

Implementasi Algoritme *K-Means++* Untuk *Clustering* Penjualan Bahan Bangunan

Mohammad Ferdiansyah , Umi Chotijah

Teknik Informatika Universitas Muhammadiyah Gresik

Jl. Sumatra No. 101 GKB Gresik 61121

Korespondensi penulis: ferdiabidzar@gmail.com

Abstract. Utilization of the *K-Means++* Clustering algorithm in the sales of building materials clustering system at UD Sumber Bangunan. Currently, the store does not use computers to run its system, resulting in transaction data being used merely as archives without being optimally utilized for marketing strategies and business decision-making. This research aims to examine whether the use of the *K-Means++* Clustering algorithm can provide advantages in forming better and more efficient clusters for building material sales data. Evaluation is conducted by comparing *K-Means++* with *K-Means* using the evaluation metrics DBI (Davies-Bouldin Index) and Silhouette Coefficient. The evaluation results show that *K-Means++* outperforms using *K-Means* alone. The DBI values are lower for *K-Means++*, and the Silhouette Coefficient is higher for *K-Means++*, indicating that *K-Means++* produces better-defined clustering. The utilization of the *K-Means++* Clustering algorithm provides benefits in business decision-making at UD Sumber Bangunan, assisting in reducing stockpiles and enhancing customer satisfaction. Additionally, the clustering system built using the Waterfall method also contributes positively to achieving the set objectives..

Keywords: *K-Means++* Clustering Algorithm; Davies-Bouldin Index (DBI); Silhouette Coefficient; Clustering; Data Mining; UD Sumber Bangunan.

Abstrak. Pemanfaatan Algoritme *K-Means++* Clustering dalam sistem clustering penjualan bahan bangunan pada UD Sumber Bangunan. Saat ini, toko tersebut belum menggunakan komputer untuk menjalankan sistemnya, sehingga data-data transaksi hanya berfungsi sebagai arsip tanpa dimanfaatkan secara optimal untuk strategi pemasaran dan pengambilan keputusan bisnis. Penelitian ini bertujuan untuk menguji apakah penggunaan algoritme *K-Means++* Clustering dapat memberikan keuntungan dalam pembentukan cluster yang lebih baik dan efisien untuk data penjualan bahan bangunan. Evaluasi dilakukan dengan membandingkan *K-Means++* dengan *K-Means* menggunakan metrik evaluasi DBI (Davies-Bouldin Index) dan Silhouette Coefficient. Hasil evaluasi menunjukkan bahwa *K-Means++* memberikan hasil yang lebih baik dibandingkan hanya menggunakan *K-Means*. Nilai DBI lebih rendah untuk *K-Means++* dan Silhouette Coefficient lebih tinggi untuk *K-Means++*, yang menandakan bahwa *K-Means++* menghasilkan clustering yang lebih baik dan lebih terdefinisi. Penggunaan Algoritme *K-Means++* Clustering memberikan manfaat dalam pengambilan keputusan bisnis di UD Sumber Bangunan, sehingga membantu mengurangi penumpukan stok dan meningkatkan kepuasan konsumen. Sistem clustering yang dibangun dengan metode Waterfall juga memberikan kontribusi positif dalam mencapai tujuan yang telah ditetapkan.

Kata kunci: Algoritme *K-Means++*; Davies-Bouldin Index (DBI); Silhouette Coefficient; Clustering; Data Mining; UD Sumber Bangunan;

LATAR BELAKANG

Di era industri 4.0 kemajuan dibidang teknologi informasi dan komunikasi saat ini berkembang pesat disegala bidang seperti pada bidang industri, ekonomi, pendidikan, konstruksi dan lain sebagainya. Sehingga dengan kemajuan teknologi manusia dimudahkan dalam hal mencari informasi yang ingin didapatkan. Teknologi informasi sangat dibutuhkan untuk pelaku usaha sebagai tolak ukur dalam memutuskan keputusan yang dapat memberikan dampak bagi keberlangsungan bisnis. Kita dituntut untuk senantiasa mengembangkan bisnis agar selalu bertahan dalam persaingan, dalam pengembangan bisnis dibutuhkan sebuah strategi

Received: Desember 20, 2023; Accepted: Januari 15, 2024; Published: Maret 30, 2024

* Mohammad Ferdiansyah, ferdiabidzar@gmail.com

untuk meningkatkan kualitas penjualan seperti meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan. Namun hal ini belum sepenuhnya dimanfaatkan oleh pemilik UD Sumber Bangunan.

