

Penerapan Metode K-Means Clustering dalam Klasterisasi Pemilihan Pasir Sesuai SNI

Ladyka Febby Olivia

Universitas Putra Indonesia” YPTK” Padang

Ladyka.febby@gmail.com

Billy Hendrik

Universitas Putra Indonesia” YPTK” Padang

billy_hendrik@upiyptk.ac.id

Alamat: Jalan. Raya Lubuk Begalung, Lubuk Begalung Nan XX, Kec. Lubuk Begalung, Kota Padang, Sumatera Barat 25145

Korespondensi penulis: ladyka.febby@gmail.com

***Abstract.** Development is a government effort to develop a prosperous society. The house is a basic need that is met by every human being as a decent place to live, which is not only for a place to rest and shelter. The method used is clustering with the K-Means algorithm against 8 sand sample data. The results obtained from the data mining process use SPSS software that produces very good, good, and not good.*

***Keywords:** K-Means algorithm, cluster, sand.*

Abstrak. Pembangunan merupakan upaya pemerintah untuk membangun masyarakat yang sejahtera. Rumah adalah kebutuhan pokok yang wajib dipenuhi setiap manusia sebagai tempat tinggal yang layak, dimana tidak hanya untuk sekedar istirahat dan bernaung. Metode yang digunakan adalah clustering dengan algoritma K-Means terhadap 8 data sampel pasir. Hasil yang didapatkan dari proses data mining menggunakan software SPSS yang menghasilkan sangat baik, baik, dan kurang baik.

Kata kunci: algoritma k-means, cluster, pasir

LATAR BELAKANG

Pembangunan merupakan upaya pemerintah untuk membangun masyarakat yang sejahtera. Pembangunan dapat diartikan sebagai kegiatan yang dilakukan oleh suatu negara untuk meningkatkan kualitas hidup masyarakatnya (Abie Rachman Muhamad & Dewi Rahmi, 2023). Pertumbuhan ekonomi adalah perkembangan kegiatan ekonomi yang mengarah pada peningkatan barang dan jasa yang diproduksi dalam masyarakat bertambah dan kemakmuran masyarakat meningkat (Rasnino et al., 2022). Rumah adalah kebutuhan pokok yang wajib dipenuhi setiap manusia sebagai tempat tinggal yang layak, dimana tidak hanya untuk sekedar tempat istirahat dan bernaung saja (wahyuwono dwi, 2022) .

Revolusi Industri 4.0 telah mampu menciptakan inovasi di segala sektor menuju era digital, hulu, tengah, dan hilir (Julianto et al., 2023). Teknologi komputer/ informasi merupakan teknologi yang paling banyak dimanfaatkan diberbagai instansi baik pemerintah

maupun swasta (Amelia Sari Lubis et al., 2023). Aplikasi komputer adalah layanan perangkat lunak yang menggunakan komputer sebagai medianya (Salim et al., 2023).

Plesteran adalah proses dari pekerjaan dinding atau campuran bahan yang digunakan untuk melapisi dinding baik batu bata atau beton yang kasar bahan campuran plesteran berupa semen, pasir dan air guna merekatkan bidang kasar yang dapat membuat suatu bidang menjadi rata (Mus et al., 2023). Pasir merupakan bahan yang berbentuk butiran. Butir dalam pasir, biasanya berukuran 0,0625-2 milimeter (Fariqie al ali, 2023).

Algoritma K-Means adalah metode data mining yang termasuk kata lain tanpa supervisi ketika melakukan proses pemodelan, algoritma ini juga merupakan algoritma dengan sistem partisi untuk melakukan pengelompokan data. Pada proses pengelompokan data, terdapat dua jenis metode, yakni Hierarchical dan Non-Hierarchical, untuk algoritma K-Means tergolong dalam metode Non Hierarchical atau dapat juga disebut dengan istilah Partitional Clustering (Febby Olivia et al., 2023).

