

Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode *Clustering K-Means* Pada Desa Kuala Dendang

Siti Amaliyah¹, Jasmir², Silvia Rianti Agustini³

¹Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

Email : sitiamaliyah3296@gmail.com, 2ijay_jasmir@yahoo.com, 3silvianti7@gmail.com

Abstrak - Program Keluarga Harapan (PKH) dimaknai sebagai program pemberian bantuan sosial bersyarat kepada Keluarga Miskin (KM) setelah ditetapkan sebagai keluarga penerima manfaat PKH. Bantuan ini sangat bermanfaat bagi warga yang membutuhkan. Namun nyatanya penerimaan bantuan PKH ini masih belum merata dan belum tepat sasaran. Dalam pengelolaan data PKH, selain tingkat akurasi data, efisiensi waktu dalam pengelolaan data juga sangat penting. Terdapat kerumitan dalam pengelolaan data yang sangat besar untuk menentukan warga yang menjadi prioritas utama untuk mendapatkan bantuan PKH pada Desa Kuala Dendang. Untuk itu penulis menggunakan *data mining* dengan metode *clustering k-means* untuk mempermudah pengelolaan data tersebut.

Dalam analisis penulis menggunakan data tahun 2020 dengan memakai 6 atribut yaitu nama KK, tunjangan, kondisi rumah, kepemilikan rumah, jumlah penghasilan, dan status kesejahteraan. Lalu data tersebut diolah dengan menggunakan *tools* SPSS. Setelah itu diperoleh hasil 3 *cluster* yaitu *Cluster 1* sebagai prioritas pertama memiliki 114 anggota dengan presentase 11%, *Cluster 2* sebagai prioritas kedua memiliki 690 anggota dengan presentase 68%, dan *Cluster 3* sebagai prioritas ketiga memiliki 218 anggota dengan presentase 21%.

Kata Kunci: Data Mining, *Clustering K-Means*, PKH

Abstract - The Family Hope Program (PKH) is a program of providing conditional social assistance to Poor Families (KM) who are designated as beneficiary families of PKH. This assistance is very useful for people in need. However, in fact, the acceptance of PKH assistance is still not evenly distributed and on target. In PKH data management, apart from the accuracy of the data, time efficiency in data management is also very important. There is a complexity in managing huge data to determine which residents are the top priority for PKH assistance in Kuala Dendang Village. For this reason, the author uses data mining with the k-means clustering method to facilitate the management of the data.

In the author's analysis, the author uses 2020 data using 6 attributes, namely the name of KK, allowances, house conditions, home ownership, income, and welfare status. Then the data is processed using SPSS tools. After that, the results of 3 clusters are obtained, namely Cluster 1 as the first priority has 114 members with a percentage of 11%, Cluster 2 as the second priority has 690 members with a percentage of 68%, and Cluster 3 as a third priority has 218 members with a percentage of 21%.

Keywords: Data Mining, *Clustering K-Means*, PKH

1. PENDAHULUAN

Kemiskinan didefinisikan sebagai salah satu permasalahan pokok yang sering kali dialami oleh sebagian besar negara di dunia, khususnya Indonesia. Saat ini kondisi kemiskinan di Indonesia diperparah dengan adanya pandemi Virus Corona (Covid-19). Menurut Badan Pusat Statistik atau BPS [1] "Angka kemiskinan di Republik tercinta ini mencapai 10,14% atau sebanyak 27,54 juta penduduk Indonesia berstatus miskin. Sedangkan pada bulan September 2019 sebelum pandemi, angka kemiskinan di Indonesia hanya sebesar 9,22%.

Untuk mengatasi kondisi kemiskinan di Indonesia, berbagai kebijakan sudah dikeluarkan oleh Pemerintah Indonesia antara lain berupa program-program bantuan untuk masyarakat yang kurang mampu atau miskin seperti program Kartu Prakerja, Subsidi Listrik, Bantuan Langsung Tunai (BLT) untuk UMKM, Program Sembako, Bantuan Sosial Tunai (BST), dan Program Keluarga Harapan (PKH).

