

Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang

M.Rizqi Sulistio^{1*}, Nana Suarna², Odi Nurdiawan³

^{1,2}Program Studi Teknik Infomatika, STMIK IKMI Cirebon, Cirebon, Indonesia

³Program Studi Manajemen Infomatika, STMIK IKMI Cirebon, Cirebon, Indonesia

Email: ^{1*}rizkysulistio15@gmail.com, ²st_nana@yahoo.com, ³odinurdiawan2020@gmail.com

Email Coresponding Author: rizkysulistio15@gmail.com

Abstrak– Banyak perusahaan menjual berbagai macam produk mulai dari accessories, fassion, sepatu, kebutuhan keluarga dan produk lainnya yang dijual belikan. Produk terdapat antara produk yang laku dengan produk yang tidak laku. Permasalahannya adalah Belum mengetahui bagaimana cara menganalisa parameter pengelompokan data penjualan menggunakan *algoritma x-means* serta belum bisa menginterpretasikan pengelompokan penjualan produk menggunakan *algoritma x-means clustering*. Tujuan penelitian ini adalah untuk memperoleh pengetahuan tingkat penjualan produk pada sebuah perusahaan untuk memberikan rekomendasi pada pemilik perusahaan tentang produk mana yang prioritas diberi perhatian demi kemajuan perusahaan tersebut. Metode yang digunakan dalam penelitian ini yaitu *algoritma clustering x-means*. Dari penelitian yang sudah dilakukan telah didapatkan hasil dari metode Clustering algoritma X-Means yang menunjukkan bahwa dari dataset 333 jenis data penjualan dan transaksi pada tahun 2019. Penelitian ini menghasilkan eksperimen terhadap dataset yang diperoleh hasil sebagai berikut. Nilai k, akan optimal pada k = 2. Penggunaan operator Normalize berpengaruh terhadap *index Davies Bouldin*. Untuk dataset ini didapat bahwa tanpa menggunakan operator Normalize menghasilkan nilai *index davies bouldin* yang lebih baik yaitu 0,574.

Kata Kunci: Klasterisasi, algoritma xmeans, produk, penjualan

Abstract– *Transmart supermarket, is selling various kinds of products ranging from accessories, fashion, shoes, family needs and other products that are traded. Products are between products that sell well and products that don't sell well. The problem is not knowing how to analyze sales data grouping parameters using the X-means algorithm and not being able to interpret product sales groupings using the X-means clustering algorithm. The purpose of this study was to obtain knowledge of the level of product sales in a company to provide recommendations to company owners about which products should be given priority attention for the progress of the company. The method used in this research is the x-means clustering algorithm. From the research that has been done, the results have been obtained from the X-Means Algorithm Clustering method which shows that from a dataset of 333 types of sales and transaction data in 2019. This research produced experiments on the dataset which obtained the following results. The k value will be optimal at k = 2. The use of the Normalize operator affects the Davies Bouldin index. For this dataset it is found that without using the Normalize operator it produces a better Davies Bouldin index value of 0.574.*

Keywords: Clustering, k-means algorithm, grades, school exams

1. PENDAHULUAN

Penjualan yang semakin tinggi tersebut membutuhkan informasi dan identifikasi produk mana yang paling potensial dan produk mana yang kurang potensial pada penjualan. Dengan adanya pengelompokan pihak perusahaan dapat mengetahui barang yang sangat laris, cukup laris dan kurang laris. Sehingga barang yang ada digudang tidak menumpuk dan sesuai dengan kebutuhan. Dengan adanya pengolahan data yang dilakukan diharapkan akan dapat memberikan solusi nyata kepada pihak perusahaan. dan diharapkan pemilik perusahaan dapat melakukan pemasaran dengan strategi yang tepat untuk melayani kebutuhan konsumen. (Herianto, Nur Syamsiyah, 2021). *Algoritma X-means* merupakan algoritma yang digunakan untuk pengelompokan data. Algoritma x-means merupakan pengembangan dari *k-means*. *X-means clustering* digunakan untuk menyelesaikan salah satunya kelemahan utama dari *k-means clustering*, yaitu perlunya pengetahuan sebelumnya tentang jumlah cluster (K). Dalam metode ini, nilai sebenarnya dari K diperkirakan dalam suatu yang tidak diawasi cara dan hanya berdasarkan set data itu sendiri. (Putra et al., 2022)

