

IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PENYEBARAN PENYAKIT DIABETES MELITUS DI KECAMATAN BAGOR

Roro Niken Enggar Tiasti¹, Abdulloh Hamid², Dian Candra Rini Novitasari³

¹UIN Sunan Ampel Surabaya, roroniken30@gmail.com

²UIN Sunan Ampel Surabaya, doelhamid@uinsby.ac.id

³UIN Sunan Ampel Surabaya, diancrini@uinsby.ac.id

Abstrak. Diabetes melitus merupakan penyakit yang ditandai dengan gangguan seleksi insulin, kerja insulin atau keduanya. Diabetes dijuluki dengan julukan “*mother of disease*” karena bisa menjadi penyebab timbulnya penyakit komplikasi lain. Penyakit diabetes melitus tersebar luas di Kecamatan Bagor Kabupaten Nganjuk. Untuk melihat kawasan penyebaran yang paling tinggi pada Kecamatan Bagor perlu dibuat sebuah pengelompokan daerah untuk mengetahui tingkat keparahan penyakit diabetes melitus yaitu dengan menggunakan Algoritma *K-Means Clustering*. Algoritma *K-Means Clustering* merupakan salah satu algoritma yang paling umum digunakan untuk mengelompokkan data yang sama ke dalam kelompok tertentu dan data yang berbeda ke dalam kelompok yang lain. Tujuan dari penelitian ini untuk mengelompokkan penyebaran diabetes melitus dan pusat penyebarannya. Hasil dari penelitian ini berupa pengelompokan daerah penyebaran penyakit diabetes melitus di Kecamatan Bagor diperoleh pusat *cluster* 1 dengan penyebaran tingkat tinggi sebanyak 6 desa meliputi desa Balongrejo, Ngumpul, Bagorkulon, Selorejo, Banarankulon, Banaranwetan, dan pusat *cluster* 0 dengan penyebaran tingkat rendah sebanyak 15 desa.

Kata kunci: Diabetes melitus, *K-Means Clustering*

Abstract. Diabetes mellitus is a disease characterized by impaired insulin selection, insulin action or both. Diabetes is nicknamed the “*mother of disease*” because it can cause other complications. Diabetes Mellitus is widespread in Bagor District, Nganjuk Regency. To see the area with the highest distribution in the District of Bagor, it is necessary to make a regional grouping to determine the severity of diabetes mellitus by using the *K-Means Clustering* Algorithm. The *K-Means Clustering* algorithm is one of the most commonly used algorithms to group the same data into certain groups and different data into other groups. The purpose of this study was to classify the spread of diabetes mellitus and its distribution center. The results of this study are grouping the distribution areas of diabetes mellitus in Bagor District. The *cluster* 1 center with a high level of distribution is 6 villages including Balongrejo, Ngumpul, Bagorkulon, Selorejo, Banarankulon, Banaranwetan, and *cluster* center 0 with a low level distribution of 15 villages.

Keywords: Diabetes mellitus, *K-Means Clustering*

1. Pendahuluan

Menurut laporan tahun 2010 oleh *American Diabetes Association* (ADA), diabetes melitus merupakan sekelompok penyakit metabolik yang ditandai dengan hiperglikemia, yang diakibatkan oleh sekresi insulin, kerja insulin, atau bahkan keduanya [1]. Diabetes melitus (DM) umumnya ditemukan di masyarakat, baik di Indonesia maupun negara-negara lain di dunia. Berdasarkan data dari *World Health Organization* (WHO) pada tahun 2012 jumlah keseluruhan kasus kematian akibat DM sebanyak 1,5 juta dan lebih dari 80 persen berasal dari negara dengan jumlah penghasilan rendah hingga menengah. Berdasarkan data dari Riskesdas tahun 2007, prevalensi kematian akibat DM pada kelompok usia 45-54 tahun di wilayah perkotaan berada di peringkat ke-2 yaitu sebanyak 14,7 persen dan wilayah pedesaan berada pada peringkat ke-6 yaitu sebanyak 5,8 persen [2]. Hasil Riset Kesehatan Dasar (Riskesdas) menunjukkan bahwa prevalensi diabetes melitus di Indonesia menurut diagnosis dokter sebesar 2 persen pada usia 15 tahun. Angka ini menunjukkan peningkatan dibandingkan 1,5 persen prevalensi diabetes pada usia 15 tahun pada hasil Riskesdas 2013. Namun menurut hasil pemeriksaan glukosa darah, prevalensi diabetes meningkat dari 6,9 persen pada tahun 2013 menjadi 8,5 persen pada tahun 2018. Angka ini menunjukkan bahwa hanya sekitar 25 persen pasien diabetes yang tahu bahwa mereka menderita diabetes [3]. Dari data yang telah diketahui diatas bahwa angka penderita diabetes melitus masih relatif tinggi dan akan terus meningkat dari tahun ke tahun. Ini menjadi masalah serius yang disebabkan oleh peningkatan jumlah kasus dan kematian [4]. Diabetes dijuluki dengan julukan “*mother of disease*” karena bisa menjadi penyebab timbulnya penyakit komplikasi lain seperti tekanan darah tinggi, penyakit jantung, pembuluh darah, stroke, gagal ginjal bahkan kebutaan dan berbagai gangguan serta kegagalan kerja organ, terutama mata, ginjal, saraf, jantung dan pembuluh darah. Diabetes mellitus adalah penyakit yang tidak bisa disembuhkan tetapi dapat dikendalikan [5].

Pada penelitian ini digunakan teknik pemrosesan dengan *data mining* yang merupakan pemrosesan dari data yang ada menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang digunakan untuk mengenali dan mengekstrak informasi yang berguna dalam berbagai pengetahuan yang terkait dengan basis data berkapasitas besar. Salah satu metode termasuk dalam data mining yaitu *clustering* [6], dengan memainkan peran penting dalam mengeksplorasi objek yang memiliki kesamaan karakteristik tertentu, yang kemudian dari karakteristik tersebut digunakan sebagai “*centroid*” dengan menggunakan algoritma k-means [7]. K-Means adalah algoritma yang digunakan untuk mengelompokkan data berdasarkan partisi dengan membagi data ke dalam kelompok yang berbeda. Algoritma ini dapat meminimalkan jarak dari data ke *cluster*. K-Means dapat membagi data menjadi *cluster*/grup, sehingga data dengan karakteristik yang sama akan dikelompokkan ke dalam *cluster* yang sama, dan data dengan karakteristik yang berbeda akan dikelompokkan ke dalam grup yang lain [8].