Data mining merupakan suatu langkah *dalam knowledge discovery in databases* (KDD) yang memiliki teknik menganalisa data untuk digali informasi tersembunyi dalam jumlah besar dan kompleks, sehingga menghasilkan *output* berupa karakteristik atau pola dari data tersebut. Salah satu teknik analisa Data Mining adalah analisis kelompok (*cluster analysis*) yang lebih dikenal dengan *Clustering* (SS lubis, 2023).

Adapun Maksud dari penelitian ini adalah mengimplementasikan konsep data mining dengan algoritma Clustering K-Means mengklasterisasi pasir yang baik sesuai dengan standar nasional indonesia yang digunakan untuk plesteran dinding dll. Penelitian ini berpotensi untuk mendapatkan sebuah ide usaha yang baik yang merubah pasir dikemas dan dapat dikirim berbagai daerah di indonesia. Pada penelitian ini menggunakan alat bantu SPSS dalam mengklasterisasi jenis pasir yang terbaik yang sesuai dengan standar Pasir SNI.

KAJIAN TEORITIS

1. Data Mining

Data mining adalah proses mengekstrak informasi yang berguna dari data yang besar dan kompleks. Ini mencakup teknik statistik, algoritma, dan proses yang digunakan untuk menemukan pola dan hubungan dalam data (Nahjan et al., 2023).

Tujuan dari data mining adalah untuk mengidentifikasi wawasan baru, tren, dan pengetahuan yang tersembunyi dalam data, yang dapat digunakan untuk mengambil keputusan dan membuat prediksi yang lebih baik (Fauzi, 2023). Clustering merupakan

suatu metode untuk mencari dan mengelompokan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain (Nasution & Rakhmawati, 2023).

2. Algoritma K-Means

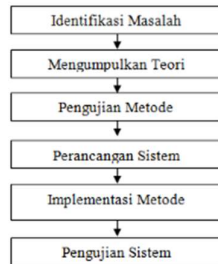
Metode K-means digunakan sebagai alternative metode klaster untuk data dengan ukuran yang lebih besar. K-Means merupakan metode pengelompokan yang paling terkenal dan banyak digunakan pada berbagai bidang karena sederhana dan mudah diimplementasikan. Langkah pada metode k-means yaitu dengan menentukan jumlah cluster secara acak, Mengalokasikan setiap data/objek ke cluster terdekat, mengulang Langkah tersebut hingga nilai cluster tidak berubah (Bambang & Margatama, 2023). Jumlah cluster K sudah ditentukan terlebih dahulu, K-Means adalah salah satu metode clustering nonhirarki yang berusaha partisi data kedalam cluster/ kelompok sehingga data memiliki karakteristik yang sama akan di kelompokkan kedalam satu cluster yang sama dan data yang memiliki karakteristik yang berbeda di kelompokkan kedalam kelompok lain (Jelita et al., 2023).

Tahapan dari algoritma K-Means sebagai proses dari metode clustering adalah sebagai berikut:

1. Pilih jumlah cluster K.
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak.
4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi

METODE PENELITIAN

Metodologi penelitian adalah proses atau cara ilmiah untuk mendapatkan data yang akan digunakan untuk keperluan penelitian. Dalam melakukan penelitian pada penelitian ini, kami mengikuti tahapan metodologi dalam penelitian ini yaitu sebagai berikut:



Gambar 1. Metodologi Penelitian

Berikut keterangan dari gambar 1:

1. Identifikasi Masalah Tahap ini merupakan tahap awal yang digunakan untuk mengidentifikasi masalah dengan tujuan untuk mengamati dan mencari permasalahan melalui data yang berdasarkan artikel terdahulu.
2. Mengumpulkan Teori Pengumpulan teori-teori yang berhubungan dengan pokok permasalahan seperti teori tentang data mining, teori dalam pemilihan karakteristik pasir yang baik dalam plasteran yang digunakan pada dinding, metode yang digunakan dan aplikasi perancangan dari sistem yang diperlukan. Dalam tahap ini, teori dikumpulkan dari beberapa sumber seperti buku-buku, jurnal, artikel dan referensi lainnya.
3. Pengujian Metode Pada tahap ini peneliti menguji metode yang digunakan dalam proses pengelompokan data yang tepat, dengan panduan yang sudah ada pada teori-teori pendukung dari buku-buku maupun jurnal terkait dengan pokok permasalahan.
4. Perancangan Sistem Pada tahap ini dilakukan perancangan sistem terhadap masalah yang sedang diteliti, bisa berupa tahap untuk merancang alur kerja dari sistem dan juga merancang desain dari tampilan tatap muka (interface) dari sistem yang dibuat. Rancangan sistem dibuat untuk mempermudah pengguna dalam memahami sistem yang dibangun.
5. Implementasi Metode Mengimplementasikan metode yang sudah diuji sebelumnya dengan rancangan sistem yang telah dibuat serta melakukan pengkodean (coding) sesuai dengan bahasa pemrograman yang digunakan untuk membuat sistem tersebut.
6. Pengujian Sistem Pada tahap akhir, dilakukan serangkaian pengujian terhadap sistem yang telah dibuat, pengujian-pengujian dilakukan agar dapat menemukan

kesalahan-kesalahan (error) pada sistem dan melakukan perbaikan-perbaikan yang diperlukan.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan salah satu teknik data mining yaitu cluster yang memiliki dapat membantu proses pengelompokan data. Dalam hal ini, algoritma K-Means digunakan pada 8 data sampel pasir. Sebelum melakukan clustering, jumlah cluster K perlu ditentukan terlebih dahulu. Pemilihan jumlah cluster didasarkan pada percobaan dengan menginputkan beberapa jumlah cluster pada software SPSS dan membandingkan hasil yang terbentuk. Sehingga diputuskan berdasarkan hasil tersebut, jumlah cluster yang diambil yaitu 7 cluster

1. Uji Multicolinearitas

H0 : tidak terdapat hubungan antar 2 variable

H1 : terdapat hubungan antar 2 variable

Tingkat signifikansi alfa =5%

		P1	P2	P3	P4	P5	P6	P7
P1	Pearson Correlation	1	.366	.188	-.659	.709*	-.234	-.127
	Sig. (2-tailed)		.372	.656	.076	.049	.577	.764
	N	8	8	8	8	8	8	8
P2	Pearson Correlation	.366	1	-.161	-.131	-.209	-.244	.098
	Sig. (2-tailed)	.372		.703	.757	.620	.560	.817
	N	8	8	8	8	8	8	8
P3	Pearson Correlation	.188	-.161	1	-.070	.654	.071	.249
	Sig. (2-tailed)	.656	.703		.869	.078	.867	.552
	N	8	8	8	8	8	8	8
P4	Pearson Correlation	-.659	-.131	-.070	1	-.323	.266	.091
	Sig. (2-tailed)	.076	.757	.869		.435	.524	.829
	N	8	8	8	8	8	8	8
P5	Pearson Correlation	.709*	-.209	.654	-.323	1	.023	-.047
	Sig. (2-tailed)	.049	.620	.078	.435		.957	.912
	N	8	8	8	8	8	8	8
P6	Pearson Correlation	-.234	-.244	.071	.266	.023	1	.718*
	Sig. (2-tailed)	.577	.560	.867	.524	.957		.045
	N	8	8	8	8	8	8	8
P7	Pearson Correlation	-.127	.098	.249	.091	-.047	.718*	1
	Sig. (2-tailed)	.764	.817	.552	.829	.912	.045	
	N	8	8	8	8	8	8	8

*. Correlation is significant at the 0.05 level (2-tailed).

