Jurnal Juita

by mutasar us

Submission date: 22-Apr-2019 12:21PM (UTC+0800)
Submission ID: 1116200888
File name: 4310-10343-1-RV_Author.docx (91.4K)
Word count: 2246
Character count: 11506

Performansi Algoritma K-Means Pada Clustering Sepeda Motor

(K-Means Algorithm Performance in Motorcycle Clustering)

Rozzi Kesuma Dinata[1], Nur Azizah[2], Safwandi[3]

[1,2,3]*Program Studi Teknik Informatika – Universitas Malikussaleh Jl. Medan-BandaAceh, Kab.AcehUtara- Aceh*

[1]rozzi@unimal.ac.id

[2]zizah2503@gmail.com

[3]safwandi@unimal.ac.id

Abstrak- Salah satu algoritma data mining yang dapat melakukan pengelompokan data atau *Clustering* adalah k-means. Penelitian ini menggunakan *k-means* untuk pengelompokan data sepeda motor berdasarkan kebutuhan konsumen kedalam tiga cluster. Adapun data yang dianalisis adalah data Sepeda Motor merek Honda dan Yamaha yang diambil dari dialer yang ada di Kecamatan Dewantara, Aceh Utara. Pengelompokkan 300 data dilakukan sebanyak 15 kali pengujian sejumlah 45 data sebagai centroid awal yang diambil secara random. Perhitungan jarak setiap data menggunakan *Euclidean Distance*. Hasil performansi dari setiap pengujian yang dilakukan diperoleh nilai *Presicion* sebesar 76%, nilai *Recall* sebesar 76% dan akurasi sebesar 81%.

Kata kunci—K-Means, Clustering, Data mining, Accuracy

Abstract- *One of the data mining algorithms that can do data clustering is k-means. This study uses k-means for grouping motorcycle data based on consumer needs into three clusters. The data analyzed are data on Honda and Yamaha brands of motorcycles taken from dialers in Dewantara District, North Aceh. Grouping 300 data was carried out 15 times testing 45 data as initial centroids are taken randomly. Calculation of the distance of each data using Euclidean Distance. The performance results of each test performed obtained a Precision value of 76%, a Recall value of 76% and an accuracy of 81%*

Keywords — K-Means, Clustering, Data mining, Accuracy

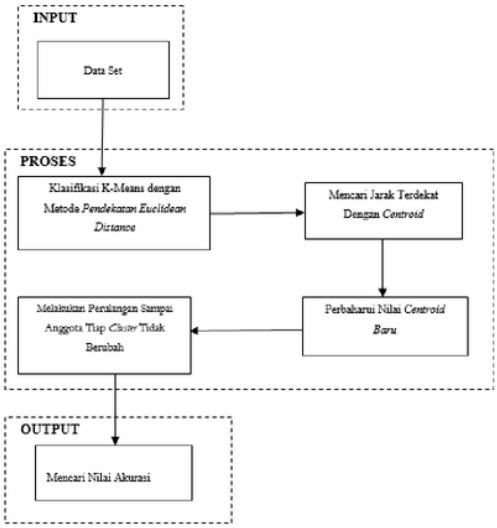
I. PENDAHULUAN

Data mining dapat mengektraksi data menjadi informasi yang sebelumnya belum mempunyai arti [2]. Data Mining juga dapat menyederhanakan data agar diperoleh informasi yang mudah dipahami [3]. Salah satu teknik pengelompokan sepeda motor berdasarkan kebutuhan konsumen yang dapat digunakan dalam *data mining* adalah metode pengelompokan *Clustering*[7]. *Clustering* mencari data dan menglompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya diperoleh [4]. Salah satu metode *Clustering* yang dapat digunakan adalah *k-means* [5][6].

Penelitian data mining tentang sepeda motor telah banyak dikembangkan oleh para peneliti. Ada beragam metode data mining yang telah diteliti, diantaranya seperti dalam peneliti yang telah diteliti oleh Putra et al. (2017) menggunakan *hadoop single node cluster* dengan metode *K-Nearest* Neighbour dalam klasifikasi sepeda motor berdasarkan karakteristik konsumen [1]. Penelitian yang dilakukan oleh Purwadi (2018) menggunakan metode algoritma c4 untuk memprediksi pola pembelian sepeda motor pada showroom cv. viva mas motors [8]. Penelitian ini mengelompokkan data sepeda motor menjadi tiga cluster yang diperoleh dari dialer yang ada di Kecamatan Dewantara, Aceh Utara. Pengujian performansi dilakukan dengan menggunakan confusion matrix, dengan menganalisis *recall, precision dan accuracy* pada setiap *initial centroid*.

