Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

Analisis Perbandingan Kinerja Clustering Data Mining Untuk Normalisasi Dataset

Siti Emalia Saqila, Intan Putri Ferina, Agus Iskandar*

Fakultas Teknologi Komunikasi dan Informatika, Program Studi Informatika, Universitas Nasional, Jakarta, Indonesia
Email: [1]emaliasaqila66@gmail.com, [2]intanputriferina09@gmail.com, [3,*]iskandaragus1005@gmail.com
Email Penulis Korespondensi: iskandaragus1005@gmail.com
Submitted: **19/10/2023**; Accepted: **26/12/2023**; Published: **26/12/2023**

Abstrak−Dimasa sekarang ini perkembangan dan pengaruh teknologi dalam kehidupan manusia sangatlah penting, dimana peran dari teknologi sangat mempengaruhi terhadap aktivitas yang dilakukan oleh manusia. Pada sebuah organisasi perusahaan teknologi bukan saja diperugnakan sebagai proses dari mempercepat proses yang dilakukan. Pemanfaatan dari teknologi yang begitu pentingnya juga meningkatkan terhadap besaran atau volume dari informasi data yang tersedia. Dataset merupakan sebuah kumpulan – kumpulan data yang didapatkan pada data warehouse. Data mining merupakan sebuah teknik yang merupakan bagian dari pada Knowledge Discovery in Database (KDD). Clustering merupakan sebuah proses pengelompokan yang dilakukan pada data mining. Permasalahan pertama yang menjadi pokok pada penelitian adalah nilai yang didapatkan dari proses clustering terkadang masih belum dianggap optimal. Hasil kinerja dari algoritma clustering data mining belum bisa sepenuhnya dipergunakan sebagai dasar dalam pengambilan keputusan. Perbandingan yang dilakukan pada clustering data mining dipergunakan untuk membantu dalam proses pengambilan keputusan. Pada penelitian ini algoritma yang akan digunakan untuk perbandingan terhadap kinerja yaitu algoritma K-Means dan K-Medoids. Permasalahan lainnya yang perlu menjadi perhatian khusus adalah permasalahan terhadap kualitas dari data. Hasil yang didapatkan dari proses data mining dapat dilihat dari kualitas data yang tersimpan atau digunakan pada proses pengolahan data tersebut. Normaslisasi merupakan bagian dari pada preprocessing data mining bertujuan untuk melakukan penalaran kembali terhadap dapat berdasarkan dengan skala baru. Z-Score merupakan normalisasi yang dilakukan pada data berdasarkan dengan fungsi statistika. Hasil yang didapatkan pada penelitian Peran normalisasi pada penelitian sangatlah penting, hal tersebut dikarenakan dengan menggunakan normalisasi Z-Score dapat meningkatkan kinerja dari algoritma K-Means dan K-Medoids, hal tersebut terlihat dari Nilai DBI yang didapatkan lebih kecil ketika normalisasi dilakukan dibandingkan sebelum dilakukan normalisasi dimana hal ini menandakan bahwasannya kinerja lebih baik setelah dilakukan normalisasi. Pada perbandingan algoritma algoritma K-Medoids mendapatkan kinerja lebih baik hal tersebut terlihat dari Nilai DBI yang didapatkan sebesar 0,773 pada K=9 setelah dilakukan normalisasi. Sedangkan pada Algoritma K-Means didapatkan nilai 0,783 pada K=9 setelah dilakukan normalisasi juga.

Kata Kunci: Perbandingan Kinerja; Clustering; Data Mining; Normalisasi; Z-Score

Abstract−Nowadays, the development and influence of technology in human life is very important, where the role of technology greatly influences the activities carried out by humans. In a company organization, technology is not only used as a process to speed up the processes carried out. The use of such important technology also increases the size or volume of available data information. A dataset is a collection of data obtained in a data warehouse. Data mining is a technique that is part of Knowledge Discovery in Database (KDD). Clustering is a grouping process carried out in data mining. The first problem that is central to the research is that the values obtained from the clustering process are sometimes still not considered optimal. The performance results of the data mining clustering algorithm cannot yet be fully used as a basis for decision making. Comparisons made in clustering data mining are used to assist in the decision making process. In this research, the algorithms that will be used for comparison of performance are the K-Means and K-Medoids algorithms. Another problem that needs special attention is the problem of data quality. The results obtained from the data mining process can be seen from the quality of the data stored or used in the data processing process. Normalization is part of preprocessing data mining which aims to re-reason it based on a new scale. Z-Score is a normalization carried out on data based on statistical functions. The results obtained in the research The role of normalization in the research is very important, this is because using Z-Score normalization can improve the performance of the K-Means and K-Medoids algorithms, this can be seen from the DBI value obtained which is smaller when normalization is carried out compared to before it is carried out normalization, which indicates that performance is better after normalization. In the comparison of algorithms, the K-Medoids algorithm gets better performance, this can be seen from the DBI value obtained at 0.773 at K=9 after normalization. Meanwhile, the K-Means algorithm obtained a value of 0.783 at K=9 after normalization as well.

Keywords: Performance Comparison; Clustering; Data Mining; Normalization; Z-Score

1. PENDAHULUAN

Dimasa sekarang ini perkembangan dan pengaruh teknologi dalam kehidupan manusia sangatlah penting, dimana peran dari teknologi sangat mempengaruhi terhadap aktivitas yang dilakukan oleh manusia. Sekarang ini peran dari pada teknologi bukan saja dimanfaatkan oleh sektor perorangan saja, tetapi sudah dapat dimanfaatkan pada sektor sekala besar secara berkelompok ataupun bahkan sebuah organisasi perusahaan.

Pada sebuah organisasi perusahaan teknologi bukan saja diperugnakan sebagai proses dari mempercepat proses yang dilakukan, tetapi pada sekarang ini perkembangan dari teknologi juga dapat dipergunakan untuk proses – proses dalam merekomendasikan pengambilan keputusan yang akan dilakukan oleh organisasi perusahaan tersebut.

