PERBANDINGAN DISTANCE MEASURES PADA K-MEANS CLUSTER DAN TOPSIS DENGAN KORELASI PEARSON DAN SPEARMAN

by Stendy Budi Hartono Sakur

Submission date: 10-May-2023 01:22PM (UTC+0700)
Submission ID: 2089269306
File name: JITEK_VOL-MARET_2023_NASKAH_JURNAL_STENDY_B_SAKUR.docx (2.26M)
Word count: 3172
Character count: 18985

JURNAL JITEK Vol 3 No. 1 Maret (2023) Hal 74-81, P-ISSN : 2809-9249 E-ISSN : 2809-9230

JURNAL INFORMATIKA DAN TEKONOLOGI KOMPUTER
Halaman Jurnal: https://journal.amikveteran.ac.id/index.php/jitek
Halaman UTAMA Jurnal : https://journal.amikveteran.ac.id/index.php

PERBANDINGAN *DISTANCE MEASURES* PADA K-MEANS CLUSTER DAN TOPSIS DENGAN KORELASI PEARSON DAN SPEARMAN

Stendy Budi Hartono Sakur
Jurusan Teknik Komputer dan Komunikasi / Program Studi Sistem Informasi, sakur.stendy@gmail.com,
Politeknik Negeri Nusa Utara - Tahuna

ABSTRACT

Clustering is a data mining method that is widely used to group data based on similarity. This clustering process can be used to streamline data so as to facilitate the data ranking process. The purpose of this study was to make comparisons of distance measurements on the K-Means and TOPSIS methods to select students who would take part in industrial visit activities. The method used in this study is the K-Means Algorithm to carry out the clustering process whose results will be processed using the TOPSIS method, both of which use Euclidean, Manhattan and Minkowsky Distance. Based on the clustering process, there were 21 respondents who were eligible to be included, then with TOPSIS a ranking process was carried out. Of the three distance measurements used based on the Pearson Euclidean distance correlation test, the highest results were 0.992, Manhattan 0.982 and Minkowsky 0.980, with ratings one, two and three respectively. For the Spearman correlation, Eculidean is 0.972, Manhattan is 0.982 and Minkowski is 0.955. Thus, Euclidean distance gives the best correlation results, while for alternatives, Manhattan distance or Minkoesky distance can be used.

Keywords: *K-Means, TOPSIS, Euclidean distance, Manhattan distance, Minkowsky distance.*

Abstrak

Clustering merupakan salah satu metode Data mining yang banyak digunakan untuk melakukan pengelompokan data berdasarkan Similarity. Proses klustering ini dapat digunakan untuk merampingkan data sehingga memudahkan proses perangkingan data. Tujuan penelitian ini adalah untuk melakukan perbandingan pengukuran jarak atau Distance Measures pada metode K-Means dan TOPSIS untuk memilih mahasiswa yang akan mengikuti kegiatan kunjungan industri. Metode yang digunakan pada penelitian ini adalah Algoritma K-Means untuk melakukan proses pengelompokan atau kulstering yang hasilnya akan di proses dengan menggunakan Metode TOPSIS yang keduanya menggunakan Euclidean, Manhattan dan Minkowsky Distance. Berdasarkan proses klusterisasi terdapat 21 responden yang layak untuk diikutsertakan kemudian dengan TOPSIS dilakukan proses perangkingan. Dari ketiga pengukuran jarak yang digunakan berdasarkan pengujian korelasi Pearson Euclidean distance memberikan hasil tertinggi 0.992, Manhattan 0.982 dan Minkowsky 0.980, dengan peringkat satu, dua dan tiga secara berurutan. Untuk korelasi Spearman, Eculidean 0.972, Manhattan 0.982 dan Minkowski adalah 0.955. Dengan demikian, Euclidean distance memberikan hasil korelasi yang terbaik, sedangkan untuk alternatif dapat digunakan Manhattan distance ataupun Minkoesky distance.

Kata Kunci: K-Means, TOPSIS, Euclidean distance, Manhattan distance, Minkowsky distance.