UD Sumber Bangunan merupakan sebuah usaha yang bergerak dalam bidang penjualan bahan bangunan, tetapi dalam hal bisnis UD Sumber Bangunan belum menggunakan komputer untuk menjalankan sistemnya, data-data yang terlibat dalam setiap transaksi pada toko ini, seperti data item yang dibeli untuk memenuhi stok, jumlah item yang dibeli oleh konsumen, data item yang terjual dan data konsumen. Karena banyaknya data yang hanya dibiarkan begitu saja dan hanya berguna sebagai arsip toko sehingga kerap timbul masalah seperti penumpukan stok dan kepuasan konsumen. Padahal dengan data yang ada, pemilik toko bisa memanfaatkan untuk strategi pemasaran. Data-data tersebut bisa dijadikan sebagai alat dalam mengambil keputusan untuk solusi bisnis dengan dukungan sebuah teknologi data mining. Data mining adalah proses penemuan pola yang tidak diketahui sebelumnya dalam data besar dan kompleks (Jiawei Han, Micheline Kamber, dan Jian Pei 2011). Informasi berupa data transaksi yang besar membutuhkan sebuah sistem yang secara otomatis mencari hubungan antar item dalam data diDatabase. Dapat dicontohkan dari toko UD Sumber Bangunan dari data yang diperoleh dari pemilik toko.

KAJIAN TEORITIS

1. Data Mining

Data mining merupakan suatu metode untuk mengumpulkan informasi yang belum ada sebelumnya dari suatu basis data dengan menganalisis pola dan karakteristik di masa mendatang. Data mining mengeksplorasi pengetahuan dan pola dalam data melalui statistik matematika dan machine learning (Sekar Setyaningtyas et al., 2022).

Terdapat empat tugas utama data mining, yaitu sebagai berikut (Adani et al., n.d.):

a. *Predictive Modelling*

Predictive modelling (pemodelan prediktif) digunakan untuk membangun model untuk target variabel sebagai fungsi dari *explanatory variable* (variabel penjelas).

b. *Association Analysis*

Association analysis (analisis asosiasi) digunakan untuk menemukan aturan asosiasi yang memperlihatkan atau menunjukkan kondisi dari nilai atribut yang sering muncul secara bersamaan dalam suatu kumpulan data.

c. *Cluster Analysis*

Berbeda dengan klasifikasi yang menganalisis kelas objek data yang berisi label. *Clustering* menganalisis objek data tanpa memeriksa label kelas yang diketahui.

d. *Anomaly Detection*

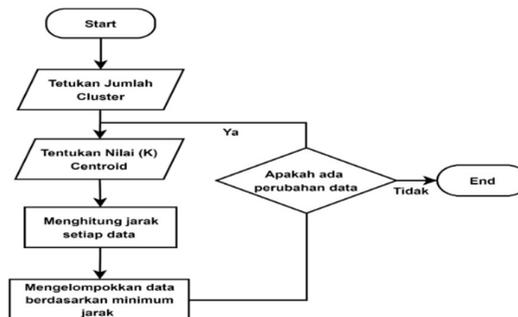
Anomaly detection (deteksi anomaly) adalah metode deteksi suatu data dimana tujuannya yaitu menemukan objek yang berbeda dari kebanyakan objek lain.

2. *Clustering*

Clustering adalah proses pembagian data kedalam kelas-kelas atau disebut *cluster* berdasarkan tingkat kesamaannya. Tujuan dari *clustering* yaitu untuk mengelompokkan karakteristik yang sama kedalam area yang sama dan data karakteristik yang tidak sama untuk menerima kelompok objek yang memiliki karakteristik yang sama (Luh et al., 2022).

3. Algoritme K-Means

K-Means merupakan salah satu Algoritme *clustering* yang digunakan dalam proses data mining. Algoritme *K-Means* bekerja dengan cara mempartisi pengelompokan yang memisahkan data kedalam kelompok-kelompok tertentu dengan meminimalkan jarak rata-rata setiap klaster data (Triyandana et al., 2022).