Pengelompokan kasus penyakit menggunakan metode *K-Means Clustering* telah dilakukan oleh beberapa peneliti seperti penelitian yang dilakukan oleh Nasari dan Sianturi pada tahun 2016 yang membahas tentang Penerapan Algoritma *K-Means Clustering* untuk Pengelompokan Penyebaran Diare di Kabupaten Langkat. Hasil penelitian tersebut menyebutkan bahwa pusat *cluster* 1 merupakan daerah dengan tingkat penyebaran penyakit diare tingkat rendah terdapat 3 kecamatan dan *cluster* 2 merupakan daerah pusat penyebaran penyakit diare dengan tingkat tinggi terdapat 3 kecamatan [9]. Adapun penelitian yang dilakukan oleh Shelly Maulina, Budi Serasi & Anton Sihombing pada tahun 2021 tentang Implementasi Data Mining Pengelompokan Jenis Penyakit Pasien Menggunakan Metode *Clustering* (Studi Kasus: Puskesmas Sambirejo). Hasil dari penelitian tersebut yaitu kelompok paling tinggi dan paling sering dialami oleh pasien adalah kelompok data pasien yaitu *cluster* 2 berjumlah 376 pasien pada umur 20-40 tahun, di desa perdamaian dengan nama penyakit hiperurisemia [10]. Penelitian lain juga dilakukan oleh Abdallah

Moubayed dkk pada tahun 2020 yang membahas tentang Tingkat Keterlibatan Siswa dalam *E-Learning* Lingkungan: Pengelompokan Menggunakan K-Means untuk mengelompokkan siswa berdasarkan 12 keterlibatan metrik dibagi menjadi 2 *cluster* atau 2 kategori yaitu interaksi terkait dan upaya terkait dengan hasil penelitian bahwa di antara dianggap antar metrik terkait tindakan dan terkait upaya, jumlah login, dan durasi rata-rata untuk mengirimkan tugas adalah yang paling mewakili tingkat keterlibatan siswa sementara jumlah konten yang dibaca/diakses tidak [11]. Penelitian yang dilakukan oleh Allwin dkk pada tahun 2021 yang membahas tentang Pengelompokan Diagnosis Diabetes Berdasarkan Rentang Usia dengan Algoritma K-Means untuk didiagnosis dengan diabetes tipe I atau tipe II dengan sampel pasien terdiagnosa diabetes tahun 2019 di RSUD Royal Prima Medan sebanyak 200 data menunjukkan bahwa *cluster* 1 merupakan kelompok pasien DM Tipe I pada usia 43-47, 56, 62, 67, dan 68 Tahun dan *cluster* 2 merupakan kelompok pasien DM Tipe II pada usia 49, 55, dan 63 Tahun [12].

Selain data mining, cabang ilmu lainnya yang memiliki metode dalam pengelompokan data yaitu logika fuzzy. Logika fuzzy dapat diterapkan pada pengelompokan data atau *clustering* yang disebut dengan *Fuzzy Clustering*. *Fuzzy Clustering* merupakan teknik untuk menemukan grup yang optimal dalam ruang vektor. Ada beberapa algoritma pengelompokan data fuzzy, salah satunya adalah *Fuzzy C-Means Clustering*. *Fuzzy C-Means Clustering* merupakan teknik pengelompokan untuk menentukan setiap titik data dalam *cluster* ada menurut derajat keanggotaan. Output dari *Fuzzy C-Means Clustering* adalah rangkaian dari beberapa derajat keanggotaan dan pusat *cluster* dari setiap titik data [13].

Pada penelitian yang dilakukan, peneliti menggunakan algoritma K-Means karena algoritma ini lebih efisien dan efektif dalam kasus pengelompokan, karena menghasilkan *cluster* dengan jumlah yang telah ditetapkan. Selain kecepatan pada proses pengelompokan dengan probabilitas *cluster* yang tinggi dan juga memiliki tingkat ketelitian yang cukup tinggi terhadap ukuran suatu objek [14]. Peneliti menggunakan model 2 *cluster* dalam pengelompokan. Dengan judul penelitian “Implementasi Algoritma K-Means *Clustering* untuk Mengelompokkan Kasus Penyakit Diabetes Melitus di Kecamatan Bagor”. Peneliti mengambil studi kasus di kecamatan Bagor Kabupaten Ngunjuk karena penyakit diabetes melitus tersebar merata bahkan menjadi salah satu penyebab kematian terbanyak pada Kecamatan Bagor. Tujuan penelitian ini untuk memahami penyebaran penyakit diabetes melitus di Kecamatan Bagor. Diharapkan penelitian ini dapat memberikan informasi tentang pengelompokan kasus diabetes kepada pemerintah dan masyarakat melalui kegiatan sosialisasi dan diharapkan pemerintah segera memberikan penanganan terhadap kasus diabetes melitus ini, sehingga dapat menekan angka kasus diabetes pada tahun-tahun yang akan datang.

2. Kajian Teori

2.1 Diabetes Melitus

Menurut WHO (*World Health Organization*) Diabetes Mellitus merupakan epidemi penyakit tidak menular tertentu yang menyebar dengan cepat di seluruh dunia dan ditandai dengan kecacatan dini dan kematian akibat komplikasi vaskular [15]. Penyakit diabetes melitus ditandai tingkat gula darah yang tinggi yaitu sama dengan atau lebih besar dari 200 mg/dl, karena kadar glukosa darah berfluktuasi sebelum dan sesudah makan [16].

Diabetes dapat dibagi menjadi empat kategori berikut: (a) Diabetes Tipe 1, (b) Diabetes Tipe 2, (c) Diabetes Gestasional, yang terjadi pada wanita hamil tetapi biasanya menghilang setelah kehamilan, (d) lainnya karena penyakit genetik, penyalahgunaan obat dan penyakit terkait penyakit. Dua bentuk khas diabetes adalah *Insulin Dependent Diabetes Mellitus (IDDM)*, biasa disebut sebagai diabetes Tipe 1 dan *Non Insulin Dependent Diabetes Mellitus (NIDDM)*, biasa disebut sebagai diabetes Tipe 2. Pada diabetes Tipe 2, individu biasanya tidak memerlukan suntikan insulin

karena dapat memproduksi insulin, tetapi tubuh tidak dapat menggunakannya secara efektif. Pada diabetes Tipe 1, individu memang membutuhkan suntikan insulin karena tubuh mereka tidak dapat memproduksi insulin. Diabetes Tipe 2 lebih umum, terhitung lebih dari 90 persen kasus diabetes di Amerika Serikat [16].