II. METODE PENELITIAN

A. Kerangka Penelitian



Gambar 1. Digram Alir Konsep K-Means [10]

Penjabaran dari diagram alir k-means sebagaimana pada Gambar 1:

1. Langkah pertama adalah menginputkan dataset
2. Langkah selanjutnya menentukan initial centroid.
3. Langkah selanjutnya adalah mencari jarak terdekat dengan *Centroid* menggunakan rumus jarak *Euclidian Distance* [10].

$$d(xi,\mu j) = \sqrt{\sum (xi-\mu j)^2} \qquad (1)$$

Keterangan: *xi* = data kriteria
μj = *centroid* pada *cluster* ke-js

4. Setelah itu memperbaharui nilai *Centroid* baru [10].

$$\mu j(t+1) = \frac{1}{Nsj}\sum_{j\in sj} xj \qquad (2)$$

Keterangan : μj(t+1) = *centroid* baru pada iterasi (t+1)
N*sj* = Data pada *cluster* Sj

5. Perhitungan akan berhenti setelah didpatkan nilai centroid baru yang konvergen. Langkah terakhir yaitu mecari nilai akurasi dari setiap pengujian yang dilakukan.

B. Confussion Matrix

Confussion matrix dapat melakukan pengujian untuk memperdiksikan suatu objek yang benar dan salah. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang diamati untuk diprediksikan [11], sebagaimana pada Tabel 1.

TABEL 1. CONFUSSION MATRIX

Nilai prediksi	**Nilai Aktual**	
	TP	TN
	FP	TN

Keterangan:

TP = *True* positif
TN= *True* negatif
FP = *False* positif
FN= *False* negatif

Rumus untuk perhitungan *confusion matrix* untuk menghitung *precision* [12], *recall* [12], dan nilai *accuracy* [12], dijabarkan pada persamaan (3),(4),dan (5):

a. *Precision*,

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

b. *Recall*,

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

c. *Acuracy*,

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

TABEL 2. DATASET SEPEDA MOTOR

No	**Sepeda Motor**	**X1**	**X2**	**X3**	**X4**	**X5**
1	Revo Fit	1	4	1	3	1
2	Revo Fit	1	2	1	3	1
3	Revo Fit	1	7	1	3	1
4	Revo X	1	8	1	3	3
5	Revo X	1	5	1	3	3
6	Blade 125 FI R	1	4	2	3	4
7	Blade 125 FI R	1	10	2	3	4
8	Blade 125 FI R	1	9	2	3	4
9	Supra X 125 SW	1	6	2	3	4
10	Supra X 125 SW	1	4	2	3	4
11	Supra X 125 SW	1	6	2	3	5
12	Supra X 125 SW	1	4	2	3	5
13	Supra X CW	1	4	2	9	5

14	Supra X CW	1	2	2	9	5
15	Supra X CW	1	11	2	9	5
…	…	…	…	…	…	…
…	…	…	…	…	…	…
300	Jupiter MX King	2	2	3	5	11

Tabel 1. merupakan dataset sepeda motor terdiri dari 300 data dengan 5 atribut yang diperoleh dari dialer yang ada di Kecamatan Dewantara, Aceh Utara.

TABEL 3. PENENTUAN *INITIAL CENTROID*

Uji ke-	Initial Centroid
1	2, 87, 100
2	1, 6, 9
3	3, 14, 23
4	61, 81, 101
5	57, 66, 84
6	37, 47, 95
7	113, 139, 150
8	117, 137, 157
9	182, 190, 210
10	202, 220,240
11	120, 142, 161
12	124, 146, 153
13	257, 263, 276
14	183, 200, 213
15	268, 274, 280

Tabel 2.menampilkan *initial centroid* yang akan diuji terhadap dataset sepeda motor dengan algoritma *k-means*.

