



Prapemrosesan klasifikasi algoritme kNN menggunakan K-means dan matriks jarak untuk dataset hasil studi mahasiswa

Preprocessing kNN algorithm classification using K-means and distance matrix with students' academic performance dataset

Sugriyono^{*)}, Maria Ulfah Siregar

Magister Informatika, Fakultas Sains dan Teknologi, UIN Sunan Kalijaga
Jl. Marsda Adisucipto, Yogyakarta, Indonesia 55281

Cara sitasi: S. Sugriyono and M. U. Siregar, "Prapemrosesan klasifikasi algoritme kNN menggunakan K-means dan matriks jarak untuk dataset hasil studi mahasiswa," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 4, pp. 311-316, 2020. doi: [10.14710/jtsiskom.2020.13874](https://doi.org/10.14710/jtsiskom.2020.13874), [Online].

Abstract – *The existence of outliers in the dataset can cause low accuracy in a classification process. Outliers in the dataset can be removed from a preprocessing stage of classification algorithms. Clustering can be used as an outlier detection method. This study applies K-means and a distance matrix to detect outliers and remove them from datasets with class labels. This research used a dataset of students' academic performance totaling 6847 instances, having 18 attributes and 3 class labels. Preprocessing applies the K-means method to get centroid in each class. The distance matrix is used to evaluate the distance of instance to the centroid. Outliers, which are a different class, will be removed from the dataset. This preprocessing improves the classification accuracy of the kNN algorithm. Data without preprocessing has 72.28 % accuracy, preprocessed data using K-means with Euclidean has 98.42 % accuracy (an increase of 26.14 %), while the K-means with Manhattan has 97.76 % accuracy (an increase of 25.48 %).*

Keywords – *preprocessing; K-means; kNN; distance matrix; Manhattan; Euclidean*

Abstrak – *Keberadaan outlier pada dataset dapat menyebabkan rendahnya hasil akurasi pada proses klasifikasi. Outlier pada dataset dapat dihilangkan pada tahapan prapemrosesan algoritme klasifikasi. Clustering dapat digunakan sebagai metode pendeteksi outlier. Kajian ini bertujuan menerapkan K-means dan matriks jarak untuk mendeteksi outlier dan menghapusnya dari dataset yang sudah memiliki kelas label. Penelitian ini menggunakan dataset hasil studi mahasiswa berjumlah 6847 instance, dengan 18 atribut dan tiga kelas. Prapemrosesan menerapkan metode K-means untuk mendapatkan pusat kluster pada tiap class, matriks jarak digunakan untuk mengevaluasi jarak instance dengan pusat kluster. Outlier, kelas baru yang berbeda dengan kelas awal, yang ditemukan akan*

dihilangkan. Prapemrosesan ini meningkatkan hasil akurasi klasifikasi algoritme kNN. Data tanpa prapemrosesan menghasilkan akurasi sebesar 72,28 %, data hasil prapemrosesan menggunakan metode K-means dan Euclidean menghasilkan akurasi hasil klasifikasi sebesar 98,42 % (meningkat 26,14 %), sedangkan metode K-means dan Manhattan menghasilkan akurasi sebesar 97,76 % (meningkat 25,48 %).

Kata kunci – *prapemrosesan; K-means; kNN; matriks jarak; Manhattan; Euclidean*

I. PENDAHULUAN

Tujuan pendidikan pada Perguruan Tinggi (PT) yang bermutu adalah menghasilkan lulusan yang sesuai dengan kebutuhan perkembangan zaman [1]. Perwalian sebagai bentuk pembinaan mahasiswa merupakan aspek yang sangat penting dalam mencapai tujuan pendidikan [2]. Perwalian ini melibatkan dosen-dosen sebagai dosen wali untuk memantau perkembangan studi mahasiswa. Perwalian harus efektif sehingga bisa mengenali atau memprediksi potensi hasil studi mahasiswa. Hal ini adalah sangat penting sehingga upaya-upaya peningkatan pembinaan mahasiswa dapat dilakukan. Sebuah metode untuk memprediksi potensi hasil studi mahasiswa sangat penting diterapkan agar proses perwalian jadi lebih efektif.

