

Penerapan Algoritma K-Means pada Ketersediaan Lapangan Olahraga Setiap Kelurahan Di Indonesia

Cristo Dhear Harafenta Lumbantobing¹, Saifullah², Fitri Rizki³
^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar – Indonesia
 Jln. Sudirman Blok A No. 1-3 Pematangsiantar, Sumatera Utara
¹christo.tobing@yahoo.com, ²saifullah@amiktunasbangsa.ac.id,
³fitri_rizki@sitkomb.ac.id

Abstract

Sports facilities are the basis of sports activities such as athletic events such as volleyball, ball, tennis and others. Each type of sports field requires advice and infrastructure to continue activities. The method used is Data Mining using the K-Means algorithm. This method data is grouped into 2 clusters. The author uses data obtained from government sites that address <http://bps.go.id>. Based on recap result data from 2014 to 2018 The final results announced are based on sports fields which are grouped into the lowest highest cluster based on the type of sport in each Kelurahan according to the provinces in Indonesia. By using the K-Means method provides grouping of an area in the Kelurahan according to the Province in Indonesia. In this study, the government provided assistance to the sports field and improved facilities and infrastructure in the field of sports for the athletic community.

Keywords : *Data mining, Clustering, K-Means, Field Sports*

Abstrak

Fasilitas olahraga merupakan dasar melakukan aktivitas olahraga misalnya pertandingan atletik seperti lapangan bola volly, sepak bola, tenis dan lain-lain. Setiap jenis lapangan olahraga membutuhkan tempat saran dan prasarana untuk berlangsungnya aktifitas. Metode yang digunakan adalah Data Mining dengan menggunakan algoritma K-Means. Metode ini data-data dikelompokkan menjadi 2 cluster. Penulis menggunakan data yang diperoleh dari situs pemerintah yang beralamatkan <http://bps.go.id>. Berdasarkan data hasil rekap mulai tahun 2014 sampai 2018 menunjukkan jumlah ketersediaan lapangan olahraga yang bervariasi jumlahnya. Hasil akhir menyebutkan bahwa ketersediaan lapangan olahraga yang dikelompokkan menjadi cluster terendah cluster tertinggi berdasarkan jenis olahraga di setiap Kelurahan menurut provinsi di Indonesia. Dengan menggunakan metode K-Means memudahkan pengelompokan suatu daerah yang ada di Kelurahan menurut Provinsi di Indonesia. Dalam penelitian ini diharapkan pemerintah setempat dapat memberi solusi terhadap ketersediaan lapangan olahraga dan meningkatkan sarana dan prasaran dibidang olahraga untuk menghasilkan atlet dari masyarakat terutama yang tinggal Kelurahan menurut Provinsi di Indonesia.

Kata Kunci : *Data mining, Clustering, K-Means, Lapangan Olahraga*

1. Pendahuluan

Lapangan olahraga merupakan suatu tempat melakukan kegiatan olahraga di ruangan terbuka untuk berlangsungnya segala aktivitas olahraga. Jenis olahraga membutuhkan sarana lapangan diantaranya sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, futsal, kolam renang. Setiap jenis lapangan memiliki ukuran yang berbeda-beda. Lapangan olahraga biasanya digunakan untuk aktivitas yang biasanya tidak dilakukan tiap harinya, yang terkadang digunakan sebagai tempat

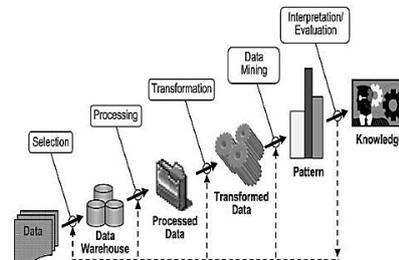
konser musik pada lapangan sepak bola. Tanpa adanya ketersediaan lapangan olahraga bisa dikatakan tidak mungkin untuk mewujudkan pembangunan keolahragaan, ada pun rendahnya minat partisipasi di suatu daerah diduga disebabkan kurangnya ketersediaan lapangan olahraga yang disediakan Pemerintah. Dalam hal ini Pemerintah Pusat dan Pemerintah Daerah dan tidak terlepas dari partisipasi masyarakat mempunyai hak dan kewajiban yang sama bila mana untuk mewujudkan dan meningkatkan hal tersebut. Agar setiap masyarakat yang tinggal di Kelurahan bisa menggunakan lapangan olahraga tersebut. Menurut data yang diperoleh dari situs resmi Pemerintah dalam laman website <https://bps.go.id> memperlihatkan sejumlah data cukup banyak ketersediaan lapangan olahraga setiap Kelurahan menurut Provinsi di Indonesia mulai dari tahun 2014 sampai 2018. Oleh karena itu, dengan penggunaan *Data Mining* dengan algoritma *K-Means* dapat mempermudah pengelompokan data yang berjumlah besar dan menentukan *cluster* setiap variabel mulai dari terendah sampai tertinggi [1]–[5].

