Research Article   **Open Access (CC–BY-SA)**

Clustering the potential bandwidth upgrade of FTTH broadband subscribers

Sasa Ani Arnomo [a, 1, *]; Yulia Yulia [b, 2]

[a] *Universitas Putera Batam, Jl. Soeprapto, Batam, Kode Pos 29422, Indonesia*
[b] *STIE Nagoya Indonesia, Batu Aji, Batam, Kode Pos 29424, Indonesia*
[1] *sasa@puterabatam.ac.id;* [2] *yuliaedwar2407@gmail.com*
* Corresponding author

Abstract

A company needs to consider determining the customers' potential before deciding to upgrade their bandwidth. It is important because, previously, determination was conducted randomly. Therefore, potential determination is necessary by grouping customers who have similar characteristics based on their data and attributes. This study employs data mining techniques using clustering method with K-means algorithm on broadband users' group of 263 FTTH. The determination was determined based on end centroid point in the grouping. The results were divided into 5 clusters consisting of 34 highly potential users (12.92%), 29 potential users (11.02%), 56 fairly potential users (21.3%), 54 less potential users (20.53%), and the remaining 90 not potential users (34.22%). The comparison of the validity of the Davies-Bouldin Index for the 5 (five) clusters is 0.538 for K-Means and 0.819 for K-Medois. This indicates that K-Means results better score. This method is useful for efficient bandwidth sharing.

Keywords: Potential Upgrades; Bandwidth; Broadband; FTTH; Cluster

Introduction

Bandwidth upgrade is taken to adjust customers' need of internet so that access between users could be stable. A wavelength-division multiplexing system has also been developed with multiple channels for fiber-to-the-home services [1]. Basically, the need for bandwidth represents connection capacity, the higher bandwidth needs the higher the performance [2]. Increasing the bandwidth will require high-capacity optical fiber [3]. Internet service users to observe in this study were taken from one internet service provider. This provider has more than 1500 active customers divided into several categories of service package types such as Dedicated, Business Broadband and Broadband FTTH (Fiber to The Home). Dedicated and business broadband services were dominated by business customers such as hotel, middle and big business sector where the setting and optimization of bandwidth are handled by IT staff or network administrator. This is different to broadband FTTH users where there is no IT staff or network administrator to manage internet network. This research emphasizes on Broadband FTTH which is a high-speed internet access distribution service through the central point of the provider to housing that uses Fiber Optic as a transmission medium. Without an IT employee or network administrator who manages the internet network, it is possible that disturbances will occur on the local network, for instance, full bandwidth that can result in unbalanced internet network usage between users, where one user has fast connection while the other might be slow or even cannot access the internet. This is the common complaints by FTTH broadband users. However, there is no proper solution for this issue by the providers.

Before upgrading the bandwidth for the customers, companies need to consider investment costs and reduce additional cost charged to customers. As a Tier 3 ISP company, some companies require Bandwidth rental to a Tier 2 ISP. If the bandwidth upgrade is done directly, the previously rented Bandwidth will automatically be reduced, and additional rental is required. Customers can possibly bear the cost of renting the bandwidth, but this will burden customers because there will be monthly additional costs. Therefore, it is necessary to select customers for Bandwidth Upgrades based on the specified customer potentials. To identify customers' attractiveness, their response against the presence of fiber optic fibers in their area could be used as reference.

Internet Service Provider (ISP) provider is a company or business organization that provides internet access and services related to the internet to individual consumers and companies [4]. This is potential because the increasing

number of smartphone users. Among the services offered by ISP is Broadband. It is an internet technology using data communication network that allows sending and receiving data at high speeds and in large quantities such as video data, images, text, and others [6]. One of the transmission media used by ISPs is FTTH (Fiber To The Home), a fiber optic network that provides direct access to home users [7].