Diabetes melitus dapat diidentifikasi dengan melakukan tes toleransi glukosa oral. Tes toleransi glukosa oral adalah tes yang lebih sensitif daripada tes toleransi intravena, hanya digunakan dalam situasi tertentu (misalnya, untuk pasien dengan riwayat penyakit hingga menjalani operasi pada lambung). Tes ini digunakan dengan memberikan larutan karbohidrat sederhana. WHO merekomendasikan pengambilan sampel beberapa jam atau minimal 2 jam setelah konsumsi glukosa. Rekomendasi dari Kelompok Data Diabetes Nasional ini juga termasuk pengumpulan sampel darah 30 dan 60 menit setelah konsumsi glukosa [17].

Tes toleransi glukosa oral tidak dapat mengkonfirmasi apakah orang itu mengidap diabetes atau tidak, tanpa ada gejala yang lebih spesifik, untuk itu perlu adanya konfirmasi gejala dan cek gula darah lagi pada kesempatan yang lain. Diabetes melitus dibedakan menjadi 2 tipe yaitu Diabetes tipe 1 atau *Insulin Dependent Diabetes Mellitus* (IDDM) yang merupakan penyakit diabetes yang diakibatkan oleh penyakit autoimun, yang mengakibatkan sistem imun akan menyerang sel beta pankreas yang memproduksi insulin. Diabetes melitus Tipe 2 atau *Non Insulin Dependent Diabetes Mellitus* (NIDDM) yang merupakan penyakit diabetes yang disebabkan oleh penolakan tubuh terhadap insulin, lemak tubuh, hati dan sel-sel otot. Sel-sel ini tidak dapat menyerap glukosa sebagai sumber energi tubuh yang akan mempengaruhi resistensi seiring bertambahnya umur [1].

Gejala penderita diabetes melitus yaitu memiliki rasa haus dan rasa lapar secara berlebihan, lebih sering buang air kecil dan menurunnya berat badan penderita secara drastis. Untuk tipe 1 penyakit diabetes melitus biasanya gejala muncul sebelum usia 30 tahun, meskipun gejala bisa timbul kapan saja penderita dengan tipe 1 ini memerlukan terapi yaitu pengobatan dengan insulin. Untuk tipe 2 penyakit diabetes melitus biasanya terjadi pada pasien usia lanjut yaitu 30 tahun atau lebih, dan pengobatan yang tepat yaitu diberikan obat oral. Jenis Diabetes yang lainnya adalah diabetes gestasional. Diabetes ini terjadi pada ibu hamil yang disebabkan oleh gangguan toleransi glukosa pada pasien ini. Biasanya terjadi pada ibu hamil pada minggu ke-24 atau bulan ke-6 [18].

2.2 K-Means Clustering

Clustering adalah teknik statistik yang dirancang untuk mengelompokkan objek-objek ke dalam suatu kelompok sehingga objek-objek dalam satu kelompok memiliki kemiripan yang lebih tinggi dibandingkan dengan objek-objek dalam kelompok lain. Tujuan dari *clustering* adalah untuk mengklasifikasikan objek berdasarkan kemiripannya dan mengumpulkan data ke dalam beberapa kelompok [19]. Teknik *clustering* memiliki dua metode pengelompokan, yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *Hierarchical Clustering* adalah suatu metode pengelompokan data yang prinsip kerjanya terstruktur dan bertahap dengan mengelompokkan dua atau lebih data yang mirip, kemudian dilanjutkan dengan proses objek lain yang dengan kedekatan yang kedua. Kemiripan sifat tersebut dapat ditentukan dengan dihitung jarak *Euclidean* atau jarak *Mahalanobis*. Proses ini berlanjut hingga *cluster-cluster* tersebut membentuk semacam pohon dimana terdapat hirarki atau tingkatan objek yang jelas mulai dari yang paling mirip sampai yang paling tidak mirip. Namun secara logika, semua objek pada akhirnya akan membentuk *cluster*. *Non-hierarchical clustering* umumnya digunakan ketika jumlah satuan dalam pengamatan besar dan jumlah *cluster* yang telah ditentukan sebelumnya [20] berdasarkan rata-rata (*mean*) yang paling dekat dengan *centroid* [21]. Salah satu metode ini juga biasa disebut sebagai metode *K-means Clustering*, metode ini mudah untuk diimplementasikan dan juga mudah diadaptasi sehingga lebih banyak digunakan dalam pengelompokan. Dalam teknik ini dimulai dengan menentukan jumlah *cluster* yang diinginkan kemudian proses *clustering* dimulai tanpa mengikuti proses hirarki [20].

K-Means clustering adalah algoritma yang membutuhkan k input untuk membagi n objek tersebut menjadi k *cluster*, sehingga kemiripan antar anggota dalam sebuah *cluster* sangat tinggi, dan kemiripan dengan anggota *cluster* lainnya sangat rendah. Kesamaan pengukuran antar anggota *cluster* diukur dari kedekatan objek dengan rata-rata *cluster* tersebut atau sering disebut sebagai *centroid cluster*. Dalam data mining metode K-Means memiliki kelebihan yaitu algoritma yang paling sederhana dan yang umum digunakan, dan kemampuannya untuk mengelompokkan jumlah data yang besar, waktu perhitungan yang relatif cepat dan efisien. Sedangkan kelemahan dari k-means yaitu dalam menganalisis distribusi data metode ini mengandalkan inisialisasi pusat. K-Means hanya melihat rentang data dari setiap *centroid* pada masing-masing *cluster* [22]. Kelebihan dari metode algoritma K-Means yaitu mampu mengklasifikasikan objek berukuran besar dengan cepat dan proses dari pengelompokan sangat mudah. Namun metode ini juga memiliki kekurangan yaitu metode ini masih cukup lemah dalam menentukan tempat penitipan anak *cluster random*, sehingga membuat hasil pengelompokan menjadi berubah-ubah [23].