Gambar 1 Flowchart Algoritme K-Means

Tahapan Algoritme K-Means:

1. Menentukan jumlah *cluster* (k) yang akan digunakan.
2. Menentukan nilai yang akan dijadikan *centroid*.
3. Menghitung jarak setiap titik data ke setiap *centroid* menggunakan *Manhattan*.
4. Mengelompokkan data berdasarkan jarak terdekat antara data dengan *centroid*.
5. Ulangi langkah menghitung jarak menggunakan teori *Manhattan*. hingga tahap penentuan nilai *cluster* baru. Tahap atau fase pengulangan ini akan dihentikan jika tidak ada perubahan data pada suatu *cluster*.

Untuk menentukan nilai centroid baru dapat dilakukan dengan menggunakan persamaan 1.1

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i$$

Keterangan:

n_k = jumlah semua data yang ada didalam *cluster* (k).

d_i = jumlah dari setiap *cluster*

4. Algoritme K-Means++

Metode K-Means++ digunakan untuk mengatasi masalah dalam pemilihan cluster yang secara acak sehingga mengakibatkan peningkatan jumlah iterasi.

Langkah – langkah untuk metode K-Means++:

1. Menentukan titik centroid secara acak.
2. Menghitung jarak semua titik dalam dataset dari centroid yang dipilih. Jarak titik titik x_i dari *centroid* terjauh dapat dihitung menggunakan persamaan 1.2

$$d_i = \max_{(j:1 \rightarrow m)} \| x_i - C_j \|^2$$

3. Menjadikan titik x_i sebagai *centroid* baru yang memiliki probabilitas maksimum.
 4. Ulangi langkah 2-3 sampai titik *centroid* pada setiap *cluster*(k) terpenuhi.
 5. Setelah mendapatkan *centroid*, langkah selanjutnya sama dengan metode *K-Means*.
- #### 5. Euclidean Distance

Euclidian Distance merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik yang dalam euclidian space (meliputi bidang euclidian dua dimensi, tiga dimensi atau bahkan lebih) (Pribadi et al., 2022).

6. Davies Bouldin Index (DBI)

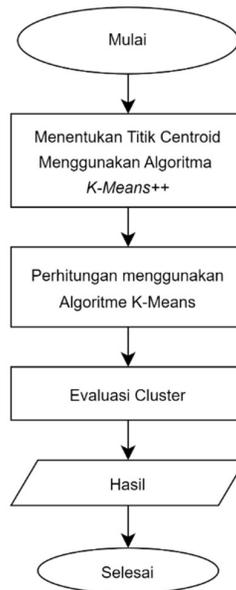
Davies Bouldin Index adalah salah satu metode evaluasi internal yang digunakan untuk mengukur evaluasi cluster yang didapatkan pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi

7. Silhouette Coefficient

Silhouette Coefficient adalah pengabungan dari metode cohesian dan separation yang merupakan metode evaluasi untuk Cluster.

METODE PENELITIAN

1. Tahapan pada perhitungan Algoritme K-Means++



Gambar 2 Tahapan Perhitungan k-Means++

Adapun tahapan perhitungan Algoritme K-Means++ sebagai berikut :

1. Menentukan Titik Centroid Menggunakan Algoritme K-Means++

Menentukan titik awal centroid sebelum dilakukan proses perhitungan.

a. Menentukan titik awal centroid secara acak

| No | Stok Awal | Jumlah Terjual | Stok Akhir |
|----|-----------|----------------|------------|
| 14 | 63 | 18 | 45 |

b. Menghitung jarak semua titik dalam dataset dari centroid yang dipilih dengan persamaan

1.2

$$d_i = (63 - 8)^2 + (18 - 3)^2 + (45 - 5)^2 = 4850$$

| No | Stok Awal | Jumlah Terjual | Stok Akhir | Hasil |
|-----|-----------|----------------|------------|-------|
| 1 | 8 | 3 | 5 | 4850 |
| 2 | 20 | 4 | 16 | 2886 |
| 3 | 23 | 19 | 4 | 3282 |
| ... | ... | ... | ... | ... |
| 40 | 30 | 24 | 6 | 2646 |

- c. Menjadikan titik x_i sebagai *centroid* baru yang memiliki probabilitas maksimum, nilai probabilitas maksimum ada pada data no 14 yaitu 5138. Maka data ke 14 menjadi *centroid* 2.

| No | Stok Awal | Jumlah Terjual | Stok Akhir |
|----|-----------|----------------|------------|
| 10 | 63 | 18 | 45 |
| 14 | 6 | 1 | 5 |

- d. Mengulangi langkah 2-3 hingga semua (k) terpenuhi.

