

LAPORAN AKHIR



**K-MEANS ALGORITHM IN CLUSTERING SALES DATA FOR
CALCULATING ESTIMATED HOUSE PRICES**

GATOT TRI PRANOTO

UNIVERSITAS TRILOGI

2024

ABSTRAK

Penetapan nilai jaminan kepada Bank dalam proses pengajuan Kredit Pemilikan Rumah (KPR) yang diajukan calon nasabah masih mengacu pada ketentuan Otoritas Jasa Keuangan, dimana penilaiannya harus mengikuti aturan yang ada dan dilakukan oleh Bank. penilai publik atau biasa disebut Kantor Pelayanan Penilai Publik (KJPP). Saat ini petugas kredit analis belum dapat memvalidasi hasil laporan penilaian dari KJPP, sehingga apabila terjadi kesalahan baik disengaja maupun tidak oleh KJPP atau pihak penilai tetap memproses sesuai nilai yang diberikan. Apabila nasabah gagal bayar karena nilai agunan pinjaman yang diberikan lebih rendah, maka bank melanggar Peraturan Bank Indonesia nomor 18/16/PBI/2016 tentang rasio pinjaman terhadap nilai. Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan penjualan rumah sehingga dapat digunakan untuk perhitungan nilai estimasi harga rumah, dan mengembangkan prototipe sistem informasi estimasi harga rumah. Pengambilan data dengan menggunakan teknik crawling atau scrapping pada website memudahkan dalam pemenuhan data pada suatu dataset. Hasil dari penelitian ini adalah data penjualan rumah kawasan Kebon Jeruk yang tersebar di internet dapat dikelompokkan menjadi 3 cluster dengan nilai David Bouldin Index di sub area Dur Kepa yaitu sebesar 0,096, di sub area Kedoya Selatan sebesar 0,087, di sub area Kedoya Utara sebesar 0,071, dan sub area Kelapa Dua sebesar 0,117. Dengan menggabungkan hasil klasterisasi menggunakan metodologi K-Means dengan rumus perhitungan harga tanah diperoleh estimasi harga tanah pada sub area.

Kata kunci: K-Means, KPR, Data Scraping, KJPP, MAPPI.

KATA PENGANTAR

Puji syukur dipanjatkan kepada Tuhan Yang Maha Esa karena atas ridho-Nya penelitian ini bisa diselesaikan pada waktunya.

Studi ini berjudul “K-MEANS ALGORITHM IN CLUSTERING SALES DATA FOR CALCULATING ESTIMATED HOUSE PRICES”. Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam mengelompokkan penjualan rumah sehingga dapat digunakan untuk perhitungan nilai estimasi harga rumah, dan mengembangkan prototipe sistem informasi estimasi harga rumah. Pengambilan data dengan menggunakan teknik crawling atau scrapping pada website memudahkan dalam pemenuhan data pada suatu dataset. Hasil dari penelitian ini adalah data penjualan rumah kawasan Kebon Jeruk yang tersebar di internet dapat dikelompokkan menjadi 3 cluster dengan nilai David Bouldin Index.

Jakarta, 31 Agustus 2024

Penulis

DAFTAR ISI

ABSTRAK.....	ii
KATA PENGANTAR.....	iii
DAFTAR ISI.....	iv
1 PENDAHULUAN	1
2 METODE PENELITIAN	3
3 HASIL DAN PEMBAHASAN	5
4 PENUTUP	10
DAFTAR PUSTAKA	11

1 PENDAHULUAN

Pertumbuhan industri keuangan saat ini sangat pesat, khususnya pada dunia industri keuangan seperti perbankan. industri keuangan seperti perbankan. Peningkatan pertumbuhan tersebut juga dibarengi dengan pertumbuhan Non Performing Loan (NPL) subsektor rumah tangga kepemilikan rumah atau biasa disebut Kredit Pemilikan Rumah (KPR).

