

Algoritma *K-Nearest Neighbor* dengan *Euclidean Distance* dan *Manhattan Distance* untuk Klasifikasi Transportasi Bus


Rozzi Kesuma Dinata^{a,1,*}, Hafizal Akbar^{a,2} dan Novia Hasdyna^{b,3}

^a Universitas Malikussaleh, Jl. Cot Tengku Nie, Reuleut, Aceh Utara, 24351, Indonesia

^b Universitas Islam Kebangsaan Indonesia, Jl. Medan B. Aceh Sp. Elak Lhokseumawe, 24352, Indonesia

¹rozzi@unimal.ac.id; ²hafizalakbar01@gmail.com; ³noviahasdyna@gmail.com

*corresponding author

INFORMASI ARTIKEL	ABSTRAK
<p>Diterima : 19 Maret 2020 Diulas : 23 Juli 2020 Direvisi : 24 Juli 2020 Diterbitkan : 27 Agustus 2020</p> <p>Kata Kunci: Komparasi Klasifikasi K-Nearest Neighbor Euclidean Distance Manhattan Distance</p>	<p><i>K-Nearest Neighbor</i> merupakan salah satu algoritma data mining yang dapat digunakan untuk proses klasifikasi. <i>K-Nearest Neighbor</i> bekerja mengklasifikasikan data berdasarkan jarak terdekat. Penelitian ini menggunakan metode <i>Euclidean distance</i> dan <i>Manhattan distance</i> untuk menghitung jarak dalam pengklasifikasian transportasi bus jalur Lhokseumawe-Medan. Data yang digunakan diperoleh dari Organisasi Angkutan Darat Kota Lhokseumawe. Berdasarkan hasil pengujian yang telah dilakukan dengan $k=3$, didapatkan performansi K-NN dengan <i>Euclidean Distance</i> dengan nilai rata-rata dari <i>Precision</i> 44,94%, <i>Recall</i> 37,06%, <i>Accuracy</i> 81,96%. Sedangkan performansi K-NN dengan <i>Manhattan Distance</i> diperoleh nilai rata-rata dari <i>Precision</i> sebesar 45,49%, <i>Recall</i> 36,39%, <i>Accuracy</i> 84,00%. Hasil analisis perbandingan perhitungan jarak didapatkan nilai akurasi tertinggi pada <i>Manhattan Distance</i>, dengan selisih 2,04% lebih tinggi daripada <i>Euclidean Distance</i>. Persentase tersebut menunjukkan bahwa <i>Manhattan Distance</i> lebih akurat dibandingkan dengan <i>Euclidean Distance</i> dalam klasifikasi transportasi bus.</p>
<p>Keywords: Comparison Classification K-Nearest Neighbor Euclidean Distance Manhattan Distance</p>	<p>ABSTRACT K-Nearest Neighbor is a data mining algorithm that can be used to classify data. K-Nearest Neighbor works based on the closest distance. This research using the Euclidean and Manhattan distances to calculate the distance of Lhokseumawe-Medan bus transportation. Data that used in this research was obtained from the Organisasi Angkutan Darat Kota Lhokseumawe. The results of the test with $k = 3$ has obtained the percentage of 44.94% for Precision, 37.06% Recall, and 81.96% Accuracy for the performance of K-NN with Euclidean Distance. Whereas by using Manhattan Distance the result obtained was 45.49% for Precision, 36.39% Recall, and 84.00% Accuracy. The result shown that Manhattan Distance obtained the highest accuracy, with the difference of 2.04% higher than Euclidean Distance. It indicates that Manhattan Distance is more accurate than Euclidean Distance to classify the bus transportation.</p> <p>This is an open access article under the CC-BY-SA license.</p> 

I. Pendahuluan

Data Mining merupakan proses ekstraksi data menjadi informasi yang memungkinkan para pengguna untuk mengakses secara cepat data dengan jumlah yang besar, dengan teknik yang tepat proses data mining akan memberikan hasil yang optimal [1]. Setiap data pada data mining terdiri dari kelas tertentu bersama dengan variabel dan faktor-faktor penentu kelas variabel tersebut. Dengan data mining, peneliti dapat menentukan suatu kelas dari variabel data yang dimiliki[2].