Teknik klasifikasi pada data mining dapat digunakan untuk melakukan prediksi [3]. Salah satu algoritme klasifikasi data mining yang dapat digunakan untuk melakukan prediksi adalah *k-Nearest Neighbor* (kNN). Algoritme kNN dipilih karena sederhana dan memberikan hasil yang kompetitif dan signifikan untuk menyelesaikan masalah klasifikasi [4], [5].

Bhattacharya dkk. [6] menyebutkan keberadaan outlier pada dataset dapat mempengaruhi akurasi algoritme kNN. Lebih lanjut, kajian tersebut menyatakan bahwa skor outlier berdasarkan perbedaan tingkat kepadatan dapat digunakan untuk memodulasi fungsi jarak potensial pada klasifikasi kNN dalam upaya meningkatkan hasil akurasi pada dataset yang memiliki outlier.

^{*)} Penulis korespondensi (Sugriyono)
Email: 18206050005@student.uin-suka.ac.id

Sinaga dkk. [7] menyatakan adanya pengaruh *outlier* pada klasifikasi algoritme kNN dan mengusulkan sebuah metode dengan cara mengkombinasikan beberapa tahapan pada vektor rata-rata lokal kNN dan jarak *harmonic* sebagai label untuk data uji. Metode tersebut dapat meningkatkan hasil akurasi klasifikasi kNN pada dataset yang memiliki *outlier*. Alternatif lain agar akurasi hasil klasifikasi meningkat adalah dengan menghilangkan *outlier* pada dataset sebelum proses (prapemrosesan) klasifikasi.

Deteksi adanya *outlier* pada dataset dapat dilakukan dengan menggunakan metode *clustering* [8], [9]. Mishra dkk. [10] menerapkan algoritme K-means untuk mendeteksi *outlier* pada dataset. Teknik yang dilakukan dengan cara menjalankan K-means tiga kali secara bersamaan. Algoritme K-means mengelompokkan data dan mengambil tiga jumlah kluster yang berbeda dari k yaitu k , $k+1$, dan $k-1$ secara bersamaan untuk setiap nilai k . *Outlier* terpisah akan dideteksi menggunakan radius dari setiap kluster.

Angelin dan Geetha [11] menerapkan algoritme K-means untuk mendeteksi *outlier* dengan menggunakan proses *clustering* dengan k-Median sebagai optimasi proses. Dalam tesisnya, Gustavsson [12] dapat mendeteksi kejadian abnormal sebagai *outlier* dalam data lalu lintas penerbangan dengan menggunakan *clustering*. Performa terbaik diperoleh dengan menggunakan spektral *clustering* yang dikombinasikan dengan sebuah kelas *support vector machine* (SVM).

Berbeda dengan [6] dan [7] dalam meningkatkan hasil akurasi klasifikasi, penelitian ini mengusulkan prapemrosesan untuk menghilangkan *outlier* pada dataset untuk meningkatkan hasil akurasi klasifikasi algoritme kNN. Berbeda dengan [10] dan [11] dalam menerapkan algoritme K-means dalam mendeteksi *outlier*, penelitian ini mengusulkan deteksi *outlier* dengan menggunakan algoritme K-means dan matriks jarak pada dataset yang sudah memiliki kelas label. Seperti yang dinyatakan [13], [14], matriks jarak memiliki peran yang sangat penting pada pemrosesan *clustering* pada algoritme K-means. Berdasarkan hal ini, penelitian ini memilih matriks jarak Euclidean dan Manhattan sebagai matriks yang berkinerja baik seperti yang dinyatakan dalam [15], [16]. Hal ini dilakukan untuk menunjukkan perbandingan hasil kinerja antar matriks jarak pada metode yang diusulkan.

