

EVALUASI KINERJA ALGORITMA K-MEANS DENGAN MATRIKS JARAK EUCLIDEAN PADA PENENTUAN SISWA BERMASALAH

Nur Aeni Widiastuti

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika

Universitas Islam Nahdlatul Ulama Jepara

Email: nuraeniwidiastuti@unisnu.ac.id

Ahmat Wakit

Fakultas Sains dan Teknologi, Program Studi Teknik Sipil

Universitas Islam Nahdlatul Ulama Jepara

Email: ahmatwakit@unisnu.ac.id

ABSTRAK

Dalam lingkup pendidikan, penilaian siswa tidak hanya berdasarkan nilai akademik maupun non-akademik tetapi juga penilaian dalam karakter siswa itu sendiri, dengan adanya catatan kejadian siswa semua aktivitas siswa akan tercatat. Namun dalam penerapannya perlu metode dalam penentuan siswa bermasalah agar cepat serta akurat dalam penentuannya. Data yang digunakan dalam penelitian ini adalah data Anecdotal Record, berjumlah 47 kategori untuk semua siswa di SMK Walisongo Pecangaan Tahun Pelajaran 2019/2020 yang berjumlah 533 siswa. Dalam pengambilan data menggunakan data kejadian siswa meliputi Aspek Sikap perilaku, Aspek Kerajinan dan Aspek Kerapian. Data diolah dengan melakukan clustering siswa dalam bermasalah, atau tidak bermasalah. Metode clustering yang digunakan dalam penelitian ini adalah metode K-Means dengan matriks jarak Euclidean. Centroid data untuk siswa tidak bermasalah 2.165 dan siswa bermasalah 7.372. Sehingga diperoleh siswa bermasalah sejumlah 87 siswa dan siswa tidak bermasalah sejumlah 446 siswa Dengan nilai Davies bouldin index mendapatkan score 0,397 dengan kategori akurat dalam clusterisasinya.

Kata kunci: clustering, euclidean distance, davies bouldin Index

ABSTRACT

In the scope of education, student assessments are not only based on academic and non-academic scores but also assessments in the character of the students themselves, with a record of student incidents all student activities will be recorded. However, in its application, a method for determining problem students is needed so that it is fast and accurate in determining it. The data used in this study were Anecdotal Record data, totaling 47 categories for all students at SMK Walisongo Pecangaan in the 2019/2020 academic year, totaling 533 students. In taking data using incident data, students include aspects of attitude behavior, aspects of craft and aspects of neatness. The data is processed by clustering students with problems or problems. The clustering method used in this research is the K-Means method with the Euclidean distance matrix. The centroid data for students with no problems is 2,165 and for students with problems of 7,372. So it is obtained that there are 87 students with problems and 446 students with no problem. With the Davies bouldin index score, it gets a score of 0.397 with the accurate category in its clustering.

Keywords: clustering, euclidean distance, davies bouldin Index

1. PENDAHULUAN

Pendidikan mempunyai peran penting dalam pembangunan suatu bangsa. Keberhasilan pendidikan dipengaruhi oleh proses belajar yang ada di lingkungan sekolah. Pembelajaran yang ada di sekolah merupakan bekal bagi peserta didik untuk bersosialisasi di lingkungan masyarakat. Oleh sebab itu, sekolah sering dijadikan media pembentuk karakter.

Pada zaman sekarang banyak remaja yang kehilangan identitas diri sehingga mengarah kepada kenakalan remaja. Kenakalan remaja merupakan salah satu perilaku menyimpang. Biasanya, anak-anak yang seperti itu cenderung menyalahgunakan kontrol diri dan meremehkan keberadaan orang lain. Kelompok yang rentan terhadap perilaku menyimpang adalah peserta didik yang biasanya dibawa ke sekolah sehingga bisa mempengaruhi peserta didik yang lain untuk masuk ke dalam siswa yang bermasalah[1]. Kenakalan remaja yang dilakukan oleh Siswa merupakan persoalan yang membutuhkan penanganan serius untuk memperbaiki perilaku sosial agar lebih mampu diterima di lingkungan masyarakat. Oleh sebab itu, melalui guru Bimbingan Konseling diharapkan dapat memberikan solusi permasalahan yang dihadapi siswa dengan memberikan layanan konseling di sekolah[2].

