complaint data available customers . The data has analyzed and grouped use device MATLAB software , so that the results displayed indicates that :

1. In cluster 1 with the number of data 125 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (2.8000) namely "Kampung Lalang" , " the most frequent customer report complaint they from source complaint as variable (Y) with value (1.9040) which is “PLN Mobile”, and the type complaint as variable (Z) with value (9.0480) namely problem “kWh Meter Shows “CHECK”.
2. In cluster 2 with the number of data 125 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (8.3571) namely “Sei Mencirim ”, the most frequent customer report complaint they from source complaint as variable (Y) with value (1.8839) namely "PLN Mobile", and the type complaint as variable (Z) with value (9.0074) is the most frequent complained about is The Problem of “ Conditions Network ".
3. In cluster 3 with the number of data 135 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (8.1704) namely “Sei Mencirim ”, the most frequent customer report complaint they from source complaint as variable (Y) with value (1.7407) which is “PLN Mobile”, and the type complaint as variable (Z) with value (9.0074) is the most frequent complained about is problem “kWh Meter Shows “CHECK”.
4. In cluster 4 with the number of data 112 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (2.5646) namely “Sei Semayang ”, the most frequent customer report complaint they from source complaint as variable (Y) with value (2.4769) which is “PLN Mobile”, and the type complaint as variable (Z) with value (4.4830) is the most frequent complained about is “ Condition Network ".

c. Results with 5 clusters

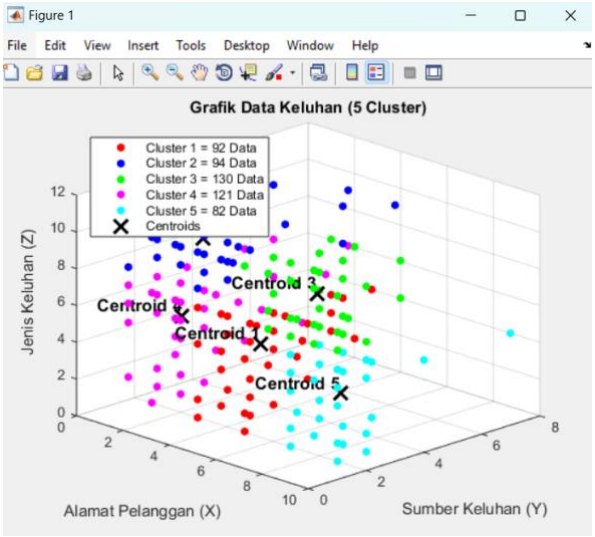

Fig. 4: Clustering results with 5 clusters

Output from testing system This is the results of the data obtained through the clustering process with using complaint data PLN ULP Binjai Timur customers who have processed in device MATLAB 2014b software . The graph displayed show results grouping based on level similarities complaints , with objective For help PLN in identify patterns frequent complaints appear so that can increase quality appropriate response and strategy target in service to customer .

Table 4: Data Cluster Center with 5 Centroids

No	Cluster Center	Variables			Amount of Data
		X	Y	Z	
1	Centroid 1	5.1957	2.1957	4.8913	92
2	Centroid 2	2,8191	2.0957	9.7234	94
3	Centroid 3	8.2538	1,7000	9.0846	130
4	Centroid 4	1.6612	2.3140	4.9339	121
5	Centroid 5	9.1341	1,8049	3.9512	82

Based on the table results classification that has been served above , information the obtained through the testing process with using complaint data available customers . The data has analyzed and grouped use device MATLAB software , so that the results displayed indicates that :

1. In cluster 1 with the number of data 116 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (5.1959) namely "Medan Krio", the most frequent customer report complaint they from source complaint as variable (Y) with value (2.1957) namely "PLN Mobile", and the type complaint as variable (Z) with value (4.8913) is the most frequent complained about is " Bill **electricity "** .
2. In cluster 2 with the number of data 105 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (2.8191) namely "Kampung Lalang", the most frequent customer report complaint they from source complaint as variable (Y) with value (2.0957) namely "PLN Mobile", and type complaint as variable (Z) with value (9.7234) is the most frequent complained about is " Disorder" installation customer ".
3. In cluster 3 with the number of data 105 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (8.2538) namely " Sei Mencirim " , the most frequent customer report complaint they from source complaint as variable (Y) with value (1.7000) which is "PLN Mobile", and the type complaint as variable (Z) with value (9.0846) is the most frequent complained about is Problem "kWh Meter Shows "CHECK".
4. In cluster 4 with the number of data 87 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (1.6612) namely "Sei Semayang ", the most frequent customer report complaint they from source complaint as variable (Y) with value (2.3140) namely " PLN Mobile" , and the type complaint as variable (Z) with value (4.9339) is the most frequent complained about is " Electricity Bill **"** .
5. In cluster 5 with the number of data 105 is dominated by complaints customers in need improvement priority service especially at the Address as variable (X) with value (9.1341) namely "Tanjung Gusta", the most frequent customer report complaint they from source complaint as variable (Y) with value (1.8049) which is "PLN Mobile", and the type as variable (Z) with value (3.9512) namely most frequent complaints complained about is " Condition Network ".