| No | Stok Awal | Jumlah Terjual | Stok Akhir |
|----|-----------|----------------|------------|
| 10 | 63 | 18 | 45 |
| 14 | 6 | 1 | 5 |
| 1 | 8 | 3 | 5 |

2. Perhitungan Menggunakan Algoritme K-Means Clustering

menghitung jarak antara data dengan pusat awal cluster digunakanlah teori *Euclidean Distance*. Berikut ini adalah contoh dari perhitungan jarak data nomor 1 terhadap pusat awal *cluster*. Disini data nomor 1 menjadi x_1 . Sementara untuk *Cluster* 1 dan 2 masing masing menjadi c_1 dan c_2 ,

$$d(c_1, x_1) = \sqrt{(63 - 8)^2 + (18 - 3)^2 + (45 - 5)^2} = 69,642$$

$$d(c_2, x_1) = \sqrt{(8 - 8)^2 + (3 - 3)^2 + (5 - 5)^2} = 0$$

Setelah dilakukan perhitungan jarak data terhadap *centroid*, selanjutnya akan dipilih jarak terkecil setiap data dengan *centroid* terdekat.

| Data No | C1 | C2 | C1 | C2 |
|---------|---------|---------|----|----|
| 1 | 69.6419 | 0 | | v |
| 2 | 53.7215 | 16.3095 | | v |
| 3 | 57.2887 | 21.9545 | | v |
| 4 | 61.1065 | 10.198 | | v |
| 5 | 54.461 | 23.3666 | | v |
| 6 | 58.1893 | 16.3095 | | v |
| 7 | 48.9285 | 38.9102 | | v |
| 8 | 50.2195 | 24.1247 | | v |
| 9 | 49.0714 | 20.9284 | | v |
| 10 | 0 | 69.6419 | v | |

| | | | | |
|-----|---------|---------|-----|-----|
| ... | ... | ... | ... | ... |
| 40 | 51.4393 | 30.4302 | | v |

Kemudian pusat cluster baru akan dihitung berdasarkan data anggota setiap cluster yang sudah didapatkan. Berikut adalah contoh untuk menghitung *centroid* baru pada cluster 1.

$$c_{1Stok\ Awal} = (63 + 65 + 75 + 55 + 42 + 45)/6 = 57.5$$

$$c_{1Jumlah\ Terjual} = (18 + 40 + 56 + 23 + 16 + 21)/6 = 29$$

$$c_{1Stok\ Akhir} = (45 + 25 + 19 + 32 + 26 + 24)/6 = 28.5$$

Dari perhitungan diatas maka dapat dilanjutkan dengan menghitung centroid cluster 2.

Hasil yang didapatkan pusat cluster baru 1\

| Data No | Stok awal | Jumlah terjual | Stok akhir |
|---------|-----------|----------------|------------|
| C1 | 57.5 | 29 | 28.5 |
| C2 | 17.44118 | 10.67647 | 6.852941 |

Iterasi selanjutnya dilakukan dengan cara yang sama sampai tidak ada perubahan data dalam suatu *cluster*.

3. Evaluasi Cluster

Pada tahap ini dilakukan tahap evaluasi *cluster* atau mempertimbangkan hasil Algoritme *clustering*, dengan menggunakan 2 metode evaluasi *cluster*.

1. *Davies Bouldin Index* (DBI)
2. *Sihouette Coefficient*

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini berupa sebuah web dan pengetahuan terkait perbedaan K-Means tanpa K-Means++ dan menggunakan K-Means++. Pembuatan web nya sendiri menggunakan bahasa program PHP didukung database MySQL.