Namun permasalahannya, terkadang proses penyaluran bantuan ini masih dianggap tidak merata atau tepat sasaran, khususnya bantuan Program Keluarga Harapan (PKH) di Desa Kuala Dendang. Adapun kondisi yang terjadi adalah banyak masyarakat yang merasa dirinya pantas mendapatkan bantuan Program Keluarga Harapan (PKH) tetapi terlihat tidak sesuai dengan kriteria yang sudah ditentukan oleh pihak pemerintahan.

Permasalahan kedua, karena banyaknya data penduduk pada desa Kuala Dendang, membuat pihak yang menangani data membutuhkan waktu lama untuk menentukan warga yang berhak mendapatkan bantuan karena masih melakukan pengecekan data secara manual, yaitu dengan melihat data satu-persatu. Apalagi data yang tersedia masih ada beberapa yang belum diperbaharui atau lengkap. Atau terkadang pihak desa hanya menerima rekomendasi dari setiap RT yang ada mengenai warga-warga yang diusulkan mendapat bantuan PKH.

Penentuan prioritas warga yang layak mendapat bantuan Program Keluarga Harapan (PKH) ini dapat dilakukan dengan dengan *data mining*, salah satunya dengan metode *Clustering K-Means*. *Data Mining* dipilih karena metodenya dapat mengekstraksi informasi prediktif dalam suatu *database* yang besar [2]. Algoritma *Clustering K-Means* dipilih karena dapat mengelompokkan data ke dalam *cluster-cluster* sehingga data yang memiliki kemiripan ditempatkan pada *cluster* yang sama dan yang tidak memiliki kemiripan di *cluster* yang lainnya [3].

Data mining dapat didefinisikan sebagai proses mencari, mengumpulkan, menyaring, dan menganalisis data dari suatu database besar untuk menemukan suatu informasi yang berharga dan bermanfaat bagi pemilik data tersebut [4]. *Clustering* merupakan pengelompokkan data-data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada [5]. Algoritma *K-Means* merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat kluster (*centroid*) terdekat dengan data. Tujuan dari *K-Means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antar kluster [6]. Program Keluarga Harapan atau yang disebut PKH adalah program pemberian bantuan sosial bersyarat pada Keluarga Miskin (KM) yang ditetapkan sebagai keluarga penerima manfaat PKH sebagai upaya percepatan penanggulangan kemiskinan [7].

Dan dalam melakukan perhitungan *Clustering K-Means* pada penelitian ini, penulis menggunakan 6 atribut yaitu: Nama Kepala Keluarga, Jumlah Tanggungan, Kondisi Rumah, Status Kepemilikan Rumah, Jumlah Penghasilan, dan Status Kesejahteraan dengan menggunakan tools SPSS. dimana SPSS merupakan sebuah aplikasi komputer yang dapat digunakan untuk mengolah dan menganalisa data secara statistik. SPSS adalah kepanjangan dari *Statistical Package for the Social Sciences*. Saat ini program SPSS sangat populer dan dapat diterapkan atau diaplikasikan pada semua bidang. SPSS adalah kepanjangan dari *Statistics Product and Service Solution* [8].

Dengan adanya penelitian ini diharapkan dapat membantu pihak Desa Kuala Dendang dalam proses penyaluran bantuan Program Keluarga Harapan (PKH), sehingga pemberian bantuan PKH ini menjadi lebih optimal dan tepat sasaran kepada masyarakat yang benar-benar membutuhkan.

1.1 Penelitian Sejenis

Referensi penelitian pertama menghasilkan sebuah rekomendasi kelompok penerima bantuan beasiswa PPA di UDINUS [9]. Referensi penelitian kedua memberi informasi mengenai pengelompokkan penduduk miskin pada Kecamatan Jambi Selatan dengan 3 kategori yaitu hampir miskin, miskin, dan sangat miskin [10]. Referensi penelitian ketiga menghasilkan suatu aplikasi yang bisa memilah data dengan cepat untuk siapa saja yang layak menerima bantuan BPNT [11]. Referensi penelitian keempat menghasilkan sistem yang dapat mengelompokkan siswa ke kelas-kelas unggulan yang ada pada sekolah tersebut [12]. Dan pada penelitian kelima menghasilkan rekomendasi siswa yang kompeten untuk mengikuti OSN dari pengolahan semua data akademik siswa tersebut [13].