II. METODE PENELITIAN

Tahapan penelitian ini meliputi penyiapan data, prapemrosesan data, klasifikasi data hasil prapemrosesan, dan evaluasi.

A. Penyiapan data

Dataset yang disiapkan adalah data yang diperoleh dari lokasi penelitian, yaitu hasil studi mahasiswa Universitas Ahmad Dahlan (UAD) Yogyakarta. Data ini diambil dari basis data sistem akademik UAD. Data hasil studi mahasiswa yang diambil memiliki 18 atribut, dengan satu atribut sebagai atribut kelas dan 17 atribut

Tabel 1. Daftar atribut, tipe, dan deskripsi dataset

No	Daftar atribut	Tipe	Deskripsi
1	ip_sem_1 s/d ip_sem_4	numerik	nilai 0,00 s.d 4,00
2	ipk_sem_1 s/d ipk_sem_4	numerik	nilai 0,00 s.d 4,00
3	sks_sem_1 s/d sks_sem_4	numerik	satuan kredit semester 1 s.d 4
4	status_sma	nominal	Negeri=1, Swasta=0
5	jalur	nominal	Non Test=1, Test=0
6	jenis_kelamin	nominal	Laki-laki=1, Perempuan=2
7	gelombang	numerik	Gelombang seleksi: 1, 2, dan 3
8	pilihan_ke	numerik	Urutan pilihan jurusan:1 atau 2
9	cuti	numerik	Jumlah cuti
10	status	nominal	1: Lulus tepat waktu 2: Lulus tidak tepat waktu 3: Drop out

sebagai fitur. Atribut kelas label dataset memiliki tiga nilai, yaitu mahasiswa lulus tepat waktu, mahasiswa lulus tidak tepat waktu, dan mahasiswa *drop out* (DO). Tabel 1 menyatakan daftar atribut-atribut, tipe data, dan deskripsi atribut pada dataset tersebut.

B. Prapemrosesan

Tujuan dari desain prapemrosesan adalah membandingkan proses metode deteksi *outlier* dari dataset hasil studi mahasiswa UAD menggunakan dua metode. Metode yang pertama adalah metode deteksi *outlier* menggunakan K-means dan matriks jarak Euclidean dan metode yang kedua adalah metode deteksi *outlier* menggunakan K-means dan matriks jarak Manhattan. Keluaran dari proses ini adalah berupa dua dataset yang sudah dibersihkan dengan metode usulan tersebut. Desain prapemrosesan ditunjukkan pada Gambar 1.

Berikut dijabarkan langkah detail pada proses algoritme deteksi dan penghapusan *outlier* dari dataset hasil studi mahasiswa. Proses pertama dimulai dengan membagi dataset menjadi subdata sejumlah kelas berdasarkan kelas pada dataset awal. Langkah kedua adalah mencari pusat kluster dari masing-masing sub dataset menggunakan K-means. Algoritme K-means ini dilakukan dengan menentukan k titik pusat kluster secara acak. Data dikelompokkan berdasarkan jarak data ke setiap pusat k kluster. Nilai titik pusat kluster diperbarui. Langkah untuk menemukan kluster dan pengelompokkan tersebut diulang sampai nilai pusat kluster tidak lagi berubah.