Salah satu tujuan yang banyak dihasilkan dalam data mining adalah klasifikasi[3]. Menurut Abdillah (2018), klasifikasi merupakan penggolongan atau pengelompokan fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui atau pembagian sesuatu menurut kelas-kelas nya. Metode-metode klasifikasi data

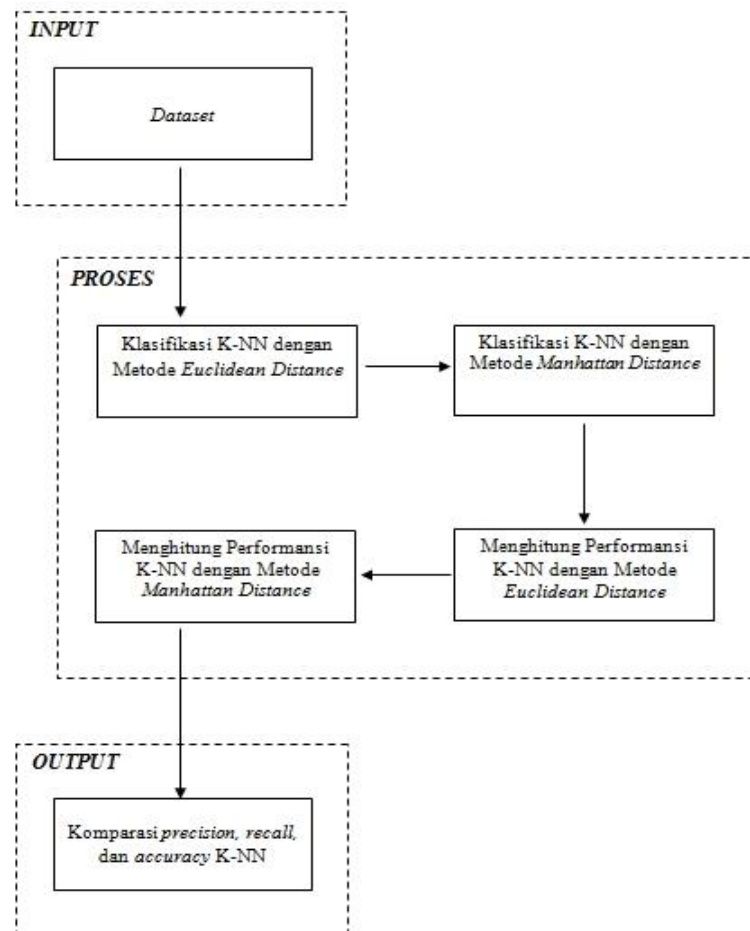
mining yang telah dilakukan oleh peneliti sebelumnya antara lain C4.5, *RainForest*, *Naïve Bayesian*, *Neural Network*, *Genetic Algorithm*, *Fuzzy*, *Case-Based Reasoning*, dan *K-Nearest Neighbor*.

Penelitian ini bertujuan untuk menganalisis performansi dari algoritma K-NN dengan membandingkan hasil akurasi dari metode *Euclidean Distance* dan *Manhattan Distance* pada klasifikasi Transportasi Bus Lhokseumawe-Medan. Data yang digunakan diperoleh dari Organisasi Angkutan Darat Kota Lhokseumawe. Adapun hasil analisis performansi terbaik dapat dijadikan rekomendasi dalam pemilihan metode perhitungan *distance* pada algoritma klasifikasi K-NN.

II. Metode

A. Kerangka Penelitian

Metode klasifikasi data mining yang digunakan adalah *K-Nearest Neighbor* dengan menggunakan *Euclidean Distance* dan *Manhattan Distance* untuk perhitungan jarak. *Framework* penelitian dapat dilihat pada Gambar 1.



Gambar 1. *Framework* Penelitian

Adapun langkah pertama adalah memilih dataset yang akan digunakan dalam proses penelitian, dalam penelitian ini menggunakan dataset Transportasi Bus Lhokseumawe-Medan. Selanjutnya pembagian dataset data training dan data testing. Setelah pembagian data menjadi dua bagian training dan testing dilakukan langkah seleksi *Class* dengan menggunakan Algoritma K-NN dengan dua metode pendekatan *Euclidean Distance* dan *Manhattan Distance*[4].

Langkah selanjutnya setelah didapatkan *Class* terbaik adalah tahapan klasifikasi data dengan metode klasifikasi data mining, sementara itu data testing yang telah terseleksi kemudian dilakukan validasi dengan data training. Setelah data training dan testing divalidasi dengan menggunakan metode klasifikasi. Langkah terakhir yaitu membandingkan hasil klasifikasi akurasi untuk mengetahui metode terbaik untuk klasifikasi dataset Transportasi Bus Lhokseumawe-Medan.