Gambar 2. Hasil Correlations

Tabel 1. Hasil dari Uji Multicolinearitas

Hubungan	p- value (sig)		Kesimpulan
Pasir 1 dan Pasir 2	0.372	Ho diterima	Tidak terdapat hubungan antar Pasir 1 dan Pasir 2
Pasir 1 dan Pasir 3	0,656	Ho diterima	Tidak terdapat hubungan antar Pasir 1 dan Pasir 3

Pasir 1 dan Pasir 4	0,076	Ho diterima	Tidak terdapat hubungan antar Pasir 1 dan Pasir 4
Pasir 1 dan Pasir 5	0,049	Ho diterima	Tidak terdapat hubungan antar Pasir 1 dan Pasir 5

Tidak ada hubungan linear diantara variabel yang dilihat dari sig pada masing-masing sampel pasir.

Final Cluster Centers

	Cluster		
	1	2	3
P1	796.00	289.00	720.75
P2	245.00	249.00	195.25
P3	256.50	141.50	259.25
P4	267.50	385.00	223.50
P5	1487.00	137.00	1473.25
P6	351.50	235.50	181.25
P7	1141.00	670.00	388.50

Gambar 2. Hasil Final Cluster Center

Tabel final cluster center adalah table yang berisi nilai tengah/ rata-rata dari setiap variabel(pasir) yang dibagi berdasarkan setiap cluster. Dari output diatas, diperoleh:

Pada variabel P1 diperoleh nilai rata-rata : Cluster 1 > Cluster 3 > Cluster 2

Pada variabel P2 diperoleh nilai rata-rata : Cluster 2 > Cluster 1 > Cluster 3

Pada variabel P3 diperoleh nilai rata-rata : Cluster 3 > Cluster 1 > Cluster 2

Pada variabel P4 diperoleh nilai rata-rata : Cluster 2 > Cluster 1 > Cluster 3

Pada variabel P5 diperoleh nilai rata-rata : Cluster 1 > Cluster 3 > Cluster 2

Dari hasil diatas diperoleh Cluster 1 > Cluster 3 > Cluster 2 atau dapat kita dikategori sebagai berikut:

Cluster 1 = Jenis Pasir yang sangat baik digunakan untuk plester dinding

Cluster 3 = Jenis Pasir yang baik digunakan untuk plester dinding

Cluster 2 = Jenis Pasir yang kurang baik digunakan untuk plester dinding

Number of Cases in each Cluster

Cluster	1	2.000
	2	2.000
	3	4.000
Valid		8.000
Missing		.000

Gambar 3. Number of Cases in each Cluster

Output diatas adalah banyak pembagian sampel pasir pada setiap cluster. Diperoleh:

Cluster 1 = Jenis Pasir yang sangat baik digunakan untuk plester dinding terdapat 2 sampel

Cluster 2 = Jenis Pasir yang baik digunakan untuk plester dinding terdapat 2 sampel

Cluster 3 = Jenis Pasir yang kurang baik digunakan untuk plester dinding terdapat 4 sampel

KESIMPULAN DAN SARAN

Melalui semua proses pengolahan serta analisis data yang telah dilakukan, hasil dari penelitian ini menunjukkan bahwa terbentuk 7 cluster dengan algoritma k-means pada metode clustering untuk mengetahui segmentasi pemilihan pasir yang baik. Cluster 1 = Jenis Pasir yang sangat baik digunakan untuk plester dinding terdapat 2 sampel, Cluster 2 = Jenis Pasir yang baik digunakan untuk plester dinding terdapat 2 sampel, Cluster 3 = Jenis Pasir yang kurang baik digunakan untuk plester dinding terdapat 4 sampel.

DAFTAR REFERENSI

Abie Rachman Muhamad, & Dewi Rahmi. (2023). Pengaruh Teknologi, Tingkat Kemiskinan, Pengeluaran Pemerintah, dan Pertumbuhan Ekonomi Terhadap Indeks Pembangunan Manusia di Provinsi Jawa Barat 2007-2021. *Jurnal Riset Ilmu Ekonomi Dan Bisnis*, 45–52. <https://doi.org/10.29313/jrieb.v3i1.1924>

Amelia Sari Lubis, F., Sahara Lubis, S., Hendrik, B., & Author, C. (2023). PERANCANGAN SISTEM INVENTORY UNTUK STOK BARANG HERBISIDA PADA UD. ANUGRAH JAYA TANI DENGAN BAHASA PEMROGRAMAN PHP DAN DATABASE MYSQL. In *Jurnal Sains Informatika Terapan (JSIT) E-ISSN* (Vol. 2, Issue 2). Bulan Juni.