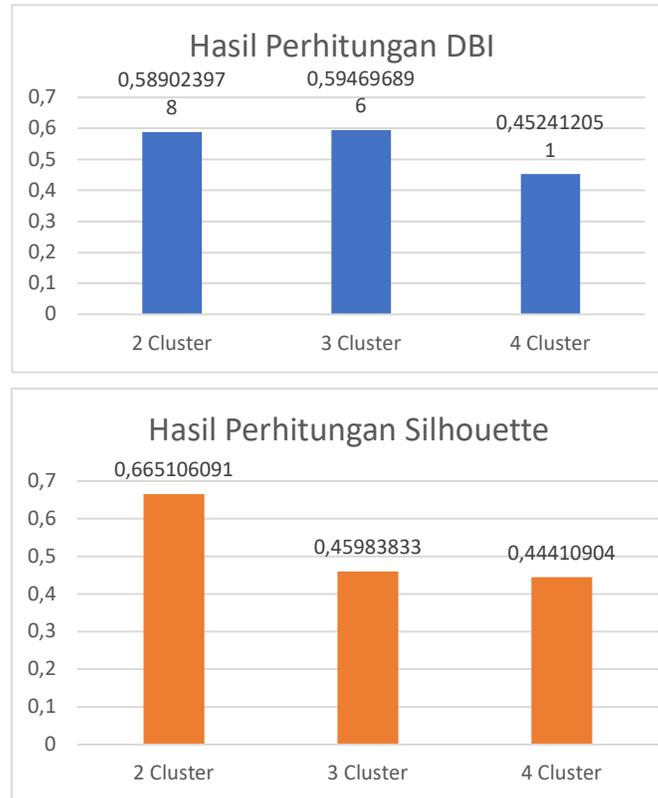
1. Halaman Login

Halaman login merupakan halaman pertama yang muncul pada sistem. Pengguna harus memasukkan *username* dan *password* yang sesuai agar dapat masuk ke halaman utama dari sistem

Pengujian Sistem

1. Grafik Penentuan Titik Centroid

Pada penentuan titik *centroid* terbaik dipilih antara 2 *cluster*, 3 *cluster* dan 4 *cluster* dilakukan dengan menggunakan *Davies Bouldin Index (DBI)* dan *Silhouette Coefficient*.



Berdasarkan hasil perhitungan *Davies Bouldin Index* didapatkan bahwa 4 *cluster* memiliki nilai terbaik (non negatif ≥ 0) dengan nilai 0.4524 dan hasil perhitungan *Silhouette Coefficient*, didapatkan 2 *cluster* karena berdasarkan tabel interpretasi nilai memiliki struktur baik dengan nilai 0.665. dan berdasarkan nilai yang didapatkan peneliti memilih 2 *cluster* menjadi *centroid* terbaik untuk digunakan diperhitungan selanjutnya.

2. Hasil Pengujian Davies Bouldin Index (DBI)

Tahapan pertama untuk mencari nilai DBI yakni dengan mencari nilai SSW, selanjutnya mencari nilai SSB, sebelum mencari nilai SSB dibutuhkan centroid pada iterasi yang didapat, setelah didapatkan nilai SSW dan SSB selanjutnya mencari nilai rasio, selanjutnya apabila telah mendapatkan nilai rasio kemudian menghitung nilai DBI

a. Hasil Pengujian Davies Bouldin Index (DBI) K-Means++

| Cluster | SSW |
|---------|----------|
| C1 | 18.52796 |
| C2 | 10.38259 |

$$SSB_{1,2} = \sqrt{(57.5 - 17.441)^2 + (29 - 10,676)^2 + (28,5 - 6,8529)^2 + \dots}$$

$$SSB_{1,2} = 49.0821$$

$$R_1 = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{18,528 + 10,382}{49,082} = 0,5890$$

$$DBI = \frac{R_1}{k} = \frac{0,5890}{1} = 0,5890$$

b. Hasil Pengujian Davies Bouldin Index (DBI) K-Means

| Cluster | SSW |
|---------|----------|
| C1 | 18.52796 |
| C2 | 10.38259 |

$$SSB_{1,2} = \sqrt{(52 - 16.3125)^2 + (28.875 - 9.5625)^2 + (23.125 - 6.8437)^2 + \dots}$$

$$SSB_{1,2} = 43.72241$$

$$R_1 = \frac{SSW_1 + SSW_2}{SSB_{1,2}} = \frac{20.53573 + 9.175693}{43.72241} = 0.67954$$

$$DBI = \frac{R_1}{k} = \frac{0,6795}{1} = 0,6795$$



Hasil dari perhitungan DBI Menggunakan Algoritma K-Means menggunakan K-Means++ dan tanpa menggunakan K-Means++ didapatkan bahwa penggunaan K-Means++ mendapatkan nilai 0.5890. angka tersebut memiliki hasil yang cukup baik karena mendekati angka 0. Dari perhitungan *Davies Bouldin Index (DBI)* dapat disimpulkan bahwa jika semakin

kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (*non negatif* ≥ 0) maka *cluster* tersebut semakin baik.