B. Algoritma K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu algoritma yang digunakan dalam pengklasifikasian[5]. Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *K-Nearest Neighbor* terdekatnya dalam data pelatihan[6]. Adapun langkah-langkah penggunaan metode K-NN ini dijelaskan sebagai berikut:

1. Tentukan parameter K
2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (urutan dari nilai yang terkecil hingga nilai yang terbesar).
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi[7].

Menurut Kustiyahningsih (2015), ada beberapa Metode Pendekatan K-NN. Dalam penelitian ini Metode Pendekatan yang digunakan adalah Metode *Euclidean Distance* dan *Manhattan Distance*. Berikut adalah rumus (1) untuk mencari nilai akar dari kuadrat dua *vector* [8].

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Keterangan:

x1 = Sampel data

x2 = Data uji atau data testing

i = Variabel data

d = Jarak

p = Jumlah data training

$$d_{ij} = \sum W_k |X_{ik} - C_{jk}| \quad (2)$$

Dimana dij dengan rumus (2) adalah jarak antara data testing dan data training dengan semua parameternya. W merepresentasikan dari jumlah bobot. X adalah data testing. C adalah training dalam case base [9].

C. Confusion Matrix

Confusion matrix melakukan pengujian untuk memperkirakan objek yang benar dan salah [10] lihat Tabel 1. Urutan pengujian ditabulasikan dalam *confusion matrix* dimana yang diprediksi ditampilkan di bagian atas matriks dan kelas yang diamati di bagian kiri[11]. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksi[12].

Tabel 1. *Confusion Matrix*

Nilai prediksi	Nilai Aktual	
	TP	TN
FP	TN	

Keterangan:

TP = *True Positive*

TN = *True Negative*

FP = *False Positive*

FN = *False Negative*

Adapun rumus perhitungan confusion matrix jika di tuliskan seperti dibawah ini[13]:

- a. *Precision* berguna untuk mengukur tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem dengan persamaan (3).

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

- b. *Recall* berguna untuk mengukur tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi, pada persamaan (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

b. Accuracy berguna untuk mengukur kinerja sebuah metode[15] dengan persamaan (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

III. Hasil dan Pembahasan

A. Data Training dan Data Testing

Adapun data training dan data testing yang digunakan pada penelitian ini dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Data Training

NO	X1	X2	X3	X4	X5	X6	Klasifikasi
1	9	1	1	11	13	11	Sempati Star, Type 1626
2	9	1	1	11	13	10	Sempati Star, Type 1626
3	7	1	1	11	13	6	Sempati Star, Type 1626
4	6	1	1	11	13	5	Sempati Star, Type 1626
5	5	1	1	11	13	4	Sempati Star, Type 1626
6	3	1	1	11	13	3	Sempati Star, Type 1626
7	1	1	1	11	13	2	Sempati Star, Type 1626
8	11	3	2	11	8	11	Sempati Star, Type K-410
9	11	3	2	11	8	10	Sempati Star, Type K-410
10	10	3	2	11	8	9	Sempati Star, Type K-410
....
176	2	3	2	11	4	2	Kurnia Type, 2542

Tabel 3. Data Testing

NO	X1	X2	X3	X4	X5	X6	Klasifikasi
1	8	1	1	11	13	9	Sempati Star, Type 1626
2	8	1	1	11	13	8	Sempati Star, Type 1626
3	7	1	1	11	13	7	Sempati Star, Type 1626
4	9	3	2	11	8	7	Sempati Star, Type K-410
5	9	3	2	11	8	6	Sempati Star, Type K-410
....
75	8	3	2	11	4	5	Kurnia Type, 2542

B. Perhitungan K-NN dan Euclidean Distance

Tabel 4. Hasil Perhitungan Jarak 1 Data Uji Dengan Data Training

No	Jarak	Class
1	2,236068	Sempati Star,Mercedes Benz Type 1626
2	1,414214	Sempati Star,Mercedes Benz Type 1626
3	3,162278	Sempati Star,Mercedes Benz Type 1626
4	4,472136	Sempati Star,Mercedes Benz Type 1626
5	5,830952	Sempati Star,Mercedes Benz Type 1626
6	7,81025	Sempati Star,Mercedes Benz Type 1626
7	9,899495	Sempati Star,Mercedes Benz Type 1626
8	6,557439	Sempati Star,Scania Type K-410
9	6,324555	Sempati Star,Scania Type K-410
10	5,830952	Sempati Star,Scania Type K-410
....
....
176	13,07669	Kurnia,Mercedes benz, Type 2542

Untuk pengujian jarak 2 sampai jarak ke 75 pengujian dilakukan dengan langkah yang sama seperti perhitungan jarak 1 pada Tabel 4.