Algoritma *K-Means* merupakan *clustering* untuk *Data Mining* yang telah diciptakan pada tahun 70an. Algoritma *K-Means* memiliki kemampuan dalam pengelompokan data dalam jumlah yang cukup besar. Diharapkan penelitian ini dapat berguna dalam penyelesaian permasalahan terhadap ketersediaan lapangan olahraga diberbagai Kelurahan menurut Provinsi di Indonesia, selalu ada aksi terbaru nyata dari pihak Pemerintah untuk mendorong lagi dalam meningkatkan kualitas lapangan olahraga yang layak dan nyaman bagi masyarakat.

2. Metodologi Penelitian

2.1. Data Mining

Data Mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Istilah lain yang sering digunakan diantaranya *knowledge discovery (mining) in databases (KDD)*, *knowledge extraction*, data atau *pattern analysis*, data *archeology*, data *dredging*, *information harvesting*, dan *business intelligence* [6]. Adapun tahapan-tahapan dalam *Data Mining* yaitu [7]:



Gambar 1. Tahapan Data Mining

2.2. Metode K-means

Metode K-Means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, di mana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya [8].

K-Means merupakan salah satu metode data *Clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *Cluster* atau kelompok. Metode ini mempartisi ke dalam *Cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu *Cluster* yang sama dan yang memiliki karakteristik yang berbeda (*Low inter class similarity*) dikelompokkan pada kelompok yang lain. Proses *Clustering* dimulai dengan mengidentifikasi data yang akan *diCluster*, X_{ij} ($i=1, \dots, n; j=1, \dots, m$)

dengan n adalah jumlah data yang akan *diCluster* dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap *Cluster* ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat *Cluster*. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat *Cluster* ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean, seperti pada persamaan (1), yaitu [6]:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (1)$$

Suatu data akan menjadi anggota dari *Cluster* ke- k apabila jarak data tersebut ke pusat *Cluster* ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat *Cluster* lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *Cluster*.

$$\text{Min } \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{ij})^2} \quad (2)$$

Nilai pusat *Cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *Cluster* tersebut, dengan menggunakan rumus pada persamaan (3):

$$c_{ij} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

Dimana $x_{ij} \in \text{Cluster ke } k$

p = banyaknya anggota *Cluster* ke k

Kelebihan Algoritma *K-Means* diantaranya adalah mampu mengelompokkan objek besar dan penciran obyek dengan sangat cepat sehingga mempercepat proses pengelompokan. Kekurangan Algoritma *K-Means* yaitu Sangat sensitive pada pembangkitan titik pusat awal secara random, hasil pengelompokan bersifat tidak unik (selalu berubah-ubah) dan proses pengerjaannya cepat tetapi keakuratannya tidak dijamin.

2.3. Clustering

Clustering data dapat dibedakan menjadi dua tujuan, yaitu *Clustering* untuk pemahaman dan *Clustering* untuk penggunaan. Jika tujuan untuk pemahaman maka *Cluster* yang terbentuk harus menangkap struktur alami data. Biasanya proses *Clustering* dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan ini seperti *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. *Clustering* dapat dibedakan menurut struktur *Cluster*, keanggotaan data dalam *Cluster* dan kekompakan data dalam *Cluster* [6].

2.4. Rapid Miner

Rapid miner merupakan aplikasi *open source* berlisensi *AGPL (GNU Affero General Public License)* yang dapat digunakan untuk mengolah *Data Mining* yang dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari University of Dortmund [9].

