



Penerapan Metode *K-Means Clustering* Untuk Menentukan Faktor Resiko Pada Penderita Diabetes Melitus

Melda Septriani¹, Pareza Alam Jusia², Rudolf Sinaga³, Shinta Renova Putri⁴, Firyal Najla 'Afifah⁵

¹⁻⁵ Sistem Informasi, Ilmu Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: mldsptrr@gmail.com¹, parezaalam@gmail.com², rudolfverdinan@gmail.com³,

Sintarenova@gmail.com⁴, firyalnajlla2005@gmail.com⁵

Alamat: Jalan Jendral Sudirman, Thehok, Jambi 36138.

*Penulis Korespondensi: mldsptrr@gmail.com

Abstract. *Diabetes Mellitus is a disease caused by the failure of the pancreas organ in producing the hormone insulin in excess causing increased blood sugar levels and resulting in a lack of insulin. This study discusses the application of the k-means clustering method to determine risk factors for diabetes mellitus. By using the clustering method, data will be grouped into several clusters or groups which in this study compare by applying several data mining tools such as RapidMiner, SPSS, WEKA, and Python. From the results of the comparison carried out resulted in 5 calculations, namely the manual calculation of cluster 1 with a ratio value of 73% being the first priority, calculations using RapidMiner resulting in cluster 3 with a ratio value of 58% being the first priority, calculations using SPSS cluster 2 with a ratio value of 34% being the first priority, and calculations using Python produce cluster 1 with a ratio value of 55% being the first priority.*

Keywords: *Diabetes Melitus; Clustering; Algoritma K-Means; Tools*

Abstrak. Diabetes Melitus adalah penyakit yang disebabkan karena gagalnya organ pankreas dalam memproduksi hormon insulin secara berkalah sehingga menyebabkan meningkatnya kadar gula dalam darah dan berakibat kurangnya insulin. Penelitian ini membahas tentang penerapan metode *k-means clustering* untuk menentukan faktor resiko pada penyakit diabetes melitus. Dengan menggunakan metode *clustering* data akan dikelompokkan ke dalam beberapa *cluster* atau kelompok yang dimana pada penelitian ini melakukan komparasi dengan menerapkan beberapa *tools* data mining seperti *RapidMiner*, SPSS, WEKA, dan *Phyton*. Dari hasil komparasi yang dilakukan menghasilkan 5 perhitungan yaitu perhitungan secara manual *cluster* 1 dengan nilai rasionya 73% menjadi prioritas pertama, perhitungan menggunakan *RapidMiner* menghasilkan *cluster* 3 dengan nilai rasionya 58% menjadi prioritas pertama, perhitungan menggunakan SPSS *cluster* 2 dengan nilai rasionya 34% menjadi prioritas pertama, dan perhitungan menggunakan *Phyton* menghasilkan *cluster* 1 dengan nilai rasionya 55% menjadi prioritas pertama.

Kata Kunci: *Diabetes Melitus; Clustering; Algoritma K-Means; Tools*

1. LATAR BELAKANG

Pada abad ini diabetes melitus menjadi salah satu penyakit darurat kesehatan global yang dimana penyakit ini dapat menyebabkan kematian 4-5 kali dibandingkan penyakit lainnya dengan mayoritas kematian pada usia 45-54 tahun (Paramitasari, 2017). Diabetes Melitus (DM) merupakan gangguan kesehatan *degenerative* yang tidak menular namun cukup serius karena gagalnya organ pankreas dalam memproduksi hormon insulin secara berkalah sehingga mengakibatkan peningkatan kadar gula dalam darah yang mengakibatkan kekurangannya insulin (Indriyani, 2021). Diabetes melitus disebut sebagai penyakit kronis karena dapat terjadi secara menahun atau akut. Diabetes melitus bersifat *muktifaktorial* dengan penyebabnya seperti genetic, gaya hidup, pola makan dan jenis makanan yang dikonsumsi (Wicaksono, 2015). Berdasarkan penyebabnya diabetes melitus dapat digolongkan menjadi beberapa jenis

diantaranya Diabetes tipe 1, Diabetes tipe 2, Diabetes gestasional dan Diabetes tipe lainnya (Apa Saja Tipe Penyakit DM?, 2018).