Pemanfaatan dari teknologi yang begitu pentingnya juga meningkatkan terhadap besaran atau volume dari informasi data yang tersedia. Besarnya volume informasi data disebabkan karena banyaknya data yang tersimpan

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

dan juga berbagai macam data yang bervariasi tersimpan pada sebuah teknologi. Seluruh data yang tersimpan pada teknologi tersebut tertumpuk pada sebuah gudang tempat penampungan data. Gudang tersebut bernama Data Warehouse dan data yang dikelola merupakan dataset.

Dataset merupakan sebuah kumpulan – kumpulan data yang didapatkan pada data warehouse, dimana pada dataset tersimpan berbagai macam informasi data dari kumpulan – kumpulan data dimasa lampau. Kumpulan informasi data yang tersimpan pada dataset merupakan data baku atau data mentah dari proses pengolahan informasi yang dilakukan, dimana data tersebut nantinya dapat dilakukan proses pengolahan data hingga didapatkan sebuah pengetahuan baru yang tersimpan berdasarkan dengan kumpulan data tersebut. Proses yang dilakukan untuk pengolahan data yaitu sebuah teknik yang bernama data mining.

Data mining merupakan sebuah teknik yang merupakan bagian dari pada Knowledge Discovery in Database (KDD). Dimana pada data mining dilakukan untuk melakukan proses pengolahan data besar dimasa lampau yang tersimpan pada kumpulan data. Proses pengolahan yang dilakukan pada data mining nantinya bertujuan untuk menemukan sebuah informasi baru dari kumpulan data tersebut, dimana informasi baru tersebut dapat dipergunakan untuk membantu dalam proses pengambilan keputusan. Dalam proses yang dilakukan pada data mining juga terdapat berbagai macam teknik penyelesaiannya, salah satu teknik yang sangat sering digunakan untuk menyelesaikan dari proses data mining merupakan clustering.

Clustering merupakan sebuah proses pengelompokan yang dilakukan pada data mining. Dimana proses yang dilakukan pada clustering untuk mengelompokan data – data yang tersimpan pada kumpulan data untuk dipindahkan ataupun dikelompokan berdasarkan dengan jumlah kelompok (cluster) yang telah ditentukan. Berbagai macam cara ataupun teknik yang dapat dilakukan pada proses clustering disesuaikan dengan algoritma yang digunakan. Pada clustering data yang biasa digunakan untuk dilakukan proses clustering merupakan data angka. Dimana pada data tersebut nantinya akan dibentuk sebuah kelompok (cluster) baru.

Permasalahan pertama yang menjadi pokok pada penelitian adalah nilai yang didapatkan dari proses clustering terkadang masih belum dianggap optimal, dimana jika proses yang dilakukan berdasarkan dengan 1 (satu) pengujian saja tentu tidak akan didapatkan hasil yang optimal. Hal ini yang menjadi dasar dari permasalahan pada penelitian dimana untuk mendapatkan hasil yang optimal maka kiranya perlu dilakukan proses pengujian yang berulang sesuai dengan kebutuhan terhadap hasil yang ingin dicapai.

Selain itu dari pada itu, permasalahan lainnya yang perlu diselesaikan bahwasannya hasil kinerja dari algoritma clustering data mining belum bisa sepenuhnya dipergunakan sebagai dasar dalam pengambilan keputusan. Maka dari itu, perlu kiranya dilakukan untuk perbandingan dari kinerja algoritma dari data mining tersebut.

Perbandingan yang dilakukan pada clustering data mining dipergunakan untuk membantu dalam proses pengambilan keputusan. Dengan melakukan perbandingan terhadap algoritma clustering data mining maka kiranya akan mendapatkan algoritma yang memiliki kinerja lebih baik dibandingkan dengan algoritma lainnya. Hal tersebut bertujuan untuk menguatkan terhadap rekomendasi yang dibarikan ataupun dilakukan untuk proses pengambilan keputusan.

Pada penelitian ini algoritma yang akan digunakan untuk perbandingan terhadap kinerja yaitu algoritma K-Means dan K-Medoids. Algoritma K-Means merupakan bagian dari proses penyelesaian dari pada algoritma K-Medoids, dimana proses penyelesaian yang dilakukan pada algoritma K-Means nantinya akan menghitungan jarak terhadap nilai centeroid dari setiap cluster di iterasi. Proses perhitungan jarak berdasarkan dengan nilai ecluiden distance. Sebelum dilakukan proses perhitungan pada iterasi, terlebih dahulu menentukan banyak dari cluster yang akan dibentuk nantinya, setelah itu menentukan nilai centeroid awal dari setiap cluster. Proses dari K-Means akan terus berulang terhadap iterasi hingga nilai pembentukan cluster tidaklah berubah.

Beberapa penelitian terdahulu seperti yang dilakukan oleh Syifa Restillah dan Ade Irma Purnamasar pada tahun 2023 dengan judul penelitian "Pengelompokan Penyebaran Virus Covid-19 Menggunakan Algoritma K-Means Clustering Yang Berada Di Wilayah Jawa Barat" dimana hasil penelitian yang didapatkan bahwasannya Dengan menerapkan algoritma K-Means pada pengelompokan covid-19 berdasarkan wilayah kabupaten atau kota didapatkan 2 cluster yaitu cluster tertinggi dan terendah[1].

Penelitian lainnya yang juga dilakukan tahun 2023 oleh Alfian Adiyanto dan Yudhistira Arie Wijaya dengan judul penelitian "Penerapan Algoritma K-Means Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023" dimana hasil penelitian yang didapatkan menghasilkan Davies Bouldin Index dari K = 2 sampai K =20, maka hasil cluster yang paling optimum adalah K = 2 dengan hasil Davies Bouldin Index sebesar 0,694[2].

Pada tahun 2023 lainnya juga telah dilakukan penelitian oleh Arisman Waruwu, dkk dengan judul penelitian "Implementasi Data Mining Dalam Mengelompokkan Data penduduk Kurang Mampu Menggunakan Metode K-Means Clustering" dimana hasil yang didapatkan dari penelitian algoritma K-Means Clustering sehingga dapat membantu pihak pemerintahan desa dalam mengelompokkan data penduduk kurang mampu di desanya[3].