Bandwidth upgrades can be triggered by several things such as requests from customer as well as full bandwidth. Full bandwidth is needed because the more devices accessing internet the greater bandwidth required [8]. In fact, whatever bandwidth used by user will be enough only for few devices. This is due to absence of restriction or bandwidth setting s for each user. One device solely may consume great bandwidth to download or upload large files. On the other hand, the increasing number of users in several agencies is not followed bandwidth upgrade [9]. Before upgrading the bandwidth, providers will surely make some considerations on the customer based on their potential.

Potential is ability that can be developed [10]. After obtaining the data of potential customers, the company may continue the decision regarding bandwidth upgrade to support users in using internet comfortably which in turn will maintain their loyalty. Customers' loyalty will benefit company in long term. Satisfied customer will remain loyal and distribute their positive experience to others.

There are several studies on internet service providers, for instance, the Quality Function Deployment method with Fuzzy TOPSIS which is used to select internet services [4]. Bandwidth management optimization has also been carried out using the queue tree method [12] or MikroTik Routers [13] to be accurate. In addition, bandwidth management could use the VyOS operating system through the VM Virtualbox application system as virtual routing [14]. Efforts to equalize bandwidth are also conducted using hierarchical token bucket method which is used for QoS optimization [15]. Even though bandwidth management is properly done, complaint from customer often occur.

The obstacle found in determining the potential customer for bandwidth management is that customer data owned by the company does not have label or target class such as number of complaints received every day. For example, if the number of complaints received is 10, company cannot classify it as high or low. This is because the number of complaints will increase or decrease in the following days. Similarly, data for other customers cannot be classified based on needs. Therefore, potential determination is necessary by identifying customers with similar characteristics to be subsequently grouped together [16].

Method

The sample data selected was based on saturated sampling technique which is the data of FTTH Broadband customer starting from initial activation until December 31, 2018. This study employed data mining as analysis tool. Before undertaking data mining step, there were several Knowledge Discovery in Database (KDD) steps that had been carried out. KDD is an organized process of identifying valid, novel, useful and understandable patterns from very large and complex data sets [17]. KDD process consists of several steps such as data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge representation [18]. This study carried out steps below.

A. Data Cleaning

Data source of this study is FTTH Broadband subscribers of PT. Solnet Indonesia, starting from initial activation until December 31, 2018. Among 274 registered customers, there were 9 duplicated data and 2 incomplete data. After cleaning the data, the remaining data to be used were 263 users.

B. Data Integration

The data to combine data is the data of FTTH broadband subscribers and data of daily complaint reports of Broadband. These data will be used in the next step.

C. Data Selection

After going through the data selection process, 6 of the 17 types of data attributes were selected to be used in this study. The six attributes are the number of complaints, the subscription package, the subscription contract, the length of the subscription, the number of upgrades during the subscription, and the payment of the initial installation fee.

D. Data Transformation

At this stage, the data transformation process will be carried out, so that the data can be processed using the K-means method. Non-numeric data will be initiated into numeric form. The initial installation payment attribute was changed from "YES" to 1 and "NO" to 2.

E. Data Mining

Data mining has been attracting a significant attention. This might due to popularity of big data concept [19]. Big data refers to large, complex, and growing data set with many independent sources [20]. Data mining is categorized into 6 modes, namely classification, estimation, affinity grouping, sequence pattern, grouping and description [21]. Clustering is part of data mining which is unsupervised. The main purpose of the unsupervised learning method is to

explore the data and some hidden structures among the data [22]. Clustering is the task to separate data or vector into several groups or clusters based on their respective characteristics. Data with similar characteristics are grouped in one cluster, while data with different characteristics will be grouped into different group or cluster. The main objective of clustering method is to group data or objects in such a way that each cluster will contain data that is as similar as possible [23]. K-Means algorithm is a well-known partition method for clustering [24]. It is a non-hierarchical approach of grouping items into various number of clusters [25]. K-Means is one of the data mining partitioning clustering where each data must be included in a certain cluster which allows each data in that cluster to pass through a stage of process, which then at the next stage moves to another cluster. K-Means separates the data into the nearest K area and then classify large data and outliers quickly [26]. However, K-Means are generally very sensitive to the initial centroid. Consequently, it can produce poor quality clustering results if the initial centroid is poor [27].