2. METODOLOGI PENELITIAN

Dalam menyusun sebuah penelitian sangat penting sekali menyusun sebuah kerangka penelitian. Kerangka penelitian digunakan sebagai acuan dalam penelitian yang akan dilakukan. Penulis menyusun beberapa tahapan yang dilakukan selama melakukan penelitian yang akan teruraikan dibawah:

1. Identifikasi Masalah

Kegiatan yang dilakukan penulis pada tahap ini adalah mengidentifikasi dan merumuskan permasalahan penelitian, hal ini dilakukan dengan tujuan untuk mengetahui masalah yang dialami Desa Kuala Dendang dalam proses mendistribusikan bantuan Program Keluarga Harapan (PKH) kepada masyarakat agar bantuan tersebut tepat sasaran.

2. Studi Literatur

Penulis melakukan kajian pustaka pada tahap studi literatur ini, yaitu mempelajari buku-buku referensi, artikel-artikel, dan hasil penelitian sejenis yang relevan dengan permasalahan yang sedang diteliti. Kegiatan ini dilakukan dengan tujuan untuk mendapatkan landasan teoritis mengenai permasalahan yang diteliti. Dengan studi literatur yang tepat, peneliti dapat memahami permasalahan yang diteliti dengan benar dan sesuai dengan pembahasan yang dilakukan.

3. Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data. Data yang dimaksud adalah penduduk Desa Kuala Dendang sesuai atribut yang dibutuhkan dalam penelitian. Untuk mengumpulkan data yang dibutuhkan, penulis menggunakan beberapa metode pengumpulan data yaitu pengamatan langsung (observasi), wawancara, dan Penelitian Kepustakaan.

4. Analisis Deskriptif

Proses analisis deskriptif dilakukan dengan beberapa tahapan. Dimana alur proses analisis deskriptif yang penulis gunakan dimulai dari *Data Selection*, *Data Processing*, *Data Transformation*, dan *Clustering* Algoritma *K-Means*.

5. Analisis Hasil

Hasil dari analisis akan dinilai keakuratan data model yang sudah didapatkan dengan perhitungan *K-Means* yang populer dan akan mengelola data penduduk mengenai kemiskinan.

6. Penulisan Laporan

Pada tahap pembuatan laporan akhir penelitian, dilakukan berdasarkan kerangka yang telah dirancang yang terdiri dari Pendahuluan, Landasan Teori, Metodologi Penelitian, Analisis Hasil, Penutup, serta lampiran bukti hasil penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Mining Menggunakan *Clustering K-Means*

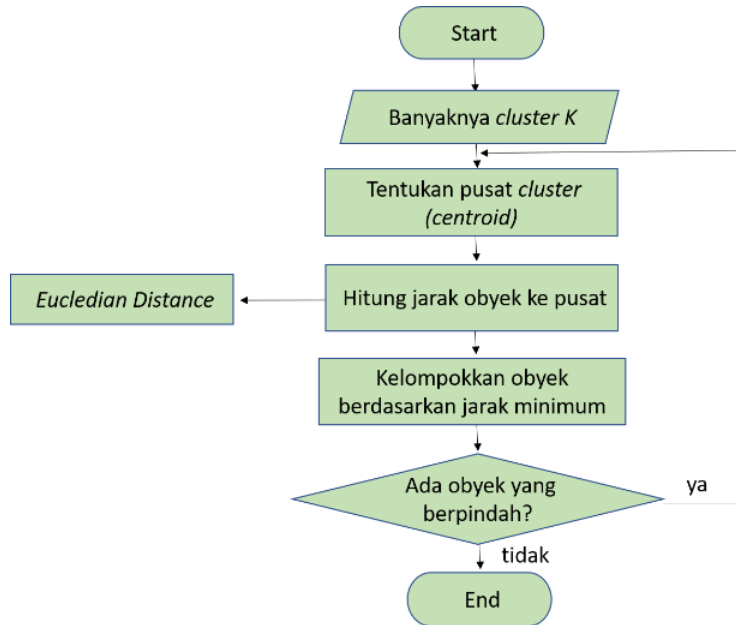
Penulis melakukan perhitungan 1022 data Kepala Keluarga pada Desa Kuala Dendang dengan menggunakan 6 atribut yaitu: Nama Kepala Keluarga, Jumlah Tanggungan, Status Kepemilikan Rumah, Jumlah penghasilan, dan Status Kesejahteraan Keluarga. Dengan menggunakan 3 *cluster*, dimana kategori *cluster* 1 dijadikan prioritas 1, *cluster* 2 dijadikan prioritas 2, dan *cluster* 3 dijadikan prioritas 3.