3. Uji Validitas Silhouette Coefficient



Berdasarkan hasil perhitungan dengan pemilihan centroid secara acak nilai *Silhouette Coefficient* mendapatkan hasil 0.6652 dari 2 *cluster*. Sesuai dengan tabel interpretasi nilai *Silhouette Coefficient* yang ditunjukkan dalam tabel 2.1 menghasilkan hasil struktur baik.

KESIMPULAN DAN SARAN

Penentuan titik *centroid* menggunakan *K-Means++* memberikan hasil yang lebih baik dibandingkan hanya dengan *K-Means*. Hal ini dapat dilihat dari nilai *DBI* lebih rendah untuk *K-Means++* dan *Silhouette Score* yang lebih tinggi untuk *K-Means++*, berdasarkan hasil evaluasi menunjukkan bahwa *K-Means++* menghasilkan yang lebih baik dan lebih terdefinisi dibandingkan *K-Means*. Sedangkan proses clustering, penentuan titik awal atau centroid sangat berpengaruh terhadap hasil dari clustering, peneliti sendiri menggunakan Algoritme *K-Means++* untuk mengatasi masalah pemilihan pusat cluster awal agar hasil menjadi optimal. dan juga dilakukan evaluasi cluster untuk menentukan jumlah centroid terbaik.

DAFTAR REFERENSI

- Annisa, K., Ginting, B. S., & Syari, M. A. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat. *Jurnal Ilmu Komputer Dan Informatika Volume*, 6(1), 113–129.
- Annur, H. (2019). Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo). *Jurnal Informatika Upgris*, 5(1). <https://doi.org/10.26877/jiu.v5i1.3091>
- Ardiansyah, A. H., Nugroho, W., Alfiyah, N. H., Handoko, R. A., & Bakhtiar, M. A. (2020). Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan Status Provinsi di Indonesia 2020. *Seminar Nasional Inovasi Teknologi*, 4(3), 329–333.

- Budiman, R., & Rudianto. (2019). Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering). *ProTekInfo(Pengembangan Riset Dan Observasi Teknik Informatika)*, 6(1), 6. <https://doi.org/10.30656/protekinfo.v6i1.1691>
- Gustrianda, R., & Mulyana, D. I. (2022). Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids. *Jurnal Media Informatika Budidarma*, 6(1), 27. <https://doi.org/10.30865/mib.v6i1.3294>
- Irawan, Y. (2019). PENERAPAN DATA MINING UNTUK EVALUASI DATA PENJUALAN MENGGUNAKAN METODE CLUSTERING DAN ALGORITMA HIRARKI DIVISIVE. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, 4(1), 13–20. <https://doi.org/10.20527/jtiulm.v4i1.34>
- Magdalena, L., & Fahrudin, R. (2020). Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM. *Jurnal Digit*, 9(2), 190. <https://doi.org/10.51920/jd.v9i2.120>
- Nasir, J. (2021a). PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(2), 690–703. <https://doi.org/10.24176/simet.v11i2.5482>
- Normah, Nurajizah, S., & Salbinda, A. (2021). Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten. *Jurnal Teknik Komputer AMIK BSI*, 7(2), 174–180. <https://doi.org/10.31294/jtk.v4i2>
- Novianto, R., & Goeirmento, L. (2019). Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(1), 85–95. <https://doi.org/10.35957/jatisi.v6i1.150>
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153. <https://doi.org/10.33060/jik/2020/vol9.iss2.181>
- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI KABUPATEN MALANG. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2).
- Ramadanti, E., & Muslih, M. (2022). Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(1), 1–7. <https://doi.org/10.36341/rabit.v7i1.2155>
- Sri Fastaf, C. A., & Yamasari, Y. (2022a). Analisa Pemetaan Kriminalitas Kabupaten Bangkalan Menggunakan Metode K-Means dan K-Means++. *Journal of Informatics and Computer Science (JINACS)*, 3(04), 534–546. <https://doi.org/10.26740/jinacs.v3n04.p534-546>
- Tamba, S. P., Kesuma, F. T., & Feryanto. (2019). PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 2(2), 67–72. <https://doi.org/10.34012/jusikom.v2i2.376>

- Triyandana, G., Putri, L. A., & Umaidah, Y. (2022). Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means. *Journal of Applied Informatics and Computing*, 6(1), 40–46. <https://doi.org/10.30871/jaic.v6i1.3824>
- Wahyuni, W. A., & Saepudin, S. (2021). Penerapan Data Mining Clustering Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci. *Seminar Nasional Sistem ...*, 306–313.