1. PENDAHULUAN

Clustering atau klusterisasi digunakan untuk mengelompokkan data yang memiliki similaritas terhadap sifat – sifat ataupun perilaku dari objek sehingga dapat memetakkan objek yang sejenis ataupun tidak sejenis pada suatu kelompok tertentu. Dalam penentuan pengelompokan, pengukuran jarak atau *measures distance* memainkan peranan penting dan sangat mempengaruhi performance dari algoritma ini [1]. Terdapat beberapa pengukuran jarak diantaranya *Eculidien distance, Manhattan distance (city block distance), Minkowsky Distance, Cosine distance, Hamming distance*[2]. Pemilihan metode pengukuran bergantung kepada tipe data

Received Januari 13, 2023; Revised Maret 3, 2023; Accepted Maret 31, 2023

yang digunakan dalam proses cluster. Proses clustering sangat dibutuhkan untuk melakukan perampingan data terhadap kasus sistem pendukung keputusan sehingga dapat meningkatkan performan dari metode tersebut. Metode TOPSIS atau *Technique for Order Preferences by Similarity to Ideal Solution* oleh Hwang dan Yoon kemudian diperbaiki oleh Hwang dan Chen [3] menggunakan perhitungan jarak untuk solusi ideal positif dan negatif. Pengukuran jarak pada metode TOPSIS ini dapat menggunakan metode seperti pada algoritma K-Means. Meminjam kasus dari penelitian, Sakur, dkk [4] peneliti melakukan proses clustering untuk merampingkan data mahasiswa dengan menggunakan tiga perbedaan pengukuran jarak yang hasil akhirnya akan diranking dengan metode TOPSIS.

Penelitian terdahulu [3] yang menggunakan metode Clustering yaitu K-Means dengan jarak Euclidien distance akan merampingkan data atau mengelompokkan data kedalam dua cluster yaitu cluster dengan nilai maksimum dan minimum. Selanjutnya dengan menggunakan metode TOPSIS dilakukan proses perangkingan untuk menentukan mahasiswa yang akan mengikuti kegiatan kunjungan industri. Hasil akhirnya dibandingkan dengan kondisi yang telah dilakukan. Metode TOPSIS tersebut membandingkan dua pengukuran jarak yaitu *Euclidean* dan *Manhattan Distance*. Beberapa penelitian terdahulu yang membandingkan pengukuran jarak yaitu Rathod [2] yang membandingkan *Euclidean, Hamming*, *Manhattan* dan *Squared distance* terhadap data permutasi string, Anggara [5] membandingkan *Euclidean*, *Manhattan* dan *Chebyshev distance* untuk pengelompokkan anggota dari alvaro members. Nishom [6] membandingkan jarak *Euclidean, Manhattan* dan *Minkowski distance* untuk menentukan disparatis kebutuhan guru dengan pengujian menggunakan *Chi-Square*.

Berdasarkan beberapa penelitian sebelumnya yang berhubungan dengan analisis perbandingan jarak (*Distance measure)* maka peneliti akan membandingkan pengukuran jarak *Euclidean, Manhattan* dan *Minkowski distance* pada K-Means Cluster untuk perampingan atau pengelompokan data yang hasilnya akan dirangking dengan menggunakan Metode TOPSIS dengan tiga model pengukuran jarak tersebut dan akan dianalisis performannya menggunakan korelasi *Pearson* untuk memperhitungkan nilai dari setiap alternatif dan *Rank Spearman* untuk menganalisis peringkat (*Rank*) yang diperoleh dari alternatif [7].

2. TINJAUAN PUSTAKA

2.1. Distance Measure

Distance measure atau pengukuran jarak memainkan peranan penting dalam menentukan kelompok dari setiap objek berdasarkan similarity sifat – sifat dari objek. Pengukuran similarity dapat dijelaskan sebagai bentuk pengukuran antara variasi titik data. Beberapa bentuk pengukuran jarak similarity diantaranya *Euclidean, Manhattan, Minkowski, Cosine, Jaccard*, *Chebyshev distance* [8] dan lain sebagainya.

