

# Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan

## *K-Means Clustering Method to Determine Customer Satisfaction*

Musthofa Galih Pradana\*<sup>1</sup>, Azriel Christian Nurcahyo<sup>2</sup>, Pujo Hari Saputro<sup>3</sup>

<sup>1</sup> Universitas Alma Ata, <sup>2</sup> STIM Shanti Bhuana, <sup>3</sup> Universitas Alma Ata

E-mail: \*<sup>1</sup>[mgalihpradana@almaata.ac.id](mailto:mgalihpradana@almaata.ac.id), <sup>2</sup>[azriel@shantibhuana.ac.id](mailto:azriel@shantibhuana.ac.id),  
<sup>3</sup>[pujo@almaata.ac.id](mailto:pujo@almaata.ac.id)

### **Abstrak**

Pengolahan data dapat dilakukan dengan banyak cara dan teknik. Peran data saat ini menjadi sangat penting bagi sebuah perusahaan atau penyedia layanan untuk pelanggan. Pentingnya data saat ini menjadikan proses pengolahan data dilakukan secara mandiri menggunakan metode-metode data mining yang ada. Beberapa metode yang dapat diterapkan diantaranya klasifikasi, prediksi maupun klustering. Masing-masing teknik tersebut memiliki hasil yang dapat dijadikan acuan evaluasi dan perencanaan yang lebih baik lagi. Penelitian ini menerapkan teknik klustering yaitu memisahkan dan mengelompokan data berdasarkan kluster. Dalam klustering ada banyak algoritma atau metode yang dapat diterapkan, salah satunya adalah K-Means Klustering. Algoritma K-Means merupakan algoritma yang banyak digunakan untuk mengelompokan data. Hasil dari penelitian ini terbagi menjadi 2 kluster yaitu Kluster 0 yaitu puas dan Kluster 1 yaitu tidak puas ataupun netral. Pengelompokan kluster tersebut berdasarkan dataset yang dimiliki dimana responden mengisi data dan menghasilkan 2 jenis kluster tersebut. Adapun hasil dari proses klustering adalah sebanyak 1303 data masuk kategori kluster 0 atau sebesar 65% dan 697 data masuk kategori kluster 1 atau sebesar 35%.

**Kata Kunci**— Data Mining, Klustering, K-Means

### **Abstract**

Data processing can be done in many ways and techniques. The role of data is now very important for a company or service provider for customers. The importance of data now makes data processing carried out independently using existing data mining methods. Some methods that can be applied include classification, prediction and clustering. Each of these techniques has results that can be used as a reference for evaluation and better planning. This study applies clustering techniques, namely separating and grouping data based on clusters. In clustering there are many algorithms or methods that can be applied, one of which is K-Means Klustering. K-Means algorithm is an algorithm that is widely used to group data. The results of this study are divided into 2 clusters, namely Cluster 0, which is satisfied and Cluster 1, which is not satisfied or neutral. Clustering is based on a dataset that is owned by where the respondent fills in data and produces 2 types of clusters. The results of the clustering process are as many as 1303 data in the category of cluster 0 or 65% and 697 data in the category of cluster 1 or 35%.

**Keywords**— Data Mining, Clustering, K-Means

---

## 1. PENDAHULUAN

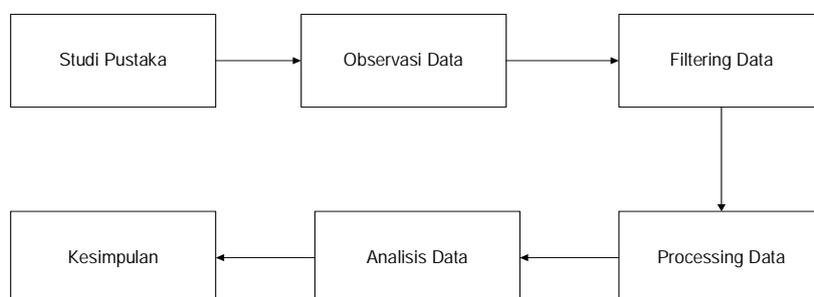
Kepuasan pelanggan adalah salah satu indikator yang dapat dilihat dan diukur dalam pelayanan sebuah perusahaan. Keberadaan pelanggan tentu adalah salah satu yang menjadikan sebuah perusahaan atau penyedia layanan tetap memiliki profit dan dapat semakin berkembang. Oleh karena itu menjadi penting untuk perusahaan tetap memberikan pelayanan yang prima kepada setiap pelanggannya. Untuk dapat tetap memberikan pelayanan yang baik, tentu perusahaan tidak hanya mengandalkan layanan yang saat ini dimiliki tanpa melakukan pengembangan dan evaluasi setiap saat. Apalagi perkembangan yang sangat dinamis dan persaingan yang juga semakin kompetitif menjadikan perusahaan harus memiliki rencana yang matang dan terukur. Untuk dapat melakukan semua hal tersebut tentunya diperlukan sebuah data. Data saat ini menjadi salah satu hal yang sangat berpengaruh dalam jalannya sebuah usaha, karena dari data dapat dilihat dan dipotret gambaran dilapangan tentang layanan yang diberikan.