Penelitian terakhir sebagai penelitian terdahulu dilakukan oleh Ahzril Pria Adistya, dkk pada tahun 2023 dengan judul penelitian "Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa" serta didapatkan hasil penelitian hasil Accuracy 81%, hal ini menunjukkan bahwa sistem dapat mengklasifikasikan data secara benar, namun dalam pengelompokkannya belum optimal,

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

karena terdapat data yang terklasifikasi benar masuk ke klasifikasi salah dan data terklasifikasi salah masuk ke klasifikasi benar[4].

Selain algoritma K-Means yang digunakan pada penelitian, juga terdapat algoritma K-Medoids. Sama halnya dengan algoritm K-Means, algoritma K-Medoids juga diperuntukan pembentukan dari cluster data. Proses pembentukan terhadap cluster data juga berdasarkan perhitungan jarak dengan menggunakan nilai Ecluiden Distance serta menentukan nilai centeroid awal. Tetapi proses akhir yang dilakukan pada algoritma K-Medoids dalam pengambilan keputusan berdasarkan dengan nilai cost. Jika nilai cost untuk hasil iterasi baru lebih besar dibandingkan dengan nilai cost iterasi lama maka proses iterasi dihentikan.

Terdapat beberapa penelitian yang digunakan sebagai penelitian terdahulu seperti yang dilakukan oleh Dela Gustiara, dkk pada tahun 2023 dimana yang menjadi judul penelitian adalah "Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Dampak Kerusakan Bencana Banjir Menggunakan K-Medoids" dengan hasil penelitian yang didapatkan diperoleh jumlah klaster optimum yang terbentuk adalah 3 klaster. Klaster 1 terdiri dari 3 kabupaten/kota dengan kategori tinggi, klaster 2 terdiri dari 23 kabupaten/kota dengan kategori sedang, dan klaster 3 terdiri dari 1 kabupaten/kota dengan kategori rendah[5].

Penelitian lainnya yang dilakukan pada tahun 2023 dilakukan oleh Freditasari Purwa Hidayat, dkk dengan judul penelitian yang dilakukan "Implementasi Clustering K-Medoids dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan" hasil yang didapatkan dari penelitian bahwasannya Berdasarkan pengelompokan analisis K-Medoids clustering diperoleh hasil yaitu cluster 1 terdapat 11 kabupaten/kota, lalu cluster 2 terdapat 12 kabupaten/kota[6].

Juga telah dilakukan penelitian oleh Elly Nur Fitriyani dan Anneke Iswani Achmad pada tahun 2023 dengan judul penelitian "Penerapan Analisis K-Medoids Cluster untuk Mengelompokkan Wilayah di Provinsi Jawa Barat Berdasarkan Fasilitas Kesehatan Tahun 2021" dimana hasil penelitian yang didapatkan Dengan mengambil 3 cluster diperoleh bahwa, untuk cluster 1 terdapat 6 wilayah dengan fasilitas kesehatan yang lengkap, cluster 3 terdapat 11 wilayah dengan fasilitas kesehatan yang sedang dan cluster 2 terdapat 10 wilayah dengan fasilitas kesehatan yang kurang lengkap[7].

Penelitian terakhir yang dilakukan sebagai penelitian terdahulu dilakukan oleh Dwi Utari Iswavigra, dkk pada tahun 2023 dengan judul penelitian "Marketing Strategy UMKM Dengan CRISP-DM Clustering & Promotion Mix Menggunakan Metode K-Medoids" dimana hasil yang didapatkan pada penelitian bahwasannya membentuk 3 cluster untuk proses pengolahan datanya, di mana pada cluster 1 sebanyak 25 UMKM, cluster 2 sebanyak 39 UMKM dan pada cluster 3 sebanyak 7 UMKM[8].

Dalam proses penyelesaian yang dilakukan pada data mining, tidak terlepas terhadap peran dari data. Kualitas data yang baik maka hasil yang akan diddapatkan juga nantinya baik pula, maka dari itu proses awal yang harus diperhatikan sebelum dilakukan proses pengolahan terhadap data mining adalah meyakini terhadap kualitas data yang tersimpan pada data mining.

Selain permasalahan terhadap proses pengolahan data, permasalahan lainnya yang perlu menjadi perhatian khusus adalah permasalahan terhadap kualitas dari data. Pada data yang tersimpan di dataset sering didapatkan data yang tersimpan secara acak. Dimana data acak tersebut seperti terdapat jarak yang cukup jauh dari data yang tesimpan pada dataset tersebut. Besarnya jarak antar setiap data tentu saja menandakan bahwasannya terjadi ketidakseimbangan data dalam proses pengoahan data terkhususnya pada proses clustering yang dilakukan.

Hasil yang didapatkan dari proses data mining dapat dilihat dari kualitas data yang tersimpan atau digunakan pada proses pengolahan data tersebut. Maka dari itu, sebelum dilakukan proses pengolahan terhadap data sebaikanya dipastikan terlebih dahulu terhadap kualitas data yang tersimpan pada dataset tersebut. Proses tersebut dapat dilakukan dengan melakukan preprocessing data. Preprocessing data sendiri pada data mining tedapat berbagai macam cara, seperti missing value, reduksi ataupun lain sebagainya. Pada penelitian ini preprocessing yang akan dibahasa yaitu normalisasi.

Normaslisasi merupakan bagian dari pada preprocessing data mining bertujuan untuk melakukan penalaran kembali terhadap dapat berdasarkan dengan skala baru. Dimana tujuan dari normaslisasi adalah untuk pembentukan nilai skala yang baru terhadap range data yang terlalu besar pada setiap kelompok data. Proses penalaran kembali atau pembentukan nilai skala baru pada normalisasi juga dapat diselesaikan dengan beberapa macam cara salah satunya dengan menggunakan Z-Score.

Z-Score merupakan normalisasi yang dilakukan pada data berdasarkan dengan fungsi statistika. Proses yang dilakukan pada Z-Score dilakukan terhadap nilai baku ataupun nilai standar yang didapatkan dari data. Pada Z-Score dilakukan transformasi atau perubahan data dengan menghasilkan range nilai yang baru bedasarkan dengan range nilai yang sudah ada sebelumnya pada dataset. Pada Z-Score nilai baru yang dihasilkan berdasarkan dengan perbedaan yang dimiliki dari nilai rata – rata dan juga nilai standar deviasi.