2.2. Clustering K-Means

Clustering merupakan metode yang berfungsi untuk mencari dan mengelompokkan data dengan karaktersitik yang sama (Similarity) antara satu data dengan data yang lain [5]. K-Means merupakan salah satu kluster yang sederhana dan banyak digunakan dalam proses pengelompokan data, metode ini ditemukan oleh Lyoid, Forgery, Friedman dan Rubin serta McQueen [9]. K-Means merupakan metode clustering *Partitioning*, selain metode tersebut terdapat metode clustering seperti *Hierarhical Method*, *Grid-Based method* dan *Density-based Method* [10]. Algoritma ini sangat bergantung kepada pengukuran jarak untuk proses pengelompokan data dan proses penentuan pusat kluster selanjutnya menggunakan nilai rata – rata dari setiap anggota yang masuk dalam kelompok tersebut.

2.3. Metode TOPSIS

Metode *Technique of Order Preferences by Similarity to Ideal Solution* atau TOPSIS merupakan salah satu Metode MultiCriteria Decision Making yang banyak digunakan. Dimana prinsip dasar adalah menentukan alternatif yang terdekat dengan jarak ideal solution positif dan terjauh dari solusi ideal negatif. Metode ini telah berevolusi dari C-TOPSIS oleh Hwang dan Yoon, A-TOPSIS sampai pada M-TOPSIS [3].

2.4. Korelasi Pearson dan Spearman

Perhitungan Korelasi dibutuhkan untuk melihat kedekatan dua nilai yang tidak saling terikat (bebas). Secara umum hubungan ini dibentuk dengan nilai 1) Negatif (-) menunjukkan tidak ada hubungan antara keduanya

dan, 2) Positif (+) menunjukkan hubungan yang sangat kuat. Korelasi pearson khusus memperhitungkan nilai yang umu dan banyak digunakan pada bidang Psikologi, sedangkan untuk korelasi spearman dikhususkan untuk melihat kondisi hubungan diantara Ranking yang diperoleh. Jika terdapat perangkingan yang sama (seringkali terjadi) maka harus dihitung koefisien koreksi (*Correction Coefisien*)[7].

3. METODOLOGI PENELITIAN

Penelitian ini akan mengelompokkan mahasiswa yang layak untuk mengikuti kegiatan dengan menggunakan algoritma K-Means, dimana proses cluster akan dilakukan sebanyak tiga kali sesuai dengan jarak yang akan digunakan yaitu *Euclidean*, *Manhattan* dan *Minkowski distance*. Dari proses tersebut akan terbentuk tiga model hasil cluster sesuai jarak yang digunakan. Selanjutnya proses perangkingan akan dilakukan pada ketiga hasil tersebut dengan metode TOPSIS yang dalam proses penentuan ideal solution akan menghitung jarak masing – masing yaitu *Euclidean, Manhattan* dan *Minkowski distance*. Proses akhir adalah menguji hasil dari ketiga metode jarak tersebut yang telah diranking menggunakan TOPSIS dengan korelasi Pearson untuk menguji kedekatan nilai C* sedangkan untuk menguji kedekatan nilai Ranking akan digunakan korelasi Spearman.

3.1. Dataset

Penelitian ini menggunakan data Mahasiswa Kunjungan Industri yang meminjam data dari Penelitian Sakur, dkk [4] dengan sumber data dapat diakses di repo [11]. Terdapat 55 mahasiswa yang mengikuti seleksi kegiatan tersebut, yang terdiri dari ujian tertulis dan wawancara. Selain itu setiap mahasiswa harus mengisi quisioner yang berhubungan dengan data penting lainnya. Beberap data yang bersifat linguistik harus diubah ke dalam nilai numerik.

3.2. Pembobotan

Penelitian menggunakan kriteria 1) Tes tertulis (20%), 2) Wawancara (20%), 3) Indeks Prestasi Kumulatif (20%), 4) Keaktifan (10%), 5) Kepemimpinan (10%), 6) UKM (10%) dan 7) Sertifikat Keahlian (10%). Jumlah responden yang digunakan adalah 55 orang dan terdapat 14 orang yang tidak mengikuti kegiatan ujian tertulis sehingga tidak akan diikutsertakan dalam perhitungan sehingga responden yang akan dihitung adalah 41 orang. Perubahan data dari nilai linguistik ke numerik dapat dilihat pada Tabel 1, yang merupakan kriteria berbentuk linguistik yang akan diubah ke nilai numerik. Kriteria Tes tertulis tidak dimasukkan karena memiliki nilai numerik.