Tabel 5. Hasil Pengurutan Jarak Data Uji Dengan Data Training

Perhitungan Jarak 1		
No	Jarak	Class
61	1	Putra Pelangi,Mercedes Benz Type 1625
2	1,414	Sempati Star,Mercedes Benz Type 1626
62	1,414	Putra Pelangi,Mercedes Benz Type 1625
60	1,732	Putra Pelangi,Mercedes Benz Type 1625

Perhitungan Jarak 1		
No	Jarak	Class
165	1,732	Anugerah,Mercedes Benz Type 1836
68	2	Putra Pelangi,Mercedes Benz Type 1626
166	2	Anugerah,Mercedes Benz Type 1836
1	2,236	Sempati Star,Mercedes Benz Type 1626
59	2,449	Putra Pelangi,Mercedes Benz Type 1625
63	2,449	Putra Pelangi,Mercedes Benz Type 1625
...
...
107	15,55	Putra Pelangi,Mercedes Benz Type MB 1836

Untuk pengurutan jarak ke 2 sampai jarak ke 75 dilakukan dengan langkah pencarian yang sama seperti pengurutan jarak ke 1 pada Tabel 5. Dengan menggunakan k=3 maka 3 jarak terdekat adalah 61,2,62. Begitu juga Untuk perhitungan nilai k yang ke 2-176.

Tabel 6. Hasil Perhitungan Nilai K antara Data Uji Dengan Data Training

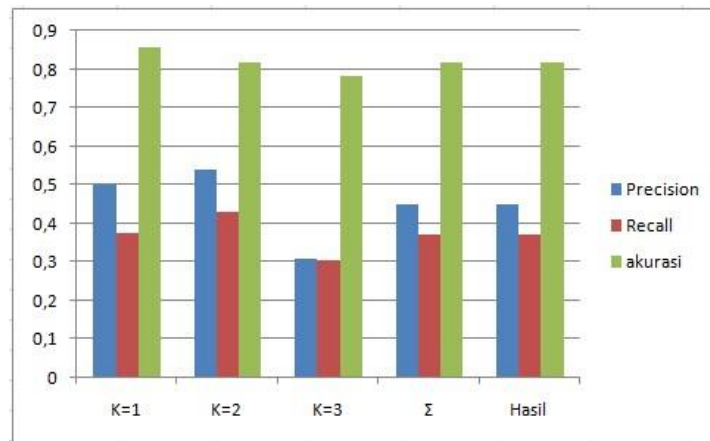
Perhitungan Jarak 1			
No	Jarak	Class	K
61	1	Putra Pelangi,Mercedes Benz Type 1625	1
2	1,4142	Sempati Star,Mercedes Benz Type 1626	2
62	1,4142	Putra Pelangi,Mercedes Benz Type 1625	3

Setelah didapatkan semua hasil perhitungan dari *Precision*, *Recall*, *Accuracy* dengan k=1, k=2 dan k=3 dengan *confusion matrix*, maka didapat persentase hasil perhitungan *Euclidean Distance* pada Tabel 7.

Tabel 7. Hasil Analisis Performansi K-NN dan *Euclidean Distance*

	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
K=1	0,502083	0,375	0,856
K=2	0,538889	0,4306	0,8187
K=3	0,3072	0,3063	0,784
	0,449391	0,3706	0,8196
	44,94%	37,06%	81,96%

Adapun hasil analisis performansi K-NN dan *Euclidean Distance* dalam bentuk grafik dapat dilihat pada Gambar 2.



Gambar 2. Grafik hasil analisis performansi K-NN dan *Euclidean Distance*

Berdasarkan hasil perhitungan diatas didapat nilai *Precision* 44,94%, *Recall* 37,06%, *Accuracy* 81,96%.