Langkah ketiga adalah menghitung jarak semua *instance* dengan masing-masing pusat kluster yang dihasilkan langkah kedua menggunakan matriks jarak. Matriks jarak yang digunakan adalah Euclidean d_1 yang dinyatakan pada Persamaan 1, sedangkan matriks Manhattan d_2 pada Persamaan 2. Parameter $d(x,y)$

menyatakan jarak antara data pada titik x dan y dimana x adalah titik data pertama (pusat kluster), y adalah titik data kedua (data dari n), dan n adalah jumlah atribut data.

$$d_1(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$d_2(x,y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

Langkah keempat adalah menemukan kelas label baru dari masing-masing metode matriks jarak berdasarkan jarak terdekat masing-masing *instance* dengan pusat-pusat kluster. Langkah kelima adalah membandingkan kelas label awal dengan kelas label baru. Data yang memiliki kelas label lama yang berbeda dengan kelas label baru dianggap sebagai *outlier* dan dihapus dari dataset. Alur proses deteksi dan penghapusan *outlier* dari dataset ditunjukkan pada [Algoritme 1](#).

C. Klasifikasi dan Evaluasi

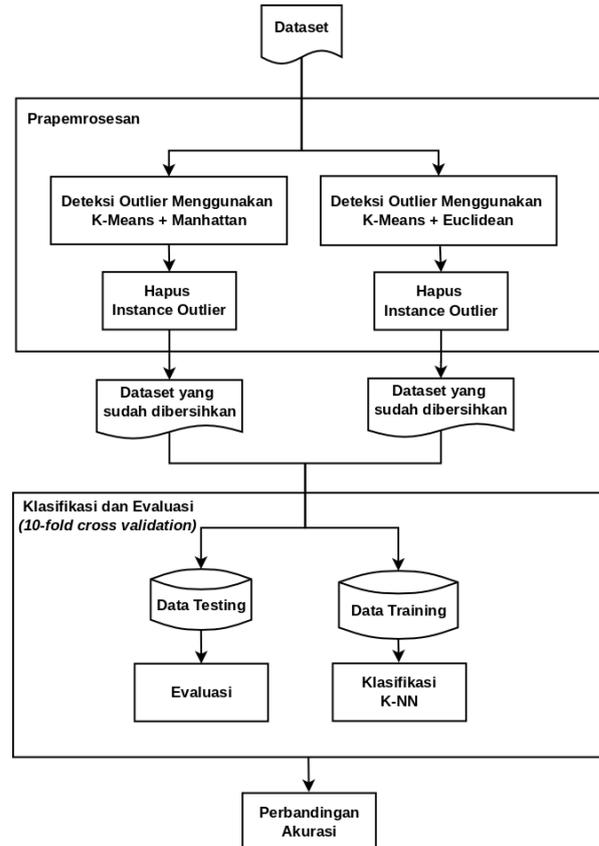
Klasifikasi dan evaluasi dilakukan pada dua dataset yang sudah dibersihkan dari *outlier*. Perbandingan dilakukan untuk menghasilkan model dan evaluasi klasifikasi hasil studi mahasiswa dengan nilai akurasi yang paling tinggi. Metode yang digunakan adalah *10-fold cross validation* yang mempunyai waktu komputasi cepat dengan tetap menjaga akurasi estimasi [17].

Metode ini membagi data menjadi sepuluh bagian data dengan ukuran sama untuk mengevaluasi kinerja model dan algoritme. Teknik ini menggunakan 9 bagian data sebagai data latih dan 1 bagian data menjadi data uji. Langkah-langkah pada tahapan klasifikasi dan evaluasi meliputi pembagian dataset menjadi data latih 90 % dan 10 % untuk data uji, pemrosesan data uji dengan algoritme klasifikasi kNN dengan parameter $k=5$, dan pengukuran performa model dalam bentuk matriks konfusi. Performa yang diukur adalah akurasi, presisi, sensitivitas, dan *F1-score*. Eksperimen ini dilakukan dengan menggunakan perangkat lunak *Python* versi 3.6 dan pustaka *Scikit-learn* versi 0.22.1.

III. HASIL DAN PEMBAHASAN

Identifikasi jumlah data dari masing-masing kelas awal yang dimiliki dilakukan setelah penyiapan dataset. Dataset memiliki jumlah data 6.847 *instance*. Distribusi data berdasarkan atribut kelas awal ditunjukkan pada [Tabel 2](#), yaitu lulus tepat waktu sejumlah 1.979, lulus tidak tepat waktu 3.183, dan *drop out* 1.685.