Tabel 1. Nilai Linguistik ke Numerik

Kriteria	Lingusitik	Numerik
Tes Wawancara	Memuaskan	1
	Cukup	0.8
	Kurang	0.6
	Tidak ada	0.4
Keaktifan	Aktif	1
	Sedikit Aktif	0.6
	Tidak Aktif	0.2
Kepemimpinan	Ketua	1
	Sekretaris	0.8
	Bendahara	0.6
	Seksi	0.4
	Tidak ada	0.1
UKM	Aktif	1
	Kurang aktif	0.6
	Tidak Aktif	0.2
Sertifikat Keahlian	Programmer	1
	Networking	0.9
	Multimedia	0.8
	Bidang lain	0.4
	Tidak ada	0.1

3.3. Proses Pengelompokan

Untuk merampingkan data agar memudahkan proses perangkingan maka digunakan algoritma K-Means dengan menggunakan nilai k = 2, dengan centroid awal menggunakan nilai Maksimun dan Minimum dari setiap data. Responden yang masuk dalam cluster Maksimum akan digunakan untuk perangkingan. Proses ini akan dilakukan sebanyak tiga kali dengan masing – masing perhitungan jarak yaitu Euclidean, Manhattan dan Minkowski distance. Untuk algoritma K-Means secara ringkas dapat diberikan berikut ini,

1. Menggunakan jumlah kluster K=2,
2. Nilai centroid awal digunakan nilai statistik yaitu maksimum dan minimum dari setiap kriteria data,
3. Menghitung jarak point data ke kedua titik centorid dengan persamaan,
   Euclidean distance,

$$Dist_{xy} = \sqrt{\sum (x_i - y_i)^2} \quad (1)$$

   Manhattan Distance,

$$Dist_{xy} = \sum |x_i - y_i| \quad (2)$$

   Minkowski Distance,

$$Dist_{xy} = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (3)$$

4. Tentukan nilai terdekat terhadap centroid,
5. Proses dilakukan sampai nilai konvergen dimana tidak ada nilai yang berpindah tempat atau kelompok.

3.4. Perangkingan

Tahapan berikutnya adalah proses perangkingan terhadap data yang telah masuk dalam kluster layak untuk dipertimbangkan sebagai calon responden yang akan mengikuti kegiatan. Dengan metode TOPSIS proses perangkingan dilakukan sebanyak tiga kali dengan menggunakan data hasil cluster yang disesuaikan dengan masing – masing pengukuran jarak yang digunakan. Secara umum algoritam Topsis dapat diruaiakan sebagai berikut,

1. Matriks yang terbentuk akan dinormalisasikan dengan vector normalisasi dengan persamaan,
   Untuk kriteria Benefit,

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \quad (4)$$

   Untuk kriteria Cost,

$$r_{ij} = \frac{\left(\frac{1}{X_{ij}}\right)}{\sqrt{\sum_{i=1}^{m} \left(\frac{1}{x_{ij}}\right)^2}} \quad (5)$$

2. Hitung bobot ternormalisasi dengan cara mengalikan bobot setiap kriteria dengan matriks normalisasi,
3. Menghitung nilai solusi ideal positif dan negatif,
4. Hitung *separation measures* alternatif dengan masing – masing jarak yang digunakan pada k-means yaitu,

$$S_i^* = \sqrt{\sum \left(v_j^* - v_{ij}\right)^2} \quad \text{untuk } \textit{Euclidean distance}, \quad (6)$$

$$S_i^* = \sum \left| v_j^* - v_{ij} \right| \quad \text{untuk } \textit{Manhattan distance}, \tag{7}$$

$$S_i^* = \left(\sum \left| v_j^* - v_{ij} \right|^p \right)^{\frac{1}{p}} \quad \text{untuk } \textit{Minkowski distance}, \tag{8}$$