Terdapat penelitian terdahulu sebagai dasar dari penelitian seperti yang dilakukan oleh Made Leo Radhitya dan I Gede Iwan Sudipa pada tahun 2020 dengan judul penelitian "Pendekatan Z-Score Dan Fuzzy Dalam Pengujian Akurasi Peramalan Curah Hujan" dimana didapatkan hasil penelitian Berdasarkan hasil uji akurasi curah hujan mulai tahun 2012 – 2016 diperoleh nilai rata-rata akurasi sebesar 85% dengan data training yaitu data tahun 2007 – 2015. Proses normalisasi sangat mempengaruhi nilai data training[9].

Penelitian lainnya yang dilakukan oleh Raditya Galih Whendasmoro dan Joseph pada tahun 2023 dengan judul penelitian "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

NN" dimana didapatkan hasil penelitian bahwa Z-Score Normalization berguna untuk meningkatkan kinerja dari pada algoritma K-NN. Hal tersebut dapat dilihat dari peningkatan nilai akurasi yang didapatkan dari proses K-NN sebelum dilakukan normalisasi pada dataset dengan sesudah dilakukan normalisasi pada dataset[10].

Penelitian lainnya yang dilakukan oleh I Wayan Pio Pratama pada tahun 2023 dengan judul penelitian "Standarisasi Z-Score sebagai Pendekatan Alternatif dalam Evaluasi Prestasi Akademik Mahasiswa : Studi Kasus di Politeknik eLBajo Commodus" didapatkan hasil penelitian bahwasannya Z-Score sebagai kriteria evaluasi prestasi akademik untuk tujuan seperti seleksi beasiswa, penghargaan akademik, dan analisis internal guna meningkatkan kualitas pendidikan[11].

Penelitian terakhir yang digunakan sebagai penelitian terdahulu yang dilakukan oleh Novi Safitri, dkk pada tahun 2023 dengan judul penelitian "Implementasi Algoritma K-Nearest Neighbor Dengan Normalisasi Z-Score Dalam Klasifikasi Penerima Bantuan Sosial Desa Serunai" dimana didapatkan hasil penelitian bahwasanya menggunakan normalisasi z-score berhasil diimplementasikan dengan baik untuk permasalahan seleksi penerima bantuan sosial di Desa Serunai pada proporsi data 90:10 dengan parameter K = 11[12].

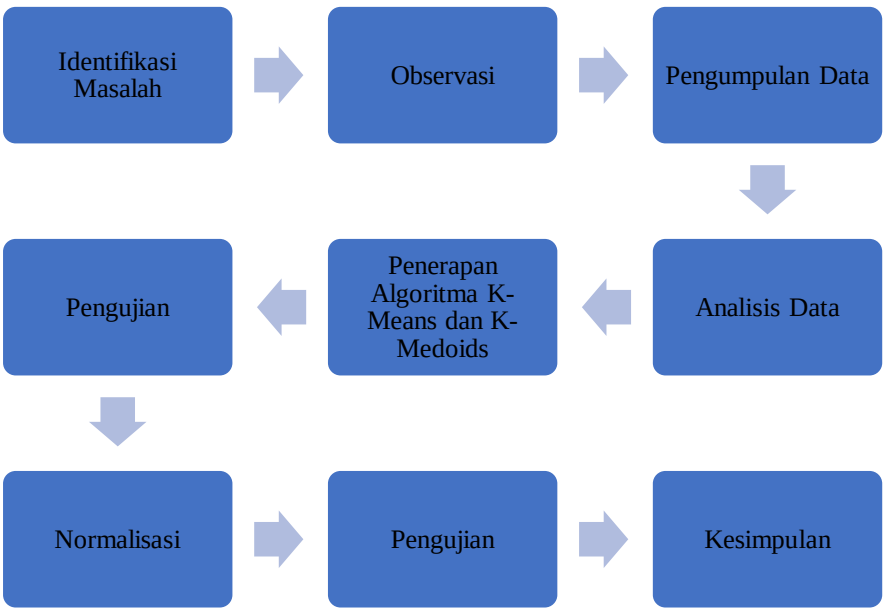
Berdasarkan dengan penjabaran terhadap permasalahan yang dihadapi diatas, maka dari itu penelitian ini bertujuan untuk melakukan perbandingan terhadap kinerja dari clustering data mining. Dimana perbandingan dilakukan dengan menggunakan algoritma K-Means dan K-Medoids untuk menemukan algoritma dengan kinerja lebih baik dan dapat dipergunakan untuk membantu dalam proses pengambilan keputusan.

Selain itu, penelitian juga bertujuan untuk mejamin terhadap kualitas data yang akan digunakan pada penelitian. Dimana cara yang digunakan dengan melakukan normalisasi data. Normalisasi data yang dilakukan bertujuan untuk mentransformasikan data dalam bentuk skala yang baru, hal tersebut dikerenakan seringkali didapatkan pada dataset range data yang cukup jauh hingga nantinya akan mempengaruhi terhadap hasil kinerja yang didapatkan dari algoritma.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian juga biasa disebut dengan metodologi penelitian. Disini akan menggambarkan setiap tahapan proses yang dilakukan pada penelitian. Dimana proses tersebut dimulai dari identifikasi masalah sampai dengan kesimpulan. Peran dari pada kerangka kerja penelitian bertujuan untuk memudahkan bagi peneliti untuk mengetahui tahapan apa saja yang harus dilakukan pada penelitian. Adapun kerangka kerja penelitian yang terdapat pada penelitian ini dapat dilihat pada gambar berikut:



Gambar 1. Kerangka Kerja Penelitian

2.2 Data Mining

Data mining merupakan sebuah proses pengolahan data. Proses pengolahan data pada data mining berdasarkan dengan data – data masa lampau yang tersimpan pada kumpulan data ataupun gudang data. Pada data mining proses pengolahan data yang dilakukan untuk mendapatkan sebuah informasi baru yang terkandung dalam kumpulan data tersebut. Informasi dari data tersebut nantinya dapat dipergunakan untuk proses pengambilan keputusan. Dimana bentuk informasi yang didapatkan dari pengolahan data dapat disesuaikan dengan hasil yang akan diperolah nantinya[13]–[15].