Data saja tidak cukup, jika hanya sekedar data perlu dilakukan pemrosesan data atau pengolahan data, teknik ini biasa disebut data mining. Pengolahan data menggunakan teknik data mining dapat memberikan banyak opsi informasi bagi perusahaan dapat berupa prediksi, klasifikasi maupun klustering. Data yang dilakukan pengolahan menggunakan teknik data mining, selanjutnya dapat dikembangkan kembali ke dalam bentuk penelitian lain yang dapat dituangkan dalam sebuah sistem pakar, sistem pendukung keputusan atau dalam bentuk sistem yang mampu membantu dalam penentuan kebijakan selanjutnya.

Pada penelitian ini akan dilakukan klustering, yaitu mengelompokkan data ke dalam kluster yang dibagi menjadi 2 kluster yaitu kluster 0 untuk pelanggan yang puas dan kluster 1 untuk pelanggan yang tidak puas atau netral.

## 2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 20. Alur Penelitian

Adapun detail dari masing-masing langkah dalam penelitian adalah sebagai berikut:

- a. Studi Pustaka  
Proses studi pustaka dilakukan dengan cara mencari referensi berupa tulisan dalam buku ataupun penelitian terdahulu dalam jurnal. Tujuan dilakukannya studi pustaka untuk mencari dan memahami penerapan penelitian terkait yang sebelumnya pernah dilakukan.
- b. Observasi Data  
Observasi data dilakukan dengan mencari dataset yang akan digunakan dalam penelitian ini. Pencarian dataset yang dilakukan adalah dengan mencari dataset yang tersedia di google dataset untuk dilakukan analisa dan pengolahan data yang lebih lanjut.
- c. Filtering Data  
Proses filtering data adalah dengan menseleksi dan memilih data yang akan dilakukan dalam penelitian. Pemilihan ini meliputi pemilihan field yang sesuai, pemilihan jumlah data yang juga disesuaikan dengan kebutuhan penelitian.

## d. Processing Data

Processing data dilakukan dengan melakukan beberapa penyesuaian dari data mentah yang ada sehingga data yang diperlukan dapat memenuhi kriteria yang sudah ditentukan sebelumnya. Dalam penelitian ini lebih spesifik bagaimana merubah tipe data yang sebelumnya masih random menjadi seragam dalam bentuk data numerical agar dapat diolah di software analisis untuk menghasilkan kesimpulan.

## e. Analisis Data

Analisis data dilakukan dengan melakukan proses klusterin, yakni melabeli dalam kluster yang tersedia untuk dicari prosentase masing-masing kluster yang ada. Kluster yang terbentuk dalam penelitian ini sebanyak 2 kluster yaitu kluster puas dan tidak puas atau netral yang dinotasikan dalam nama kluster 0 dan kluster 1.

## f. Kesimpulan

Penarikan kesimpulan dapat dilakukan setelah proses klustering selesai dilakukan, dan didapatkan hasil pengkulusteran data dalam bentuk jumlah data dan bentuk prosentase masing-masing kluster yang terbentuk.

## 2.1.1 Rumus Matematika

Menurut Hasn & Kamber algoritma KMeans bekerja dengan membagi data ke dalam k buah cluster yang telah ditentukan. Beberapa cara penghitungan jarak yang biasa digunakan yaitu [1]:

## 1. Euclidean distance

Formula jarak antar dua titik dalam satu, dua dan tiga dimensi secara berurutan ditunjukkan pada formula 1, 2, 3 berikut ini:

$$\sqrt{(x - y)^2} = |x - y| \quad (1)$$

$$d(p, q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2} \quad (2)$$

$$d(p - q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \quad (3)$$

## 2. Manhattan Distance

Manhattan distance disebut juga taxicab distance dengan menggunakan persamaan 4.