III. HASIL DAN PEMBAHASAN

Adapun hasil perhitungan jarak data ke-1 pada pengujian pertama dengan data uji masing-masing *cluster* yang ada di tabel 3 adalah:

$$d_{(2.1)} = \sqrt{(1-1)^2+(2-4)^2+(1-1)^2+(3-3)^2+(1-1)^2}$$
$$= \sqrt{4}$$
$$= 2$$
$$d_{(87.1)} = \sqrt{(1-1)^2+(6-4)^2+(3-1)^2+(15-3)^2+(19-1)^2}$$
$$= \sqrt{476}$$
$$= 21{,}81742423$$
$$d_{(100.1)} = \sqrt{(1-1)^2+(5-4)^2+(3-1)^2+(18-3)^2+(38-1)^2}$$
$$= \sqrt{1599}$$
$$= 39{,}98749805$$

Hasil perhitungan jarak pada 300 data dapat dilihat pada tabel 4.

TABEL 4. HASIL PERHITUNGAN JARAK PADA TIAP *CLUSTER*

No	A2	A87	A100	Cluster
1	2	21,81742	39,9875	1
2	0	22,09072	40,0874	1
3	5	21,74856	40,02499	1
4	6,324555	20,19901	38,24918	1
5	3,605551	20,12461	38,13135	1
6	3,741657	19,33908	37,18871	1
7	8,602325	19,64688	37,51	1
8	7,681146	19,46792	37,38984	1
9	5,09902	19,23538	37,18871	1
10	3,741657	19,33908	37,18871	1
11	5,744563	18,46619	36,27671	1
12	4,582576	18,57418	36,27671	1
13	7,549834	15,3948	34,23449	1
14	7,28011	15,77973	34,35113	1
15	11,57584	16,06238	34,74191	1
…	…	…	…	…
…	…	…	…	…
300	10,44031	13,45362	30,13304	1

Adapun hasil mengelompokkan data sesuai clusternya dijabarkan pada tabel 5.

TABEL 5. CLUSTER DENGAN JARAK TERDEKAT

No	C1	C2	C3
1	*		
2	*		
3	*		
4	*		
5	*		
6	*		
7	*		
8	*		
9	*		
10	*		
11	*		
12	*		
13	*		
14	*		
15	*		

…	…	…	…	
…	…	…	…	
13	*			

Nilai centroid baru yang terbentuk dari pengelompokkan cluster dapat dilihat pada tabel 6.

TABEL 6. NILAI *CENTROID* BARU

C1	1,41059	6,0331	1,95364	4,45695	4,69536
C2	1,67021	9,436	3,07446	15,7766	15,4042
C3	1,23636	11	4,67272	22,10909	34,47273

Pada pengujian ke-1 iterasi belum konvergen, konvergensi data berhenti pada iterasi ke-4. Pada masing-masing pengujian data yang sudah sama tidak semua berhenti pada iterasi yang sama tetapi berhenti dibeda iterasi, untuk pemberhentian data pada iterasi berapa pada setiap pengujian bisa dilihat pada tabel 7 dibawah ini.

TABEL 7. ITERASI PENGUJIAN

Uji ke-	**Data Uji**	**Berhenti**
1	2, 87, 100	di iterasi ke-4
2	1, 6, 9	di iterasi ke-4
3	3, 14, 23	di iterasi ke-6
4	61, 81, 101	di iterasi ke-6
5	57, 66, 84	di iterasi ke-9
6	37, 47, 95	di iterasi ke-7
7	113, 139, 150	di iterasi ke-6
8	117, 137, 157	di iterasi ke-6
9	182, 190, 210	di iterasi ke-7
10	202, 220,240	di iterasi ke-8
11	120, 142, 161	di iterasi ke-8
12	124, 146, 153	di iterasi ke-7
13	257, 263, 276	di iterasi ke-6
14	183, 200, 213	di iterasi ke-8
15	268, 274, 280	di iterasi ke-6

Hasil pengelompokan data ke 300 data sepeda motor dengan 15 kali pengujian dengan setiap pengujian menggunakan 3 *cluster* data. Selanjutnya perbandingan kelompok data yang di uji dengan kelompok hasil pengujian, seperti pada tabel 8.