Perilaku yang tidak sehat serta perubahan gaya hidup menjadi salah satu faktor resiko dari penderita diabetes melitus (Nasution et al., 2021). Perubahan gaya hidup bisa seperti diet yang tidak sehat, kurang aktifitas fisik, obesitas (mempunyai berat badan lebih), hipertensi, konsumsi alkohol dan kebiasaan merokok. Terdapat faktor-faktor resiko lainnya juga misalnya usia, jenis kelamin, dan keturunan (*Faktor Risiko Penyakit Diabetes Melitus (DM)*-, 2019). Untuk menarik suatu kesimpulan dalam menentukan diagnosis faktor resiko yang paling prediktif pada resiko penderita Diabetes melitus maka diperlukan suatu metode diagnosis yang dapat mengatasi yang sesuai dengan faktor resikonya dengan memanfaatkan teknologi yang berkembang yaitu data mining.

Data mining adalah suatu proses *ekstraksi* data menjadi informasi atau pengetahuan maupun pola dari data dengan jumlah data yang banyak menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Suntoro, 2019). Teknik *data mining* mampu menggali, mendapatkan informasi dari kumpulan data dengan melakukan teknik *data mining*, yaitu salah satunya menggunakan *clustering* (Susanti & Widodo, 2017). *Clustering* merupakan teknik pengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam suatu *cluster* yang memiliki tingkat kemiripan yang maksimum dan antara *cluster* tersebut memiliki kemiripan yang minimum, atau suatu metode yang mencari atau mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara data dengan data lain (Irwansyah & Faisal, 2015).

Pada penelitian ini metode clustering digunakan untuk melakukan pengelompokan faktor resiko penderita diabetes melitus, yang kemudian akan dianalisis pola pengelompokkannya. Algoritma klasterisasi bermacam-macam sebagai contoh *K-Means clustering*, *Fuzzy C-Means clustering*, *Particle Swarm Optimization* dan lain sebagainya (Bulolo, 2020). Algoritma klastering yang sangat umum digunakan adalah klastering k-means. K-Means mampu mengelompokkan data dalam jumlah yang besar atau banyak dalam waktu komputasinya yang relatif cepat dan hasil yang efisien (Danny & Yusup, 2020). K-Means juga memiliki ketelitian yang cukup tinggi terhadap ukuran objek sehingga algoritma k-means relatif lebih terukur, selain itu *k-means* tidak terpengaruh oleh urutan objek (Prasetyo, 2012). Sedangkan setelah melakukan studi literatur pada permasalahan ini, peneliti mendapatkan solusi yang telah dilakukan oleh penelitian sebelumnya sebagai berikut:

Akmal Rustandi melakukan penerapan algoritma *k-means clustering* dalam mengelompokkan tingkat resiko penyakit diabetes melitus menggunakan data mining dengan melalui beberapa tahapan dan menghasilkan cluster dengan nilai yang optimal. Metode clustering menggunakan *k-means* mendapatkan sebuah wawasan baru yaitu pengelompokkan tingkat resiko penyakit diabetes melitus dengan 13 atribut dan 5 cluster yang dimana menghasilkan bahwa penderita yang paling berpengaruh terhadap penyakit diabetes melitus adalah cluster 1 dengan 31 anggota kelompok sebagai prioritas pertama, cluster 4 dengan 28 anggota kelompok sebagai prioritas kedua, cluster 3 dengan 21 anggota kelompok sebagai prioritas ketiga, cluster 5 dengan 15 anggota kelompok sebagai prioritas keempat, dan cluster 2 dengan 6 anggota kelompok sebagai prioritas terakhir (Rustandi, 2020).

Ade Bastian dkk melakukan penggunaan algoritma *k-means cluster analysis* dalam studi kasus menular pada suatu objek yang berhasil mengelompokkan objeknya dengan cepat dan sehingga mempercepat proses pengelompokkan namun sensitif pada penbangkitan titik pusat dan berakibat ketidak akuratan (Bastian et al., n.d.).