Bambang, M., & Margatama, L. (2023). 3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta (Vol. 2, Issue 2).

Fariqie al ali. (2023). PENGARUH VARIASI CAMPURAN PASIR ZEOLIT SEBAGAI SUBSTITUSI AGREGAT DALAM PEMBUATAN BATAKO TERHADAP KUAT TEKAN. *Renovasi*.

Fauzi, R. (2023). PENERAPAN DATA MINING PADA PENJUALAN PRODUK ELEKTRONIK. *JURNAL COMASIE*.

Febby Olivia, L., Abdi Juliantho, D., & Hendrik, B. (2023). *Komprasi Perbandingan Algoritma K-Means dan K-Medoids Dalam Clustering Penyebaran Kasus Covid 19*. 1(2), 30–32. <https://doi.org/10.31849/digitalzone.v12i1.6572ICCS>

Jelita, T., Buaton, R., Simajuntak, M., & Kaputama, S. (2023). Pengelompokan Bidang Usaha Terhadap Bantuan Produktif Usaha Mikro (BPUM) Berdasarkan Wilayah Deli Serdang Menggunakan Metode Clustering K-Means (Studi Kasus: Dinas Koperasi Dan UMKM Kabupaten Deli Serdang). In *Journal of Computer Science and Information Technology E-ISSN* (Vol. 3, Issue 2).

Julianto, D. A., Febby Olivia, L., & Hendrik, B. (2023). PREDIKSI VOLATILITAS HARGA JUAL PRODUK PADA E-COMMERCE UNTUK INDEPENDENT STOCKASHTIC DATA. In *Jurnal Sains Informatika Terapan (JSIT) E-ISSN* (Vol. 2, Issue 2). Bulan Juni.

Mus, Z., Maulana, R., Hermawan, A., & Sari, N. (2023). Analisis Aktualisasi Penggunaan Material Plesteran Terhadap Koefisien AHSP Pada Proyek Pembangunan Rumah Maguwoharjo. In *Jurnal Ilmiah Teknik sipil* (Vol. 4, Issue 1).

Nahjan, M. R., Heryana, N., & Voutama, A. (2023). IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).

Nasution, N., & Rakhmawati, F. (2023). *METODE DBSCAN (DENSITY BASED SPATIAL CLUSTERING APPLICATION WITH NOISE) DI KOTA MEDAN*. 4(2). <https://doi.org/10.46306/lb.v4i2>

Rasnino, C. A., Nuryadin, D., & Suharsih, S. (2022). Pengaruh Angka Harapan Hidup, Rata-rata Lama Sekolah dan Konsumsi Rumah Tangga Terhadap Pertumbuhan Ekonomi di Kabupaten/Kota Provinsi Lampung, 2014-2019. *Jurnal Impresi Indonesia*, 1(3), 191–200. <https://doi.org/10.36418/jii.v1i3.29>

Salim, A., Rijal, K., & Hendrik, B. (2023). Studi Literatur Sistem Penilaian Esai Otomatis Pada E-Learning Dengan Algoritma Winnowing. *Jurnal Sistem Informasi Dan Ilmu Komputer*, 1(3), 163–172. <https://doi.org/10.59581/jusiik-widyakarya.v1i3.1227>

SS lubis. (2023). *SS Lubis - Customer Relationship Management Dalam Meningkatkan Loyalitas Pelanggan Pada Doorsmeer Keluarga Nasution Menggunakan Metode Algoritma K-Means*.

wahyuwono dwi. (2022). *Pembangunan Rumah tidak layak huni dari perspektif hukum dan permukiman*.