Penelitian terdahulu yang dilakukan oleh Yogiswara Dharma Putra dkk dalam jurnalnya yang berjudul : “*Clustering History Data Penjualan Menggunakan Algoritma K-Means*” menyimpulkan bahwa algoritma K-Means pada 23 item data penjualan dari persediaan barang yang telah dilakukan dari tahun 2019 sampai 2020 dengan kategori barang laptop, komputer dan printer yang menghasilkan 3 cluster yang di interpretasikan sangat diminati, diminati dan kurang diminati. Dari 3 cluster yang terbentuk dengan titik centroid awal yaitu C1 (800, 580,200), C2 (500, 450, 300) dan C3 (100, 190, 95) maka proses akhir berhenti pada iterasi ke 4 yang mendapatkan data pada cluster 1 atau sangat diminati memiliki 5 barang, cluster 2 atau diminati memiliki 4 barang dan cluster 3 atau kurang diminati memiliki 14 barang yang memiliki titik centroid akhir yaitu C1 (911.8, 853, 58.8), C2 (442.5, 426.25, 16.25) dan C3 (97.85, 86, 11.85). (Yogiswara Dharma Putra1, Made Sudarma2, 2021). Usulan penelitian ini yaitu

menggunakan *Clustering* yang digunakan pada penelitian ini adalah *algoritma X-Means* yaitu sebuah algoritma yang dapat mendefinisikan objek dalam suatu pusat kelompok data yang biasanya menjadi titik-tengah dari kelompok data tersebut. Algoritma ini akan melakukan perulangan untuk mendapatkan titik-tengah yang optimal. *Algoritma X-Means* ini akan membagi klaster sesuai dengan jumlah klaster yang telah ditentukan atau diinisiasi diawal saat menjalankan algoritma ini. Dengan *algoritma Clustering* konstruksinya terhadap perusahaan yaitu dapat mengetahui pembagian klaster berdasarkan produk yang banyak laku dan kurang laku. (Fintri Indriyani1 & Irfiani, 2019)

Kajian yang diusulkan yaitu klasterisasi penjualan produk menggunakan algoritma x-means, yaitu untuk mengetahui informasi dan identifikasi produk mana yang paling potensial dan produk mana yang kurang potensial pada penjualan. Dengan adanya pengelompokan pihak perusahaan dapat mengetahui barang yang sangat laris, cukup laris dan kurang laris. Sehingga barang yang ada digudang tidak menumpuk dan sesuai dengan kebutuhan. Dengan adanya pengolahan data yang dilakukan diharapkan akan dapat memberikan solusi nyata kepada pihak perusahaan. dan diharapkan pemilik perusahaan dapat melakukan pemasaran dengan strategi yang tepat untuk melayani kebutuhan konsumen. (Ahmad Jurnaidi Wahidin1, 2021)

Tujuan dari penelitian ini adalah untuk mengetahui bagaimana cara menganalisa parameter pengelompokan produk penjualan menggunakan algoritma X-means dan untuk menginterpretasikan pengelompokan data produk penjualan menggunakan algoritma X-means clustering. Berdasarkan permasalahan data dan latarbelakang yang telah diutarakan maka penulis membuat penelitian dengan judul “Analisa Penerapan Metode *Clustering X-Means* Dalam Pengelompokan Penjualan Produk” dengan alasan yaitu ingin mengetahui dan memperoleh pengetahuan tingkat penjualan produk pada sebuah Transmart supermarket dalam memberikan rekomendasi pada pemilik perusahaan tentang produk mana yang prioritas diberi perhatian demi kemajuan perusahaan tersebut.