Sofia Defiyanti mengintegrasikan metode *clustering* dengan metode klasifikasi untuk data numerik yang menghasilkan peningkatan setelah dilakukan integrasi metode klasifikasi dan clustering dengan hasil yang baik dan nilai akurasi, waktu yang digunakan untuk pembangunan model, RMSE dan ROC *curve* mengalami peningkatan mencapai lebih dari 50%(Defiyanti, 2017).

Andi Maulida Argina melakukan pengukuran performa metode klasifikasi dalam mengelolah data set penderita diabetes yang menghasilkan akurasi sebesar 39% yang dimana nilai yang dihasilkan masih tidak cukup baik karena jumlah data yang digunakan cukup kecil(Argina, 2020).

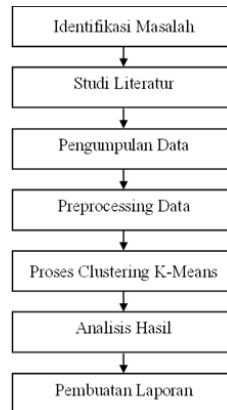
Muhammad Ezar Al Rivan dkk menerapkan algoritma *genetic* untuk optimasi *fuzzy c-means* dan *k-means* dalam melakukan pengklasteran data set yang menghasilkan pada *fuzzy c-means* terjadinya penurunan iterasi sebanyak 61,11% sedangkan untuk optimasi pada *k-means* yaitu 32,25%(Al Rivan et al., 2020).

Berdasarkan penjelasan tersebut, maka menggunakan Algoritma K-Means untuk mengelompokkan faktor resiko penyakit diabetes melitus dan melakukan diagnosa yang diharapkan mampu mendapatkan nilai yang optimal dan hasil akurasi yang baik. Hal ini jugalah yang melatar belakangi peneliti untuk melakukan penelitian dengan judul “**Penerapan Metode K-Means Clustering Untuk Menentukan Faktor Resiko Pada Penderita Diabetes Melitus**”

2. METODE PENELITIAN

Tahapan Penelitian

Pada penelitian dibutuhkan kerangka kerja penelitian yang berisi langkah-langkah terstruktur agar penelitian dapat berjalan dengan baik. Adapun kerangka kerja penelitian dapat dilihat pada gambar 1 :



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan Gambar 1, Berikut ini merupakan penjelasan dari setiap bagian yang tercantum pada kerangka kerja penelitian:

Identifikasi Masalah

Tahapan ini akan melakukan identifikasi dengan melihat data penderita diabetes melitus setelah itu dapat menentukan rencana kerja dan menentukan data apa saja yang dibutuhkan dalam penelitian ini.

Studi Literatur

Pada tahapan ini akan melakukan kajian pustaka dengan mempelajari buku-buku referensi, artikel-artikel, dan hasil penelitian sejenis yang relevan dengan permasalahan yang sedang diteliti. Studi literatur ini bertujuan untuk mendapatkan landasan teoritis mengenai permasalahan yang akan diteliti seperti teknik data mining, proses klasterisasi menggunakan algoritma *k-means*, *tool rapidminer* yang akan digunakan, data penderita penyakit diabetes melitus. Hal ini bertujuan agar peneliti dapat memahami permasalahan yang diteliti dengan benar dan sesuai dengan pembahasan yang dilakukan.

Pengumpulan Data

Dalam tahapan pengumpulan data yang dibutuhkan dalam penelitian ini yaitu data penderita diabetes melitus yang diambil dari dataset yang bersumber dari <https://www.kaggle.com/> data terkait dengan penyakit diabetes melitus dengan nama dataset BRFSS2015. *Behavioral Risk Factor Surveillance System (BRFSS)* yang merupakan data

survey telepon terkait kesehatan terutama diabetes melitus yang dikumpulkan setiap tahun oleh CDC dengan berisikan tanggapan sebanyak 70.692 dan 23 variabel fitur yang telah dilakukan pembersihan menggunakan teknik machine learning dengan memiliki 2 kelas yaitu tanpa diabetes atau prediabetes dan diabetes serta datasetnya yang seimbang, data ini disebut sebagai data sekunder. Data sekunder merupakan data yang telah diolah lebih lanjut dan disajikan dengan baik oleh pihak pengumpul data primer atau pihak lain yang dapat diperoleh melalui media perantara atau tidak langsung yang dapat berupa buku, catatan, bukti yang telah ada, maupun dari arsip yang telah dipublikasikan secara umum.