5. Bagian akhir menghitung nilai *Closeness Coefficient* pada solsusi ideal C_i^* dengan persamaan berikut ini,

$$C_i^* = \frac{S_i'}{\left(S_i^* + S_i'\right)} \tag{9}$$

dimana, $0 < C_i^* < 1$

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Berdasarkan skenario yang telah diusulkan maka hasil penelitian dapat diberikan secara detail sebagai berikut, data responden awal akan diperiksa dan menghilangkan data responden yang tidak memiliki data yang lengkap (termasuk responden dengan kriteria yang tidak memiliki nilai). Hal ini berbeda dengan penelitian yang diusulkan oleh Sakur, dkk [3] yang menambahkan nilai terendah bagi kriteria yang tidak memiliki nilai. Hasil konversi nilai terlihat seperti pada Tabel 2,

Tabel 2. Konversi Nilai Kriteria

No.	NIM	TT	TW	IPK	A	P	UKM	SK
1	1705001	10	0.4	3.49	0.6	0.1	0.2	0.4
2	1705002	50	0.8	3.82	0.9	0.4	0.6	0.4
3	1705004	33	0.8	3.71	1	0.4	0.2	0.90
4	1705005	46	0.6	3.54	0.2	0.1	0.2	0.1
5	1705006	48	0.6	3.69	0.2	0.1	0.6	0.4
…	…	…	…	…	…	…	…	…
35	1705068	65	1	3.93	1	0.4	1	0.90
36	1705073	20	0.8	3.72	1	0.4	0.6	0.4
37	1705077	38	0.6	3.82	0.9	0.4	0.6	0.4
38	1705079	25	0.6	3.29	0.2	0.1	0.2	0.1
39	1705080	35	0.4	3.65	0.2	0.8	0.2	0.4
40	1705082	71	1	3.95	1	0.6	1	0.4
41	1705083	5	0.4	3.10	0.2	0.1	0.2	0.4

Ket: TT(Tes Tertulis), TW(Tes Wawancara), IPK, A (Keaktifkan), P (Kepemimpinan), UKM, SK (Sertfifkat Keahlian)

Dari Tabel 2, dilakukan proses clustering menggunakan k-means dengan tiga jarak yang akan digunakan. Dengan mengguanan K = 2 maka ditentukan centroid untuk ketiga jarak k-means yaitu nilai maksimum dan minimum seperti Tabel 3,

Tabel 3. Centroid awal yang digunakan

		TT	TW	IPK	A	P	UKM	SK
C1	Maks	75.00	1.00	3.95	1.00	1.00	1.00	1.00
C2	Min	5.00	0.40	3.04	0.20	0.10	0.20	0.10

Kemudian hitung jarak dari data ke centroid dengan menggunakan *Euclidean distance*, lalu periksa konvergensinya apakah nilai kelas mengalami perpindah kluster atau tidak, jika masih terjadi perpindahan kluster maka proses pehitungan dilanjutkan dengan mengambil nilai rata – rata dari setiap data yang masuk

dalam kluster tersebut. Proses ini dilakukan sampai seluruh perhitungan jarak pada *Manhattan* dan *Minkowski distance* selesai dilakukan. Hasil akhirnya dari ketiga perhitungan tersebut terlihat pada Tabel 4,

Tabel 4. Daftar jarak D1 dan D2 serta Cluster yang terbentuk

No.	NIM	Euclidean Distance			Manhattan Distance			Minkowsky Distance		
		D1	D2	Cluster	D1	D2	Cluster	D1	D2	Cluster
1	1705001	39.39	4.77	2	39.94	5.09	2	39.94	5.09	2
2	1705002	0.65	35.25	1	0.95	35.95	1	0.95	35.95	1
3	1705004	16.40	18.27	1	16.89	19.58	1	16.89	19.58	1
..	…	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…	…
39	1705080	14.40	20.26	1	15.25	21.00	1	15.25	21.00	1
40	1705082	21.63	56.26	1	22.48	57.48	1	22.48	57.48	1
41	1705083	44.39	9.77	2	45.73	10.52	2	45.73	10.52	2

Dari Tabel 4 akan terbentuk dua cluster yang terdiri dari clusert 1 dan 2 dimana anggota dari cluster 1 akan digunakan untuk proses perangkingan yang berjumlah 21 responden. Dengan menggunakan metode TOPSIS dan solusi ideal positif dan negatif menggunakan tiga jarak yang sama dengan k-means maka akan diperoleh hasil ranking seperti pada Tabel 5 berikut,