In general, the basic algorithm of K-Means clustering is as follows:
1. Determine the number of clusters
2. Allocate the data into clusters randomly
3. Calculate the centroid /data average in each cluster
4. Allocate each data to the nearest centroid/average
5. Return to step 3, if there is data that moves to different cluster or if the centroid value changes that could be above the specified threshold value or change in the value of the objective function used which is above the specified threshold value.

Distance space is used to calculate the distance between the data and the centroid. One of the equations that can be used is Euclidean Distance Space as it gives results of shortest distance between the two calculated points [28]. The formula is shown in equation (1).

$$d_{ij} = \sqrt{\sum_{k=1}^{p}\{x_{ik} - x_{jk}\}^2} \tag{1}$$

d_{ij}	= The distance between object i and j
P	= Data dimension
X_{ik}	= The coordinate of object i in dimension k
X_{jk}	= The coordinate of object j in dimension k

F. Pattern Evaluation

In evaluating the results of the K-Means algorithm in the pattern of determining the number of clusters, the Davies Bouldin Index (DBI) method is used. The first step to calculating DBI is to find the Sum of Square Within (SSW) value using the equation (2). The ratio value can be calculated after obtaining the SSW and SSB values. The DBI value is the average of the highest ratio values. The smaller the DBI value, the better the spread of the data in each cluster

$$SSW_I = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_i) \tag{2}$$

From the equation (2), *mi* is the number of data in the cluster *i*; *ci* is the centroid of cluster *i*; and d() is the distance of each data to the centroid calculated using Euclidean distance. *Sum of Square Between Cluster* (SSB) is an equation used to determine the separation between clusters which is calculated using the equation (3).

$$SSB_{i,j} = d(c_i, c_j) \tag{3}$$

After obtaining the value of cohesion and separation, ratio measurement (Rij) is carried out to determine the comparison value between cluster i and cluster j. A good cluster is a cluster with smallest possible cohesion and maximum separation. The ratio value is calculated using the equation (4).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{4}$$

The ratio value can be used to calculate the value of *Davies-Bouldin Index* (DBI) using the equation (5).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \tag{5}$$

From the equation above, k is the number of clusters used. The smaller the DBI value obtained (non-negative>=0), the better the cluster obtained from the K-Means grouping used.

G. *Knowledge Representation*

After the grouping result is obtained, the potential determination is carried out by summing up the centroids in each cluster and labelling category of: 'Not Potential', 'Less Potential', 'Fairly Potential', 'Potential', and 'Highly Potential'.

Results and Discussion

Data that is combined in this research is data on FTTH Broadband subscribers and daily complaint report. These data would be used in the next steps. The data of FTTH broadband subscribers is shown in **Table 1**.

Table 1. The data of FTTH broadband subscribers

No.	Item
1.	Name
2.	Activation date
3.	Address
4.	Subscription package
5.	Connection
6.	Subscription contract
7.	The length of subcription
8.	Number of upgrades during subscription
9.	Payment of the initial installation fee

The data of daily complaint report is shown in **Table 2**.

Table 2. Data of Daily Complaint Report

No.	Item
1.	Name
2.	Time span
3.	Number of complaints
4.	Repairment
5.	PIC
6.	Phone
7.	NOC
8.	Fixed by

The source of data used in this study is FTTH broadband subscriber data. Of the 274 registered customers, there were 9 duplicated data and 2 incomplete data. Therefore, after conducting data cleaning, the remaining data used is 263 users. Data cleaning is shown in **Table 3**.

Table 3. Data Cleaning Results

Item	No. of Customers
FTTH broadband subscribers	274
Duplicated data	(9)
Incomplete data	(2)
Number of data for research	263

At this stage, the data transformation process will be carried out, so that the data can be processed using the K-Means method. Non-numeric data will be initiated into numeric form. The data on the 'payment of initial installation fee' attribute is non-numeric data that requires initiation. **Table 4** is sample data that will be clustered.