Tahapan algoritma K-Means yaitu [24]:

1. Menentukan jumlah *cluster* yang ingin dibentuk.
2. Memilih *centroid* awal secara acak sebanyak angka *cluster*.
3. Menghitung jarak setiap data terhadap terhadap *centroid* awal menggunakan rumus *Euclidean Distance*, sebagai berikut:

$$J(a_i, b_i) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{1}$$

Keterangan:

$J(a_i, b_i)$ = jarak antara nilai data dan *centroid* ($i = 1, 2, 3, \dots, n$)

a_i = nilai data ukuran

b_i = nilai pusat dari *centroid*

n = banyaknya jumlah data

4. Menggabungkan data berdasarkan jarak terpendek dengan *centroid*.
5. Menentukan nilai *centroid* baru menggunakan rumus berikut:

$$W = \frac{1}{n_k} \sum_{i=1}^n j_i \tag{2}$$

Keterangan:

W = Nilai *centroid* baru

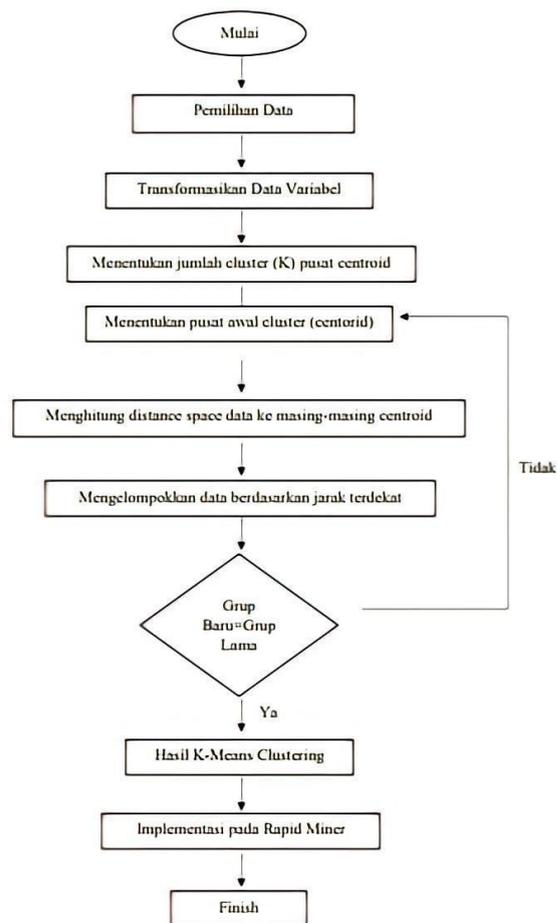
n_k = jumlah data di setiap *cluster*

j_i = Jumlah nilai jarak yang terdapat di setiap *cluster* ($i = 1, 2, 3, \dots, n$)

6. Lakukan tahap ketiga hingga tahap kelima sampai keanggotaan setiap *cluster* tetap.

3. Metode Penelitian

Pada penelitian ini dilakukan dengan menggunakan metode *K-Means Clustering* dimana data yang digunakan merupakan data sekunder berupa data jumlah penderita diabetes melitus pada tahun 2019 dan 2020 dan data jumlah dusun di setiap desa di Kecamatan Bagor yang diperoleh dari bagian rekam medis Puskesmas Bagor Kabupaten Nganjuk. Pada penelitian ini dilakukan dengan bantuan aplikasi *RapidMiner*. Diagram alir tahapan penelitian dapat dilihat pada gambar berikut:



Gambar 1: Diagram Alir Tahapan Penelitian

4. Hasil dan Pembahasan

Data dari penelitian yang sedang dilakukan diperoleh dari data PTM puskesmas Bagor Kabupaten Nganjuk pada tahun 2019 dan 2020 yaitu data penyakit diabetes melitus pada Kecamatan tersebut sebanyak 2666 data. Data yang diperoleh kemudian akan diolah menggunakan metode K-Means dengan mengambil nilai-nilai dari setiap atribut yang ada pada data, untuk mengelompokkan data penyakit diabetes melitus. Data penderita diabetes melitus di kecamatan Bagor dapat dilihat pada Tabel 1.

Tabel 1: Data Penderita Diabetes Melitus di Kecamatan Bagor

No.	Desa	Jumlah Dusun	Penderita Diabetes Melitus (2019)	Penderita Diabetes Melitus (2020)
1	Balongrejo	4	59	97
2	Girirejo	5	61	81
3	Pesudukuh	3	58	70
4	Sekar putih	2	67	55
5	Buduran	2	73	35
6	Ngumpul	8	87	121
7	Bagorkulon	5	37	91
8	Petak	2	36	49

9	Paron	2	32	50
10	Karangtengah	1	31	55
11	Selorejo	4	20	113
12	Gandu	4	33	64
13	Guyangan	2	31	45
14	Kedondong	2	31	47
15	Sugihwaras	2	83	78
16	Kutorejo	3	72	44
17	Kerepkidul	2	70	61
18	Gemenggeng	3	81	59
19	Kendalrejo	3	96	46
20	Banarankulon	4	53	133
21	Banaranwetan	5	57	104

Setelah itu data tersebut akan masuk ke tahap *clustering* dengan menerapkan algoritma K-Means untuk mengelompokkan data :

1. Menentukan Jumlah *Cluster*
 Penentuan jumlah *cluster* dilakukan sebagai inisialisasi/langkah awal pada metode ini, yang nantinya akan dibagi menjadi data yang akan dibuat atau dihasilkan. Pada penelitian ini jumlah *cluster* yang akan dibentuk adalah 2 *cluster*.
2. Penentuan *Centroid* Awal
 Pada penentuan *Centroid* awal dilakukan dengan mengambil data/nilai secara random. *Centroid* awal dari penelitian ini adalah:
 Maka diperoleh *Centroid* awal yang digunakan untuk perhitungan adalah:

Tabel 2: *Centroid* Awal

Data No.	Desa	Jumlah Dusun	Jumlah Penderita (2019)	Jumlah Penderita (2020)
6	Ngumpul	8	87	121
12	Gandu	4	33	64

Centroid-1 (8, 87, 121)

Centroid-2 (4, 33, 64)

3. Menghitung *Distance Space* Data ke Masing-masing *Centroid*
 Gunakan rumus *Euclidean Distance* untuk mengukur jarak ke pusat *cluster* seperti pada persamaan (1). Sebagai contoh, perhitungan jarak dari data ke-1 yaitu data desa Balongrejo terhadap pusat *cluster* adalah sebagai berikut :

$$C1 = \sqrt{(8 - 4)^2 + (87 - 59)^2 + (121 - 97)^2} = 37,09$$

$$C2 = \sqrt{(4 - 4)^2 + (33 - 59)^2 + (64 - 97)^2} = 42,01$$

Kemudian lanjutkan pada data kedua hingga data ke-n. Sehingga akan diperoleh matriks jarak dan pilihlah jarak terpendek dari masing-masing data. Hasil perhitungan pusat jarak *cluster* ditunjukkan pada Tabel 3.