Tabel 5. Nilai C_i^* dan Rank

No.	Euclidean			Manhattan			Minkowsky (p=4)		
	Nim	C_i^*	Rank	Nim	C_i^*	Rank	Nim	C_i^*	Rank
1	1705002	0.4615	9	1705002	0.4744	8	1705002	0.4640	10
2	1705004	0.4795	7	1705004	0.4864	7	1705004	0.4700	8
3	1705005	0.1984	20	1705005	0.1427	20	1705005	0.2315	20
4	1705006	0.2914	18	1705006	0.2749	17	1705006	0.2898	18
5	1705009	0.2533	19	1705009	0.2269	19	1705009	0.2647	19
6	1705010	0.4291	12	1705010	0.4123	12	1705010	0.4641	9
7	1705014	0.4733	8	1705014	0.4598	9	1705014	0.4618	11
8	1705018	0.8213	1	1705018	0.8781	1	1705018	0.7785	1
9	1705020	0.1691	21	1705020	0.1159	21	1705020	0.2070	21
10	1705022	0.6576	5	1705022	0.6865	5	1705022	0.6480	3
11	1705025	0.3664	14	1705025	0.3725	14	1705025	0.3800	15
12	1705026	0.5961	6	1705026	0.6513	6	1705026	0.5736	6
13	1705038	0.3272	17	1705038	0.2931	16	1705038	0.3670	16
14	1705040	0.4068	13	1705040	0.4273	11	1705040	0.3928	14
15	1705052	0.4440	10	1705052	0.4429	10	1705052	0.4384	12
16	1705057	0.4353	11	1705057	0.4100	13	1705057	0.4715	7
17	1705062	0.6931	4	1705062	0.7557	4	1705062	0.6415	4
18	1705068	0.7145	2	1705068	0.8163	2	1705068	0.6390	5
19	1705077	0.3463	15	1705077	0.3643	15	1705077	0.3307	17
20	1705080	0.3273	16	1705080	0.2399	18	1705080	0.3939	13
21	1705082	0.7044	3	1705082	0.7946	3	1705082	0.6555	2

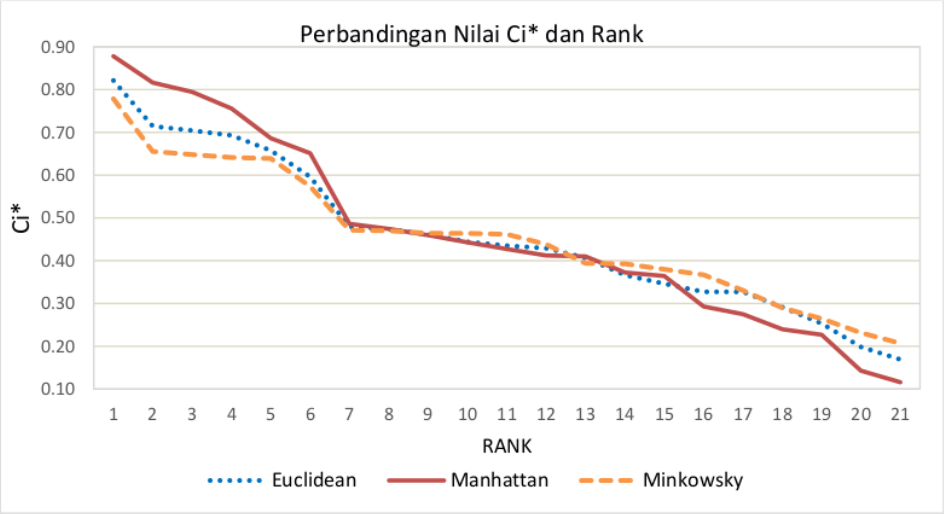
Tabel 5, menunjukkan responden dengan nilai C_i^* dan rank. Bagian terakhir adalah proses pengujian dengan menggunakan korelasi *Pearson* dan *Rank Spearman*. Hasil korelasi terlihat pada Tabel 6, seperti berikut ini