Table 4. Research Data after Transformation

No.	Customer Name	No. of Complaint (2018)	Subscription Package (Mbps)	Subscription Contract (Day)	Length of Subscription (Day)	No. of Upgrade during Subscription	Payment of the intl inst. fee
1	SIBTTH-01	4	15	365	206	0	1
2	SIBTTH-02	3	15	365	189	0	2
3	SIBTTH-03	1	5	365	180	0	1
4	SIBTTH-04	9	15	365	298	0	1
5	SIBTTH-05	2	15	365	140	1	2
…	…	…	…	…	…	…	…
259	SIBTTH-259	5	15	365	224	2	1

No.	Customer Name	No. of Complaint (2018)	Subscription Package (Mbps)	Subscription Contract (Day)	Length of Subscription (Day)	No. of Upgrade during Subscription	Payment of the intl inst. fee
260	SIBTTH-260	13	15	180	119	0	2
261	SIBTTH-261	14	20	365	222	0	2
262	SIBTTH-262	22	5	365	177	0	1
263	SIBTTH-263	1	10	365	33	0	2

Data will be grouped into 5 clusters. Five initial centroid points will be randomly selected from the data. This is because the best value after comparison is at k=5. The iteration process can be stopped if at the iteration n there is no change in the members of each cluster. In other words, the members in each cluster are the same as the previous iteration. The sequence of iterations is shown in **Table 5**.

Table 5. Number of Cluster's Member for Each Iteration

Iteration	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
1	35	10	137	46	35	263
2	40	21	103	34	65	263
3	43	31	90	24	75	263
4	54	33	90	20	66	263
5	56	33	90	21	63	263
6	56	33	90	30	54	263
7	56	34	90	29	54	263
8	56	34	90	29	54	263

After conducting several iterations, the data is stable at iteration 8, where there is no change in the members if each cluster. **Table 6** is the final centroid point in the clustering process.

Table 6. Final Centroid Point

Centroid 0	Centroid 1	Centroid 2	Centroid 3	Centroid 4
13.821	14.676	11.600	4.724	6.944
15.179	15.882	15.444	15.690	15.463
383.750	679.706	132.000	625.172	365.000
266.464	518.824	72.922	58.862	79.093
0.357	0.471	0.289	0.276	0.185
1.500	1.471	1.567	1.517	1.481

After obtaining the centroid, the determination of the potential category is sorted by the largest centroid point in cluster 1, followed by cluster 3, cluster 0, cluster 4 and cluster 2 is shown in **Table 7**.

Table 7. Potential Category

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Not Potential	Less Potential	Fairly Potential	Potential	Highly Potential

The first step to calculating DBI is to find the Sum of Square Within (SSW) value. This SSW value can be obtained from the sum of the distances of each data to the centroid divided by the amount of data in the cluster is shown in **Table 8**.

Table 8. SSW Value

	Cluster 0	*Cluster 1*	*Cluster 2*	*Cluster 3*	*Cluster 4*
Amount	3.886,91	4.047,95	5.669,08	3.112,77	2.257,12
n	56	34	90	29	54
SSW	69,409	119,057	62,990	107,337	41,799

SSW is an equation used to determine the calculated separation between clusters. The results obtained are used for the next calculation, namely SSB is shown in **Table 9**.

Table 9. SSB Value

$SSB_{0,1}$	$SSB_{0,2}$	$SSB_{0,3}$	$SSB_{0,4}$	$SSB_{1,2}$
388,943	317,556	318,538	188,433	706,271
$SSB_{1,3}$	$SSB_{1,4}$	$SSB_{2,3}$	$SSB_{2,4}$	$SSB_{3,4}$
463,290	540,799	493,421	233,128	260,967

The Ratio value can be calculated after the SSW and SSB values are found. A good cluster is a cluster that has the smallest possible cohesion value and the largest possible separation with the DBI value as shown in **Table 10**.