Data sekunder yang dikumpulkan akan dilakukan penyeleksian data dengan mengambil 593 data dan 16 atribut sebagai sampel data. Dalam penyeleksian dataset ini akan dilakukan dengan metode clustering karena pada dasarnya clustering merupakan metode menganalisis pengelompokan terhadap suatu objek yang digunakan sesuai dengan tujuan analisis data.

Pre-Processing

Pada tahap ini peneliti melakukan persiapan data penyakit diabetes melitus dengan melakukan seleksi atribut yang berpengaruh dalam klasterisasi untuk menentukan faktor resiko pada penderita diabetes melitus.

1) Represtasi Data

Pada tahap ini peneliti melakukan persiapan data penyakit diabetes melitus dengan melakukan seleksi atribut yang berpengaruh dalam klasterisasi untuk menentukan faktor resiko pada penderita diabetes melitus.

2) Data Cleaning

Tahapan ini melakukan pembersihan atau pemisahan data dari noise data dan data yang tidak konsisten. Kemudian hasil dari cleaning data akan disusun dalam bentuk csv atau arff.

3) Data Transformation

Transformasi merupakan metode yang digunakan dalam cleaning data pada tahap ini data yang telah dipisah dan dipilih lalu diubah atau diinisialisasikan data dari data berbentuk nominal ke dalam bentuk data angka.

4) Data Selection

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam data mining dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis

data operasional. Pada tahapan ini melakukan seleksi data penderita diabetes melitus yang mempunyai 70.692 data diambil 593 data dan 16 atribut sebagai data sampel.

Proses Clustering K-Means

Pada tahapan ini penulis melakukan dari perhitungan manual metode clustering k-means dan perhitungan menggunakan beberapa tools.

Analisis Hasil

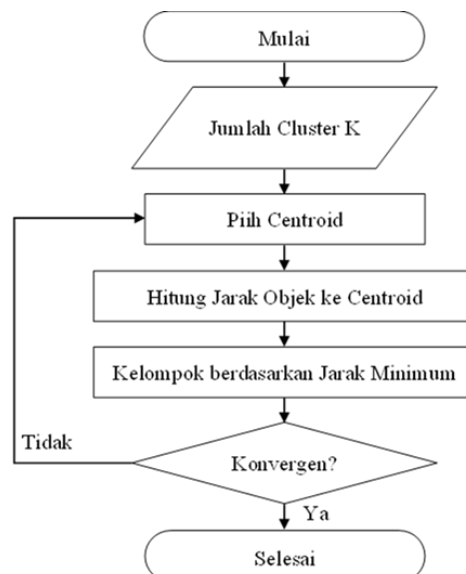
Pada tahapan ini akan dilakukan penganalisan data penderita penyakit diabetes melitus hasil dari perhitungan menggunakan metode k-means clustering dan mengelolah data diabetes melitus yang kemudian akan dijadikan penentuan faktor resiko pada penderita, hasil analisis perhitungan penulis dengan hasil perhitungan menggunakan tools RapidMiner, SPSS, dan Phyton.

Pembuatan Laporan

Pada tahapan pembuatan laporan akhir penelitian dilakukan berdasarkan kerangka yang telah dirancang dimulai dari tahapan identifikasi masalah, studi literatur terkait masalah yang diteliti, pengumpulan data, analisis prediktif yang terdiri dari data selection, data preprocessing, data transformation, menganalisis pengelompokan untuk penderita penyakit diabetes melitus dan pattern evaluation, kemudian dilakukan analisis hasil keakuratan data dengan clustering. Pembuatan laporan yang bertujuan untuk sebagai dokumentasi dari penelitian yang telah dilakukan.

Tahapan Clustering K-Means

Berikut adalah langkah-langkah dari tahapan klasterisasi menggunakan algoritma K-means.