2. KERANGKA TEORI

2.1 Algoritma X-Means

Algoritma X-Means merupakan algoritma yang digunakan untuk pengelompokan data dan merupakan pengembangan dari *k-means*. *X-means clustering* digunakan untuk menyelesaikan salah satunya kelemahan utama dari *K-means clustering*, yaitu perlunya pengetahuan sebelumnya tentang jumlah cluster (K). (Achmad Fikri Sallaby1, Rizka Tri Alinse2, Venny Novita Sari3 & 123, 2022)

2.2 Klasterisasi

Clustering algoritma dapat diterapkan dalam berbagai bidang, misalnya:

- Pemasaran: mencari kelompok pelanggan yang mirip dengan perilaku, diberikan database yang besar berisi data pelanggan mereka memperoleh properti dan catatan masa lalu;

- a. Biologi: klasifikasi tanaman dan binatang;
- b. Perpustakaan: katalog buku;
- c. Asuransi: mengidentifikasi kelompok pemegang polis asuransi motor dengan rata-rata klaim biaya tinggi;
- d. Perencanaan kota: mengidentifikasi kelompok rumah sesuai dengan tipe rumah, nilai dan lokasi geografis; (Wakhidah, n.d.)

2.3 Data mining

Menurut Tan (2006) Dikutip Oleh Prasetyo (2012:2), mendefinisikan data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam data mining, pengelompokan data juga bisa dilakukan (Achmad Fikri Sallaby1, Rizka Tri Alinse2, Venny Novita Sari3 & 123, 2022)

2.4 Pengelompokan

Pengelompokan atau grouping adalah pengelompokan peserta didik berdasarkan karakteristik-karakteristiknya. Karakteristik demikian perlu digolongkan, agar mereka berada dalam kondisi yang sama. Adanya kondisi yang sama ini bisa memudahkan pemberian layanan yang sama. Oleh karena itu, pengelompokan (*grouping*) ini lazim dengan istilah pengklasifikasian (*clasification*). Pengelompokan siswa diadakan dengan maksud agar pelaksanaan kegiatan proses belajar mengajar di sekolah bisa berjalan lancar, tertib, dan bisa tercapai tujuan-tujuan pendidikan yang telah diprogramkan. (Harli et al., 2016).

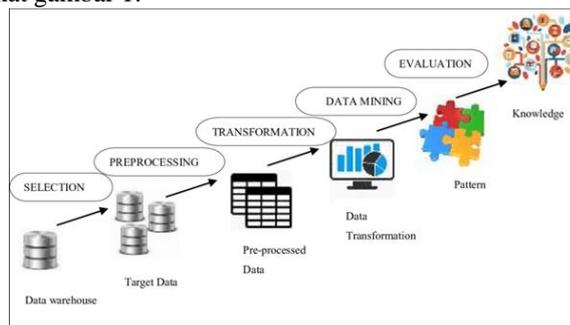
3. METODE PENELITIAN

Metode penelitian ini dalam menyusun skripsi menggunakan *algoritma X-means* sebagai berikut : *Algoritma klusterisasi* yang digunakan pada penelitian ini adalah *X- Means* yaitu *algoritma X-means* merupakan pengembangan dari *k-means cluster*. *X-means* membentuk cluster awal menggunakan *k-means*. Setiap cluster awal yang terbentuk dibagi menjadi dua cluster berdasarkan kriteria BIC. Proses ini berulang hingga tidak ada lagi cluster yang dapat dibagi. *X-means* membutuhkan komputasi yang lebih sedikit daripada *k-means* dan mampu mengoptimalkan jumlah cluster yang terbentuk. (Nurlita Kusuma Dewi, Dr. Gunardi, M.Si). *Algoritma X-Means bersumber dari jurnalourn Informatics, Information System, Software Engineering and Applications* merupakan algoritma yang menyempurnakan kekurangan dari algoritma *K-Means*, dimana algoritma *K-means* memiliki kelemahan dalam komputasi yang relative lama dan nilai kluster *K* yang harus ditentukan oleh user. Algoritma pembelajaran ini termasuk dalam kategori *unsupervised learning* dimana komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. *Algoritma x-means* mengoptimalkan nilai Bayesian Information Criterion (BIC). Komputasi pada pengoptimalan nilai BIC jauh lebih efisien daripada satu kali iterasi pada *algoritma k-means* (Pelleg and Moore, 2000).