Table 10. DBI Value

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	R Max	DBI
Cluster 0	0	0,485	0,417	0,555	0,590	0,590	
Cluster 1	0,485	0	0,258	0,489	0,297	0,489	
Cluster 2	0,417	0,258	0	0,345	0,449	0,449	0,538
Cluster 3	0,555	0,489	0,345	0	0,571	0,571	
Cluster 4	0,590	0,297	0,449	0,571	0	0,590	

Next, the evaluation and comparison were conducted using another clustering method, namely K-Medoids. The results of the evaluation using the Davies-Bouldin Index are as **Table 11**.

Table 11. Comparison of DBI Value

	2 cluster	3 cluster	4 Cluster	5 Cluster	6 Cluster
K-Means	0,775	0,649	0,640	0,538	0,618
K-Medoids	0.568	1,017	1,122	0,819	0,876

The DBI value is the average of the highest ratio values. The smaller the DBI value, the better the spread of the data in each cluster. The results show that clustering using the K-Means algorithm provided better results than K-Medoids. For 5 Clusters on K-Means resulted in the smallest value among other clusters so it can be concluded that 5 Clusters are better. The DBI value is 0.538.

Conclusion

The process of determining the potential bandwidth upgrade of FTTH Broadband subscribers can be carried out by applying the K-Means algorithm data mining. In this algorithm, grouping is done by identifying similarities between data, so that groups can be formed to allow determining potential easily. The results were obtained based on potential determination of 263 FTTH Broadband subscribers. It indicates that there are 34 highly potential subscribers (12.92%). The remaining 29 (11.02%), 56 (21.30%), 54 (20.53%), and 90 (34.22%) subscribers are potential, fairly potential, less potential, and not potential respectively. This method is useful for customer determination.

Reference

[1]   T. Hao, A. Sanchez-Postigo, P. Cheben, A. Ortega-Monux, and W. N. Ye, "Dual-Band Polarization-Independent Subwavelength Grating Coupler for Wavelength Demultiplexing," *IEEE Photonics Technol. Lett.*, vol. 32, no. 18, pp. 1163–1166, 2020.

[2]   D. Kuswoyo and N. Agani, "Model Perhitungan Kebutuhan Bandwidth Jaringan Komputer menggunakan Sistem Pakar Fuzzy dengan Metode Adaptive Neuro Fuzzy Inference System(ANFIS) : Studi Kasus PT.GMF Aero Asia Cengkareng," *J. TICOM*, vol. 3, no. 3, pp. 1–15, 2015.

[3]   M. Yang, Y. Lian, J. Wang, and Y. Zhang, "Dual-Mode Large-Mode-Area Multicore Fiber With Air-Hole Structure," *IEEE Photonics J.*, vol. 11, no. 4, pp. 1–10, 2019.

[4]   N. D. Prasongko and R. Gernowo, "Metode Quality Function Deployment Dan Fuzzy Topsis Untuk Sistem Pendukung Keputusan Pemilihan Perusahaan Penyedia Jasa Internet," *J. Sist. Inf. BISNIS*, vol. 5, no. 2, pp. 137–144, 2015.

[5]   S. A. Arnomo and H. Hendra, "Perbandingan Fitur Smartphone, Pemanfaatan Dan Tingkat Usability Pada Android Dan iOS Platforms," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 3, no. 2, pp. 184–192, 2019.

[6]   A. Haque, Z. Wang, S. Chandra, B. Dong, L. Khan, and K. W. Hamlen, "FUSION - An online method for multistream classification," *Int. Conf. Inf. Knowl. Manag. Proc.*, vol. Part F131841, no. 2, pp. 919–928, 2017.

[7]   A. Fitriyani, T. N. Damayanti, and M. S. Yudha, "Perancangan Jaringan Fiber To The Home (FTTH) Perumahan Nataendah Kopo," *e-Procceding Appl. Sci. Vol.1, No.2 Agustus 2915*, vol. 4, no. 3, pp. 3565–3572, 2017.

[8]   F. Fitriastuti and D. P. Utomo, "Implementasi Bandwith Management Dan Firewall System Menggunakan Mikrotik OS 2.9.27," *J. Tek.*, vol. 4, no. 1, pp. 1–9, 2014.