TABEL 8. PERBANDINGAN DATA UJI DENGAN DATA HASIL

Data Uji	**Data Yang di Uji**	**Data Hasil**
1	Murah	Murah
2	Murah	Murah
3	Murah	Murah
4	Murah	Murah
5	Murah	Murah
6	Murah	Murah
7	Murah	Murah

8	Murah	Murah
9	Murah	Murah
10	Murah	Murah
11	Murah	Murah
12	Murah	Murah
13	Murah	Murah
14	Murah	Murah
15	Murah	Murah
16	Murah	Murah
....	....	....
....	....	....
300	Standar	Murah

Perbandingan data uji dengan data hasil dikelompokkan berdasarkan status klasifikasi *clustering* yaitu dengan melakukan pengelompokan 300 data sepeda motor dengan atribut yang berbeda ke dalam 3 *Cluster* status klasifikasi yaitu *cluster* 1 - murah, *cluster* 2 - standar, dan *cluster* 3 - mahal. Setelah dikelompokkan semua datanya, langkah berikutnya adalah menghitung nilai dengan *Confusion Matrix*.

TABEL 9. PREDICTED CLASS

	Actual Class			
	Murah	**Standar**	**Mahal**	
Murah	107	0	0	107
Standar	50	82	1	133
Mahal	0	4	56	60
Rata-rata	157	86	57	300

Untuk menghitung nilai *Confusion Matrix* menggunakan nilai prediksi TP, FP,FN,TN sebagaimana pada Tabel 10.

TABEL 10. CONFUSION MATRIX

Confusion Matrik	**Class**	**TP**	**FP**	**FN**	**TN**	**Total**
	Murah	107	50	0	143	300
	Standar	82	4	51	163	300
	Mahal	56	1	4	239	300

Langkah selanjutnya menghitung nilai *presicision*, *recall*, dan *accuracy,* masing-masing *class*, kemudian menghitung nilai rata-rata semua *class* sebagaimana pada tabel 11.

TABEL 11. PERFORMANSI K-MEANS

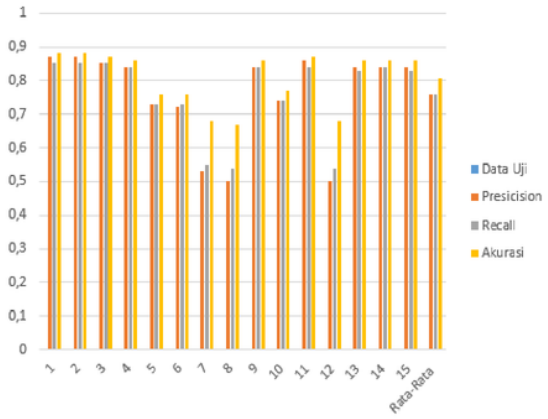
Confusion Matrik	**class**	**Precision**	**Recall**	**Akurasi**
	Murah	0,681529	1	0,833333
	Standar	0,953488	0,616541	0,816667
	Mahal	0,982456	0,933333	0,983333
	Jumlah	87%	85%	88%

Setelah hasil *Confusion Matrix* sudah selesai dicari semua, langkah selanjutnya yaitu menggabungkan semua nilai *Confusion Matrix* untuk mencari nilai rata-rata dari nilai *presicision*, *recall*, dan *accuracy,* dari semua ke-15 kali pengujian yang sudah dicari sebelumnya. Hasilnya dijabarkan pada table 12.

TABEL 12. HASIL PERFORMANSI K-MEANS UNTUK SEMUA PENGUJIAN

Uji ke	Data Uji	Precision	Recall	Akurasi
1	2, 87, 100	87%	85%	88%
2	1, 6, 9	87%	85%	88%
3	3, 14, 23	85%	85%	87%
4	61, 81, 101	84%	84%	86%
5	57, 66, 84	73%	73%	76%
6	37, 47, 95	72%	73%	76%
7	113, 139, 150	53%	55%	68%
8	117, 137, 157	50%	54%	67%
9	182, 190, 210	84%	84%	86%
10	202, 220,240	74%	74%	77%
11	120, 142, 161	86%	84%	87%
12	124, 146, 153	50%	54%	68%
13	257, 263, 276	84%	83%	86%
14	183, 200, 213	84%	84%	86%
15	268, 274, 280	84%	83%	86%
Rata-Rata		76%	76%	81%

Nilai akurasi dari metode *K-Means* dalam bentuk grafik disajikan pada gambar 2.



Gambar 2. Grafik Hasil Performansi *K-Means*

Berdasarkan gambar 2, dapat dilihat bahwa nilai rata-rata dari Precision adalah 76%, Recall 76%, dan Acuracy 81%.