5. Conclusions and Suggestion

5.1. Conclusions

As part closing from study this is interesting a number of conclusions based on the application K-Means Clustering method . This method used For grouping complaint data PLN ULP East Binjai customers based on address , source complaints , as well as type complaints . From the results analysis and data grouping process, obtained conclusion as following :

1. This research succeed classify complaint customer to in a number of segment based on similarities pattern emergence complaints . Through implementation K-Means Clustering method , complaints the divided become a number of group with characteristics similar problem , so make things easier in set scale priority handling .
2. From the results implementation Clustering method with The K-Means algorithm uses 3 clusters with a total of 519 grouped data , which has been identified complaint data group as following :
    a. The results of the clustering process with MATLAB R2014b show data in cluster 1 with total data 125 centers centoid (7.9840; 1.8960; 3.8960). dominated by complaints customers in need improvement priority service especially in the Sei Mencirim area , with Most frequent customers report complaint they via PLN Mobile, and the most frequent types of complaints complained about is Condition Network .
    b. The results of the clustering process with MATLAB R2014b show data in cluster 2 with total data 225 centers centoid (2.2844; 2.2311; 6.3156). dominated by complaints customers in need improvement priority service specifically from the Sei Semayang area , with most frequent customers report complaint they via PLN Mobile. And the type most frequent complaints complained about is Token Issues .
    c. The results of the clustering process with MATLAB R2014b show data in cluster 3 with total data 169 centers centoid (7.4201; 1.8284; 9.2189). dominated by complaints customers in need improvement priority service specifically from the Purwodadi area , with most frequent customers report complaint they via PLN Mobile, and types most frequent complaints complained about is kWh Meter problem shows “CHECK”.

3. Tests conducted with 5 clusters resulting in conclusion as following :
    a. Cluster 1 has mark variance of 5.2377, which shows that the data in this cluster own level variation moderate . With mark *Vmin* of 0.0884 and *Vmax* of 24.5228 shows range sufficient data distribution wide , with real difference between the data in it
    b. Cluster 2 has mark variance amounting to 4.7965, which is a little more low from Cluster 1 but still show existence variation in the data. With mark *Vmin* of 0.1183 and *Vmax* of 28.8884 shows sufficient data distribution wide , with a number of data points have level high variation .
    c. Cluster 3 has mark variance of 3.8155, which is mark lowest compared to other clusters . This shows that the data in this cluster tend more uniform . With mark *Vmin* of 0.1615 and *Vmax* this is 11.2075 show narrow spread of data , with difference between relative data small .
    d. Cluster 4 has mark variance highest of 5.9392, which indicates that the data in it is the most diverse . With mark *Vmin* 0.2177 and *Vmax* 31.0035 shows wide distribution of data , with a number of data points have level high variation .
    e. Cluster 5 has mark variance of 5.2952, which shows level sufficient data variation high . With mark *Vmin* 1.1560 and *Vmax* 28.2952 shows wide distribution of data , although A little more narrow compared to Cluster 4.
4. Test results against cluster data shows that existence variation in each cluster provides clear understanding about distribution of complaint data customers in it . Testing with 3 to 5 clusters producing variations that reflect level accuracy tall as well as effectiveness in the grouping process . Complaint data succeed grouped based on relevant characteristics . This proves that clustering method used capable recognize and group complaint data optimally , producing information that can trusted For analysis advanced as well as support determination priority handling and improving service complaint customers at PLN ULP East Binjai .
5. Grouping data in study This produce segmentation that can utilized For understand difference between group complaint customer information This give outlook important related variation type complaint in each cluster, so that PLN can develop a more effective handling strategy right and appropriate with need specific to each group . These results show that although part big complaint can completed through procedures that have been there is , but Still there is a number of issues that require attention special and handling priority based on characteristics from the formed cluster .

With Thus , the clustering results are not only beneficial For analysis descriptive , but also becomes base strategic in taking decision priority handler complaints and improvements services at PLN East Binjai .

5.2. Suggestion

Based on results and findings in study Here are some suggestions that can be given For study continuation and development system to front and author give some suggestions between other :

1. For study Next , it is recommended that the data used covers more Lots period report complaint as well as covers a wider work area area below under the auspices of PLN ULP East Binjai . This will give a clearer picture complete in grouping pattern complaint customer , help in determine handling strategies priorities and improve quality service complaint customer .
2. Focus study next should directed at analysis trends and evolution pattern complaint customer in period time certain . Through observation said , PLN can develop handling strategies more disturbances responsive , optimize service technical , as well as strengthen satisfaction customer .
3. To in the future , it is hoped system this data grouping can developed become more applications efficient and capable used by more Lots user .

References

[1] IKS Hartawan, JA Bisnis, P. Studi, A. Bisnis, and PN Bali, "IMPLEMENTATION OF CUSTOMER COMPLAINT SERVICES AT PT PLN (PERSERO) ULP IMPLEMENTATION OF CUSTOMER COMPLAINT SERVICES AT PT PLN (PERSERO) ULP," 2024.
[2] S. Nasional and D. Mengabdi, "SENADA:," vol. 5, no. 1, pp. 41–48, 2024.
[3] YI Handayani, V. Rahmawati, M. Junaedi, and MA Erwita, "Service Excellence for Small Business Owners in Surabaya," *Peka J. Pengabdi. Kpd. Masy.* , vol. 4, no. 2, pp. 64–72, 2021, doi: 10.33508/peka.v4i2.3554.
[4] K. Annisa, BS Ginting, and MA Syar, "Application of Data Mining to Group Clean Water User Data Based on Complaints Using the Clustering Method at Pdam Langkat," *J. Sist. Inf. Kaputama* , vol. 6, no. 2, pp. 165–179, 2022, doi: 10.59697/jsik.v6i2.167.
[5] BL Hasibuan, Sofiah, and E. Yolanda, "Classification of Urine Test Patient Data Using the Clustering Method at the North Sumatra Province National Narcotics Agency (BNNP SUMUT) Office," *JUKI J. Comput. and Inform.* , vol. 4, no. 2, pp. 183–193, 2022.
[6] A. Alhamad and M. Hasan, "K-Means Method for Determining Priority of Customers Receiving Promotions," *J. Minfo Polgan* , vol. 12, no. 1, pp. 804–811, 2023, doi: 10.33395/jmp.v12i1.12502.