[9]   F. N. Khasanah, "Performa Kecepatan Akses Internet Dengan Squid Proxy Server Pada Ubuntu Server 10.10 Fata," *Informatics Educ. Prof.*, vol. 2, no. 1, pp. 11–18, 2017.

[10] A. Naleng, H. Manossoh, and S. S. Tangkuman, "Analisis Potensi Dan Efektivitas Pemungutan Retribusi Pasar Di Kabupaten Kepulauan Siau Tagulandang Biaro," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 5, no. 2, pp. 2240–2249, 2017.

[11] S. A. Arnomo, "Pengaruh Sistem Informasi Pemasaran dan Loyalitas Konsumen Terhadap Kinerja Pemasaran," *J. Ekon. Bisnis*, vol. 1, no. 1, pp. 1–16, 2012.

[12] R. Hadi, Soiful., Wibowo, "Implementasi Manajemen Bandwidth menggunakan Metode Queue Tree (Studi Kasus pada Universitas Pancasila)," *J. Teknol. Inform. dan Komput.*, vol. 5 No. 1, no. 1, pp. 19–23, 2019.

[13] E. A. Darmadi, "Manajemen Bandwidth Internet Menggunakan Mikrotik Router Di Politeknik Tri Mitra Karya Mandiri," *IKRA-ITH Teknol. J. Sains Teknol.*, vol. 3, no. 3, pp. 7–13, 2019.

[14] M. Fajri, R. Munadi, and T. Y. Arif, "Manajemen Bandwidth Pada Jaringan Lokal Menggunakan Sistem Operasi Vyos," *J. Karya Ilm. Tek. Elektro*, vol. 5, no. 1, pp. 1–5, 2020.

[15] P. Ferdiansyah, R. Indrayani, and S. Subektiningsih, "Analisis Manajemen Bandwidth Menggunakan Hierarchical Token Bucket Pada Router dengan Standar Deviasi," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 1, pp. 38–45, 2020.

[16] D. P. Hidayatullah, R. I. Rokhmawati, and A. R. Perdanakusuma, "Analisis Pemetaan Pelanggan Potensial Menggunakan Algoritma K-Means dan LRFM Model Untuk Mendukung Strategi Pengelolaan Pelanggan (Studi Pada Maninjau Center Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2406–2415, 2018.

[17] A. Darmawan, N. Kustian, and W. Rahayu, "Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 2, no. 3, p. 299, 2018.

[18] S. Beniwal and J. Arora, "Classification and Feature Selection Techniques in Data Mining," *Int. J. Eng. Res. Technol.*, vol. 1, no. 6, pp. 1–6, 2012.

[19] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: Privacy and data mining," *IEEE Access*, vol. 2, pp. 1149–1176, 2014.

[20] R. Sowmya and K. R. Suneetha, "Data Mining with Big Data," in *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*, 2017, vol. 26, no. 1, pp. 246–250.

[21] S. Li, D. C. Yen, W. Lu, and C. Wang, "Identifying the signs of fraudulent accounts using data mining techniques," *Comput. Human Behav.*, vol. 28, no. 3, pp. 1002–1013, 2012.

[22] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[23] 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.

[24] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," in *Physics Procedia*, 2016, vol. 78, no. December 2015, pp. 507–512.

[25] S. Gopalani and R. Arora, "Comparing Apache Spark and Map Reduce with Performance Analysis using K-Means," *Int. J. Comput. Appl.*, vol. 113, no. 1, pp. 8–11, 2015.

[26] K. Handoko and L. S. Lesmana, "Data Mining Pada Jumlah Penumpang Menggunakan Metode Clustering," *Snistek*, no. 1, pp. 97–102, 2018.

[27] M. A. Rahman and M. Z. Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-Means," *Knowledge-Based Syst.*, vol. 71, no. August, pp. 345–365, 2014.

[28] P. S. Hasugian, "Penerapan Data Mining untuk Klasifikasi Produk Menggunakan Algoritma K-Means (Studi Kasus : Toko Usaha Maju Barabai)," *J. Mantik Penusa*, vol. 2, no. 2, pp. 191–198, 2018.