Dampak NPL dapat mengakibatkan berkurangnya pendapatan bank, pengurangan tersebut timbul karena adanya tambahan biaya yang timbul. Pengurangan timbul karena adanya tambahan biaya yang timbul akibat kredit bermasalah (Sari, 2012). (Sari, 2012). Dampak dari kredit bermasalah tentunya juga akan menimbulkan permasalahan bagi nasabah bank, karena rumah yang dijadikan agunan bank akan dilelang oleh bank untuk menutup kerugian bank. bank untuk menutupi kerugian bank.

Penyebab terjadinya NPL banyak sekali, karena kondisi perekonomian negara, nasabah mengalami kesulitan keuangan atau kesalahan proses di bank yang menyebabkan kesalahan proses di bank yang menyebabkan kesalahan dalam pemberian nilai pinjaman atau agunan.

Salah satu kesalahan yang dilakukan pihak bank adalah pada proses analisis kredit dimana ditemukan permasalahan oleh analis kredit dalam memvalidasi nilai pinjaman atau agunan. permasalahan yang dihadapi para analis kredit dalam memvalidasi nilai agunan yang diberikan oleh KJPP (Kantor Jasa Penilai Publik), karena analis kredit tidak memiliki data untuk memastikan keabsahan nilai agunan yang diberikan KJPP.

Memastikan keabsahan nilai agunan yang diberikan oleh KJPP sehingga berpotensi terjadi kesalahan dalam pemberian batasan pinjaman yang dapat menimbulkan kerugian bagi bank, dan kerugian tersebut akan berdampak pada bank. bank, dan kerugian tersebut akan berdampak pada risiko reputasi Bank serta dapat melanggar ketentuan rasio pinjaman terhadap nilai dalam Peraturan Bank Indonesia Nomor 18/16/PBI/2016.

Selain permasalahan NPL juga terdapat persaingan antar bank dimana semua bank berlomba-lomba meningkatkan kemampuan sistem informasinya dalam memberikan kecepatan dalam proses pengajuan KPR khususnya aplikasi yang dapat membantu sales untuk dapat memperkirakan nilai agunan. calon pelanggan.

Algoritma k-means memiliki kinerja yang lebih baik dibandingkan k-medoids, baik dari segi rata-rata dalam nilai jarak centroid maupun kompleksitas waktu (Mediana et al., 2018), sehingga penulis menggunakan algoritma k-means, 2018), sehingga Penulis menggunakan algoritma K-Means untuk mengelompokkan kelompok mana yang paling dominan dengan menghilangkan data anomali untuk dijadikan acuan atau nilai wajar nilai tanah pada suatu wilayah di suatu kota.

2 METODE PENELITIAN

Pada bagian metodologi penelitian ini akan diuraikan langkah-langkah sistematis dan terarah yang akan dijadikan acuan sebagai kerangka penelitian untuk menentukan kesamaan permasalahan dalam menentukan estimasi harga appraisal, dengan menggunakan K-means sehingga dapat diketahui metode yang mana. menghasilkan hasil cluster terbaik dan mendapatkan estimasi harga terdekat, metode mana yang menghasilkan hasil cluster terbaik dan mendapatkan estimasi harga terdekat.

2.1 Penemuan Pengetahuan Dalam Basis Data (KDD)

KDD merupakan suatu metode untuk memperoleh pengetahuan dari database yang ada. Di dalam database terdapat tabel-tabel yang saling berhubungan. Hasil pengetahuan yang diperoleh dari proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan. Proses KDD sebagian besar adalah seleksi data, pra-pemrosesan/pembersihan, transformasi, penambahan data, evaluasi, dan pengetahuan (Riadi et al. 2020).