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

2.3 Algoritma K-Means

Algoritma K-Means merupakan bagian dari pada proses clustering pada data mining. K yang dimaksud pada algortima yaitu nilai konstanta atau banyaknya cluster yang diinginkan untuk dibentuk. Nilai cluster dibentuk berdasarkan dengan niali rata – rata dari kelompok cluster. Algoritma K-Means melakukan pengelompokan data dalam bentuk beberapa kelompok dimana pada kelompok tersebut memiliki karakterisitk data yang sama. Proses pada algoritma K-Means berdasarkan dengan parameter nilai k, proses pembentukan cluster berdasarkan dengan nilai kedekatan antar data dengan mencari nilai masing – masing centroid[16]–[18]:

$$d(P,Q) = \sqrt{\sum_{j=1}^{p}(x_j(p) - x_j(q))^2} \quad (1)$$

2.4 Algoritma K-Medoids

Algoritma K-Medoids atau Partitioning Around Medoids (PAM) adalah algoritma *clustering* yang mirip dengan K-Means. Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids atau PAM menggunakan objek sebagai perwakilan (medoid) sebagai pusat *cluster* untuk setiap *cluster*, sedangkan K-Means menggunakan nilai rata-rata (mean) sebagai pusat *cluster*. Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset. Langkah-langkah algoritma K-Medoids[19]–[21]:

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster)
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$dij = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \quad (2)$$

3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoi dbaru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distancelama. Jika S< 0, maka tukar objek dengan data cluster untuk membentuk sekumpulan objek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

2.5 Z-Score Normalization

Z-score normalization adalah suatu metode normalisasi yang hasilnya didapatkan dari nilai rata-rata dan standar deviasi dari data. Metode ini mempunyai nilai yang stabil terhadap outlier maupun adanya nilai yang lebih besar dari maksA atau lebih kecil dari minA. Zscore normalization dapat dihitung menggunakan rumus berikut[22]–[24]:

$$Z = \frac{x - \bar{x}}{\sigma} \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dalam penelitian data mining, data merupakan hal yang paling penting untuk diperhatikan. Pada penelitian ini sampel data yang digunakan sebagai dataset merupakan data tentang udara. Dimana data tersebut terdapat 13 atribut didalamnya dengan 9357 record. Adapun data tersebut dapat dilihat berikut:

Tabel 1. Sampel Data

No	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)
1	2,6	1360	150	11,9	1046	166
2	2	1292	112	9,4	955	103
3	2,2	1402	88	9,0	939	131
4	2,2	1376	80	9,2	948	172
5	1,6	1272	51	6,5	836	131
6	1,2	1197	38	4,7	750	89
7	1,2	1185	31	3,6	690	62
8	1	1136	31	3,3	672	62
9	0,9	1094	24	2,3	609	45
10	0,6	1010	19	1,7	561	-200
…	…	…	…	…	…	…
…	…	…	…	…	…	…

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

No	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)
9357	2,2	1071	-200	11,9	1047	265

Lanjutan Tabel 1. Sampel Data

No	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	AH
1	1056	113	1692	1268	13,6	48,9	0,7578
2	1174	92	1559	972	13,3	47,7	0,7255
3	1140	114	1555	1074	11,9	54,0	0,7502
4	1092	122	1584	1203	11,0	60,0	0,7867
5	1205	116	1490	1110	11,2	59,6	0,7888
6	1337	96	1393	949	11,2	59,2	0,7848
7	1462	77	1333	733	11,3	56,8	0,7603
8	1453	76	1333	730	10,7	60,0	0,7702
9	1579	60	1276	620	10,7	59,7	0,7648
10	1705	-200	1235	501	10,3	60,2	0,7517
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
9357	654	168	1129	816	28,5	13,1	0,5028

Setelah diketahui data tersebut, maka sebelum dilakukan proses pengujian dengan menggunakan algoritma K-Means dan K-Medoids tersebut terlebih dahulu dilakukan proses normalisasi data dengan menggunakan Z-Score. Adapun hasil normalisasi yang didapatkan pada data dapat dilihat pada tabel berikut:

Tabel 2. Normalisasi Data

No	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)
1	-0,6898	1,674522015	-0,433078118	-0,67366	1,126720162	-0,40521	1,145444709
2	-0,6909	1,556513825	-0,499267213	-0,67798	0,968650152	-0,51494	1,350108358
3	-0,6905	1,747678383	-0,541070852	-0,67868	0,941651969	-0,46617	1,291321991
4	-0,6905	1,701520198	-0,555005398	-0,67828	0,957328333	-0,39476	1,207714713
5	-0,6916	1,52167746	-0,605518129	-0,683	0,760938321	-0,46617	1,404540179
6	-0,6923	1,390605633	-0,628161767	-0,68609	0,612448311	-0,53933	1,633589285
7	-0,6923	1,369703814	-0,640354495	-0,68804	0,50663285	-0,58636	1,851752026
8	-0,6926	1,284790172	-0,640354495	-0,68856	0,47615103	-0,58636	1,83694657
9	-0,6928	1,211198349	-0,652547223	-0,69028	0,365545569	-0,61597	2,05598022
10	-0,6933	1,064450158	-0,661256314	-0,6914	0,282373745	-1,04271	2,275449325
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
9357	-0,6905	1,170701074	-1,04271452	-0,67357	1,129768344	-0,23242	0,444798301