Langkah pertama dari tahapan metode yang diusulkan adalah membagi dataset menjadi sub dataset berdasarkan kelas label dan menemukan pusat kluster dari masing-masing sub dataset menggunakan K-means. Kluster data terhadap pusat kluster yang ditemukan ditunjukkan pada [Gambar 2](#). Kluster tersebut menjelaskan adanya *outlier*, yaitu sejumlah data yang



Gambar 1. Desain metode yang diusulkan

Algoritme 1. Alur deteksi dan penghapusan outlier

Begin

- 1: data_split = split_by_initial_class(data)
- 2: centroid = get_centroid_use_k_mean(data_split)
- 3: **for** i := 1 **to** count(data) **do**
- 4: data_distance = distance_matrix(centroid,data[i])
- 5: new_class = nearest_centroid(data_distance)
- 6: **if** (data[i][inisial_class] <> new_class)
- 7: remove_from_data(data[i])

end for

- 8: **return** data

Finish

Tabel 2. Jumlah distribusi data tiap kelas

No	Kelas	Jumlah data
1	Lulus tepat waktu (1)	1.979
2	Lulus tidak tepat waktu (2)	3.183
3	Drop out (3)	1.685
<i>Total data</i>		6.847

memiliki jarak lebih dekat terhadap pusat data kluster yang lain.

Evaluasi jarak semua *instance* pada dataset terhadap pusat kluster dilakukan untuk menentukan bahwa semua *instance* pada dataset telah masuk pada kelas yang benar yang ditunjukkan dengan jarak *instance* yang lebih dekat dengan pusat kluster. Jika *instance* tersebut

Tabel 3. Distribusi jumlah data tiap kelas dari hasil prapemrosesan data

No	Kelas	Jumlah data awal	K-means+Euclidean		K-means+Manhattan	
			Jumlah	Data outlier	Jumlah	Data outlier
1	Lulus tepat waktu (1)	1.979	1.522	457 23,09%	1.528	451 22,79%
2	Lulus tidak tepat waktu (2)	3.183	1.364	1.819 57,15%	1.337	1.846 57,99%
3	Drop out (3)	1.685	1.486	199 11,81%	1.460	225 13,35%
<i>Total data</i>		<i>6.847</i>	<i>4.372</i>	<i>2.475 36,15%</i>	<i>4.325</i>	<i>2.522 36,83%</i>

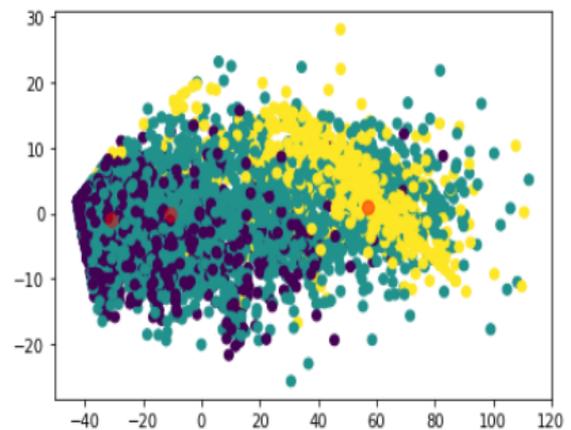
terdeteksi lebih dekat dengan pusat kluster yang lain, maka data tersebut teridentifikasi sebagai *outlier*.

Matriks jarak yang digunakan adalah Euclidean dan Manhattan. Hasil evaluasi matriks jarak menggunakan Euclidean berhasil mendeteksi data *outlier* sebanyak 2.475 *instance* atau sebesar 36,15 % dari keseluruhan *instance* pada dataset, Matriks jarak Manhattan bisa mendeteksi data *outlier* sebanyak 2.522 *instance* atau sebesar 36,83 %. Rincian data *outlier* pada masing-masing kelas dari evaluasi menggunakan matriks jarak Euclidean dan Manhattan dinyatakan dalam Tabel 3.