Gambar 2. Tahapan clustering K-means

Berdasarkan Gambar 2, tahapan-tahapan dalam menggunakan algoritma k-means adalah sebagai berikut:

1. Langkah 1: Tentukan jumlah clustering yang diinginkan (misalkan: k4).
2. Langkah 2: Pilih centroid awal secara acak.
3. Langkah 3: Hitung jarak dengan centroid.
4. Langkah 4: Pembaharuan centroid dengan menghitung rata-rata nilai pada masing-masing cluster, setelah itu maka akan didapatkan centroid baru.
5. Langkah 5: (iterasi ke-2) Kembali kelangkah ke-3, jika masih ada data yang berpindah cluster atau jika nilai centroid diatas ambang, atau jika fungsi objektif yang digunakan masih diatas ambang. Jika tidak maka iterasi dihentikan.

3. HASIL DAN PEMBAHASAN

Pre-Processing

Data yang digunakan berupa data diabetes melitus dengan jumlah data sebanyak 70.692 data dengan 23 atribut. Tabel data diabetes melitus dapat dilihat pada Tabel 1.

Table 1. Data penderita diabetes melitus

No	ID	Diabetes_Binary	Hg@BP	Hg@Chol	Ch@Chol	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	any@healthcare	Nob@doctor	Gen@hth	Ment@hth	Phys@hth	Diff@Walk	Sex	Age	Education	Income
1	00001	No	Yes	No	Yes	26	No	No	No	Yes	No	Yes	No	Yes	No	Good	5	30	No	M	35	6	8
2	00002	No	Yes	Yes	Yes	26	Yes	Yes	No	No	Yes	No	No	No	Yes	Good	0	0	No	M	77	6	8
3	00003	No	No	No	Yes	26	No	No	No	Yes	Yes	Yes	No	Yes	No	Excellent	0	10	No	M	81	6	8
...
70690	70690	Yes	Yes	Yes	Yes	25	No	No	Yes	No	Yes	No	No	Yes	No	Poor	15	0	Yes	F	81	6	4
70691	70691	Yes	Yes	Yes	Yes	18	No	No	No	No	No	No	No	Yes	No	Fair	0	0	Yes	F	71	2	4
70692	70692	Yes	Yes	Yes	Yes	25	No	No	Yes	Yes	Yes	No	No	Yes	No	Very Good	0	0	No	F	60	6	2

Data diabetes melitus pada Tabel 1. akan dilakukan transformasi data dengan menginisialisasi data dari data nominal ke dalam bentuk angka. Berikut transformasi teks ke dalam angka pada penyakit diabetes melitus yang dapat dilihat pada Tabel 2.

Table 2. Data Awal yang Telah Ditransformasi

No	ID	Diabetes_Binary	Hg@BP	Hg@Chol	Ch@Chol	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	any@healthcare	Nob@doctor	Gen@hth	Ment@hth	Phys@hth	Diff@Walk	Sex	Age	Education	Income
1	00001	No	Yes	No	Yes	26	No	No	No	Yes	No	Yes	No	Yes	No	Good	5	30	No	M	35	6	8
2	00002	No	Yes	Yes	Yes	26	Yes	Yes	No	No	Yes	No	No	No	Yes	Good	0	0	No	M	77	6	8
3	00003	No	No	No	Yes	26	No	No	No	Yes	Yes	Yes	No	Yes	No	Excellent	0	10	No	M	81	6	8
...
70690	70690	Yes	Yes	Yes	Yes	25	No	No	Yes	No	Yes	No	No	Yes	No	Poor	15	0	Yes	F	81	6	4
70691	70691	Yes	Yes	Yes	Yes	18	No	No	No	No	No	No	No	Yes	No	Fair	0	0	Yes	F	71	2	4
70692	70692	Yes	Yes	Yes	Yes	25	No	No	Yes	Yes	Yes	No	No	Yes	No	Very Good	0	0	No	F	60	6	2

Setelah dilakukan transformasi selanjutnya data akan di seleksi dengan memfokuskan pada atribut yang akan digunakan dalam perhitungan yaitu ID, *Diabetes binary*, *high blood pressure*, *High Cholesterol*, *BMI (Body Mass Index)*, *Smoker*, *Stroke*, *Heart Diseaseor Attack*, *Phys Activity*, *Hvy Alcohol Consumption*, *General Health*, *Mental Health*, *Physical Health*, *Difficulty Walk*, *Sex*, dan *Age*.