$$d1(p, q) = ||p - q||_1 = \sum_{i=1}^n |pi - q1| \quad (4)$$

## 3. Chebichev Distance

Di dalam Chebichev distance atau Maximum Metric jarak antar titik didefinisikan dengan cara mengambil nilai selisih terbesar dari tiap koordinat dimensinya, dengan menggunakan persamaan 5.

$$Dcheb(p, q) = \max (|pi - q1|) \quad (5)$$

## 2.1.2. Pengacuan Pustaka

Penelitian rujukan yang dijadikan dasar dalam penelitian ini yang pertama adalah penelitian dari Dian Eka Ratnawati dkk yang menuliskan dengan menggunakan dataset car evaluation rata-rata hasil akurasi dimulai yang tertinggi adalah K-Means LVQ(86,50%), K-Means berbasis OWA(86,16%) kemudian K-Means konvensional(56,50%) [2].

Selanjutnya tulisan dari Suhardi Rustam dkk yang dipublikasikan di ILKOM Jurnal ilmiah. Penelitian ini menuliskan bahwa dilakukan pengujian model dengan menggunakan algoritma k-means dengan optimasi PSO (particle swarm optimization) untuk data endemik. Model yang dihasilkan diuji untuk mendapat nilai SSE (sum of square error) dan dari nilai ini digunakan sebagai bobot awal optimasi untuk mendapatkan Nilai IDB (davies bouldin index) sebagai nilai ukur terhadap validasi data endemic [3].

Penelitian berikutnya dari Suhardi Rustam dkk yang menyimpulkan Penerapan algoritma K-Means dan K-Means KNN dimana  $K=2$  menghasilkan kluster untuk mengelompokkan Kelas Mata Kuliah Kosentrasi mahasiswa semester akhir dan masing-masing kluster tersebut memiliki nilai prediksi untuk kedua klustering tersebut [4].

Selain dilakukan klustering, hal lain yang dapat dieksplorasi lebih dalam adalah menuangkannya ke dalam sebuah sistem pakar yang akan lebih membantu dalam memberikan kepakaran seperti yang dituliskan Musthofa Galih Pradana dkk di CCIT Jurnal yang dapat dijadikan pertimbangan dalam penelitian selanjutnya [5].

Cara lain yang dapat dilakukan juga dengan melakukan prediksi menggunakan data mining seperti yang dituliskan kaharudin dkk di Jurnal Informasi Interaktif yang melakukan prediksi terhadap kemungkinan pelanggan menjadi churn [6].

Kemudian teknik lain yang dapat dikembangkan pada penelitian lanjutan adalah dengan membuat sebuah sistem pendukung keputusan yang dapat membantu dalam pengambilan keputusan berdasarkan pengolahan data yang sudah dilakukan oleh teknik data mining seperti yang dituliskan Musthofa Galih Pradana dkk di Jurnal Informasi Interaktif [7].

Berikutnya yang bisa dilakukan dengan pengolahan data mining yang berkelanjutan seperti pemanfaatan fitur word cloud dan document term matrix untuk dapat mengidentifikasi kata apa yang paling sering muncul dalam pengolahan data mining [8].

Penelitian rujukan berikutnya adalah tulisan dari Dong Su dkk yang menuliskan bahwa telah meningkatkan state-of-the-art pada pribadi berbeda k-berarti pengelompokan dalam beberapa cara. Metode noninteraktif diperkenalkan untuk mengelompokkan k-means pribadi yang berbeda dan meningkatkan satu metode interaktif berdasarkan kesalahan sistematis analisis [9].

Robust and Sparse Fuzzy K-Means Clustering dituliskan oleh Jinglin Xu dkk yang menuliskan bahwa mengusulkan metode baru, yang disebut robust dan algoritma pengelompokan K-Means fuzzy yang jarang, untuk mendapatkan hasil pengelompokan yang lebih akurat [10].

Xinwang Liu juga menuliskan tentang klustering dimana hasilnya menyimpulkan bahwa mengusulkan algoritma MKKM dengan regularisasi yang diinduksi oleh matriks - sebuah algoritma yang secara konseptual sederhana namun efektif yang juga menangani redundansi dan 1893 keragaman lambat dari algoritma MKKM yang ada [11].

Penelitian dari Jinglin Xu yang menghasilkan kesimpulan bahwa mengusulkan kerangka kerja pengelompokan tanpa pengawasan yang menyematkan banyak diskriminasi subruang ke multi-view K-Means clustering untuk membangun kerangka kerja terpadu, dan secara adaptif mengendalikan interkoordinasi antara banyak pandangan melalui bobot [12].