Tabel 6. Korelasi Pearson dan Spearman

		D1		D2		D3		Means		Rank	
		P	S	P	S	P	S	P	S	P	S
Euclidean	D1			0.994	0.990	0.990	0.955	0.992	0.972	1	1
Manhattan	D2	0.994	0.990			0.970	0.927	0.982	0.958	2	2
Minkowsky	D3	0.990	0.955	0.970	0.927			0.980	0.955	3	3

4.2 Pembahasan

Dari penelitian diatas, terlihat bahwa proses klustering yang menggunakan jarak euclidean, manhattan dan minkowsky memiliki hasil pengelompokan yang sama sehingga tidak terjadi masalah pada saat proses perangkingan, walaupun memiliki nilai jarak yang berbeda. Dengan metode TOPSIS dilakukan perhitungan perangkingan dimana proses perbedaan perangkingan dapat dilihat pada Gambar 1. Pada Gambar tersebut terlihat bahwa manhattan distance memberikan nilai yang lebih tinggi untuk bagian awal perangkingan dan memberikan hasil yang kecil pada bagian akhir dari perangkingan. Berdasarkan hasil pengujian dengan menggunakan korelasi Pearson dan Spearman dan mengambil nilai rata – rata atau mean maka Euclidean distance merupakan jarak yang ideal untuk dapat digunakana pada k-means ataupun TOPSIS dengan nilai korelasi pearson 0.992 (99.2%) dan rank spearman sebesar 0.972 (97.2%), sedangkan sebagai alternatif dapat menggunakan pengukuran jarak *Manhattan distance*.



Gambar 1. Perbandingan jarak *Distance Measures*

5. KESIMPULAN DAN SARAN

Penelitian ini telah memberikan dasar pengujian bahwa penggunaan Euclidean distance memberikan hasil yang sangat baik sedangkan untuk manhattan distance dapat dijadikan sebagai alternatif perhitungan jarak lainnya. Dengan menggunakan penggabungan k-means dan TOPSIS dapat terlihat kedua metode tersebut dapat digunakan untuk memercepat proses perhitungan ranking karena proses klustering ataupun pengelompokan dapat merampingkan data yang ada. Penggunaan kedua data ini dapat dijadikan sebagai standar perhitungan baru bagi institusi dalam melakukan pemilihan kandidat untuk kegiatan kunjungan industri. Sebagai saran, perlu dilakukan analisis terhadap algoritma pengukuran jarak (distance measures) lainnya yang ada untuk membantuk peneliti lainnya memilih algoritma yang tepat dengan kasus yang berbeda.

Ucapan Terima Kasih

Terima kasih kepada tim peneliti Miske Silangen, S.Kom., M.Cs dan Desmin Tuwohingide, S.Pd., M.Kom yang telah mengizinkan penggunaan dataset ini, dan terima kasih kepada bapak Dr. Wendy Alexander Tanod, S.Kel., M.Si. yang telah memotivasi untuk selalu berkarya dan melakukan berbagai penelitian.

DAFTAR PUSTAKA

[1] D. J. Bora and D. A. K. Gupta, "Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab," vol. 5, 2014.

[2] A. B. Rathod, S. M. Gulhane, and S. R. Padalwar, "A comparative study on distance measuring approches for permutation representations," in *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*, Pune, India: IEEE, Dec. 2016, pp. 251–255. doi: 10.1109/ICAECCT.2016.7942593.

[3] S. B. H. Sakur, M. Silangen, and D. Tuwohingide, "Penerapan Algoritma K-Means Cluster dan Metode TOPSIS pada Pemilihan Mahasiswa kunjungan Industri," *Jutisi J. Ilm. Tek. Inform. Dan Sist. Inf.*, vol. 11, no. 3, Art. no. 3, Dec. 2022, doi: 10.35889/jutisi.v11i3.1045.

[4] S. B. H. Sakur, M. Silangen, and E. H. Israel, "Penggunaan Metode Technique For Order Performance Of Similarity To Ideal Solution (TOPSIS) Dan Vector Normalization Pada Pemilihan Mahasiswa Kunjungan Industri," Politeknik Negeri Nusa Utara, Tahuna, Laporan Penelitian Unggulan Perguruan Tinggi 461/Sistem Informasi, Nov. 2021.