- 1) Seleksi Data
- 2) Pra-Pemrosesan
- 3) Transformasi Data
- 4) Seleksi Data

2.2 Metode Pemilihan Populasi dan Sampel

Pengambilan sampel data adalah teknik yang digunakan untuk secara sistematis memilih sejumlah kecil perwakilan data dari populasi yang telah ditentukan untuk dijadikan sebagai sumber data observasi atau eksperimen sesuai dengan tujuannya, dan sesuai dengan ilmu pengetahuan dan statistik penelitian, prosedur pengambilan sampel harus dilakukan beberapa kali. faktor penting (Sharma, 2017).

2.3 Metode Pengumpulan Data

Pengumpulan data menggunakan teknik pengumpulan data yang bersifat kuantitatif untuk menguji hipotesis yang telah ditentukan dengan teknik seperti berikut:

- 1) Perayapan Data
- 2) Pembersihan Data

2.4 Pengelompokan K-Means

Algoritma K-Means merupakan algoritma pengelompokan berulang yang mempartisi kumpulan data menjadi sejumlah K cluster yang telah ditentukan (Parlina et al., 2018). Metode tersebut akan membagi data menjadi beberapa kelompok yang mana kelompok-kelompok tersebut mempunyai sifat atau karakteristik yang sama. Langkah-langkah dalam tahapan clustering:

- 1) Tentukan k sebagai jumlah cluster yang akan dibentuk

- 2) Tentukan pusat cluster
- 3) Hitung jarak setiap data ke pusat cluster menggunakan persamaan Euclidean.

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

- 4) Kelompokkan data ke dalam cluster dengan jarak terpendek menggunakan persamaan

$$\text{Min } \sum_{k=1}^K d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

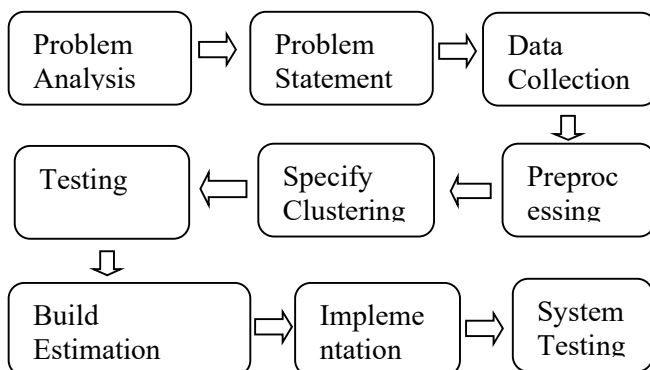
- 5) Hitung pusat cluster baru menggunakan persamaan

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p}$$

- 6) Ulangi langkah 2 hingga 4 hingga tidak ada lagi data yang berpindah ke cluster lain.

2.5 Langkah Penelitian

Langkah-langkah yang akan dilakukan dalam penelitian ini secara umum seperti terlihat pada gambar di bawah ini.



3 HASIL DAN PEMBAHASAN

Algoritma K-Means adalah algoritma pengelompokan berulang yang mempartisi kumpulan data menjadi sejumlah kluster K yang telah ditentukan sebelumnya (Parlina et al., 2018). Metode ini akan membagi data menjadi beberapa kelompok di mana kelompok tersebut memiliki sifat atau karakteristik yang sama. Tujuan dari pengelompokan adalah untuk meminimalkan keragaman dalam suatu kelompok dan memaksimalkan tipe dalam kelompok (Agustin et al., 2015).

3.1 Akuisisi Data

Untuk akuisisi data awal, yang dilakukan adalah crawling data dari website e-commerce sesuai dengan kebutuhan penelitian. Proses pengikisan web mengambil data dari situs web penjualan rumah seperti rumah123.com dan rumah.trovit.co.id. Dalam hal ini, alat yang digunakan dalam pengikisan web adalah bahasa pemrograman python dan jupyter notebook. Data tersebut diambil dari salah satu website jual beli rumah, yaitu www.rumah123.com dan www.trovit.rumah.co.id, yang merupakan salah satu website e-commerce yang menyediakan penjualan properti di Indonesia.