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Dataset digunakan sebagai data acuan yang akan diolah dan di analisa dalam penelitian ini. Dataset pelanggan perusahaan yang digunakan berjumlah 2.000 record. Adapun beberapa variable yang masuk ke dalam dataset ditunjukkan pada Tabel 1.

Tabel 7. Variable Dataset

No	Nama Variable
1	id
2	Satisfaction
3	Gender
4	Customer Type
5	Age
6	Type of Travel
7	Class

No	Nama Variable
8	Flight Distance
9	Seat comfort
10	Departure/Arrival time convenient
11	Food and drink
12	Gate location
13	Inflight wifi service
14	Inflight entertainment
15	Online support
16	Ease of Online booking
17	On-board service
18	Leg room service
19	Baggage handling
20	Check in service
21	Online boarding
22	Departure Delay in Minutes
23	Arrival Delay in Minutes

### 3.2. Pengolahan Data Awal

Data yang ada perlu dilakukan pengolahan awal seperti perubahan data kategorikal menjadi data numerical agar dapat diproses dan dilakukan pengolahan data dalam klustering. Berikut ini adalah beberapa tahapan yang dilakukan dalam pengolahan data awal:

### 3.3. Identifikasi data

Data awal yang akan diolah masih merupakan data campuran antara data numerical dan data kategorikal. Data kategorikal perlu diubah menjadi data numerical, beberapa data kategorikal dalam data ini ditunjukkan pada Tabel 2.

Tabel 8. Field Data

No	Nama Field
1	Satisfaction
2	Gender
3	Customer Type
4	Type of Travel
5	Class

Dari masing-masing field tersebut detail data dan isinya ditunjukkan pada Tabel 3.

Tabel 9. Detail Data Field

No	Satisfaction	Gender	Customer Type	Type of Travel	Class
1	Satisfied	Female	Loyal Customer	Personal Travel	Eco
2	Satisfied	Male	Loyal Customer	Personal Travel	Business
3	Satisfied	Female	Loyal Customer	Personal Travel	Eco
4	Satisfied	Female	Loyal Customer	Personal Travel	Eco
5	Satisfied	Female	Loyal Customer	Personal Travel	Eco
6	Satisfied	Male	Loyal Customer	Personal Travel	Eco
7	Satisfied	Female	Loyal Customer	Personal Travel	Eco
8	Satisfied	Female	Loyal Customer	Personal Travel	Eco Plus
9	Satisfied	Female	Loyal Customer	Personal Travel	Business

10	Satisfied	Male	Loyal Customer	Personal Travel	Eco
.....					
2000	Neutral or Dissatisfied	Male	Loyal Customer	Personal Travel	Eco

### 3.4. Konversi Data

Data kategorikal menjadi numerical diubah dengan ketentuan pada Tabel 4 sampai dengan Tabel 8. Untuk hasil konversi ditunjukkan pada tabel 9.

1. Satisfaction dengan detail konversi field satisfaction dibagi menjadi dua jenis dan detailnya ditunjukkan pada Tabel 4.

Tabel 10. Field Satisfaction

No	Kategori
1	Satisfied
2	Neutral or Dissatisfied

2. Gender detail konversi field gender dibagi menjadi dua jenis dan ditunjukkan pada Tabel 5.

Tabel 11. Field Gender

No	Kategori
1	Male
2	Female

3. Customer Type, detail konversi field Customer Type dibagi menjadi dua jenis dan ditunjukkan pada Tabel 6.

Tabel 12. Field Customer Type

No	Kategori
1	Loyal Customer
2	Disloyal Customer

4. Type of Travel, detail konversi field dibagi menjadi dua jenis dan Customer Type ditunjukkan Tabel 7.

Tabel 13. Field Type of Travel

No	Kategori
1	Personal Travel
2	Public Travel

5. Class, detail konversi field Class dibagi menjadi dua jenis dan ditunjukkan pada Tabel 8.

Tabel 14. Field Class

No	Kategori
1	Business
2	Eco
3	Eco Plus

Tabel 15. Hasil Konversi Data

No	Satisfaction	Gender	Customer Type	Type of Travel	Class
1	1	2	1	1	1
2	1	1	1	1	2
3	1	2	1	1	1
4	1	2	1	1	1
5	1	2	1	1	1
6	1	1	1	1	1
7	1	2	1	1	1
8	1	2	1	1	3
9	1	2	1	1	2
10	1	1	1	1	1
.....					
2000	1	2	1	1	1

### 3.5. Proses Klustering

Algoritma yang digunakan adalah K-Means Clustering adapun hasil dari klustering adalah dilakukan iterasi sebanyak 6 kali dengan nilai cluster sum of squared errors sebesar 3632.7459128990963.