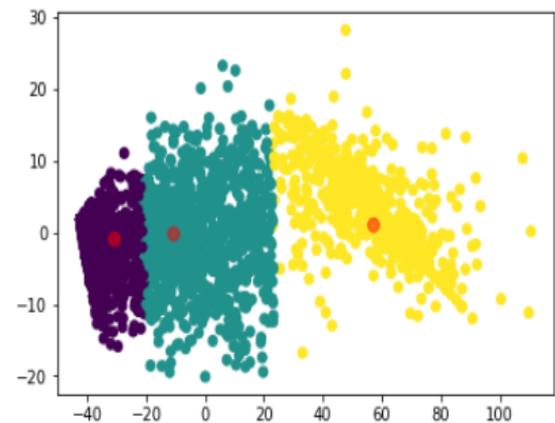
Langkah terakhir dari metode prapemrosesan yang diusulkan adalah menghapus data-data *outlier* yang ditemukan. Dataset yang dibersihkan dengan menggunakan metode K-means dan evaluasi matriks jarak Euclidean (K-means+Euclidean) menghasilkan dataset seperti pada Tabel 3, yaitu sejumlah 4.372 *instance*. Metode K-means dan matriks jarak Manhattan (K-means+Manhattan) menghasilkan dataset sejumlah 4.325 *instance*. Gambar 3 menunjukkan kondisi kluster dataset prapemrosesan menggunakan metode K-means+Euclidean, sedangkan Gambar 4 menunjukkan kondisi kluster dataset prapemrosesan menggunakan K-means+Manhattan.

Tahapan klasifikasi untuk memprediksi hasil studi mahasiswa UAD menggunakan algoritme klasifikasi kNN ini membandingkan pengujian dengan tiga metode, yaitu tanpa prapemrosesan, K-means+Euclidean, dan K-means+Manhattan. Perbandingan hasil klasifikasi ditunjukkan pada Tabel 4. Hasil prapemrosesan pembersihan *outlier* dari dataset pada klasifikasi algoritme kNN menunjukkan peningkatan akurasi yang signifikan dibandingkan sebelum dibersihkan dari *outlier*, yaitu peningkatan sebesar 26,14 % untuk metode K-means+Euclidean dan peningkatan sebesar 25,48 % untuk metode K-means+Manhattan.

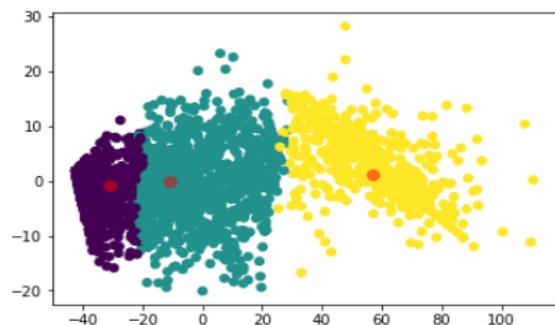
Metode penghapusan *outlier* pada prapemrosesan klasifikasi memberikan kinerja lebih baik dibandingkan dengan optimasi algoritme klasifikasi pada dataset yang memiliki *outlier*. Hal ini dibuktikan dengan hasil metode usulan memiliki hasil akurasi jauh lebih besar dibandingkan dengan metode dalam [6] yang hanya dapat menaikkan 1 % sampai dengan 2 % pada tiap matriks jarak yang digunakan. Lebih lanjut, metode dalam [7] hanya dapat meningkatkan akurasi tertinggi sampai 16,18 %. Performa kinerja matriks jarak metode usulan ini selaras dengan [15] yang menunjukkan bahwa matriks jarak Euclidean memberikan kinerja yang lebih baik dibandingkan dengan Manhattan, namun bertentangan dengan [16] dalam pengklasteran ringkasan teks. Hal tersebut menunjukkan bahwa kinerja



Gambar 2. Kluster data tanpa prapemrosesan



Gambar 3. Kluster data dengan prapemrosesan K-means+Euclidean



Gambar 4. Kluster data dengan prapemrosesan K-means+Manhattan

matriks jarak ini ditentukan kelas dan dataset yang digunakan dalam aplikasi tertentu [18].