SMK Walisongo Pecangaan Jepara merupakan sekolah menengah kejuruan yang menerapkan jika ada siswa yang melakukan pelanggaran akan dicatat oleh wali kelas di buku pelanggaran kemudian akan disetorkan ke guru bimbingan konseling untuk ditindaklanjuti dalam perbaikan perilaku dan kedisiplinan siswa. Siswa yang melakukan pelanggaran akan diberikan sanksi sesuai dengan pelanggaran yang dibuat. Pelanggaran yang dilakukan oleh siswa didokumentasikan dalam *anecdotal record*, yaitu sebuah cara pengumpulan data dengan melalui pengamatan secara langsung mengenai sikap dan perilaku yang muncul dengan tiba-tiba atau peristiwa yang terjadi secara insidental[3]. Akan tetapi dalam pelaksanaannya pihak sekolah masih terkendala dalam menentukan siswa-siswi yang bermasalah, dikarenakan jumlah siswa yang banyak dan banyaknya pelanggaran siswa-siswi yang sering terjadi. Berdasarkan permasalahan yang muncul di SMK Walisongo adalah Masih menggunakan metode perkiraan dalam menentukan siswa yang bermasalah dikarenakan jumlah pelanggaran ada 47 kategori. Oleh sebab itu, pihak sekolah memerlukan sebuah metode untuk menyelesaikan masalah tersebut dengan tepat, cepat, dan akurat.

Berdasarkan permasalahan yang ada di SMK Walisongo akan diselesaikan menggunakan algoritma data mining K-means dengan evaluasi matriks jarak Euclidean untuk mengetahui ke akurasi dalam menggolongkan siswa bermasalah atau tidak. Beberapa penelitian tentang algoritma k-means diantaranya penelitian tentang penilaian dosen berdasarkan indeks kepuasan mahasiswa. Data didapat dari mahasiswa dengan pembagian kuesioner atau angket sesuai aspek *Empathy*, *Assurance*, *Responsiveness*, dan *Reliability*. Data diolah untuk mendapatkan nilai dari tiap dosen. Data kemudian diolah menggunakan Software RapidMiner guna menentukan nilainya dalam cluster baik dan cluster kurang baik dengan algoritma K-Means[4]. Kemudian penelitian yang dilakukan oleh Fauziah,dkk (2017) tentang penerapan algoritma K-Means pada siswa baru sekolah menengah kejuruan untuk *clustering* jurusan, dengan menggunakan data nilai tes masuk sekolah, tanggungan anak orang tua, dan penghasilan orang tua. Berdasarkan hasil *cluster* dengan menerapkan beberapa kriteria dari calon siswa menggunakan K-Means dapat diambil pengelompokan dengan rata – rata jurusan yang diambil adalah rekayasa perangkat lunak dan sedikit jumlah siswa yang tidak lulus. Bahkan ada beberapa jurusan yang tidak dibuka dikarenakan kriteria – kriteria siswa tidak dapat lulus dalam jurusan tersebut [5] . Penelitian lain yang membahas implementasi Algoritma K-Means *clustering* untuk menentukan kelas kelompok bimbingan belajar tambahan (studi kasus : Siswa SMA Negeri 1 Ranah Pesisir), dengan menggunakan data nilai UAS Semester IV SMAN 1 Ranah Pesisir jurusan IPA, data tersebut akan dikategorikan menjadi tiga kelompok yaitu kemampuan siswa kurang pintar, siswa sedang, dan siswa pintar. Sehingga dari perhitungan tersebut dapat mengetahui siswa yang berhak diberi pelajaran tambahan supaya dapat mencapai nilai standar kelulusan Ujian Nasional[6]. Kemudian penelitian yang membandingkan matriks jarak Euclidean, Manhattan dan Haversine diperoleh Haversine paling akurat dalam perhitungan penentuan posisi karyawan [7]. Penelitian yang dilakukan oleh We Gie dan Deny Jollyta (2020) tentang perbandingan Euclidean dan Manhattan untuk optimasi status cluster covid -19 di Riau yang dievaluasi dengan Davies Bouldin Index menghasilkan Euclidean paling akurat dengan jarak 0,394 [8].