3. Hasil dan Pembahasan

Tahap ini menjelaskan pengolahan data mengenai jumlah ketersediaan lapangan olahraga setiap Kelurahan di Indonesia mulai dari tahun 2014 dan 2018 yang diperoleh dari situs resmi Pemerintah dalam laman website <http://bps.go.id>. Data selanjutnya akan dijadikan menjadi 2 *Cluster*, mulai dari tertinggi sampai terendah. Alat analisis data digunakan dalam penelitian ini aplikasi *Rapidminer 5.3*. Berikut merupakan uraian menjelaskan tahap perhitungan manual dari proses Metode Algoritma *K-Means Clustering*.

a) Menentukan Data Yang akan di *Cluster*

Berikut data telah diproses *Clustering* yang merupakan jumlah ketersediaan lapangan olahraga setiap Kelurahan menurut Provinsi di Indonesia. Data dapat dilihat pada tabel 1 berikut :

Tabel 1. Data Jumlah Ketersediaan Lapangan Olahraga

No	Provinsi	Sepak Bola	Bola Volly	Bulu Tangkis	Bola Basket	Tennis Lapangan	Futsal	Kolam Renang
1	Aceh	2587	3735,5	959	160,5	154,5	249,5	66
2	Sumatera Utara	2018,5	3057	2086	201,5	218	567	172,5
3	Sumatera Barat	823,5	981,5	749,5	221,5	109	353,5	64,5
4	Riau	1530	1757,5	1364,5	170	103,5	450,5	90,5
5	Jambi	1141	1314	1126	129,5	88	272	33
6	Sumatera Selatan	1652	2791	2110	204	139,5	337	93
7	Bengkulu	596,5	1064,5	519,5	109,5	43	195,5	31
8	Lampung	1846,5	2159,5	1902,5	196,5	95,5	319,5	98,5
9	Kep. Bangka Belitung	350	350	294	74,5	43,5	175,5	29,5
10	Kep. Riau	352	404,5	183	67	54,5	85	36
11	Dki Jakarta	138,5	205,5	248,5	181	137	178,5	81
12	Jawa Barat	4232,5	4983	4412,5	1003,5	608,5	1654,5	713
13	Jawa Tengah	6495	6242,5	5305	779	761,5	1020,5	551
14	Di Yogyakarta	381	407,5	383,5	121	100,5	89,5	90
15	Jawa Timur	5337,5	6112	3258	778,5	651,5	1042,5	693
16	Banten	946,5	1095,5	852,5	205,5	136,5	365	156,5
17	Bali	267,5	551,5	423,5	123,5	72	155	55,5
18	Nusa Tenggara Barat	652,5	630,5	523,5	99,5	65,5	120	60,5
19	Nusa Tenggara Timur	1632	2343,5	337	99,5	50,5	122	28,5
20	Kalimantan Barat	1763,5	1945,5	1008,5	127,5	94,5	198,5	63
21	Kalimantan Tengah	950	1402	914,5	95	51	128,5	20,5
22	Kalimantan Selatan	903,5	910	1028	164,5	99,5	172	43
23	Kalimantan Timur	828	916	779,5	125	84	147	66
24	Kalimantan Utara	280	341,5	175	32,5	22	52,5	11
25	Sulawesi Utara	819,5	776	599	96,5	54,5	83,5	48,5
26	Sulawesi Tengah	1532,5	1691,5	600	49,5	48,5	128,5	14
27	Sulawesi Selatan	1775,5	1831,5	1215	290,5	266,5	385	137
28	Sulawesi Tenggara	1093	1681,5	711,5	75,5	39,5	264,5	23
29	Gorontalo	374,5	426	238,5	43	14,5	27,5	9,5
30	Sulawesi Barat	372,5	529	243,5	39,5	40,5	44,5	22,5
31	Maluku	648,5	798	115,5	51	26	95	13
32	Maluku Utara	734,5	621	128	15,5	19,5	30,5	4
33	Papua Barat	375,5	789	96	45	33	64,5	22
34	Papua	1327,5	2534,5	195,5	118,5	74,5	136	61

b) Menentukan Jumlah *Cluster*

Pada tahap ini menentukan jumlah *Cluster* sebanyak 2 (dua) mulai dari *Cluster* tertinggi (C1) sampai *Cluster* terendah (C2) dalam perhitungan manual Metode Algoritma *K-Means*.

c) Menentukan Nilai *Centroid*

Langkah berikutnya yaitu mencari nilai *Cluster* tertinggi (C1) dan *Cluster* terendah (C2) pada tabel 1. Berikut merupakan tabel *centroid* data pada tabel 2.