Table 3. Data Diabetes Melitus Telah Diseleksi

No	ID	Diabetes_binary	HighBP	HighChol	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	HvyAlcoholConsump	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age
1	00001	0	1	0	26	0	0	0	1	0	3	5	30	0	1	35
2	00002	0	1	1	26	1	1	0	0	0	3	0	0	0	1	77
3	00003	0	0	0	26	0	0	0	1	0	1	0	10	0	1	81
...
70690	70690	1	1	1	25	0	0	1	0	0	5	15	0	1	0	81
70691	70691	1	1	1	18	0	0	0	0	0	4	0	0	1	0	71
70692	70692	1	1	1	25	0	0	1	1	0	2	0	0	0	0	60

Perhitungan Clustering K-Means

Perhitungan clustering k-means dibagi menjadi beberapa kelompok perhitungan yaitu perhitungan secara manual dan menggunakan tools data mining seperti RapidMiner, SPSS, dan Phytion.

Perhitungan Manual Clustering K-Means

Data yang awalnya 70.692 data dengan 16 atribut akan diseleksi untuk dijadikan sebagai data sampel, yang dimana teknik yang digunakan berupa teknik pengambilan sampel dengan taraf keyakinan penelitian 95% dan kesalahan toleransi nya yaitu 5% dan 10 % . beberapa teknik yang digunakan yaitu rumus slovin, rumus isaac & michael, rumus krejcie & morgan, rumus cohen manion & marrison yang dimana dari ke 4 teknik ini untuk dijadikan data sampel dapat disimpulkan pada Tabel 4. Sebagai berikut :

Table 4. Menentukan Jumlah Sampel

Teknik Pengambilan Sampel	Jumlah data	
	10%	5%
Slovin	100	398
Isaac & Michael	270	348
Krejcie & Morgan	-	381
Cohen Manion & Marrison	381	593

Pada tabel manion & marrison memperlihatkan jumlah data paling banyak sehingga akan digunakan sebagai pengambilan data sampel sehingga data yang digunakan sebanyak 593 data dengan 16 atribut yang akan dilakukan perhitungan secara manual. Hasil dari perhitungan manual dengan menggunakan data sampel dilihat pada Tabel 5.

Table 5. Hasil Manual (Excel)

Jumlah Iterasi	Jumlah Cluster	Jumlah data tiap cluster	Centroid	Nilai ratio	Performa
9	C1	432	4	73%	0,892
	C2	35		6%	
	C3	43		7%	
	C4	83		14%	

Perhitungan Tools Clustering K-Means

Pada penelitian ini menggunakan beberapa tools data mining yaitu RapidMiner, SPSS, dan Phytion untuk mengklasterisasi penderita diabetes melitus. Hasil komperasi kinerja dari masing-masing tools tersebut ditampilkan pada Tabel 6.

Table 6. Hasil Tools

Perbandingan	Data	Jumlah Iterasi	Jumlah Cluster	Jumlah Data tiap Cluster	Centroid	Nilai Ratio	DBI
Perhitungan RapidMiner	Data Set (70.692)	9	C1	10.389	4	15%	1,027
			C2	4.749		7%	
			C3	41.155		58%	
			C4	14.399		20%	
		Max	C1	38.623		55%	1,038
			C2	16.832		24%	
			C3	8.672		12%	
			C4	6.565		9%	
Perhitungan SPSS	Data Set (70.692)	9	C1	12.034	4	17%	3,892
			C2	24.147		34%	
			C3	20.269		29%	
			C4	14.242		20%	
		Max	C1	12.034		17%	3,892
			C2	24.147		34%	
			C3	20.269		29%	
			C4	14.242		20%	
Perhitungan Python	Data Set (70.692)	9	C1	38.619	4	55%	1,308
			C2	6.566		9%	
			C3	16.832		24%	
			C4	8.675		12%	
		Max	C1	38.632		55%	1,309
			C2	8.669		12%	
			C3	16.834		24%	
			C4	6.566		9%	

Setelah dilakukan perhitungan dari masing-masing tools, maka diperoleh performa terbaik pada tools RapidMiner sebesar 1,027 pada jumlah iterasi 9 dengan prioritas pertama pada cluster 3 dengan nilai rasionya 58% yang ditampilkan pada Tabel 6.