Dari beberapa penelitian yang pernah dilakukan oleh [4] dan [8] algoritma yang digunakan adalah k-means dengan perbandingan matriks jaraknya menggunakan *euclidean distance* dengan metode evaluasi dengan menggunakan DBI. Pada penelitian ini data Anecdotal Record yang ada di SMK Walisongo Pecangaan Jepara. Data yang digunakan adalah data kejadian siswa meliputi Aspek

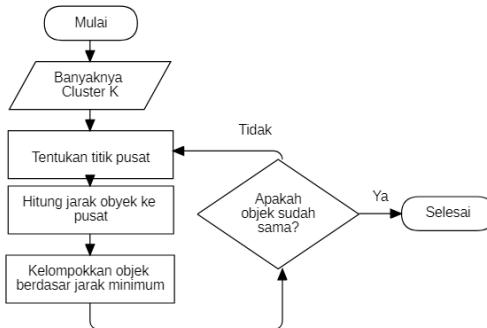
Sikap perilaku, Aspek Kerajinan dan Aspek Kerapian. Data akan di clusterkan menjadi 2 yaitu siswa bermasalah, atau tidak bermasalah. *Tools* yang digunakan untuk pengolahan data awal adalah MS Excel yang kemudian akan dimasukkan ke RapidMiner. Diharapkan dari penelitian dihasilkan pola dalam penentuan prediksi siswa bermasalah atau tidak bermasalah yang akurat.

2. METODOLOGI PENELITIAN

Desain penelitian yang digunakan dalam penelitian ini adalah analisis kuantitatif. Metode ini merupakan metode penelitian yang menekankan analisisnya terhadap data berupa *numerical* atau angka sebagai alat penunjang tentang apa yang ingin dicapai dalam penelitian. Metode pengumpulan data yang digunakan metode observasi, dokumentasi dan studi pustaka. Dari pengumpulan data diperoleh data *anecdotal record* tahun pelajaran 2019/2020 dengan 47 kategori dan jumlah siswa 533 orang.

Setelah data diperoleh selanjutnya dilakukan *preprocessing* data, agar tidak ada duplikasi data, tidak *missing values* dan memperbaiki kesalahan-kesalahan yang ada pada data set baru dalam format *excel*. Selanjutnya data akan dilakukan pembobotan sesuai dengan poinnya dan dijumlah sehingga pada tahapan ini sudah diperoleh perhitungan nilai berdasarkan Anecdotal Record pada data pelanggaran.

Metode yang diusulkan adalah Algoritma K-Means *Clustering*. Data yang sudah di pre prosesing akan diolah dengan algoritma k-means dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma dengan rumus *Euclidean Distance*[4]. Cara kerja Algoritma K-Means adalah algoritma yang mempartisi data ke dalam cluster – cluster sehingga data yang mempunyai kemiripan berada pada satu cluster yang sama sedangkan data yang memiliki ketidaksamaan berada pada cluster yang lain[9]. Analisis cluster digunakan sebagai metode untuk membagi rangkaian data menjadi beberapa grup/kategori berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Pada Gambar 1 merupakan langkah-langkah untuk clustering sesuai dengan metode K-Means sebagai berikut :



Gambar 1 Langkah-langkah Clustering

Langkah-Langkah Clustering:

- 1) Menentukan K sebagai jumlah cluster yang ingin dibentuk.
- 2) Membangkitkan nilai acak untuk pusat cluster awal (centroid) sebanyak Kluster yang telah ditentukan.
- 3) Menghitung jarak setiap data input terhadap masing – masing centroid menggunakan rumus jarak Euclidean (*Euclidean Distance*) sampai ditemukan jarak yang paling terdekat dari setiap data dengan centroid. Berikut adalah persamaan *Euclidian Distance* :

$$d(c_j, x_i) = \sqrt{\sum (x_i - c_j)^2} \quad (i)$$

Dimana :

d : titik dokumen

xi : data kriteria,

cj : centroid pada cluster ke-j

- 4) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terpendek).
- 5) Setelah didapat hasil perhitungan data dengan centroid awal setelah itu dicari nilai rata-rata tiap anggota cluster yang nantinya akan digunakan sebagai penentuan centroid baru.
- 6) Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terpendek).
- 7) Setelah didapat hasil perhitungan data dengan centroid awal setelah itu dicari nilai rata-rata tiap anggota cluster yang nantinya akan digunakan sebagai penentuan centroid baru. Melakukan pengulangan dari langkah 2 sampai 5, sampai anggota tiap cluster tidak ada yang berubah. Jika langkah 6 telah terpenuhi, maka nilai pusat cluster (cj) pada literasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data[10].
- 8) Kemudian melakukan evaluasi cluster dengan metode *Davis Bouldin Index* (DBI) [8].