C. Perhitungan K-NN dan *Manhattan Distance*

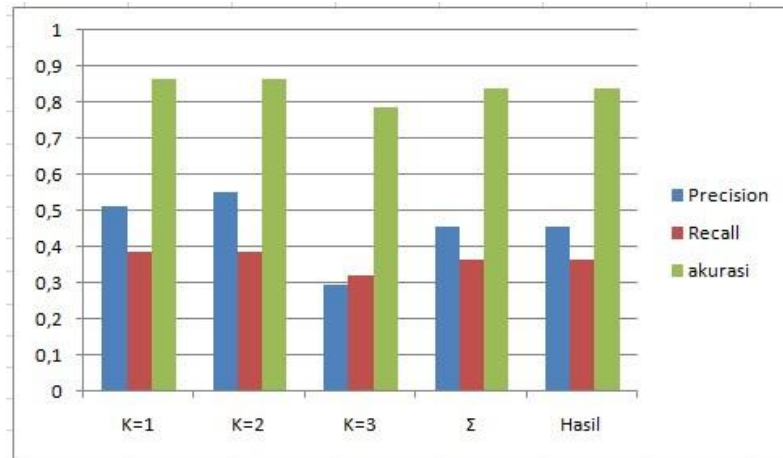
Adapun hasil analisis performansi K-NN dengan menggunakan *Manhattan Distance* dapat bdilihat pada Tabel 8.

Tabel 8. Hasil Analisis Performansi K-NN dan *Manhattan Distance*

	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
K=1	0,515212	0,3861	0,8666
K=2	0,553142	0,3861	0,8666
K=3	0,296212	0,3199	0,7866

<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
0,454855	0,363888	0,84
45,49%	36,39%	84,00%

Adapun hasil analisis performansi K-NN dan *Manhattan Distance* dalam bentuk grafik dapat dilihat pada Gambar 3.



Gambar 3. Grafik hasil analisis performansi *Manhattan Distance*

Berdasarkan hasil perhitungan diatas didapat nilai *Precision* 45,49%, *Recall* 36,39%, *Accuracy* 84,00%.

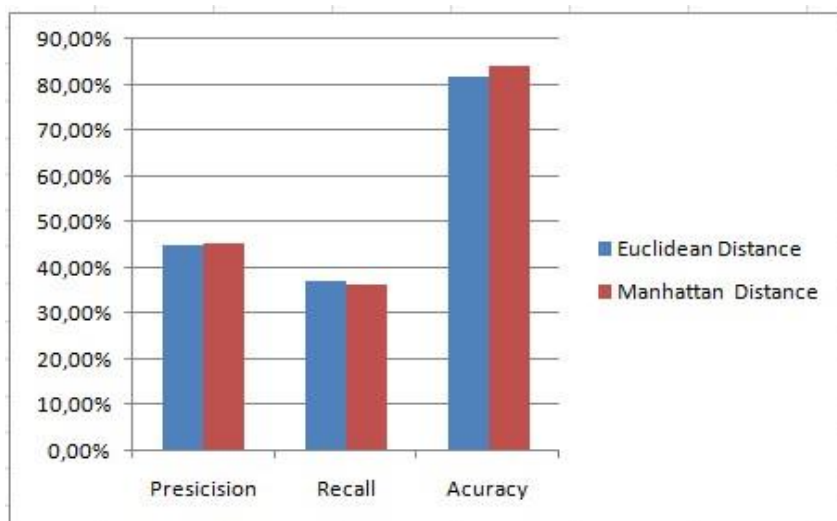
D. Perbandingan Performansi K-NN dan Euclidean Distance dengan K-NN dan Manhattan Distance

Berdasarkan hasil analisis performansi dengan perhitungan Euclidean Distance dan Manhattan Distance disajikan seperti pada Tabel 9.

Tabel 9. Komparasi K-NN dan Euclidean Distance dan Manhattan Distance

	<i>Euclidean Distance</i>	<i>Manhattan Distance</i>
<i>Precision</i>	44,94%	45,49%
<i>Recall</i>	37,06%	36,39%
<i>Accuracy</i>	81,96%	84,00%

Adapun hasil komparasi dalam bentuk grafik dapat dilihat pada Gambar 4.