3.2 Pemrosesan Data

Tahap selanjutnya adalah memproses data yang telah diambil melalui crawling. Data atribut yang dipilih untuk dimasukkan dalam kategori penelitian adalah judul, luas tanah, luas bangunan, harga, luas dan juga sub area. Dalam hal ini, penulis mengambil data luas Kebon Jeruk sebagai simulasi perhitungan dan sub area yang diperoleh di kawasan Kebon Jeruk adalah Duri Kepa, Kelapa Dua, Kedoya Selatan, Kedoya Utara.

3.3 Perhitungan Harga Tanah

Setelah transformasi data, langkah selanjutnya adalah menghitung harga tanah dengan rumus berikut:

Harga bangunan = Luas bangunan $\times$ harga bangunan per m²

Harga Tanah = (Harga Rumah - Harga Bangunan) / Luas Tanah

3.4 Centroid Data

Dalam penerapan algoritma K-means, nilai titik tengah atau centroid data dihasilkan. Proses pencarian nilai titik tengah dilakukan dengan mengambil nilai terbesar (maksimum) untuk kluster tingkat tinggi (C1), nilai rata-rata (rata-rata) untuk kluster tingkat menengah (C2) dan nilai terkecil (minimum) untuk kluster tingkat rendah (C3).

TABEL 1. ENTROID DATA

Cluster	Centroid	Centroid	Centroid	Centroid
1	8.111.047	9.878.908	9.878.908	9.878.908
2	24.648.910	22.040.079	22.431.854	22.431.854
3	33.009.776	69.067.415	78.081.825	78.081.825

3.5 Pengklusteran Data

Proses pengelompokan menggunakan nilai centroid awal yang terdapat pada tabel 9, akan memperoleh hasil pengelompokan pada iterasi 1 yang dapat dilihat pada tabel di bawah ini:

TABEL 2. PENGELOMPOKAN DATA

Kedoya Selatan - 2					
Initial Iteration		1st Iteration	2nd Iteration	3rd Iteration	
Cluster	Centroid	Centroid	Centroid	Centroid	Number Of Houses
1	8.111.047	9.878.908	9.878.908	9.878.908	25
2	24.648.910	22.040.079	22.431.854	22.431.854	28
3	33.009.776	69.067.415	78.081.825	78.081.825	4

3.6 Perhitungan Estimasi

Melanjutkan proses pengelompokan, data mentah yang telah terbentuk digunakan sebagai sumber data untuk menghitung harga tanah. Sementara itu, harga bangunan mengacu pada perhitungan harga bangunan berdasarkan hasil tabel matriks MAPPI.

Harga bangunan = Luas bangunan $\times$ harga bangunan per m²

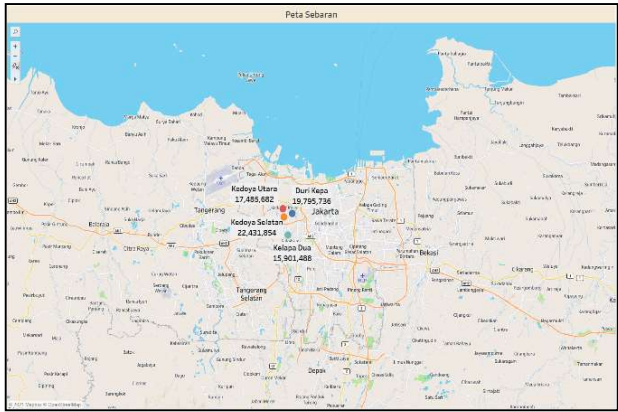
Harga Tanah = Harga Tanah di Luas $\times$ Luas Tanah