$$DBI = \frac{1}{M} \sum_{j=1}^M \text{Max}_{j \neq k} R_{j,k} \quad (\text{ii})$$

Penelitian ini akan menghitung nilai siswa berdasarkan data Anecdotal Record pada Data pelanggaran dengan menggunakan *tools* RapidMiner. Rapid Miner sebagai salah satu *software* untuk pengolahan data mining yang menyediakan berbagai tool untuk membuat sebuah *decision tree*[11]. Evaluasi yang digunakan dengan *Davies Bouldin Index* (DBI) sebagai *Performance*. DBI menginginkan nilai dengan sekecil (non-negatif ≥ 0) untuk menentukan baik atau tidaknya cluster yang didapat. semakin kecil nilai indeksnya semakin bagus cluster. Diharapkan dengan menggunakan algoritma k-means dan di validasi dengan DBI dapat bekerja dengan baik sehingga dapat meningkatkan akurasi dalam penentuan siswa bermasalah.

3. HASIL DAN PEMBAHASAN

Dalam melakukan clustering, data yang diperoleh akan dihitung terlebih dahulu berdasarkan bobot yang telah ditentukan. Data akan dihitung sehingga didapatkan data yang siap dicluster.

Sumber data utama yang digunakan dalam penelitian ini adalah *database* Anecdotal Record SMK Walisongo Pecangaan Jepara Tahun Pelajaran 2019/2020 dengan jumlah keseluruhan data ada 533 record.

Gambar 2 Gambar 1* ARABIC 1.
Flowchart Algoritma K-MeansADDIN
CSL_CITATION ("citationItems" : [{ "id" :
"ITEM-1" , "itemData" : { "ISBN" :
"978-623-54-7336-0" , "NAME" : "BAB
I'U00a0u00a0PPI DAHULUANu00a001.1
PENGERTIAN DATA MININGDefinisi
sederhana dari data mining adalah
ekstraksiu00a0informasi atau pola yang
penting atau berharga dari data yang
ada di dalam database yang besar. Dalam jumlah limiah,
data mining juga dikenal dengan nama
Knowledge Discovery in
Database(KDD).u00a0data mining
didefinisikan sebagai suatu teknik
yangdigunakan secara otomatis untuk
mengeksplorasi secara menyeluruh dan
membawa ke perumkaan relasi-relasi
yangkompleks pada set data yang sangat
besar. Data mining merupakan proses analisis
set data yang berbentuk tabulasi, seperti yang
banyak diimplementasikan dalam
teknologi manajemen basis data relasional.
Akhir tetapi teknik-teknik data mining dapat
juga dipemasukkan pada representasi data yang
lain, seperti domain data spatial, berbasis

Gambar 2 Anecdotal Record

Setelah data diperoleh dilakukan pembersihan data terhadap data yang memiliki *redundant*. Dalam dataset terdapat atribut NIS, Nama Siswa, Kelas dan Pelanggaran Siswa, atribut yang dihapus yaitu jenis kelamin dan Kelas. Hal itu dilakukan karena data tersebut tidak akan dipakai dalam perhitungan.

ID	NIS	NAMA	KERAJINAN														
			A16	A17	A18	A19	A20	A21	B01	B02	B03	B04	B05	B06	B07	B08	B09
242	2074	DIMAS ARIFIN IHAM	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
243	2078	EKO LORIJAWAN	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0
244	2080	FANYA TRISYAWATI	0	0	0	0	0	0	2	1	0	1	0	0	0	0	0
245	2081	FAARAH DIAN PUSPITA	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
246	2086	IHAM SUYADI	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
247	2088	KOMARIATUN AFIFA NURCAHYANI	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
248	2089	LUCKY ARDIANSYAH	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
249	2091	LUTFIYATUL MALIKHAH	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
250	2092	M. AZKA TASLIMI CHOREI	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
251	2093	M. SALMAN FARISH FAISAL FAIS	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
252	2094	MAFTUH AHNAN	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
253	2095	MAFTUKHATU RIZQIYAH	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
254	2096	MAHMUD SUFAT'AT	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
255	2097	MARIA ANA ELINA	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
256	2098	MARIA AUFIA	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
257	2099	MAULIA YUSUF WAIFI UDDIN	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
258	2100	MOHAMAD AZKA FAHMI	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
259	2101	MUHAMMAD ALFIN SYAIFUDIN	0	0	0	0	0	0	1	0	1	0	0	3	0	0	0
260	2113	MUHAMMAD UMAR FARUQ	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0
261	2115	NIAM KHOLDIN GHOFUR	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0
262	2119	RAFLY YUSUF	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0