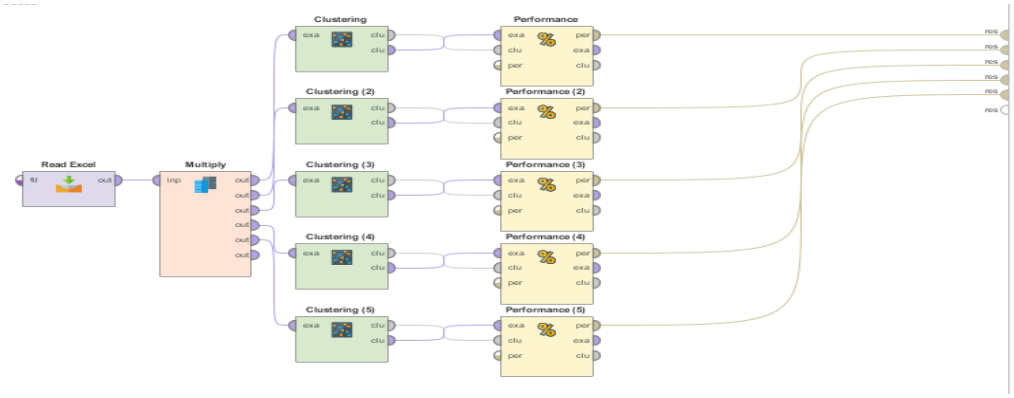
Lanjutan Tabel 2. Normalisasi Data

No	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	AH
1	-0,49753	2,252805687	1,513403823	-0,670662133	-0,609219491	-0,693030992
2	-0,5341	2,0207084	0,999131972	-0,671184678	-0,611266129	-0,693087194
3	-0,49578	2,013305672	1,176361983	-0,673623224	-0,600336219	-0,693044082
4	-0,48185	2,064253857	1,401491997	-0,67519086	-0,589841764	-0,692980551
5	-0,4923	1,900958392	1,239067442	-0,674929588	-0,590582036	-0,692976925
6	-0,52714	1,732002018	0,959070151	-0,674886042	-0,591278764	-0,692983932
7	-0,56023	1,627057466	0,581531037	-0,67462477	-0,595459129	-0,693026536
8	-0,56197	1,627057466	0,576305582	-0,675756951	-0,589841764	-0,693009246
9	-0,58984	1,528209278	0,38470557	-0,675800497	-0,590407856	-0,693018686
10	-1,04271	1,456359274	0,178735557	-0,676497224	-0,589493399	-0,693041611
…	…	…	…	…	…	…
…	…	…	…	…	…	…
9357	-0,40225	1,27129108	0,726972864	-0,64470904	-0,671489497	-0,693475069

3.2 Pengujian Algoritma K-Means

Pengujian yang pertama dilakukan terhadap dataset adalah pengujian sebelum dilakukan normalisasi data. Pada pengujian ini pertama dilakukan dengan menggunakan algoritma K-Means. Proses pengujian yang dilakukan pada penelitian ini dengan menggunakan tools rapid miner. Dimana pada pengujian yang dilakukan terdapat beberapa cluster yang dibentuk yaitu K=7, K=8, K=9, K=10 dan K=11. Adapun proses pengujian dapat dilihat pada gambar berikut:

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919



Gambar 2. Proses Pengujian Algoritma K-Means

Dari gambar 2 diatas, selanjutnya dapat dilakukan proses pengujian dan juga mendapatkan hasil dari proses pengujian. Adapun hasil pengujian yang dilakukan didapatkan hasil sebagai berikut:

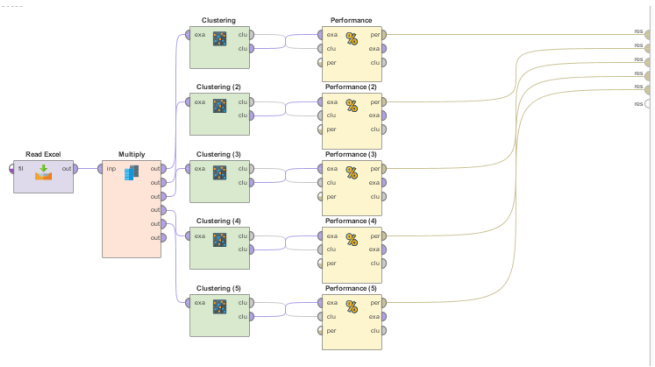
Tabel 3. Hasil Pengujian Algoritma K-Means Sebelum Normalisasi

No	Jumlah Klaster	Nilai DBI
1	K=7	1,124
2	K=8	1,117
3	K=9	1,082
4	K=10	1,073
5	K=11	1,142

Pada tabel 3 diatas dapat dilihat dari hasil pengujian yang dilakukan. Pada algoritma K-Means hasil pengujian berdasarkan dengan Nilai DBI. Dari proses pengujian yang dilakukan bahwasannya Nilai DBI yang paling kecil merupakan hasil dengan kinerja terbaik. Dalam hal tersebut dapat dilihat bahwasannya nilai K=10 merupakan hasil kinerja terbaik.

3.3 Pengujian Algoritma K-Medoids

Setelah pengujian yang dilakukan dengan algoritma K-Means, selanjutnya pengujian yang sama dengan data sebelum dilakukan normalisasi dengan menggunakan algoritma K-Medoids. Pengujian dengan menggunakan K-Medoids juga dilakukan menggunakan tools rapidminer studio . Dimana pada pengujian yang dilakukan terdapat beberapa cluster yang dibentuk yaitu K=7, K=8, K=9, K=10 dan K=11. Adapun proses pengujian dapat dilihat pada gambar berikut:



Gambar 3. Proses Pengujian Algoritma K-Medoids

Dari gambar 3 diatas, selanjutnya dapat dilakukan proses pengujian dan juga mendapatkan hasil dari proses pengujian. Adapun hasil pengujian yang dilakukan didapatkan hasil sebagai berikut:

Tabel 4. Hasil Pengujian Algoritma K-Medoids Sebelum Normalisasi

No	Jumlah Klaster	Nilai DBI
1	K=7	0,876
2	K=8	0,772
3	K=9	1,107
4	K=10	1,876
5	K=11	1,111

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

Pada tabel 4 diatas dapat dilihat dari hasil pengujian yang dilakukan. Pada algoritma K-Medoids hasil pengujian berdasarkan dengan Nilai DBI. Dari proses pengujian yang dilakukan bahwasannya Nilai DBI yang paling kecil merupakan hasil dengan kinerja terbaik. Dalam hal tersebut dapat dilihat bahwasannya nilai K=7 merupakan hasil kinerja terbaik.