V. KESIMPULAN

Outlier pada dataset dapat terdeteksi dan dihapus dengan menggunakan metode K-means dengan jarak matriks pada tahap prapemrosesan klasifikasi. Hasil akurasi klasifikasi pada algoritme kNN dapat meningkat dibandingkan dengan sebelum adanya prapemrosesan. Prapemrosesan dengan menggunakan metode K-means+Euclidean menunjukkan kinerja lebih baik, yaitu mendapatkan hasil akurasi 98,42 % (meningkat 26,14 %), jika dibandingkan dengan metode K-means+Manhattan yang mendapatkan akurasi 97,76 % (meningkat 25,48 %).

UCAPAN TERIMA KASIH

Terima kasih ditujukan kepada Biro Akademik dan Admisi, Universitas Ahmad Dahlan Yogyakarta yang telah mengizinkan pengambilan data untuk penelitian ini.

DAFTAR PUSTAKA

- [1] M. Maisah, F. Hairul, A. Iwan, A. Amiruddin, and Zulqarnain, "Strategi pengembangan mutu perguruan tinggi," *Jurnal Ilmu Manajemen Terapan*, vol. 1, no. 5, pp. 416-424, 2020. doi: [10.31933/jimt.v1i5.202](https://doi.org/10.31933/jimt.v1i5.202)
- [2] I. P. Darmawan and D. Triastanti, "Pola perwalian sebagai pembinaan akademik, kerohanian dan karakter mahasiswa," *Jurnal Ilmiah Religiosity Entity Humanit*, vol. 2, no. 1, pp. 13-26, 2020. doi: [10.37364/jireh.v2i1.32](https://doi.org/10.37364/jireh.v2i1.32)
- [3] J. Han, M. Kamber, and J. Pei, "8 - Classification: Basic Concepts," in *Data Mining*, Third Edition, J. Han, M. Kamber, and J. Pei, Eds. Boston: Morgan Kaufmann, 2012, pp. 327-391.
- [4] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, vol. 12, no. 1, pp. 90-108, 2016. doi: [10.1016/j.aci.2014.10.001](https://doi.org/10.1016/j.aci.2014.10.001)
- [5] M. Hijazi, J. Khalife, H. El Ghor, and J. Verdejo, "Network traffic classification based on class weight based k-NN classifier (CWK-NN)," in *2nd International Conference on Big Data and Cyber-Security Intelligence*, Versailles, France, Dec. 2019, pp. 105-112.
- [6] G. Bhattacharya, K. Ghosh, and A. S. Chowdhury, "kNN classification with an outlier informative distance measure," in *Pattern Recognition and Machine Intelligence*, Kolkata, India, Dec. 2017, pp. 21-27. doi: [10.1007/978-3-319-69900-4_3](https://doi.org/10.1007/978-3-319-69900-4_3)
- [7] D. C. P. Sinaga, T. Tulus, and P. Sihombing, "Performance of distance-based k-nearest neighbor classification method using local mean vector and harmonic distance," *IOP Conference Series:*

Tabel 4. Hasil perbandingan performa klasifikasi kNN (dalam %)

Kelas	Accuracy	Precision	Recall	F1-score
Tanpa Prapemrosesan	72,28	74,80	72,70	73,74
K-Means+ Euclidean	98,42	98,40	98,43	98,42
K-Means+ Manhattan	97,76	97,72	97,78	97,75