Hasil dari proses klustering akan diolah dan menghasilkan 2 kluster yang sudah terbentuk sebelumnya yaitu kluster 0 untuk pelanggan yang puas dan kluster 1 untuk pelanggan yang netral ataupun tidak puas. Dari hasil tersebut masing-masing field akan memiliki rata-rata nilai dari masing-masing cluster, jadi rata-rata yang masuk ke dalam kedua kluster. Sebagai contoh pada class memiliki nilai rata-rata sebesar 1.27 untuk kluster 0 dan nilai 1.29 untuk nilai kluster 1. Adapun detail dari nilai final cluster centroids yang terbentuk pada masing-masing field kategori ditunjukkan pada Tabel 10.

Tabel 16. Final Cluster Centroid

Attribute	Cluster		
	Full Data	Cluster 0	Cluster 1
id	65762.4645	65021.1266	67148.3515
satisfaction	1.3105	1.0023	1.8867
Gender	1.5085	1.769	1.0215
Customer Type	1	1	1
Age	38.2585	38.0752	38.6011
Type of Travel	1	1	1
Class	1.283	1.2763	1.2956
Flight Distance	1839.8615	1671.4712	2154.6571
Seat comfort	1.0065	1.0706	0.8867
Departure/Arrival time convenient	2.7835	3.6147	1.2296
Food and drink	1.0055	1.0507	0.9211
Gate location	2.908	2.9018	2.9197
Inflight wifi service	3.0255	3.0714	2.9397
Inflight entertainment	1.6655	1.9747	1.0875
Online support	3.1875	3.2886	2.9986
Ease of Online booking	2.984	2.9708	3.0086
On-board service	3.0345	3.2041	2.7174
Leg room service	2.67	2.5096	2.9699
Baggage handling	3.356	3.5027	3.0818

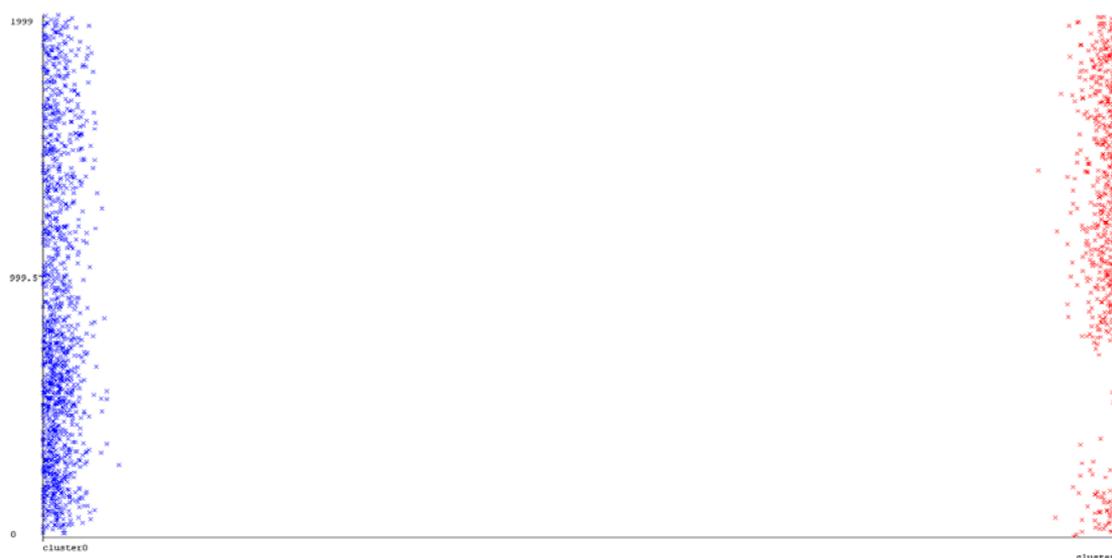
Attribute	Cluster		
	Full Data	Cluster 0	Cluster 1
Checkin service	3.1645	3.3952	2.7331
Cleanliness	3.3745	3.4965	3.1463
Online boarding	3.089	3.109	3.0516
Departure Delay in Minutes	11.4385	8.934	16.1205
Arrival Delay in Minutes	11.7891	9.1312	16.7578

Hasil dari eksekusi klustering menggunakan K-Means sebanyak 2.000 data adalah dalam waktu 0.11 detik dan menghasilkan clustered instances yang ditunjukkan pada Tabel 11.