3.4 Pengujian Algoritma K-Means dengan Normalisasi

Setelah dilakukan proses pengujian dengan algoritma K-Means tanpa normalisasi data, maka selanjutnya proses yang dilakukan dengan cara yang sama tetapi data yang digunakan merupakan data yang memiliki normalisasi terhadap data. Adapun hasil yang didapatkan dapat dilihat berikut:

Tabel 5. Hasil Pengujian Algoritma K-Means + Normalisasi

No	Jumlah Klaster	Nilai DBI
1	K=7	0,796
2	K=8	0,871
3	K=9	0,783
4	K=10	0,993
5	K=11	1,004

Pada tabel 5 diatas dapat dilihat dari hasil pengujian yang dilakukan. Pada algoritma K-Means hasil pengujian berdasarkan dengan Nilai DBI. Dari proses pengujian yang dilakukan bahwasannya Nilai DBI yang paling kecil merupakan hasil dengan kinerja terbaik. Dalam hal tersebut dapat dilihat bahwasannya nilai K=9 merupakan hasil kinerja terbaik.

3.5 Pengujian Algoritma K-Medoids dengan Normalisasi

Setelah dilakukan proses pengujian dengan algoritma K-Medoids tanpa normalisasi data, maka selanjutnya proses yang dilakukan dengan cara yang sama tetapi data yang digunakan merupakan data yang memiliki normalisasi terhadap data. Adapun hasil yang didapatkan dapat dilihat berikut:

Tabel 6. Hasil Pengujian Algoritma K-Medoids + Normalisasi

No	Jumlah Klaster	Nilai DBI
1	K=7	0,787
2	K=8	0,984
3	K=9	0,773
4	K=10	1,233
5	K=11	0,889

Pada tabel 6 diatas dapat dilihat dari hasil pengujian yang dilakukan. Pada algoritma K-Medoids hasil pengujian berdasarkan dengan Nilai DBI. Dari proses pengujian yang dilakukan bahwasannya Nilai DBI yang paling kecil merupakan hasil dengan kinerja terbaik. Dalam hal tersebut dapat dilihat bahwasannya nilai K=9 merupakan hasil kinerja terbaik.

3.6 Pembahasan

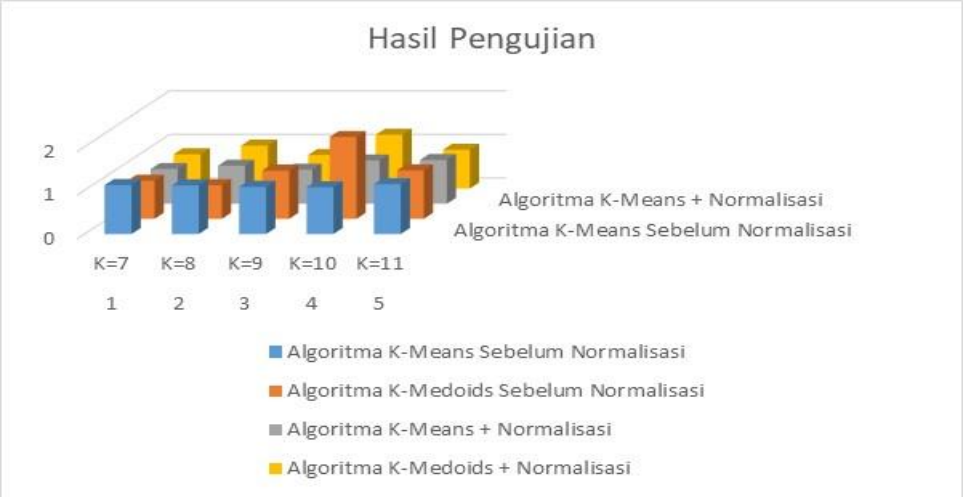
Setelah dilakukan pengujian dengan algoritma K-Means dan K-Medoids baik sebelum normalisasi dan sesudah normalisasi, maka untuk mempermudah dalam proses penyajian hasil dapat dilihat pada tabel berikut:

Tabel 6. Hasil Pengujian Keseluruah

No	Jumlah Cluster	Algoritma K-Means Sebelum Normalisasi	Algoritma K-Medoids Sebelum Normalisasi	Algoritma K-Means + Normalisasi	Algoritma K-Medoids + Normalisasi
1	K=7	1,124	0,876	0,796	0,787
2	K=8	1,117	0,772	0,871	0,984
3	K=9	1,082	1,107	0,783	0,773
4	K=10	1,073	1,876	0,993	1,233
5	K=11	1,142	1,111	1,004	0,889

Pada tabel 6 diatas dapat dilihat hasil dari proses pengujian secara keseluruhan. Dimana pada tabel tersebut dapat dilihat bahwasannya dengan dilakukannya normalisasi terhadap dataset maka akan membantu dalam peningkatan kinerja yang dilakukan. Hal tersebut dapat dilihat dari hasil yang didapatkan bahwasannya nilai DBI setiap algoritma yang didapatkan lebih kecil setelah dilakukan normalisasi. Hal tersebut menandakan bahwasannya pentingnya dilakukan normalisasi terhadap data.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919



Gambar 4. Hasil Pengujian Keseluruhan

Berdasarkan dengan hasil yang didapatkan pada gambar 4 diatas, maka secara keseluruahn untuk algoritma K-Means didapatkan kinerja terbaik setelah dilakukan normalisasi pada K=9 dengan nilai DBI 0,783 dan pada algoritma K-Medoids didapatkan kinerja terbaik juga setelah dilakukan normalisasi pada K=9 dengan nilai DBI 0,773. Hal ini menandakan bahwasannya algoritma K-Medoids memiliki kinerja lebih baik dibandingkan dengan algoritma K-Means

4. KESIMPULAN

Setelah rangkaian proses yang dilakukan, tahapan akhir yang dilakukan pada penelitian merupakan penarikan kesimpulan. Pada penelitian ini didapatkan kesimpulan bahwasannya proses pengolahan data dengan data mining dapat dipergunakan untuk menyelesaikan permasalahan dalam pengambilan keputusan. Peran normalisasi pada penelitian sangatlah penting, hal tersebut dikarenakan dengan menggunakan normalisasi Z-Score dapat meningkatkan kinerja dari algoritma K-Means dan K-Medoids, hal tersebut terlihat dari Nilai DBI yang didapatkan lebih kecil ketika normalisasi dilakukan dibandingkan sebelum dilakukan normalisasi dimana hal ini menandakan bahwasannya kinerja lebih baik setelah dilakukan normalisasi. Pada perbandingan algoritma algoritma K-Medoids mendapatkan kinerja lebih baik hal tersebut terlihat dari Nilai DBI yang didapatkan sebesar 0,773 pada K=9 setelah dilakukan normalisasi. Sedangkan pada Algoritma K-Means didapatkan nilai 0,783 pada K=9 setelah dilakukan normalisasi juga.