Tabel 3: Perhitungan Jarak Antar *Cluster*

No.	Desa	C1	C2	Jarak Terdekat	Cluster
1	Balongrejo	37,09	42,01	37,09	1
2	Girirejo	47,80	32,77	32,77	2
3	Pesudukuh	58,88	25,73	25,73	2
4	Sekar putih	69,22	35,23	35,23	2
5	Buduran	87,34	49,45	49,45	2
6	Ngumpul	0	78,62	0	1
7	Bagorkulon	58,39	27,31	27,31	2
8	Petak	88,44	15,43	15,43	2
9	Paron	90,01	14,18	14,18	2
10	Karangtengah	86,84	9,70	9,70	2
11	Selorejo	67,59	50,70	50,70	2
12	Gandu	78,62	0	0	2
13	Guyangan	94,59	19,21	19,21	2
14	Kedondong	92,99	17,23	17,23	2
15	Sugihwaras	43,60	51,96	43,60	1
16	Kutorejo	78,61	43,84	43,84	2
17	Kerepkidul	62,65	37,18	37,18	2
18	Gemenggeng	62,49	48,27	48,27	2
19	Kendalrejo	75,70	65,53	65,53	2
20	Banarankulon	36,28	71,84	36,28	1
21	Banaranwetan	34,61	46,66	34,61	1

4. Mengelompokkan Data Berdasarkan Jarak Terdekat
 Setelah memilih jarak terpendek dari masing-masing data yang diperoleh, kemudian, kelompokkan data tersebut sesuai dengan posisi jarak terpendek dari masing-masing data. Hasil pengelompokan data dapat dilihat pada Tabel 4.

Tabel 4: Hasil Pengelompokan Data

No.	Desa	C1	C2	Grup Lama	Grup Baru
1	Balongrejo	I		0	1
2	Girirejo		I	0	2
3	Pesudukuh		I	0	2
4	Sekar putih		I	0	2
5	Buduran		I	0	2
6	Ngumpul	I		0	1
7	Bagorkulon		I	0	2
8	Petak		I	0	2
9	Paron		I	0	2
10	Karangtengah		I	0	2
11	Selorejo		I	0	2
12	Gandu		I	0	2
13	Guyangan		I	0	2
14	Kedondong		I	0	2

15	Sugihwaras	I	0	1
16	Kutorejo	I	0	2
17	Kerepkidul	I	0	2
18	Gemenggeng	I	0	2
19	Kendalrejo	I	0	2
20	Banarankulon	I	0	1
21	Banaranwetan	I	0	1

5. Jika grup lama dan grup baru memiliki nilai yang sama atau nilai *centroid* sudah optimal dan posisi *cluster* pada data tersebut tidak lagi berubah, maka proses iterasi dihentikan. Namun jika grup lama berbeda dengan grup baru, maka proses iterasi dilanjutkan dengan menentukan pusat *cluster* baru dengan menggunakan rumus yang sesuai dengan pusat untuk menghitung keanggotaan *cluster* berdasarkan data setiap *cluster* yang telah diperoleh dan proses iterasi akan terus berlanjut sampai *cluster* memiliki nilai yang sama. Untuk perhitungan pusat *cluster* baru menggunakan persamaan (2) sebagai berikut:

Untuk grup 1 = 5 Data

$$C1_1 = (4 + 8 + 2 + 4 + 5) : 5 = 4,60$$

$$C1_2 = (59 + 87 + 83 + 53 + 57) : 5 = 67,80$$

$$C1_3 = (97 + 121 + 78 + 133 + 104) : 5 = 106,60$$

Untuk grup 2 = 16 Data

$$C2_1 = (5+3+2+2+5+2+2+1+4+4+2+2+3+2+3+3):16 = 2,81$$

$$C2_2 = (61+58+67+73+37+36+32+31+20+33+31+31+72+70+81+96):16 = 51,81$$

$$C2_3 = (81+70+55+91+49+50+55+113+64+45+47+44+61+59+46):16 = 60,31$$

Berdasarkan hasil perhitungan di atas didapatkan nilai cluster baru 1 sebagai berikut:

Tabel 5: Pusat *Cluster* Baru 1

C1	4,60	67,80	106,60
C2	2,81	51,81	60,31

Tabel 6: Perhitungan Jarak Antar *Cluster* Iterasi 1

No.	Desa	C1	C2	Jarak Terdekat	<i>Cluster</i>
1	Balongrejo	13,04	37,41	13,04	1
2	Girirejo	26,49	22,74	22,74	2
3	Pesudukuh	37,92	11,50	11,50	2
4	Sekar putih	51,67	16,11	16,11	2
5	Buduran	71,84	33,02	33,02	2
6	Ngumpul	24,24	70,35	24,24	1
7	Bagorkulon	34,53	34,15	34,15	2
8	Petak	65,85	19,46	19,46	2
9	Paron	67,02	22,35	22,35	2
10	Karangtengah	63,48	21,55	21,55	2
11	Selorejo	48,23	61,56	48,23	1

12	Gandu	55,01	19,21	19,21	2
13	Guyangan	71,80	25,85	25,85	2
14	Kedondong	70,09	24,72	24,72	2
15	Sugihwaras	32,49	35,87	32,49	1
16	Kutorejo	62,76	25,96	25,96	2
17	Kerepkidul	45,73	18,22	18,22	2
18	Gemenggeng	49,42	29,22	29,22	2
19	Kendalrejo	66,86	46,45	46,45	2
20	Banarankulon	30,27	72,71	30,27	1
21	Banaranwetan	11,12	44,05	11,12	1

Berikutnya dilakukan kembali tahap-3 dan seterusnya. Pada iterasi ke-3, nilai *centroid* 1 dan *centroid* 2 sudah bernilai sama begitupun dengan grup lama dan grup baru sudah tidak ada perubahan, maka dari itu proses pengolahan data dihentikan dan didapatkan hasil perhitungan pada iterasi ke-3 sebagai berikut:

Tabel 7: Pusat *Cluster* Baru 3

C1	5,00	52,17	109,83
C2	2,53	57,00	55,93

Tabel 8: Perhitungan Jarak Antar *Cluster* Iterasi 3

No.	Desa	C1	C2	Jarak Terdekat	<i>Cluster</i>
1	Balongrejo	14,57	41,14	14,57	1
2	Girirejo	30,15	25,51	25,51	2
3	Pesudukuh	40,30	14,11	14,11	2
4	Sekar putih	56,88	10,06	10,06	2
5	Buduran	77,73	26,35	26,35	2
6	Ngumpul	36,70	71,86	36,70	1
7	Bagorkulon	24,18	40,45	24,18	1
8	Petak	63,01	22,12	22,12	2
9	Paron	63,21	25,70	25,70	2
10	Karangtengah	58,91	26,06	26,06	2
11	Selorejo	32,34	68,03	32,34	1
12	Gandu	49,69	25,36	25,36	2
13	Guyangan	68,26	28,21	28,21	2
14	Kedondong	66,37	27,50	27,50	2
15	Sugihwaras	44,41	34,11	34,11	2
16	Kutorejo	68,78	19,17	19,17	2
17	Kerepkidul	52,07	13,96	13,96	2
18	Gemenggeng	58,47	24,20	24,20	2
19	Kendalrejo	77,46	40,25	40,25	2
20	Banarankulon	23,21	77,19	23,21	1
21	Banaranwetan	7,57	48,13	7,57	1

Tabel 9: Hasil Pengelompokan Data

No.	Desa	C1	C2	Jarak Terdekat	Cluster
1	Balongrejo	14,57	41,14	14,57	1
2	Girirejo	30,15	25,51	25,51	2
3	Pesudukuh	40,30	14,11	14,11	2
4	Sekar putih	56,88	10,06	10,06	2
5	Buduran	77,73	26,35	26,35	2
6	Ngumpul	36,70	71,86	36,70	1
7	Bagorkulon	24,18	40,45	24,18	1
8	Petak	63,01	22,12	22,12	2
9	Paron	63,21	25,70	25,70	2
10	Karangtengah	58,91	26,06	26,06	2
11	Selorejo	32,34	68,03	32,34	1
12	Gandu	49,69	25,36	25,36	2
13	Guyangan	68,26	28,21	28,21	2
14	Kedondong	66,37	27,50	27,50	2
15	Sugihwaras	44,41	34,11	34,11	2
16	Kutorejo	68,78	19,17	19,17	2
17	Kerepkidul	52,07	13,96	13,96	2
18	Gemenggeng	58,47	24,20	24,20	2
19	Kendalrejo	77,46	40,25	40,25	2
20	Banarankulon	23,21	77,19	23,21	1
21	Banaranwetan	7,57	48,13	7,57	1

Maka diperoleh nilai *Centroid* 1 (C1) dengan **pusat cluster 1 (5; 52,17; 109,83)** dan *Centroid* 2 dengan titik **pusat cluster 2 (2,53; 57; 55,93)**. Grup terakhir yang dihasilkan kemudian digambarkan dalam sebuah grafik *cluster* data dengan nilai *centroid* terakhir menjadi titik pusat *cluster*.

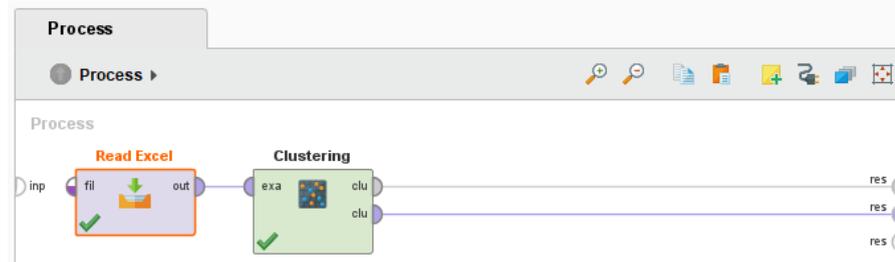
Tabel 10: Hasil Akhir *Clustering*

No.	Desa	Cluster
1	Balongrejo	1
2	Girirejo	2
3	Pesudukuh	2
4	Sekar putih	2
5	Buduran	2
6	Ngumpul	1
7	Bagorkulon	1
8	Petak	2
9	Paron	2
10	Karangtengah	2
11	Selorejo	1
12	Gandu	2
13	Guyangan	2
14	Kedondong	2

15	Sugihwaras	2
16	Kutorejo	2
17	Kerepkidul	2
18	Gemenggeng	2
19	Kendalrejo	2
20	Banarankulon	1
21	Banaranwetan	1

6. Implementasi *RapidMiner*

Berikut adalah pengolahan data dengan menggunakan k-means pada *RapidMiner*:



Gambar 2: Pemodelan *Clustering K-Means* pada *RapidMiner*

Cluster Model

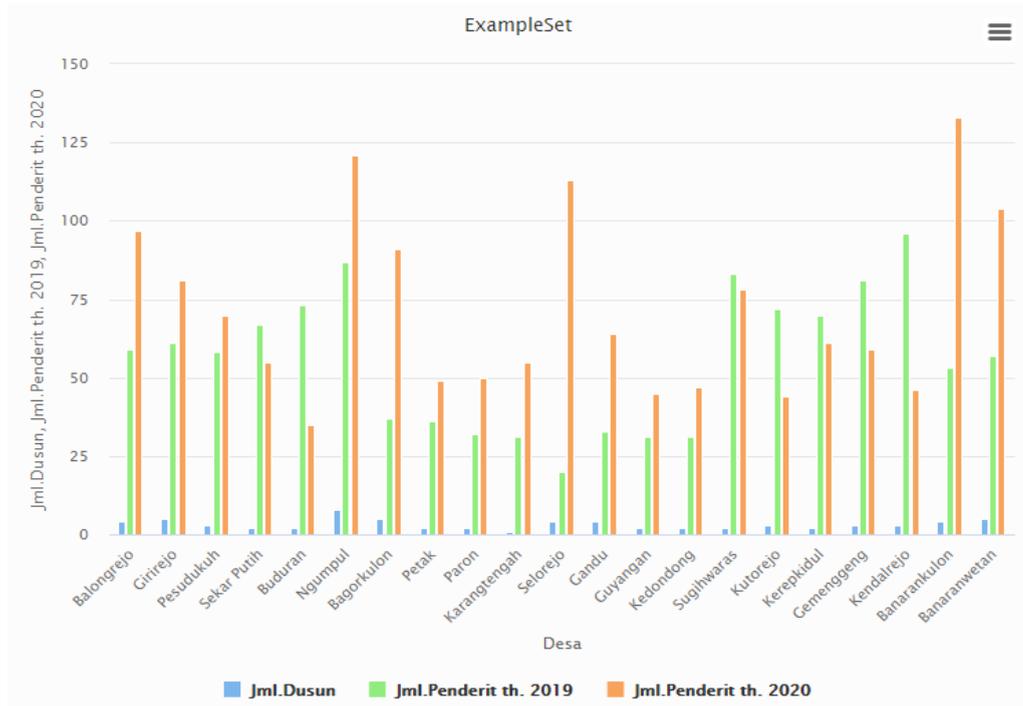
```
Cluster 0: 15 items
Cluster 1: 6 items
Total number of items: 21
```

Gambar 3: Hasil Data *Cluster K-Means* dalam Implementasi *RapidMiner*

Dengan menggunakan pemodelan *K-Means Clustering* yang ditunjukkan pada Gambar 2, dan jumlah *cluster* yang telah diinisialisasi dari proses awal terbentuk menjadi 2 *cluster* (dapat dilihat pada Gambar 3) yaitu *cluster 0* terdapat 15 data dan *cluster 1* terdapat 6 data dengan total jumlah data sebanyak 21 data.