Gambar 3. Hasil Cleaning Data

Data tersebut kemudian dibobotkan sesuai dengan poinnya kemudian diakumulasi. Berikut Tabel 1. Data bobot pelanggaran dan Tabel 2. Hasil Pembobotan:

Tabel 1. Data Bobot Pelanggaran

Kode	Nama pelanggaran	Poin
A01	Berkuku Panjang	2
A02	Mengganggu ketenangan Kegiatan Belajar Mengajar	2
...
A21	Berzina / hamil	100
B01	Datang terlambat <10 menit	1
B02	Datang terlambat <30 menit	2
...
B12	Tidak membawa mukena	3
C01	Tidak Memasukan Baju	1
C02	Tidak memakai kaoskaki	2
...
C14	Pakaian transparan (bagi siswa putri)	30

Tabel 2. Hasil Pembobotan

No	NIS	Nilai
1	2322	0
2	2323	9
3	2325	9
4	2326	8
5	2327	9
6	2358	5
7	2361	4
8	2328	15
9	2213	8
10	2206	8
...
533	2366	0

Data yang telah melalui tahap *Pre-processing* selanjutnya akan dibagi menjadi dua cluster, sebelumnya tentukan centroid awal terlebih dahulu. Ambil data terkecil dan terbesar sebagai perhitungan pertama, *cluster 0* sebagai Siswa tidak bermasalah dan *cluster 1* sebagai siswa

bermasalah. Perhitungan jarak minimum data terhadap centroid menggunakan persamaan *Euclidean Distance*.

Tabel 3. Nilai Centroid Awal

Cluster	x
C ₀	0
C ₁	18

Tabel 3 adalah nilai centroid awal yang akan digunakan untuk perhitungan k-means pada literasi ke -1 dihasilkan data 516 siswa dinyatakan masuk C₀ dan 17 masuk C₁, sesuai hasil nilai yang terkecil sesuai dengan Tabel 4.

Tabel 4. Pengelompokkan Data Pada Literasi ke- 1

NO	ID	C ₀	C ₁	Cluster
1	2322	0	18	C ₀
2	2323	9	9	C ₀
3	2325	9	9	C ₀
4	2326	8	10	C ₀
5	2327	9	9	C ₀
6	2358	5	13	C ₀
7	2361	4	14	C ₀
8	2328	15	3	C ₁
9	2213	8	10	C ₀
10	2206	8	10	C ₀
11	2207	6	12	C ₀
12	2208	11	7	C ₁
13	2209	4	14	C ₀
14	2210	10	8	C ₁
15	2211	8	10	C ₀
16	2212	4	14	C ₀
17	2214	12	6	C ₁
18	2215	8	10	C ₀
19	2216	4	14	C ₀
20	2217	9	9	C ₀
....
533	2366	0	18	C ₀

Setelah mendapatkan label cluster untuk masing-masing data maka dicari nilai rata-ratanya dengan menjumlahkan seluruh anggota masing-masing cluster dan dibagi jumlah anggotanya.untuk menentukan nilai centroid literasi selanjutnya. Pada data ini literasinya sampai literasi ke 3. Karena *centroid* tidak mengalami perubahan (sama dengan centroid sebelumnya) maka proses *clustering* selesai, dan hasil yang dipakai pada literasi terakhir sesuai dengan Tabel 5. siswa yang dinyatakan tidak bermasalah (C₀) ada 446 siswa dan siswa yang bermasalah (C₁) berjumlah 87 orang.

Tabel 5. Perbandingan Nilai Literasi

Cluster	Literasi ke -				
	1	2	3	4	5
C ₀	516	471	446	446	446
C ₁	17	62	87	87	87

Hasil pada proses *Clustering* dalam *K-Means* sebelumnya akan dilakukan validasi dengan *Davies-Bouldin Index*. Data hasil perhitungan DBI dilihat pada Gambar 6. Dari hasil perhitungan nilai *Davies Bouldin Index* cluster k-means yang sudah diolah dinyatakan akurat/valid . Sedangkan Hasil dari perhitungan dengan Rapid Miner dihasilkan 446 siswa dalam cluster tidak bermasalah dan 87 siswa dalam cluster bermasalah. Dengan nilai Davied Bouldin Index : 0.397. Hasil Evaluasi Cluster dengan menggunakan Davies bouldin indexdari perhitungan yang telah dilakukan terlihat pada Gambar 6.