IV. KESIMPULAN

Dari hasil 15 kali pengujian yang dilakukan dengan menggunakan 3 *cluster* dari data pengujian yang berbeda setiap pengujiannya, iterasi paling sedikit berhenti di iterasi ke-4 pada pengujian 1 dengan data uji 2, 87, 100 dan pada pengujian 2 dengan data uji 1, 6, 9. Dengan iterasi yang paling banyak berhenti di iterasi ke-9 pada pengujian ke-5 dengan data uji 57, 66, 84. Hasil performansi k-means dari 15 pengujian dari setiap uji coba yang dilakukan, diperoleh nilai rata-rata *Precision* sebesar 76%, nilai *Recall* sebesar 76% dan *Accuracy* sebesar 81%.

DAFTAR PUSTAKA

[1] Putra, N. A., Putri, A. T., Prabowo, D. A., Surtiningsih, L., Arniantya, R., & Cholissodin, I. 2017. "Klasifikasi Sepeda Motor Berdasarkan Karakteristik Konsumen Dengan Metode K-Nearest Neighbour Pada Big Data Menggunakan Hadoop Single Node Cluster". *Jurnal Teknologi Informasi dan Ilmu Komputer*, *4*(2), 81-86.

[2] Abdillah, G., Putra, F. A., & Renaldi, F. 2016. "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means". In *Seminar Nasional Teknologi Informasi dan Komunikasi* (pp. 498-506).

[3] Abdurrahman, G. 2017. "Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means". *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*,*1*(2).

[4] W. Wardhani, Anindya Khrisna. 2016. "Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, no. 1, pp. 30–37.

[5] Nur, F., Zarlis, M., & Nasution, B. B. 2017. "Penerapan Algoritma k-means pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan". *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, *1*(2), 100-105.

[6] Windha Mega Pradnya Dhuhita. 2016. "Clustering Menggunakan Metode K-Means untuk Menentukan Status Gizi Balita," *J. Inform.*, vol. 15, no. 2, pp. 160–174.

[7] L. Maulida. 2018. "Penerapan Data Mining Dalam Mengelompokan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta Dengan K-Means," *JISKa*, vol. 2, no. 3, pp. 167–174.

[8] Purwadi, P. 2018. "Implementasi Data Mining Untuk Memprediksi Pola Pembelian Sepeda Motor pada showroom cv. viva mas motors dengan metode algoritma c4. 5". *JSIK (Jurnal Sistem Informasi Kaputama)*,*2*(2).

[9] B. M. Metisen and H. L. Sari. 2015. "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila," vol. 11, no. 2, pp. 110–118.

[10] N. Rohmawati, S. Defiyanti, and M. Jajuli, 2015."Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa," *J. Ilm. Teknol. Inf. Terap.*, vol. I, no. 2, pp. 62–68.

[11] I. Menarianti. 2015. "Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi," *J. Ilm. Teknosains*, vol. 1, no. 1, pp. 1–10.

[12] J. Riany, M. P. Lukman, and M. Fajar, 2017. "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," *Sisfo*, vol. 06, no. 01, pp. 147–156.

Jurnal Juita

ORIGINALITY REPORT

19%	17%	2%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Brawijaya Student Paper	3%
2	docobook.com Internet Source	2%
3	jtiik.ub.ac.id Internet Source	1%
4	hildemihosverden.blogspot.com Internet Source	1%
5	ejournal.uin-suka.ac.id Internet Source	1%
6	jurnal.unimed.ac.id Internet Source	1%
7	eprints.dinus.ac.id Internet Source	1%
8	eprints.umm.ac.id Internet Source	1%
9	Submitted to Universitas Nasional Student Paper	1%

10	docplayer.info Internet Source	1%
11	Submitted to Universitas Dian Nuswantoro Student Paper	1%
12	www.scribd.com Internet Source	1%
13	jurnal.unmuhjember.ac.id Internet Source	1%
14	www.golftracer.com Internet Source	1%
15	ejournal.bsi.ac.id Internet Source	1%
16	Submitted to Universitas Siswa Bangsa Internasional Student Paper	1%
17	ejournal.poltektegal.ac.id Internet Source	<1%
18	simki.unpkediri.ac.id Internet Source	<1%
19	Submitted to Universitas Pendidikan Indonesia Student Paper	<1%
20	adoc.tips Internet Source	<1%
	jurnal.pcr.ac.id	

21	Internet Source	<1%
22	de.scribd.com Internet Source	<1%
23	ojs3.unpatti.ac.id Internet Source	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		