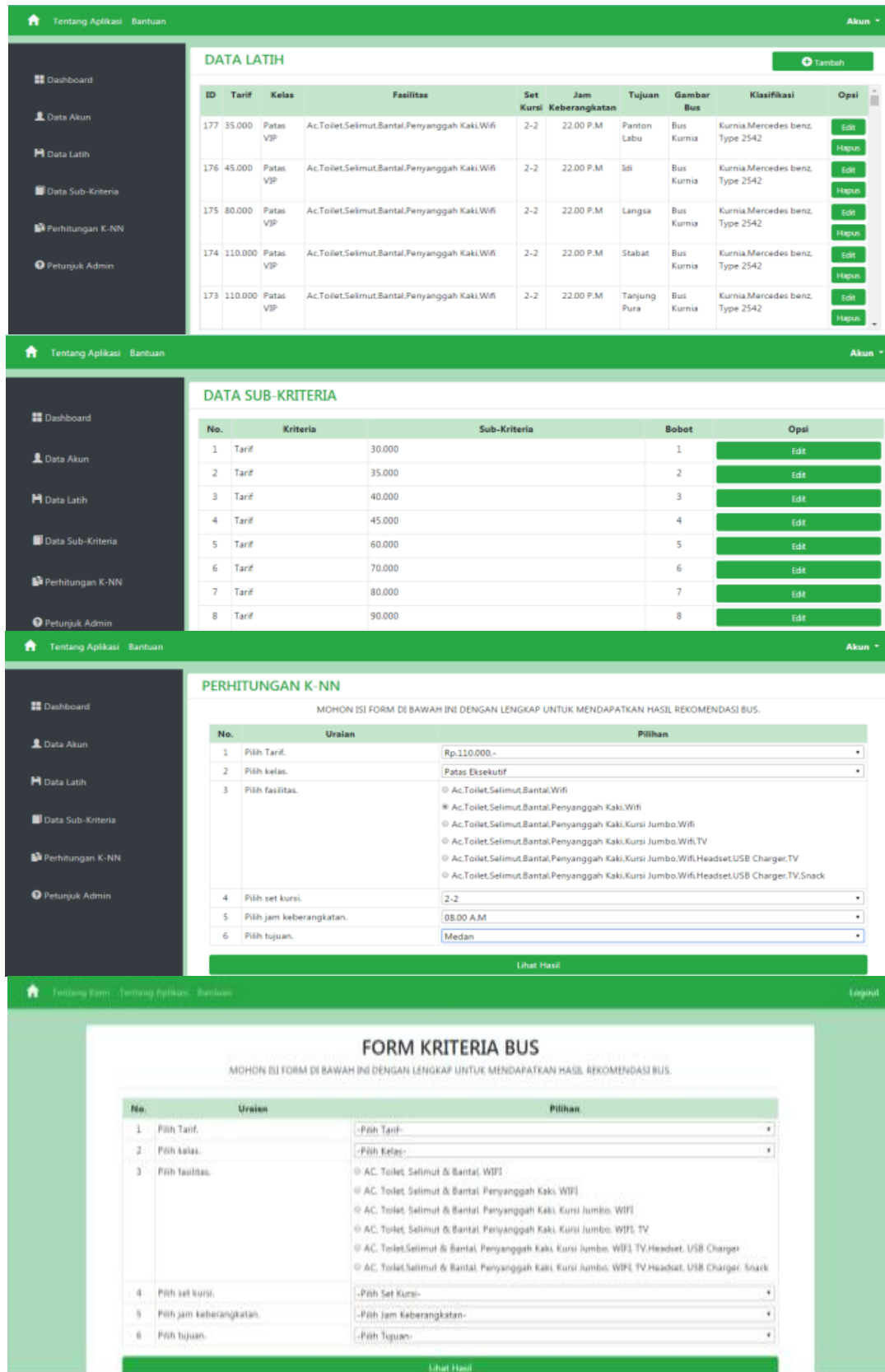


Gambar 4. Grafik Hasil Komparasi K-NN dengan *Euclidean Distance* dan *Manhattan Distance*

Nilai akurasi yang tertinggi di dapat oleh *Manhattan Distance* selisih 2,04% lebih tinggi dibandingkan dengan *Euclidean Distance*. Persentase tersebut menunjukkan bahwa *Manhattan Distance* lebih akurat dibandingkan dengan *Euclidean Distance*.

E. Implementasi Sistem

Pada penelitian ini sistem klasifikasi transportasi bus dibangun berdasarkan kebutuhan penumpang dengan menggunakan bahasa pemrograman PHP. Adapun tampilan sistem dapat dilihat pada Gambar 5.