Attribute	cluster_0	cluster_1
Jml.Dusun	2.533	5
Jml.Penderit th. 2019	57	52.167
Jml.Penderit th. 2020	55.933	109.833

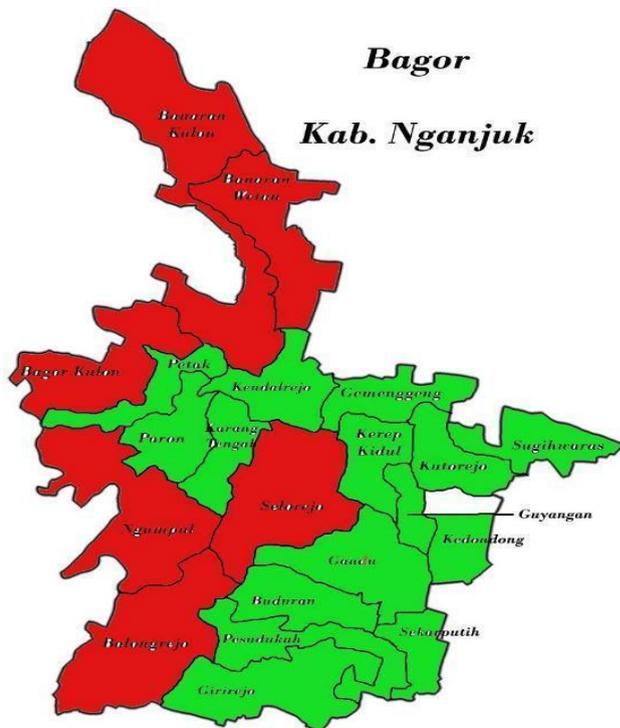
Gambar 4: Hasil Perhitungan Rata-rata Jarak antar *Cluster*



Gambar 5: Grafik Data Hasil K-Means Clustering pada RapidMiner

Keterangan:

- = cluster 1 (Penyebaran Diabetes Melitus Tingkat Tinggi)
- = cluster 2 (Penyebaran Diabetes Melitus Tingkat Rendah)



Gambar 6: Peta Pengelompokan Penyebaran Diabetes Melitus Kecamatan Bagor

Untuk melihat interpretasi *cluster* mana yang diartikan sebagai pusat *cluster* (penyebaran

penyakit dengan tingkat tinggi) dapat dilihat pada hasil akhir perhitungan manual dan perhitungan pada *RapidMiner*. Pada perhitungan manual diatas di iterasi ke-1 dihasilkan untuk desa yang masuk dalam *cluster* 1 sebanyak 6 desa dan *cluster* 2 sebanyak 15 desa, sementara itu nilai titik pusat *cluster* 1 lebih tinggi daripada *cluster* 2, sehingga dapat disimpulkan bahwa untuk *cluster* 1 merupakan pusat *cluster* atau penyebaran dengan tingkat tinggi dan *cluster* 2 merupakan penyebaran dengan tingkat rendah. Kemudian karena interpretasi pada *RapidMiner* berbeda, maka kita lihat berdasarkan kesamaan jumlah data akhir, yaitu untuk *cluster* 0 sebanyak 15 data dan *cluster* 1 sebanyak 6 data. Dengan melihat kesamaan dari 2 perhitungan diatas dapat disimpulkan bahwa “*cluster* 1 merupakan pusat *cluster* atau penyebaran penyakit dengan tingkat tinggi dan *cluster* 0 merupakan penyebaran penyakit dengan tingkat rendah”. Pada penelitian sebelumnya yang dilakukan oleh Mentari Tri Indah Rahmayani pada tahun 2018 yaitu mengenai pengelompokan tingkat keparahan pasien dengan Algoritma K-Means untuk pengolahan data secara manual dengan sampel 15 pasien diperoleh 3 kelompok *cluster* yaitu *cluster* 0 (C0) sebagai kelompok penyakit tingkat sedang sebanyak 5 orang, *cluster* 1 (C1) sebagai kelompok penyakit tingkat berat sebanyak 2 orang dan *cluster* 2 (C2) sebagai kelompok penyakit tingkat ringan sebanyak 8 orang. Pengolahan data dengan menggunakan *software RapidMiner* dengan jumlah pasien sebanyak 278 orang. Dimana *cluster* 0 (C0) sebagai kelompok penyakit tingkat berat sebanyak 47 orang, *cluster* 1 (C1) sebagai kelompok penyakit tingkat ringan sebanyak 82 orang dan *cluster* 2 (C2) sebagai kelompok penyakit tingkat sedang sebanyak 149 orang. Dapat disimpulkan untuk perbandingan hasil dari penelitian sebelumnya yaitu terletak pada jumlah *cluster* yang digunakan, namun untuk pengolahan data secara manual dan menggunakan *software RapidMiner* sama-sama memiliki hasil nilai *centroid* yang sama [25].

5. Simpulan

Dari hasil dan pembahasan yang telah di atas dapat disimpulkan bahwa pusat *cluster* yang diperoleh yaitu untuk *cluster* 0 meliputi desa Girirejo, Pesudukuh, Sekar Putih, Buduran, Petak, Paron, Karangtengah, Gandu, Guyangan, Kedondong, Sugihwaras, Kutorejo, Kerepkidul, Gemenggeng, Kendalrejo dan *cluster* 1 meliputi desa Balongrejo, Ngumpul, Bagorkulon, Selorejo, Banarankulon, dan Banaranwetan. Pusat *cluster* 1 merupakan daerah-daerah pusat penyebaran penyakit diabetes melitus dengan tingkat tinggi, untuk itu pada daerah-daerah tersebut harus mendapatkan perhatian yang lebih dari pemerintah untuk penanganan penyakit diabetes melitus. Pusat *cluster* 0 merupakan daerah penyebaran diabetes melitus dengan jumlah penderita tingkat rendah sampai menengah dan bukan merupakan pusat penyebaran penyakit diabetes melitus.