4. KESIMPULAN DAN SARAN

Penelitian ini menggunakan perhitungan secara manual dengan data sampel berjumlah 593 data penderita diabetes melitus sedangkan perhitungan dengan bantuan *tools* menggunakan dataset berjumlah 70.693 data penderita diabetes melitus dengan atribut yang digunakan 15 atribut yaitu ID, *Diabetes binary*, *high blood pressure*, *High Cholesterol*, *BMI (Body Mass Index)*, *Smoker*, *Stroke*, *Heart Disease or Attack*, *Phys Activity*, *Hvy Alcohol Consumption*, *General Health*, *Mental Health*, *Physical Health*, *Difficulty Walk*, *Sex*, dan *Age*. Dari hasil perhitungan, perhitungan manual memiliki performa terbaik sebesar 0,892 namun data yang digunakan hanya data sampel sedangkan untuk operforma terbaik dengan dataset adalah rapidminer dengan performanya 1,027 yang terletak pada 9 iterasi dengan nilai ratio tertingginya 58% yaitu cluster 3 yang menjadi prioritas pertama.

DAFTAR REFERENSI

- Al Rivan, M. E., Steven, S., & Tanzil, W. (2020). Optimasi Fuzzy C-Means dan K-Means Menggunakan Algoritma Genetika untuk Pengklasteran Dataset Diabetic Retinopathy. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(5), 993. <https://doi.org/10.25126/jtiik.2020711872>
- Apa Saja Tipe Penyakit DM?* (2018). P2PTM Kemenkes Ri.
- Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29–33. <https://doi.org/10.33096/ijodas.v1i2.11>
- Bastian, A., Sujadi, H., & Febrianto, G. (n.d.). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)*. 1, 26–32.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. Deepublish.
- Danny, M., & Yusup, D. (2020). ANALISIS DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES DALAM MEMPREDIKSI PEMBELIAN MATERIAL PLASTIK INJECTION (STUDI KASUS: PT SURYA TECHNOLOGY INDUSTRI). Vol. 11 No. (2407–3903).
- Defiyanti, S. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. *Citee*, July, 256–261.
- Faktor Risiko Penyakit Diabetes Melitus (DM)-*. (2019). P2PTM Kemenkes RI.
- Indriyani, N. P. A. (2021). GAMBARAN KADAR GULA DARAH PADA PASIEN DIABETES MELITUS DI BRUS TABANAN TAHUN 2021.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi*. Deepublish.
- Nasution, F., Andilala, A., & Siregar, A. A. (2021). FAKTOR RISIKO KEJADIAN DIABETES MELLITUS.
- Paramitasari, M. (2017). HUBUNGAN CEMAS DENGAN KADAR GULA DARAH PADA PENDERITA DIABETES MELITUS TIPE II DI DESA MORODEMAK.
- Prasetyo, E. (2012). Data Mining-Konsept dan Aplikasi menggunakan Matlab. *Nikodemus WK, Ed*.
- Rustandi, A. (2020). PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENENTUKAN PENDERITA DIABETES. Universitas Dinamika Bangsa.
- Suntoro, J. (2019). DATA MINING: Algoritma dan Implementasi dengan Pemrograman php. PT Elex Media Komputindo.
- Susanti, Y. H., & Widodo, E. (2017). Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015. *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai Islami)*, 1(1), 116–122.
- Wicaksono, A. P. (2015). Pengaruh Pemberian Ekstrak Jahe Merah (Zingiber Officinale) terhadap Kadar Glukosa Darah Puasa. *Majority*, 4(7), 97–102.