Tabel 17. Clustered Instances

Kluster	Jumlah Data	Prosentase
0	1303	65%
1	697	35%

Setelah data clustered terbentuk pada Tabel 11 dengan masing-masing prosentasesnya, selanjutnya data akan divisualisasikan kedalam bentuk diagram atau gambar untuk melihat sebaran cluster. Adapun hasil visualisasi sebaran data dalam setiap kluster ditunjukkan pada Gambar 2.



Gambar 21. Visualisasi Sebaran Kluster

Cluster 0 diwakili warna biru dengan jumlah rentang data paling pada angka 0-999.5 sedangkan Cluster 1 diwakili warna merah memiliki data yang berkebalikan pada range 0-999.5 memiliki rata-rata data yang lebih sedikit.

#### 4. KESIMPULAN

Kesimpulan pada penelitian ini adalah sebagai berikut:

1. Dengan data pelanggan sebanyak 2.000 data dihasilkan kluster sebanyak 2 dengan rincian kluster 0 sebanyak 1303 data atau sebesar 65% dan kluster 2 sebanyak 697d data atau sebesar 35%.
2. Jumlah Iterasi yang dilakukan sebanyak 6 kali
3. Nilai cluster sum of squared errors sebesar 3632.7459128990963.

## 5. SARAN

Adapun saran untuk penelitian yang selanjutnya adalah:

1. Dapat dilakukan menggunakan metode lain selain K-Means.
2. Dapat dikembangkan dan dibuat dalam model klasifikasi.
3. Dapat dikembangkan menjadi sebuah expert system.
4. Dapat dikembangkan menjadi sebuah decision support system.

## DAFTAR PUSTAKA

- [1] Larose, D. T., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley, New Jersey
  - [2] Ratnawati, D. E., Indriati., 2015, Klustering Dengan K-Means Berbasis OWA, *JTIK Journal*, No. 1, Vol. 2, Hal. 29–33
  - [3] Rustam, S., Santoso, H. A., Supriyanto, C., 2018, Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle, *ILKOM Jurnal Ilmiah*, Vol. 10, Hal. 251–259.
  - [4] S. Rustam and H. Annur, “akademik data mining (adm) k-means dan k-means k-nn untuk mengelompokan kelas mata kuliah kosentrasi,” vol. 11, no. 28, pp. 260–268, 2019.
  - [5] Pradana, M. G., Pamekas, B. W., Kusri., 2018, Penyakit Diabetes Mellitus Menggunakan Metode Certainty Factor Design Expert System for Diagnosing Diabetes, *CCIT Journal*, No. 2, Vol. 11, Hal. 182–191.
  - [6] Kaharudin., Pradana, M. G., 2019, Prediksi Customer Churn Perusahaan Telekomunikasi Menggunakan Naïve Bayes Dan K-Nearest Neighbor, No. 3, Vol. 4.
  - [7] Pradana, M. G., Kusri., Luthfi, E. T., Perbandingan Metode Weighted Product Dan Simple Additive Weighting Dalam Seleksi Pengurus Forum Asisten (Studi Kasus: Universitas Amikom Yogyakarta), *Jurnal Informasi Interaktif*, No. 2, Vol. 4.
  - [8] Pradana, M. G., 2020, Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining, *Jurnal Ilmiah Informatika*, No. 1, Vol. 8, Hal. 38-43.
  - [9] Su, D., Cao, J., Li, N., Bertino, E., Jin, H., 2016, Differentially Private K-Means Clustering, *CODASPY '16: Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy*, New Orleans, 9 - 11 Maret
  - [10] Xu, J., Han, J., Xiong, K., Nie, F., 2016, Robust and Sparse Fuzzy K-Means Clustering, *International Joint Conference on Artificial Intelligence 2016*, New York, 9 - 15 Juli.
  - [11] Liu, X., Dou, Y., Yin, J., Wang, L., Zhu, En., 2016, “Multiple Kernel k -Means Clustering with Matrix-Induced Regularization, *Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, 12 - 17 Februari
  - [12] Xu, J., Han, J., Nie, F., 2016, Discriminatively Embedded K-Means for Multi-view Clustering, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27 - 30 Juni
-