Gambar 5. Implementasi Sistem

IV. Kesimpulan dan Saran

Dari hasil pengujian yang telah dilakukan dapat disimpulkan bahwa penerapan metode K-NN dengan metode *Euclidean Distance* dan *Manhattan Distance* diperoleh hasil akurasi tertinggi dengan nilai sebesar 84%, dengan k=3. Metode Pendekatan *Manhattan Distance* memiliki nilai selisih 2,04% lebih tinggi dibandingkan dengan *Euclidean Distance*. Persentase tersebut menunjukkan bahwa *Manhattan Distance* lebih akurat dibandingkan dengan *Euclidean Distance* sehingga *Manhattan Distance* bekerja dengan baik dalam memberikan rekomendasi untuk klasifikasi transportasi bus. Adapun saran peneliti untuk penelitian lebih lanjut adalah perlu penerapan metode lainnya dalam analisis performansi klasifikasi data mining seperti algoritma SVM dan *Naïve Bayes*.

Daftar Pustaka

- [1] G. Abdillah *et al.*, “Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means,” *Sentika 2016*, vol. 2016, no. Sentika, pp. 18–19, 2016.
- [2] W. Yustanti, “Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah,” *J. Mat. Stat. dan komputasi*, vol. 9, no. 1, pp. 57–68, 2012.
- [3] S. A. Novarina, “Klasifikasi Jenis Infeksi Berdasarkan Hasil Pemeriksaan Leukosit Menggunakan K-Nearest Neighbor (KKN),” 2018.
- [4] S. A. Naufal, A. Adiwijaya, and W. Astuti, “Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 162, 2020, doi: 10.30865/jurikom.v7i1.2014.
- [5] R. Siringoringo, “Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan k-Nearest Neighbor,” *J. ISD*, vol. 3, no. 1, pp. 44–49, 2018.
- [6] S. R. Andani, “Penerapan Metode SMART dalam Pengambilan Keputusan Penerima Beasiswa Yayasan AMIK Tunas Bangsa,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 3, p. 166, 2019, doi: 10.26418/justin.v7i3.30112.
- [7] N. H. A. Sari, M. A. F. Fauzi, and P. P. Adikara, “Klasifikasi Dokumen Sambat Online Menggunakan Metode K-Nearest Klasifikasi Dokumen Sambat Online Menggunakan Metode K-Nearest Neighbor dan Features Selection Berbasis Categorical Proportional Difference,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. August, pp. 2449–2454, 2018.
- [8] M. Reza Noviansyah, T. Rismawan, and D. Marisa Midyanti, “Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya),” *J. Coding, Sist. Komput. Untan*, vol. 06, no. 2, pp. 48–56, 2018, [Online]. Available: <http://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672>.
- [9] D. Sebastian, “Implementasi Algoritma K-Nearest Neighbor untuk Melakukan Klasifikasi Produk dari beberapa E-marketplace,” *J. Tek. Inform. dan Sist. Inf.*, vol. 5, no. 1, pp. 51–61, 2019, doi: 10.28932/jutisi.v5i1.1581.
- [10] R. Latifah, E. Susilowati, and W. Febriyanti, “Sistem Pendukung Keputusan Penentuan Calon Penerima Kartu Jakarta Pintar (KJP) Menggunakan K-Nearest Neighbor,” *J. Sist. Informasi, Teknol. Inform. dan Komput.*, vol. 8, pp. 97–104, 2017.
- [11] J. Riany, M. Fajar, and M. P. Lukman, “Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor,” *Sisfo*, vol. 06, no. 01, pp. 147–156, 2016, doi: 10.24089/j.sisfo.2016.09.011.
- [12] F. Liantoni, “Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor,” *J. Ultim.*, vol. 7, no. 2, pp. 98–104, 2016, doi: 10.31937/ti.v7i2.356.
- [13] N. Hasdyna, “Information Gain dalam Reduksi Dimensi Dataset untuk Peningkatan Kinerja Algoritma K-Nearest Neighbor,” TESIS. Universitas Sumatera Utara. 2019.
- [14] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik20185983.
- [15] I. Hasimah and H. Yasin, “Klasifikasi Calon Debitur Kredit Pembiayaan Rumah (KPR) Multiguna Take Over Menggunakan Metode k Nearest Neighbor Dengan Pembobotan Global Diversity Index 1,2,3,” *J. Gaussian*, vol. 8, pp. 407–417, 2019, [Online]. Available: <http://ejournal3.undip.ac.id/index.php/gaussian>.