Tabel 2. Nilai *Centroid* Data

	Sepak Bola	Bola Volly	Bulu Tangkis	Bola Basket	Tennis Lapangan	Futsal	Kolam Renang
<i>Cluster</i> I	6495	6242,5	5305	1003,5	761,5	1654,5	713
<i>Cluster</i> II	138,5	205,5	96	15,5	14,5	27,5	4

d) Menghitung Jarak *Centroid*

$$D_{Aceh,C1} = \sqrt{(2587 - 6495)^2 + (3735,5 - 6242,5)^2 + (959 - 5306)^2 + (160,5 - 1003,5)^2 + (154,5 - 761,5)^2 + (249,5 - 1664,5)^2 + (66 - 713)^2} = 6627,0$$

$$D_{Aceh,C2} = \sqrt{(2587 - 138,5)^2 + (3735,5 - 206,5)^2 + (959 - 96)^2 + (160,5 - 15,5)^2 + (154,5 - 14,5)^2 + (249,5 - 27,5)^2 + (66 - 4)^2} = 4392,6$$

Tabel 3. Nilai Jarak Terdekat Pada Cluster Iterasi 1

No	Provinsi	C1	C2	Jarak Terpendek
1	Aceh	6627,0	4392,6	4392,6
2	Sumatera Utara	6554,5	4002,6	4002,6
3	Sumatera Barat	9151,3	1288,3	1288,3
4	Riau	7953,5	2484,4	2484,4
5	Jambi	8602,9	1836,8	1836,8
6	Sumatera Selatan	6980,1	3631,3	3631,3
7	Bengkulu	9402,9	1079,6	1079,6
8	Lampung	7287,7	3183,1	3183,1
9	Kep. Bangka Belitung	10081,0	362,9	362,9
10	Kep. Riau	10116,9	318,3	318,3
11	Dki Jakarta	10298,7	307,2	307,2
12	Jawa Barat	2743,2	7918,1	2743,2
13	Jawa Tengah	691,8	10315,5	691,8
14	Di Yogyakarta	9985,1	460,5	460,5
15	Jawa Timur	2446,4	8626,0	2446,4
16	Banten	8950,0	1485,1	1485,1
17	Bali	9946,3	526,8	526,8
18	Nusa Tenggara Barat	9621,5	805,6	805,6
19	Nusa Tenggara Timur	8226,3	2622,5	2622,5
20	Kalimantan Barat	7941,4	2559,8	2559,8
21	Kalimantan Tengah	8808,7	1666,8	1666,8
22	Kalimantan Selatan	9043,3	1414,9	1414,9
23	Kalimantan Timur	9213,5	1217,5	1217,5
24	Kalimantan Utara	10213,4	214,0	214,0
25	Sulawesi Utara	9406,0	1027,4	1027,4
26	Sulawesi Tengah	8468,2	2101,9	2101,9
27	Sulawesi Selatan	7820,2	2619,2	2619,2
28	Sulawesi Tenggara	8653,0	1878,6	1878,6
29	Gorontalo	10078,9	354,1	354,1
30	Sulawesi Barat	10013,3	427,8	427,8
31	Maluku	9753,2	785,9	785,9
32	Maluku Utara	9810,8	727,3	727,3
33	Papua Barat	9938,7	632,1	632,1
34	Papua	8399,1	2622,4	2622,4

Dari table diatas dapat dilihat bahwa jumlah dari cluster tinggi (C1) 3 dan cluster rendah (C2) 31 record. Penentuan cluster dapat dilihat dari nilai jarak. Jika hasil jarak pusat cluster C1 lebih rendah dari C2 maka berda pada cluster tinggi, begitu juga sebaliknya.

e) Menghitung Centroid baru dari hasil setiap anggota masing-masing *Cluster*. Setelah didapat hasil jarak dari setiap objek pada iterasi ke-1 maka lanjut ke iterasi ke-2 pada perhitungan dan tabel dibawah ini :

$$D_{Sepak Bola,c1} = \frac{4232,5+6495+5337,5}{3} = 5355,0$$

$$D_{\text{Sepak Bola},c2} = \frac{2587+2018,5+823,5+1530+1141+1652+596,5+1846,5+350+352+138,5+381+946,5+267,5+652,5+1632+1763,5+950+903,5+828+280+819,5+1532,5+1775,5+1093+374,5+372,5+648,5+734,5+375,5+1327,5}{31} = 930,1$$