Algoritma X-Means terdiri dari dua operasi yang berulang hingga eksekusi berakhir, yaitu:

1. *Improve – Params*
2. *Improve – Structure*
3. If , maka proses eksekusi berhenti dan mengeluarkan model *k* terbaik yang ditemukan.

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan *Knowledge Discovery in Databases* (KDD) yang terdiri dari, *Data Cleaning*, *Data transformation*, *Data mining*, *Pattern evolution*, *knowledge*, seperti terlihat gambar 1:



Gambar 1. Knowledge Discovery in Databases (KDD)

1. Selection
Untuk membaca dataset dalam bentuk file excel, menggunakan operator Read Excel dari tempat kerja. Data yang digunakan pada penelitian ini ialah sebanyak 333 dataset.
2. Preprocessing
Proses cleansing atau pembersihan data yang missing atau memiliki nilai yang tidak konsisten pada langkah preprocessing. Sebelum melakukan proses ini, dilakukan analisa terlebih dahulu apakah dataset memiliki nilai missing atau tidak serta memiliki data yang konsisten atau tidak
3. Data transformation
adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Pada langkah transformation dilakukan untuk data bertipe polynominal menjadi numerik agar data dapat diolah berdasarkan jarak. Untuk mengubah data bertipe polynominal menjadi numerik menggunakan operator Nominal to Numerical.
4. Data mining
Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses knowledge discovery in databases (KDD) secara keseluruhan. Tahap ini merupakan inti dari tahapan knowledge discovery in databases (KDD) yang dilakukan untuk menganalisis data yang telah dibersihkan.
5. Evaluation
Penerjemah alur yang dihasilkan dari data mining, alur informasi yang dihasilkan dari proses perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Proses ini merupakan proses KDD yang mencakup pemeriksaan apakah alur atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada pada sebelumnya.

4. HASIL

Populasi dan sampel penelitian ini yaitu bagaimana untuk mengelompokan data barang. Dalam pengujian pengklasifikasian ini yaitu menggunakan *Software RapidMiner Studio* dengan menggunakan pengujian metode *x-means clustering*. Untuk sampel yang digunakan pada penelitian ini adalah data barang pada tahun 2019 yang di jual di Transmart supermarket. Penelitian ini menggunakan teknik pengambilan dataset dengan total adalah 333 record data yang bersumber dari data hasil survey dan analisa dilapangn.

4.1 Dataset

Dalam pengujian pada machine learning ini sumber data utama yang digunakan dalam penelitian ini adalah dataset persentase dan akurasi data barang berdasarkan sebaran di di lapangan yang terdiri dari 333 record untuk data pada tahun 2019, Data yang diterima dalam bentuk dokumen *soft file* yang di sajikan seperti pada table 1.

Tabel 1: Data Set

NO	TGL	NOTA	PRODUCT	QTY	HARGA	SALES
1	03/04/2019	755	Udv miss cotton mask edp 75 ml	2	207000	Dian
2	03/04/2019	756	AB queen seduction edt 80 ml	1	352000	Dian
3	05/04/2019	757	Udv D'orient rubis edp 100 ml	2	228000	Rizqi
4	05/04/2019	401	Udv for man edt 100 ml	3	200000	Rizqi
5	07/04/2019	758	Aigner x-limited edt 100 ml	1	551000	Rizqi
6	12/04/2019	759	Udv for man edt 100 ml	4	200000	Dian
7	13/04/2019	760	Beyonce heat rush edt 100 ml	1	585000	Dian
8	13/04/2019	402	Aigner x-limited edt 100 ml	3	551000	Dian
9	13/04/2019	762	Aigner blue emotion edt 100 ml	2	805000	Rizqi
10	13/04/2019	403	Ab blue seduction women 80 ml	2	352000	Rizqi
100	25/05/2019	432	Addidas UEFA champion ede 100ml	3	100000	Dian
.....
333	25/05/2019	431	Gucci mtm edt 90 ml Free	3	1980000	Rizqi

4.2 Selection

Untuk membaca dataset dalam bentuk *file excel*, menggunakan operator *Read Excel* seperti tampak pada gambar 2: dibawah ini.