[5] M. Anggara, H. Sujiani, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness," vol. 1, no. 1, Art. no. 1, 2016.

[6] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, Jan. 2019, doi: 10.30591/jpit.v4i1.1253.

[7] S. B. H. Sakur and M. Silangen, "ANALISIS PERBANDINGAN NORMALISASI DARI METODE ANALYTICAL HIERARCHY PROCESS TERHADAP METODE SIMPLE MULTI ATTRIBUTE RATING TECHNIQUE UNTUK PEMILIHAN MAHASISWA BERPRESTASI," Politeknik Negeri Nusa Utara, Tahuna, Laporan Penelitian Unggulan Perguruan Tinggi 461/Sistem Informasi, Nov. 2022.

[8] J. Irani, N. Pise, and M. Phatak, "Clustering Techniques and the Similarity Measures used in Clustering: A Survey," *Int. J. Comput. Appl.*, vol. 134, no. 7, Art. no. 7, Jan. 2016, doi: 10.5120/ijca2016907841.

[9] Haviluddin *et al.*, "A Performance Comparison of Euclidean, Manhattan and Minkowski Distances in K-Means Clustering," in *2020 6th International Conference on Science in Information Technology (ICSITech)*, Palu, Indonesia: IEEE, Oct. 2020, pp. 184–188. doi: 10.1109/ICSITech49800.2020.9392053.

[10] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm For Big Data," *Procedia Comput. Sci.*, vol. 78, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.

[11] S. B. H. Sakur, "Data Excel Proses Analisis Kunjungan Industri Metode TOPSIS." Data Excel, Google Drive, Nov. 06, 2021. Accessed: Apr. 05, 2023. [Excel]. Available: https://drive.google.com/file/d/1Z8UmqyJ7v5Nc4Y42zLWOv7kxjzoI78ny/view

PERBANDINGAN DISTANCE MEASURES PADA K-MEANS CLUSTER DAN TOPSIS DENGAN KORELASI PEARSON DAN SPEARMAN

ORIGINALITY REPORT

14%	14%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	journal.amikveteran.ac.id Internet Source	5%
2	ojs.stmik-banjarbaru.ac.id Internet Source	3%
3	journal.lembagakita.org Internet Source	1%
4	sisfotenika.stmikpontianak.ac.id Internet Source	1%
5	www.sistemphp.com Internet Source	1%
6	jurnal.darmajaya.ac.id Internet Source	<1%
7	Ni Luh Putu Purnama Dewi, I Nyoman Purnama, Nengah Widya Utami. "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)", Jurnal Ilmiah Teknologi Informasi Asia, 2022	<1%

	Publication	
8	Submitted to Universitas Trunojoyo Student Paper	<1 %
9	adoc.pub Internet Source	<1 %
10	download.garuda.ristekdikti.go.id Internet Source	<1 %
11	etds.lib.ntnu.edu.tw Internet Source	<1 %
12	Luo, Yu, and Gang Guo. "Method of extension classification configuration for product family case", IEEE 10th International Conference on Cognitive Informatics and Cognitive Computing (ICCI-CC 11), 2011. Publication	<1 %
13	docslide.us Internet Source	<1 %
14	www.infopmb.web.id Internet Source	<1 %
15	doku.pub Internet Source	<1 %
16	e-journal.unair.ac.id Internet Source	<1 %
17	journal.uny.ac.id Internet Source	<1 %

18	publikasi.dinus.ac.id Internet Source	<1 %
19	repository.its.ac.id Internet Source	<1 %
20	www.researchgate.net Internet Source	<1 %
21	Amir Ali. "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2019 Publication	<1 %
22	Yudistira Ergha Riandana, Muhammad Hamka. "Sistem Pendukung Keputusan Penerima Pembiayaan Akad Multijasa Menggunakan Metode Analitycal Hierarchy Process Dan Technique For Order Preference By Similarity To Ideal Solution", Techno (Jurnal Fakultas Teknik, Universitas Muhammadiyah Purwokerto), 2020 Publication	<1 %
23	ejournal.poltektegal.ac.id Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		