3.1.1 Transformasi Nilai Data

Tabel 3.1 Inisialisasi Nilai Data

1	JUMLAH TANGGUNGAN	
	1	1 ANAK
	2	2 ANAK
	3	3 ANAK
	4	>3 ANAK
2	KONDISI RUMAH	
	1	BAIK
	2	CUKUP
	3	KURANG BAIK
3	STATUS KEPEMILIKAN	
	1	HAK MILIK
	2	NUMPANG
	3	SEWA
4	JUMLAH PENGHASILAN	
	1	1500000 - 2000000
	2	2000000 - 3000000
	3	3000000 - 4000000
	4	>4000000
5	STATUS KESEJAHTERAAN	
	1	BAIK
	2	CUKUP
	3	KURANG

3.1.2 Flowchart Langkah-Langkah *Clustering K-Means*



Gambar 3.1 Flowchart *Clustering K-Means*

3.1.3 Nilai *Centroid* Awal

Pemilihan nilai *Centroid* awal ini dilakukan secara acak oleh Tools SPSS.

Tabel 3.2 Nilai *Centroid* Awal

<i>Centroid 1 (M1)</i>	4	2	1	1	3
<i>Centroid 2 (M2)</i>	0	2	1	3	2
<i>Centroid 3 (M3)</i>	4	1	3	4	1

3.1.4 Rumus *Euclidian Distance*, *BCV*, *WCV*, dan *Ratio*

a. *Euclidian Distance*

$$Euclidian Distance = d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

b. *BCV (Between Cluster Variation)*

$$D(M1, M2) = \sqrt{(4 - 0)^2 + (2 - 2)^2 + (1 - 1)^2 + (1 - 3)^2 + (3 - 2)^2} = 4.5826$$

$$D(M1, M3) = \sqrt{(4 - 4)^2 + (2 - 1)^2 + (1 - 3)^2 + (1 - 4)^2 + (3 - 1)^2} = 4.2426$$

$$D(M2, M3) = \sqrt{(0 - 4)^2 + (2 - 1)^2 + (1 - 3)^2 + (3 - 4)^2 + (2 - 1)^2} = 4.7958$$

$$\text{Jadi BCV} = D(M1, M2) + D(M1, M3) + D(M2, M3) = 4.5826 + 4.2426 + 4.7958 = 13.6210$$

c. *WCV (Within Cluster Variation)*

$$WCV = \text{Total min } \{d(X_i, Y_i)^2\}$$

d. *Ratio*

$$\text{Ratio} = \text{BCV}/\text{WCV}$$

3.1.5 Nilai Centroid Final

Tabel 3.3. Nilai Centroid Final

Centroid 1	2.0	2.0	1.0	1.7	2.6
Centroid 2	1.1	1.0	1.0	3.3	1.0
Centroid 3	3.4	1.1	1.0	3.3	1.1