Untuk menentukan Centroid baru digunakan cara yang sama untuk seluruh criteria berdasarkan hasil cluster dari iterasi 1. Berikut adalah hasil dari Centroid baru:

Tabel 4. Centroid Baru

	Sepak Bola	Bola Volly	Bulu Tangkis	Bola Basket	Tennis Lapangan	Futsal	Kolam Renang
Cluster I	5335,0	5779,2	4325,2	853,7	673,8	1239,2	652,3
Cluster II	930,1	1213,9	670	113,1	78,2	181,6	52,9

- f) Bagian ini dilakukan iterasi selanjutnya dikarenakan nilai *centroid* tidak sama. Selanjutnya menghitung perhitungan jarak terhadap data jumlah Desa/Kelurahan menurut Provinsi di Indonesia yang memiliki ketersediaan lapangan olahraga dari data *centroid* baru seperti pada perhitungan dan tabel 5 berikut :

$$D_{\text{Aceh},C1} = \sqrt{(2587 - 5335,0)^2 + (3735,5 - 5779,2)^2 + (959 - 4325,2)^2 + (160,5 - 853,7)^2 + (154,5 - 673,8)^2 + (249,5 - 1239,2)^2 + (66 - 662,3)^2} = 5013,3$$

$$D_{\text{Aceh},C2} = \sqrt{(2587 - 930,1)^2 + (3735,5 - 1213,4)^2 + (959 - 670)^2 + (160,5 - 113,1)^2 + (154,5 - 78,2)^2 + (249,5 - 181,6)^2 + (66 - 52,9)^2} = 3033,6$$

Begitu seterusnya hingga diperoleh hasil dari seluruh record untuk masing-masing cluster. Berikut adalah hasil dari iterasi 2.

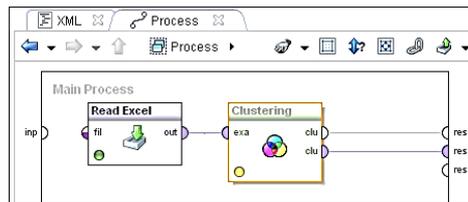
Tabel 5. Hasil Dari Perhitungan Iterasi Ke-2

Provinsi	C1	C2	Jarak Terpendek
Aceh	5013,3	3033,6	3033,6
Sumatera Utara	4973,8	2603,6	2603,6
Sumatera Barat	7616,2	337,4	337,4
Riau	6414,7	1102,7	1102,7
Jambi	7067,5	521,0	521,0
Sumatera Selatan	5407,0	2263,0	2263,0
Bengkulu	7848,1	397,5	397,5
Lampung	5749,8	1811,7	1811,7
Kep. Bangka Belitung	8549,9	1107,5	1107,5
Kep. Riau	8578,5	1112,7	1112,7
Dki Jakarta	8768,2	1352,4	1352,4
Jawa Barat	5124,1	6542,3	5124,1
Jawa Tengah	1609,9	8922,2	1609,9
Di Yogyakarta	8452,1	1021,5	1021,5
Jawa Timur	1138,5	7214,0	1138,5
Banten	7414,5	322,2	322,2
Bali	8405,9	968,9	968,9
Nusa Tenggara Barat	8088,9	665,2	665,2
Nusa Tenggara Timur	6634,5	1373,2	1373,2
Kalimantan Barat	6383,3	1160,2	1160,2
Kalimantan Tengah	7253,8	317,3	317,3
Kalimantan Selatan	7518,9	473,5	473,5
Kalimantan Timur	7678,0	335,3	335,3
Kalimantan Utara	8675,5	1206,6	1206,6
Sulawesi Utara	7871,7	468,1	468,1

Provinsi	C1	C2	Jarak Terpendek
Sulawesi Tengah	6908,3	778,2	778,2
Sulawesi Selatan	6277,4	1228,5	1228,5
Sulawesi Tenggara	7085,1	508,0	508,0
Gorontalo	8540,4	1072,1	1072,1
Sulawesi Barat	8469,9	993,9	993,9
Maluku	8202,4	758,3	758,3
Maluku Utara	8270,4	849,2	849,2
Papua Barat	8380,1	915,7	915,7
Papua	6790,4	1459,6	1459,6