Estimasi Harga Rumah = Harga Tanah + Harga Bangunan

3.7 Visualitasi Data

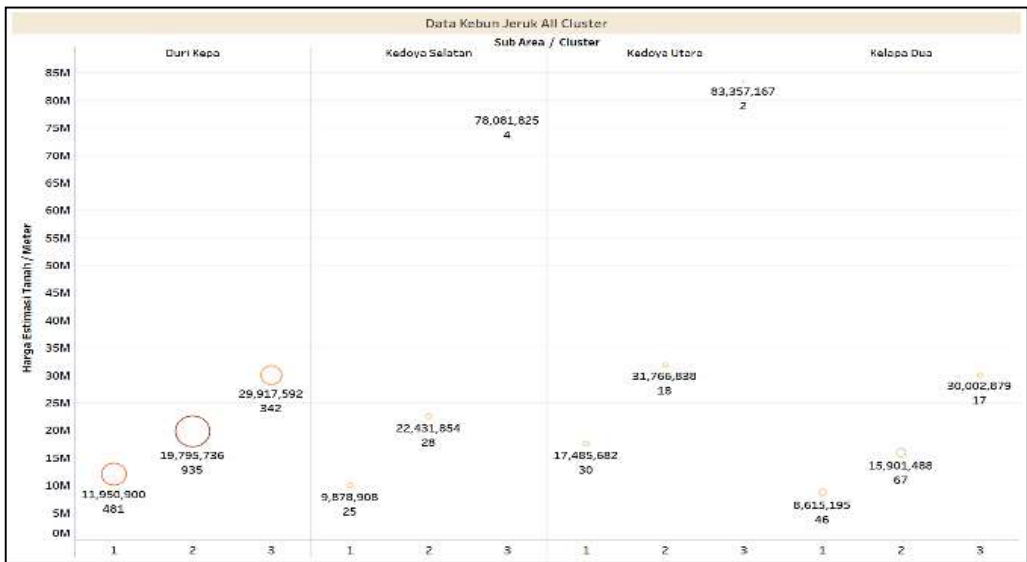
Visualisasi pertama adalah untuk mengilustrasikan distribusi data harga tanah karena pengelompokan dan perhitungan estimasi harga tanah yang dikombinasikan dengan informasi

geotagging untuk posisi bujur dan lintang di masing-masing sub-area, sehingga mendapatkan sebaran hasil estimasi harga tanah dengan hasil seperti yang ditunjukkan pada Gambar 1 di bawah ini:



Gambar 1. Distribusi Data

Selain menggunakan visualisasi distribusi estimasi harga tanah berdasarkan geotagging, berikut ini adalah hasil pengelompokan atau pengelompokan harga tanah berdasarkan masing-masing kelompok yang dapat memberikan gambaran perkiraan harga tanah di setiap klaster di suatu sub-area sesuai Gambar 2 di bawah ini:



Gambar 2. Distribusi Kecamatan

3.8 Hasil Data

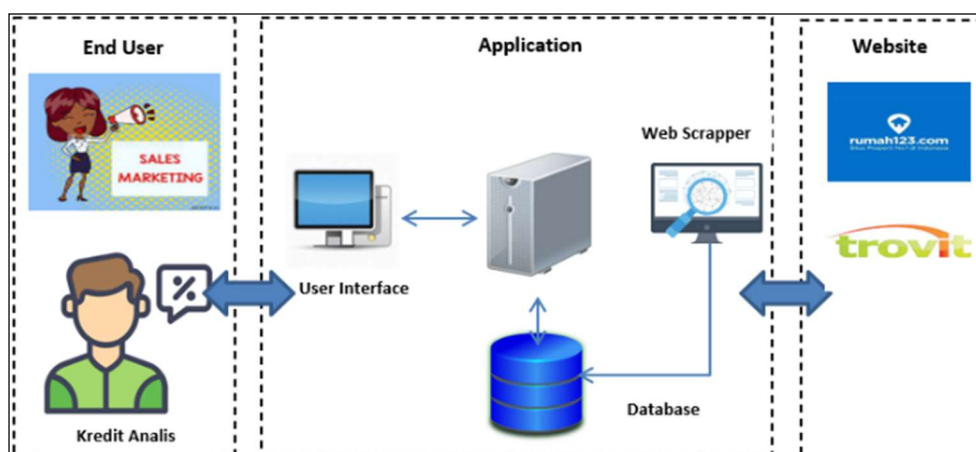
Hasil penerapan metodologi K-Mean dikombinasikan dengan proses perhitungan harga tanah, dimana kelompok dominan akan dirata-rata dan menghasilkan data seperti pada tabel 4.21 di bawah ini sebaran hasil estimasi harga tanah dengan hasil seperti gambar sebagai berikut:

TABEL 3. DATA RESULT

Province	City	District	Sub District	cluster_0	cluster_1	cluster_2	Land Price
DKI Jakarta	Jakarta Barat	Kebon Jeruk	Duri Kepa	265	110	1390	19.693.632
DKI Jakarta	Jakarta Barat	Kebon Jeruk	Kedoya Selatan	7	39	11	14.022.951
DKI Jakarta	Jakarta Barat	Kebon Jeruk	Kedoya Utara	36	12	2	22.484.850
DKI Jakarta	Jakarta Barat	Kebon Jeruk	Kelapa Dua	42	86	2	13.324.546

3.9 Desain Sistem Terintegrasi

Secara umum, persyaratan sistem adalah memberikan solusi yang dihadapi Bank dalam proses penerimaan pengajuan kredit oleh tim pemasaran dan proses analisis kredit oleh tim analisis kredit. Berikut ini adalah deskripsi sistem yang dibutuhkan, sesuai dengan Gambar 3 di bawah ini:

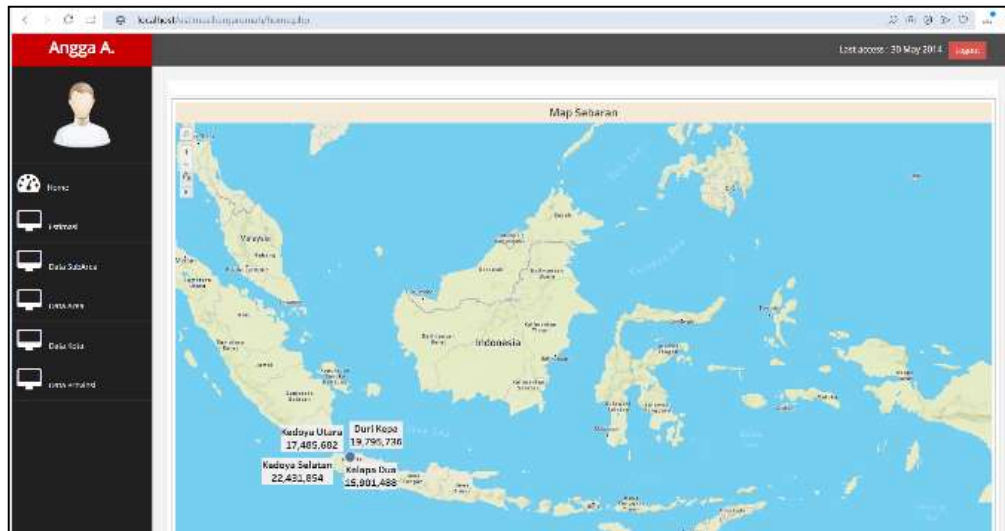


Gambar 3. Distribusi Kecamatan

Gambar 3 adalah gambaran desain aplikasi estimasi harga rumah yang menggunakan data yang dikelompokkan yang akan digunakan oleh analisis penjualan dan kredit.