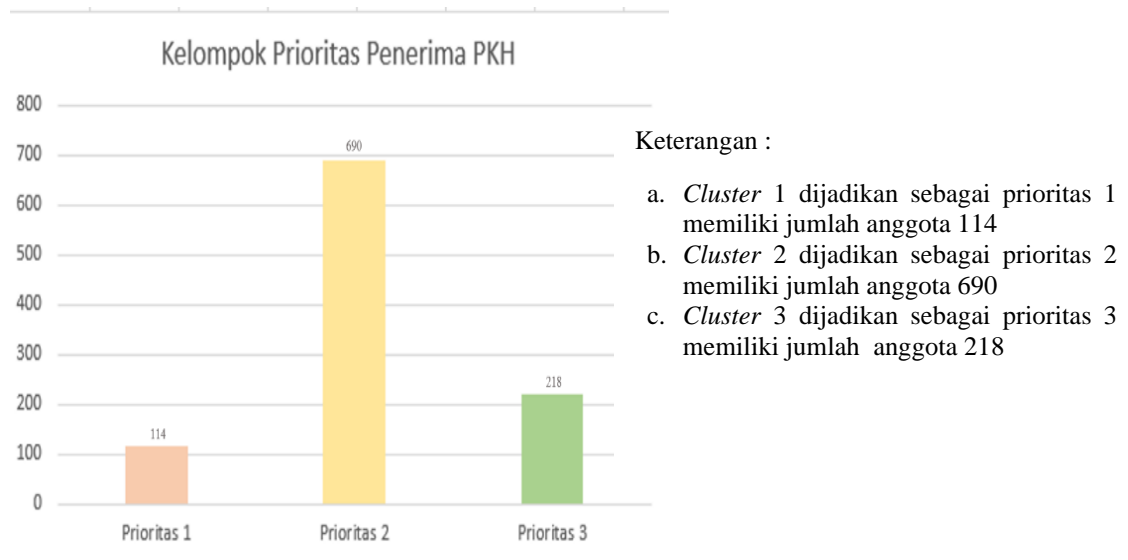
3.1.6 Perbandingan Nilai BCV, WCV, dan Ratio Pada Iterasi 1,2,3,4,5

Tabel 3.4 Perbandingan Nilai BCV, WCV, dan Ratio

Nilai	Iterasi 1	Iterasi 2	Iterasi 3	Iterasi 4	Iterasi 5
BCV	13.621	7.6833	7.8659	7.7371	7.7685
WCV	4327	1070.4448	991.8422	979.9861	978.8095
Rasio	0.0031	0.0072	0.0079	0.0079	0.0079

Perhitungan *Clustering K-Means* dari data warga Desa Kuala Dendang berhenti pada iterasi ke 5 (lima) karena tidak ada lagi data yang berpindah tempat dan nilai ratio juga sudah lebih kecil atau sama dengan nilai ratio pada iterasi sebelumnya.

3.1.7 Hasil Akhir Perhitungan



Gambar 3.2 Kelompok Prioritas Penerima PKH

3.1.8 Perbandingan Hasil Penerima PKH Pada Desa Kuala Dendang Dan Hasil Penerima PKH Yang Penulis Rekomendasikan

Setelah proses perhitungan data untuk prioritas penerima Bantuan PKH selesai penulis lakukan. Penulis melakukan pengecekan untuk membandingkan hasil dari data yang penulis dapatkan. Didapatkan data dari Desa Kuala Dendang yang menerima PKH sebanyak 74 Kepala Keluarga. Sedangkan dari perhitungan *Clustering K-Means* dengan atribut yang penulis gunakan ada sebanyak 114 kepala keluarga (yang berada pada Cluster 1 atau prioritas 1) yang berhak mendapat bantuan PKH. Itu berarti ada 40 kepala keluarga yang menurut rekomendasi penulis berhak mendapat bantuan PKH tetapi dalam data pada Desa Kuala Dendang tidak tercatat.

4. KESIMPULAN

Kesimpulan yang dapat disampaikan dari hasil penelitian yang telah dilakukan pada data penduduk Desa Kuala Dendang tahun 2020 dengan jumlah data 1022 kepala keluarga, adalah Penerapan data mining untuk menentukan prioritas penerima bantuan PKH dilakukan menggunakan metode *Clustering K-Means* dengan cara membagi 3 *cluster* dengan kategori prioritas 1, prioritas 2, dan prioritas 3. Dengan iterasi yang dilakukan sebanyak 5 iterasi. Hasil dari perhitungan *Clustering K-Means* menghasilkan 3 kelompok prioritas penerima bantuan Program Keluarga Harapan (PKH) yaitu: *Cluster 1* sebagai prioritas 1 terdapat 114 data kepala keluarga dengan hasil presentase 11%. *Cluster 2* sebagai prioritas 2 terdapat 690 data kepala keluarga dengan hasil persentase 68%. *Cluster 3* sebagai prioritas 3 terdapat 218 data kepala keluarga dengan hasil persentase 13%. Hasil dari data penduduk menggunakan algoritma *Clustering K-Means* menghasilkan bahwa penduduk yang paling diprioritaskan atau paling membutuhkan bantuan PKH berada pada *Cluster* pertama.