Gambar 2. Operator Read Excel pada Rapidminer

4.3. Preprocessing

Proses *cleansing* atau pembersihan data yang *missing* atau memiliki nilai yang tidak konsisten pada langkah *preprocessing*. Sebelum melakukan proses ini, dilakukan analisa terlebih dahulu apakah dataset memiliki nilai *missing* atau tidak serta memiliki data yang konsisten atau tidak. Dari hasil *result* dari statistik dataset seperti tampak pada gambar 3 dibawah ini, diketahui bahwa tidak ada atribut yang memiliki nilai *missing* karena memiliki nilai 0

pada kolom *missing*. Untuk memeriksa konsisten atau tidak konsistennya dataset yang digunakan diperiksa per-record secara langsung dan menunjukkan bahwa dataset memiliki data yang konsisten terhadap nilainya.

Field	Type	Min	Max	Average
NOTA	Integer	0	401	710.511
TGL	Date	0	Earliest date: Apr 3, 2019 Latest date: May 31, 2019	Duration: 58 days
PRODUCT	Polynomial	0	Least: ck eupho [.] 100 ml (1) Most: Adidas g [.] 10 ml (13)	Values: Adidas g [.] di 100 ml (13), Adidas p [.] di 100 ml (11), ... [203 more]
QTY	Integer	0	1	6
HARGA	Integer	0	450000	4350000
SALES	Polynomial	0	Least: Rizq (148) Most: Dam (185)	Values: Dian (185), Rizq (148)
TOTAL	Integer	0	190000	4702000

Gambar 3. Result dari statistik dataset

Pada langkah *Preprocessing*, tidak dilakukan apa-apa karena data dianggap sudah konsisten dan tidak memiliki nilai yang *missing*.

4.4 Transformation

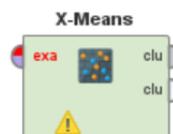
Pada langkah *transformation* dilakukan untuk data bertipe *polynomial* menjadi numerik agar data dapat diolah berdasarkan jarak. Untuk mengubah data bertipe *polynomial* menjadi numerik menggunakan operator *Nominal to Numerical* seperti pada gambar 4 dibawah ini.



Gambar 4. Operator *Nominal to Numerical* pada Rapidminer

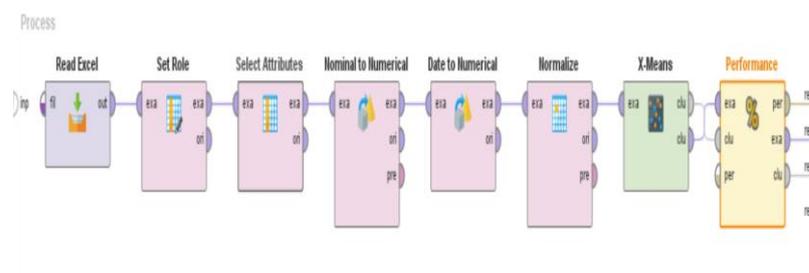
4.5 Data mining

Pada langkah data mining, dilakukan pengelompokan data menggunakan algoritma x-means sehingga menggunakan operator x-means seperti tampak pada gambar 5 dibawah ini.



Gambar 5. Operator X-Means pada Rapidminer

Model proses pada rapidminer sampai dengan langkah *Data Mining* tampak pada gambar 6 dibawah ini.



Gambar 6. Model proses sampai dengan langkah Data Mining

4.6 Evaluation

Evaluasi yang dilakukan terhadap hasil kegiatan eksperimen terhadap dataset yang diperoleh hasil sebagai berikut:

- Nilai k, akan optimal pada k = 2

- b. Penggunaan operator Normalize berpengaruh terhadap *index Davies Bouldin*. Untuk dataset ini didapat bahwa tanpa menggunakan operator Normalize menghasilkan nilai *index Davies Bouldin* yang lebih baik yaitu 0,574.
- c. Pada hasil eksperimen diperoleh bahwa untuk mendapatkan nilai Dbi yang konfiden (tetap) dimulai pada iterasi 8 dengan nilai Dbi sebesar 0,574