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 3.015  
Avg. within centroid distance_cluster_0: 2.165  
Avg. within centroid distance_cluster_1: 7.372  
Davies Bouldin: 0.397
```

Gambar 6. Hasil Performance

Dari hasil penelitian dapat disimpulkan bahwa penelitian ini lebih baik dari penelitian [8] dimana dihasilkan nilai euclidean yang paling bagus di k 8 dengan nilai 0,39 dan nilai manhattan di k 7 dengan nilai 0,434. Sedangkan pada penelitian ini bisa menghasilkan nilai 0,39 di k 3. Sedangkan untuk nilai Manhattan pada data ini sama dengan nilai euclidean. Kemudian jika dibandingkan dengan penelitian [5] pada penelitian ini menggunakan metode evaluasi DBI untuk mengetahui keakuratan dalam pengclusterannya. Sehingga dapat disimpulkan dari beberapa penelitian yang dijadikan referensi bahwa perhitungan Euclidean pada data anacdetal record bisa lebih bagus dan bekerja dengan efektif dan hasil evaluasinya hasil clustenya dengan nilai yang baik dengan nilai 0.39.

4. KESIMPULAN

Penerapan algoritma k-means dalam menentukan siswa bermasalah dengan jumlah data 533 siswa dan di-cluster-kan menjadi 2 yaitu siswa bermasalah dan tidak bermasalah menghasilkan penelitian dengan Centroid data untuk siswa tidak bermasalah 2.165 dan siswa bermasalah 7.372. Sehingga diperoleh siswa bermasalah sejumlah 67 siswa dan siswa tidak bermasalah sejumlah 446 siswa nilai DBI 0.39 bahwa hasil clustenya akurat. Sehingga algoritma ini dapat diterapkan di sekolah untuk penentuan siswa bermasalah terutama di SMK Walisongo Jepara.

Saran untuk penelitian selanjutnya dapat dilakukan penelitian lanjutan dengan menghitung jarak matriks Euclidean, Manhattan atau Haversine dan dievaluasi validasi dengan metode yang lain untuk menghasilkan cluster yang akurat.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada bapak Ardana Hermawan, S.T. Selaku Kepala Sekolah SMK Walisongo Pecangaan Jepara atas izin dan bantuan dalam melakukan penelitian ini.

DAFTAR PUSTAKA

- [1] L. W. Etsukrilasa, Yuline, "Persepsi terhadap perilaku bermasalah pada siswa kelas viii smpn 19 pontianak.," 2013.
- [2] W. N. Wendari, A. Badrujaman, and A. Sismati S., "Profil Permasalahan Siswa Sekolah Menengah Pertama (Smp) Negeri Di Kota Bogor," *Insight J. Bimbing. Konseling*, vol. 5, no. 1, p. 134, 2016.
- [3] "Kamus Bahasa Indonesia." Jakarta : Pusat Bahasa, 2008.
- [4] A. Nurzahputra, M. A. Muslim, and M. Khusniati, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," *Techno.Com*, vol. 16, no. 1, pp. 17–24, Jan. 2017.
- [5] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, Mar. 2017.
- [6] M. Mardalius, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Kelas Kelompok Bimbingan Belajar Tambahan (Studi Kasus: Siswa SMA Negeri 1 Ranah Pesisir)," Open Science Framework, preprint, Aug. 2018.
- [7] Y. Miftahuddin, S. Umaroh, and F. R. Karim, "Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan," *J. Tekno Insentif*, vol. 14, no. 2, pp. 69–77, 2020.
- [8] W. Gie *et al.*, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index : Status Covid-19 Wilayah Riau," *Pros. Semin. Nas. Ris. Dan Inf. Sci. 2020*, vol. 2, no. April, pp. 187–191, 2020.
- [9] E. Prasetyo, *Data Mining mengolah data menjadi informasi menggunakan Matlab*. Yogyakarta: Andi, 2018.
- [10] Kusrini and E. T. Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: Andi, 2009.
- [11] A. Puspabhuana and A. M. Siregar, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Surakarta: CV Kekata Group, 2017.