Materials Science and Engineering, vol. 725, 12122, 2020. doi: [10.1088/1757-899X/725/1/012122](https://doi.org/10.1088/1757-899X/725/1/012122)

- [8] F. Yoseph and M. Heikkilä, "A clustering approach for outliers detection in a big point-of-sales database," in *International Conference on Machine Learning and Data Engineering*, Taipei, Taiwan, Dec. 2019, pp. 65-71. doi: [10.1109/iCMLDE49015.2019.00023](https://doi.org/10.1109/iCMLDE49015.2019.00023)
- [9] V. Bhatt, M. Dhakar, and B. Chaurasia, "Filtered clustering based on local outlier factor in data mining," *International Journal of Database Theory and Application*, vol. 9, pp. 275-282, 2016. doi: [10.14257/ijda.2016.9.5.28](https://doi.org/10.14257/ijda.2016.9.5.28)
- [10] G. Mishra, S. Agarwal, P. K. Jain, and R. Pamula, "Outlier detection using subset formation of clustering based method," in *International Conference on Advanced Computing Networking and Informatics*, West Bengal, India, Dec. 2019, pp. 521-528. doi: [10.1007/978-981-13-2673-8_55](https://doi.org/10.1007/978-981-13-2673-8_55)
- [11] B. Angelin and A. Geetha, "Outlier detection using clustering techniques - k-means and k-median," in *4th International Conference on Intelligent Computing and Control Systems*, Madurai, India, May 2020, pp. 373-378. doi: [10.1109/ICICCS48265.2020.9120990](https://doi.org/10.1109/ICICCS48265.2020.9120990)
- [12] H. Gustavsson, "Clustering based outlier detection for improved situation awareness within air traffic control," *Thesis*, KTH Royal Institute of Technology, Sweden, 2019.
- [13] T. Nizam and S. I. Hassan, "Exemplifying the effects of distance metrics on clustering techniques: f-measure, accuracy and efficiency," in *7th International Conference on Computing for Sustainable Global Development*, New Delhi, India, Mar. 2020, pp. 39-44. doi: [10.23919/INDIACom49435.2020.9083687](https://doi.org/10.23919/INDIACom49435.2020.9083687)
- [14] S. Aggarwal, N. Agarwal, and M. Jain, "Performance analysis of uncertain k-means clustering algorithm using different distance metrics," in *Computational Intelligence: Theories, Applications and Future Directions - Volume I*, 2019, pp. 237-245. doi: [10.1007/978-981-13-1132-1_19](https://doi.org/10.1007/978-981-13-1132-1_19)
- [15] S. Kapil and M. Chawla, "Performance evaluation of k-means clustering algorithm with various distance metrics," in *1st International Conference on Power Electronics, Intelligent Control and*

- Energy Systems*, New Delhi, India, Jul. 2016, pp. 1–4. doi: [10.1109/ICPEICES.2016.7853264](https://doi.org/10.1109/ICPEICES.2016.7853264)
- [16] S. A. Salihu, I. P. Onyekwere, M. A. Mabayoje, and H. A. Mojeed, “Performance evaluation of manhattan and euclidean distance measures for clustering based automatic text summarization,” *Journal of Engineering and Technology*, vol. 4, no. 1, pp. 135-139, 2019. doi: [10.46792/fuoyejet.v4i1.316](https://doi.org/10.46792/fuoyejet.v4i1.316)
- [17] Y.-D. Zhang and S. Wang, “Detection of Alzheimer’s disease by displacement field and machine learning,” *PeerJ*, vol. 3, E1251, 2015. doi: [10.7717/peerj.1251](https://doi.org/10.7717/peerj.1251)
- [18] Y. S. Thakare and S. B. Bagal, “Performance evaluation of k-means clustering algorithm with various distance metrics,” *International Journal of Computer Application*, vol. 110, no. 11, pp. 12-16, 2015.