REFERENCES

- [1] Badan Pusat Statistik, "2 Februari 2021," *Jumlah Pendud. Miskin Pada Sept. 2021*, p. 57, 2021, [Online]. Available: <https://jambi.bps.go.id/pressrelease/2022/02/02/596/jumlah-penduduk-miskin-pada-september-2021-sebesar-279-86-ribu-orang.html>.
- [2] L. R. Angga Ginanjar Mabur, "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit," *J. Komput. dan Inform.*, vol. 1, no. 1, pp. 53–57, 2012, [Online]. Available: <http://komputa.if.unikom.ac.id/jurnal/penerapan-data-mining-untuk.8>.
- [3] T. Khotimah, "Pengelompokan Surat Dalam Al Qur'an Menggunakan Algoritma K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 5, no. 1, pp. 83–88, 2014, doi: 10.24176/simet.v5i1.141.
- [4] B. A. R. Hassan, Z. B. M. Yusoff, M. A. H. Othman, S. Bin, A. information is available at the end of the Chapter, and [Http://dx.doi.org/10.5772/55358](http://dx.doi.org/10.5772/55358), *We are IntechOpen , the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %*. 2012.
- [5] E. Prasetyo, "Data Mining Konsep dan Aplikasi Menggunakan Matlab," 2012, Accessed: Oct. 16, 2021. [Online]. Available: https://scholar.google.co.id/citations?view_op=view_citation&hl=id&user=9N8JRC4AAAAJ&citation_for_view=9N8JRC4AAAAJ:u5HHmVD_uO8C.
- [6] T. Noviana, Jasmir, and Y. Novianto, "Bab II Landasan Teori," *Penerapan Data Min. Untuk Menentukan Kelompok Prioritas Penerima Bantu. Beras Rastra Menggunakan Metod. Clust. K-Means Pada Kec. Muara Papalik*, vol. 53, no. 9, pp. 1689–1699, 2019, [Online]. Available: <http://digilib.stikom-db.ac.id/gdl42/gdl.php?mod=browse&op=read&id=gdlhub-gdl-tianoviana-5792>.
- [7] D. J. P. dan J. Sosial and K. S. RI, *Pedoman operasional kelembagaan PKH tahun 2013*. Direktorat Jaminan Sosial : Jakarta.,2013, 2013.
- [8] C. Gunawan, *Mahir Menguasai SPSS: (Mudah mengolah Data Dengan IBM SPSS Statistic 25)*. CV BUDI UTAMA, 2018.
- [9] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [10] D. Sunia, Kurniabudi, and P. A. Jusia, "Bab II Landasan Teori," *PENERAPAN DATA Min. UNTUK Clust. DATA Pendud. MISKIN MENGGUNAKAN Algoritm. K-MEANS*, vol. 53, no. 9, pp. 1689–1699, 2019, [Online]. Available: <http://digilib.stikom-db.ac.id/gdl42/files/disk1/119/gdlhub-gdl-dinasunia-5916-1-bab2.pdf>.
- [11] S. Ghousi, P. Ali, I. S. Achmadi, and M. Si, "Klasifikasi Penentuan Penerima Bantuan Pangan Non Tunai Menggunakan Metode K-Means Clustering," pp. 1–8, 2020, [Online]. Available: <http://eprints.itn.ac.id/4547/9/Jurnal.pdf>.
- [12] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
- [13] W. Sirait, S. Defit, and G. W. Nurcahyo, "Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA," *J. Sistim Inf. dan Teknol.*, vol. 1, no. 3, pp. 25–30, 2019, doi: 10.35134/jsisfotek.v1i3.6.