5. KESIMPULAN

Berdasarkan hasil penelitian pengelompokan Dataset barang menggunakan algoritma *X-Means*, maka dapat disimpulkan sebagai berikut :

- a. Dapat menganalisa pengelompokan barang menggunakan algoritma *x-means*, Dengan Cluster 0 memiliki jumlah anggota sebanyak 271 item
- b. Dapat mengimplementasikan pengelompokan dataset barang menggunakan *analisis x-means clustering*. Cluster 1 memiliki jumlah anggota sebanyak 62 item.

DAFTAR PUSTAKA

- Achmad Fikri Sallaby¹, Rizka Tri Alinse², Venny Novita Sari³, T. R., & 123. (2022). PENGELOMPOKAN BARANG MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN HASIL PENJUALAN DI TOKO WIDYA BENGKULU. *Jurnal Media Infotama*, 18(1), 99–104.
- Ahmad Jurnaidi Wahidin¹, D. I. S. (2021). Perbandingan Algoritma K-Means , X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *Jurnal ICT : Information Communication & Technology Vol.*, 20(2), 298–302.
- Bibit Sudarsono, Umi Faddillah, Ayuni Asistiyasari, Y. N. (2022). *Klasifikasi Data Penjualan Menggunakan Algoritma K-Means Dan Analytic Hierarchy Process*. 2(1), 23–28.
- Dr. Ratna Wijayanti Daniar Paramita, S.E.,M.M. Noviansyah Rizal, S.E.,M.M.,Ak, CA, CFrA. Riza Bahtiar Sulistyan, S.E., M. M. (2021). *Metode penelitian kuantitatif* (p. 167).
- Fintri Indriyani¹, E. I., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K-Means Method). *JUITA: Jurnal Informatika*, 7(November), 109–113.
- Harli, E., Fauzi, A., & Kusmanto, T. H. (2016). *Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok*. 2(2), 90–95.
- Herianto, Nur Syamsiyah, S. A. (2021). ANALISA TINGKAT PENJUALAN PRODUK MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN) DAN K-MEANS. *Program Studi Sistem Informasi Universitas Darma Persada*, XI(1), 8–18.
- Putra, P. H., Hasibuan, A., & Marpaung. (2022). Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 19(2), 424–429.
- Rifki Adhitama #1, Auliya Burhanuddin *2, A. F. #3, & 1Program. (2021). *Penerapan X Means Clustering Pada UMKM Kab Banyumas Yang Mendukung Mega Shifting Consumer Behavior Akibat Covid- 19*. 8106, 71–80.
- Rifki Adhitama¹, Auliya Burhanuddin², R. A. (2020). *PENENTUAN JUMLAH CLUSTER IDEAL SMK DI JAWA TENGAH DENGAN METODE X-MEANS CLUSTERING DAN K-MEANS CLUSTERING*. 3(1), 1–5. <https://doi.org/10.33387/jiko>
- Riza, A. A., Retno, D., Saputro, S., & Maret, S. (2022). CLUSTERING DATA NUMERIK MENGGUNAKAN ALGORITME X-MEANS. *Seminar Nasional Matematika, Geometri, Statistika, Dan Komputasi SeNa-MaGeStiK 2022*.
- Sulistiyawati, A., & Supriyanto, E. (2020). *Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan*. 15(2), 25–36.
- Wakhidah, N. (n.d.). *CLUSTERING MENGGUNAKAN K-MEANS ALGORITHM (K-MEANS ALGORITHM CLUSTERING)*.
- Wijayanto, A., & Adhitama, R. (2019). *Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas*. 8106.
- Wiwik Dian Astuti^{1*}, Saleh Hidayat², Haryadi³. (2020) Hubungan Nilai Ujian Nasional, Nilai Tes Ujian Masuk Sekolah, Dan Tingkat Perekonomian Keluarga Dengan Hasil Belajar Biologi Siswa
- Yusma Elda¹, Dkk. (2021). Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means. *Jurnal Informasi dan Teknologi*, 103-108.