REFERENCES

[1] S. Restillah and A. I. Purnamasari, “PENGELOMPOKAN PENYEBARAN VIRUS COVID-19 MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING YANG BERADA DI WILAYAH JAWA BARAT,” vol. 7, no. 3, pp. 1953–1957, 2023.

[2] A. Adiyanto and Y. Arie Wijaya, “Penerapan Algoritma K-Means Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1344–1350, 2023, doi: 10.36040/jati.v7i2.6849.

[3] A. Waruwu, M. Yetri, and F. Setiawan, “Implementasi Data Mining Dalam Mengelompokkan Data penduduk Kurang Mampu Menggunakan Metode K-Means Clustering,” *J. Sist. Inf. TGD*, vol. 2, no. 6, pp. 945–955, 2023.

[4] A. P. Adistya, N. Lutfiyani, P. Tara, Rifaldi, R. Adriyan, and P. Rosyani, “Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa,” *OKTAL J. Ilmu Komput. dan Sci.*, vol. 2, no. 8, pp. 2301–2306, 2023.

[5] D. Gustiara, A. D. Mulyaningsih, R. Anadra, and E. Widodo, “Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Dampak Kerusakan Bencana Banjir Menggunakan K-Medoids,” *Indones. J. Appl. Stat.*, vol. 5, no. 2, pp. 109–120, 2023.

[6] F. P. Hidayat, R. P. Putra, M. D. Alfitrah, and E. Widodo, “Implementasi Clustering K-Medoids dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan,” *Indones. J. Appl. Stat.*, vol. 5, no. 2, pp. 121–130, 2023.

[7] E. N. Fitriyani and A. Iswani Achmad, “Penerapan Analisis K-Medoids Cluster untuk Mengelompokkan Wilayah di Provinsi Jawa Barat Berdasarkan Fasilitas Kesehatan Tahun 2021,” *Bandung Conf. Ser. Stat.*, vol. 3, no. 2, pp. 283–293, 2023, doi: 10.29313/bcss.v3i2.8080.

[8] D. U. Iswavigra, L. E. Zen, Okfalisa, and H. Hanim, “Marketing Strategy UMKM Dengan CRISP-DM Clustering &Promotion Mix Menggunakan Metode K-Medoids,” *J. Inf. dan Teknol.*, vol. 5, no. 1, pp. 45–54, 2023, doi: 10.37034/jidt.v5i1.260.

[9] M. L. Radhitya and G. I. Sudipa, “Pendekatan Z-Score Dan Fuzzy Dalam Pengujian Akurasi Peramalan Curah Hujan,” *SINTECH (Science Inf. Technol. J.*, vol. 3, no. 2, pp. 149–156, 2020, doi: 10.31598/sintechjournal.v3i2.567.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 5, Nomor 2, Desember 2023
e-ISSN 2685-998X
DOI 10.30865/json.v5i2.6919

[10] R. G. Whendasmoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 872, 2022, doi: 10.30865/jurikom.v9i4.4526.
[11] I. W. P. Pratama, "Standarisasi Z-Score sebagai Pendekatan Alternatif dalam Evaluasi Prestasi Akademik Mahasiswa : Studi Kasus di Politeknik eLBajo Commodus," *JPTM J. Penelit. Terap. Mhs.*, vol. 1, no. 2, pp. 77–85, 2023.
[12] N. Safitri, D. Kusnandar, and S. Martha, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR DENGAN NORMALISASI Z-SCORE DALAM KLASIFIKASI PENERIMA BANTUAN SOSIAL DESA SERUNAI," *Bul. Ilm. Math. Stat. dan Ter.*, vol. 13, no. 1, pp. 99–106, 2023.
[13] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.
[14] S. Dewi, "Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 59–65, 2019, doi: 10.31294/jki.v7i1.5744.
[15] H. Gunawan and V. Purwayoga, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 1–8, 2022, doi: 10.32736/sisfokom.v11i1.1316.
[16] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI.
[17] T. Hartati, O. Nurdiawan, and E. Wiyandi, "Analisis Dan Penerapan Algoritma K-Means Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari," *J. Sains Teknol. Transp. Marit.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.51578/j.sitektransmar.v3i1.30.
[18] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
[19] J. Faran and R. T. Aldisa, "Penerapan Data Mining Untuk Penjurusan Kelas dengan Menggunakan Algoritma K-Medoids," *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, pp. 543–552, 2023, doi: 10.47065/bits.v5i2.4313.
[20] N. Widiawati, B. N. Sari, and T. N. Padilah, "Clustering Data Penduduk Miskin Dampak Covid-19 Menggunakan Algoritma K-Medoids," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 55–63, 2022, doi: 10.30871/jaic.v6i1.3266.
[21] Y. Diana and F. Hadi, "Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang," *J. Inf. Syst. Informatics Eng. Vol.*, vol. 7, no. 1, pp. 97–103, 2023.
[22] A. Masitha and M. K. Biddinika, "Preparing Dual Data Normalization for KNN Classfication in Prediction of Heart Failure," vol. 4, no. 3, pp. 1227–1234, 2023, doi: 10.30865/klik.v4i3.1382.
[23] M. Qori'atunnadyah, "Pengelompokkan Wilayah Berdasarkan Rasio Guru-Murid Pada Jenjang Pendidikan Menggunakan Algoritma K-Means," *J. Informatics Dev.*, vol. 1, no. 2, pp. 33–38, 2022.
[24] A. sami Jaddoa, S. J. Saba, and E. A.Abd Al-Kareem, "Liver Disease Prediction Model Based on Oversampling Dataset with RFE Feature Selection using ANN and AdaBoost algorithms," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 4, no. 2, pp. 85–93, 2023, doi: 10.36805/bit-cs.v4i2.5565.