3.10 Implementasi Prototipe

Pada tahap implementasi prototipe ini, prototipe dibangun dengan bahasa pemrograman PHP sebagai tampilan antarmuka dan MYSQL sebagai penyimpanan database hasil pemrosesan, berikut adalah tampilan prototipe:



Gambar 4. Distribusi Kecamatan

4 PENUTUP

Hasil penelitian yang telah penulis lakukan adalah sebagai berikut:

1. Data penjualan yang tersebar di internet dapat kita bawa pulang dengan menggunakan web scrapping untuk dijadikan sebagai sumber data.
2. Pengujian yang dilakukan pada penelitian ini menghasilkan 3 cluster dengan nilai Indeks David Bouldin yang besar yaitu sub area Duri Kepa sebesar 0,096, sub area Kedoya Selatan sebesar 0,087, sub area Kedoya Utara sebesar 0,071, dan sub area Kelapa Dua luas 0,117.
3. Dengan menggabungkan hasil clustering menggunakan metodologi K-Means dengan rumus perhitungan harga tanah maka diperoleh estimasi harga tanah pada sub areal tersebut.

DAFTAR PUSTAKA

- Ahmad Syaripul, N., & Mukharil Bachtiar, A. (2016). Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi Dki Jakarta: Topik Ekonomi Dan Keuangan Daerah. *Jurnal Sistem Informasi*, 12, 15–29
- Agustin, F. E. M. (2015). Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta). *Jurnal Teknik Informatika*, 8(1), 73–78. <https://doi.org/10.15408/jti.v8i1.1938>.
- Bhatia, P. (2019). Introduction to Data Mining. *Data Mining and Data Warehousing*, 17–27. <https://doi.org/10.1017/9781108635592.003>.
- Dogan, O., Aycon, E., & Bulut, Z. A. (2018). Customer Segmentation By Using Rfm Model and Klasterisasi Methods: a Case. *International Journal of Contemporary Economics and Administrative Sciences*, 8(July), 1–19.
- Edy Irwansyah, S.T., M. S. (2017). KLASTERISASI. <https://socs.binus.ac.id>. <https://socs.binus.ac.id/2017/03/09/Klasterisasi/Klasterisasi> merupakan metode segmentasi data, computer vision dan image processing.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>.
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125. <http://garuda.ristekdikti.go.id/documents/detail/1051741>.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Mediana, Madyatmadja, E. D., & Miranda, E. (2018). Application of K-Means and K-Medoids Klasterisasi Pada Data Internet Banking Di Bank Xyz. 349–356.
- Metisen, B. M., & Sari, H. L. (2015). Analisis Klasterisasi menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118.
- Muhammad, A. F. (2015). Pengelompokkan Proses Seleksi Pemain Menggunakan Algoritma K-Means (Study Kasus: Tim Hockey Kabupaten Kendal). *Jurusan Teknik Informatika FIK UDINUS*, 1–5.
- Nurul rohmawati, sofi defiyanti, mohamad jajuli. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jitter* 2015, 1(2), 62–68.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteksi*, 4(1), 85–92. <https://doi.org/10.33330/jurteksi.v4i1.28>.
- S.C.M. de S Sirisuriya. (2015). A Comparative Study on Web Scraping. 8th International Research Conference KDU, November, 135–140. <http://www.kdu.ac.lk/proceedings/irc2015/2015/com-020.pdf>.
- Steinbach, M., Tan, P., & Boriah, S. (2006). The Application of Klasterisasi to Earth Science Data: Progress and Challenges. *The Application of Klasterisasi to Earth Science Data: Progress and Challenges Michael*, 1–6.
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14(1), 30–37.

- Windarto, A. P. (2017). Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Klasterisasi Method. *Techno.Com*, 16(4), 348–357. <https://doi.org/10.33633/tc.v16i4.1447>.
- Sari, T. M., Syam, D., & Ulum, I. (2012). Pengaruh Non-Performing Loan Sebagai Dampak Krisis Keuangan Global terhadap Profitabilitas Perusahaan Perbankan. *Jurnal Akuntansi & Investasi*, 13(2), 83–98.