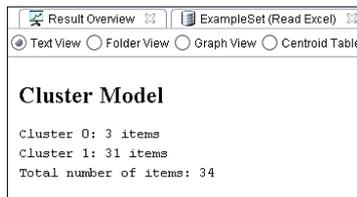
Hasil perhitungan manual yang telah dilakukan diatas tersebut menunjukkan hasil akhir pada iterasi 1 dan iterasi 2 didapatkan hasil yang sama. Nilai iterasi 1 dan iterasi 2 bernilai C1 =3 dan C2 =31 pada posisi data tiap *Cluster*. Jika proses telah dilakukan menghasilkan iterasi yang sama, maka proses iterasi berhenti sampai iterasi 2.

Selanjutnya pada bagian ini menjelaskan tentang proses sistem pada *RapidMiner*, data yang telah *diimport* kemudian ketik “*K-Means*” dan pilih “*Clustering and Segmentation*” pada bagian pencarian tersebut. Data yang di import adalah data yang sama dengan data yang digunakan saat pengolahan data secara manual. Lalu hubungkan *Read Excel* ke bagian *Clustering*, seperti gambar dibawah ini :



Gambar 2. Pemrosesan Data

Untuk melihat hasil pemrosesan data tersebut, kemudian klik *run* (tanda panah biru) pada bagian *tool bar*. Hasil akhir dari pemrosesan data dapat dilihat pada gambar dibawah :



Gambar 3. Cluster Model Pada Akhir Pemrosesan Data

Keterangan :

1. *Cluster 0* (Tinggi) = 3 items
2. *Cluster 1* (Rendah) = 31 items

Berdasarkan hasil percobaan dalam pengujian data menggunakan aplikasi *RapidMiner* berhasil menunjukkan hasil yang sama dengan penerapan perhitungan manual *algoritma K-Means* dari *Microsoft excel* dengan hasil *Cluster* tertinggi memiliki nilai 3 sedangkan *Cluster* terendah memiliki nilai 31.

4. Kesimpulan

Berdasarkan hasil pengujian pengolahan data menggunakan aplikasi *RapidMiner* pada *Clustering* data ketersediaan lapangan olahraga di Kelurahan menurut Provinsi di Indonesia, maka dapat disimpulkan hasil perhitungan metode *algoritma K-Means* yang sebelumnya dilakukan secara manual dan diimplementasikan ke dalam pengujian data

menggunakan aplikasi *RapidMiner*. Hasil yang diperoleh memiliki nilai yang sama yaitu menghasilkan 2 *cluster*, Mulai dari *cluster* tertinggi bernilai 3 yaitu Provinsi Jawa Barat, Jawa Tengah dan Jawa Timur . Sedangkan *cluster* terendah bernilai 31 yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Dki Jakarta, Di Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Hasil tersebut sama dengan hasil yang telah diimplementasikan ke dalam *Software RapidMiner*. Penelitian ini berguna untuk masukan kepada Pemerintah Daerah maupun Pusat untuk meningkatkan jumlah ketersediaan lapangan maupun kualitas fasilitas lapangan olahraga tersebut.

Daftar Pustaka

- [1] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [2] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.
- [3] M. G. Sadewo, A. Eriza, A. P. Windarto, and D. Hartama, "Algoritma K-Means Dalam Mengelompokkan Desa / Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," *Semin. Nas. Teknol. Komput. Sains SAINTEKS 2019*, pp. 754–761, 2019.
- [4] A. P. Windarto *et al.*, "Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012001.
- [5] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- [6] M. A. W. K. MURTI, "Penerapan metode K-means clustering untuk mengelompokan potensi produksi buah – buahan di provinsi daerah istimewa yogyakarta," 2017.
- [7] T. Taslim and F. Fajrizal, "Penerapan algorithm k-mean untuk clustering data obat pada puskesmas rumbai," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 108–114, 2016, doi: 10.31849/digitalzone.v7i2.602.
- [8] P. Sari, B. Pramono, and L. ode H. S. Sagala, "Improve K-Means Terhadap Status Nilai Gizi Pada Balita," *semanTIK*, vol. 3, no. 1, pp. 143–148, 2017, doi: 10.1063/1.2957900.
- [9] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam," *Y. Darmi, A. Setiawan*, vol. 12, no. 2, pp. 148–157, 2016.