Berdasarkan kesimpulan tersebut, maka saran yang dapat diberikan setelah penelitian ini dilakukan yaitu diharapkan untuk penelitian selanjutnya dapat menggunakan algoritma yang lain selain Algoritma K-Means Clustering agar didapatkan perbandingan diantara sekian banyak algoritma yang ada dan dapat digunakan dalam metode pengelompokan dan diharapkan dapat digunakan data terbaru atau yang telah di-*update* tentang jumlah kasus penyakit yang diteliti.

Referensi

- [1] M. G. Pradana, B. W. Pamekas, dan K. Kusriani, “Perancangan Sistem Pakar Untuk Mendiagnosa Penyakit Diabetes Mellitus Menggunakan Metode Certainty factor Design Expert System for Diagnosing Diabetes Mellitus Using Certainty factor Method”, *Creative Communication and Innovative Technology Journal*, vol. 11, no. 2, pp. 182-191, 2018.
- [2] E. Kurniaty, dan M. R. Saraswati, “Tingkat Pengetahuan Tentang Diet pada Penderita Diabetes Melitus dengan Komplikasi Chronic Kidney Disease Di RSUP Sanglah Denpasar Tahun 2016”.

- [3] Kementerian Kesehatan RI, "Infodatin pusat data dan informasi kementerian kesehatan RI Diabetes Melitus", 2020.
- [4] F. R. Mahmud, S. Sudirman, dan N. Afni, "Faktor-Faktor yang Berhubungan Dengan Penyakit Diabetes Melitus di Ruang Poli Interna RSUD Mokopido Kabupaten Tolitoli", *Jurnal Kolaboratif Sains*, vol. 1, no. 1, 2018.
- [5] A. Kantono, I. Y. Purbasari, dan F. T. Anggraeny, "Penerapan Pruning Pada Algoritma C5. 0 Untuk Mendiagnosis Penyakit Diabetes Melitus", *In Prosiding Seminar Nasional SANTIKA Ke-1 2019*, pp. 184-189, 2019.
- [6] H. Sugara dan M. A. Hanafah, "Book Data Grouping in the Library Using the K-Means Clustering Method", *Jurnal Mantik*, vol.4, no. 3, pp. 2225-2231, 2020.
- [7] I. Kamila, U. Khairunnisa, dan M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau", *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119-125, 2019.
- [8] V. Febriyanti, H. S. Tambunan, I. S. Saragih, I. S. Damanik, dan H. Okprana, "Implementasi Algoritma K-Means Dalam Pengelompokan Kasus Penyakit Tuberkulosis Paru Berdasarkan Provinsi", *In Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 2, pp. 450-456, 2020.
- [9] F. Nasari, dan C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare di Kabupaten Langkat", *Cogito Smart Journal*, vol. 2, no. 2, pp. 108-119, 2016.
- [10] S. Maulia, B. S. Ginting, A. Sihombing, "Implementasi Data Mining Pengelompokan Jenis Penyakit Pasien Menggunakan Metode Clustering (Studi Kasus: Puskesmas Sambirejo)", *Jurnal Informatika Kaputama*, vol. 5, no. 1, 2021.
- [11] M. Abdallah, I. Mohammadnoor, S. Abdallah dan L. Hanan, "Student Engagement Level in e-Learning Environment: Clustering Using K-means", *American Journal of Distance Education*, vol. 34, no. 2, pp. 137-156, 2020.
- [12] A. M. Simarmata, M. A. P. Siamipar, S. Singh, I. I. M. Gulo dan J. B. R. Purba, "Grouping Diabetes Diagnosis Based on Age Range with K-Means Algorithm", *Jurnal Mantik*, vol. 5, no. 2, pp. 1408-1412, 2021.
- [13] A. Praja, C., Lubis, dan D. E. Herwindiati, "Deteksi Penyakit Diabetes dengan Metode Fuzzy C-Means Clustering dan K-Means Clustering", *Computation: Journal of Computer Science and Information Systems*, vol. 1, no. 1, pp. 15-24, 2017.
- [14] M. Fauzi dan Yudi, "Penerapan Algoritma K-Means Untuk Mendeteksi Penyebaran Penyakit TBC (Studi Kasus: Di Kabupaten Deli Serdang)", *Jurnal Teknik Informatika Kaputama (JTik)*, vol. 1, no. 2, 2017.
- [15] A. T. Nadezhda, I. A. Marina, I. Anna, Rybakova, I. S. Larisa dan B. K. Tatiana, "Characteristics of Social Psychological Adaptation and Self Regulation in Patient with Diabetes Mellitus", *International Journal of Environmental and Science Education*, vol. 11, no. 14, pp. 6616-6634, 2016.
- [16] R. W. Leung, J. Kamla, M. C. Lee, J. Y. Mak, "Preventing and Treating Type 2 Diabetes through a Physically Active Lifestyle", *Journal of Physical Education, Recreation dan Dance*, vol. 78, no. 4, April 2007.
- [17] J. A. Wicaksana, H. Yasin, dan S. Sudarno, "Probabilistic neural network berbasis gui matlab untuk klasifikasi data rekam medis (studi kasus penyakit diabetes melitus di balai kesehatan kementerian perindustrian jakarta)", *Jurnal Gaussian*, vol. 5, no. 3, pp. 427-436, 2016.
- [18] D. Saryanti, dan D. Nugraheni, "Peningkatan Pengetahuan Masyarakat Tentang Penyakit Diabetes Melitus", *JPPM (Jurnal Pengabdian dan Pemberdayaan Masyarakat)*, vol. 3, no. 1, pp. 111-116, 2019.
- [19] R. Silvi, "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means

- Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia”, *Jurnal Matematika “MANTIK”*, vol. 4, no. 1, pp. 22-31, 2018.
- [20] U. Ma’rifatin, “Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng”, *In Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 4, no. 1, pp. 285-291, 2020.
- [21] M. M. Khairunnisa, A. Triayudi dan E. T. E. Handayani, “Application of K-Means *Clustering* on the Performance Evaluation of Lecturers Based on Student Questionnaire”, *Jurnal Mantik*, vol. 4, no. 1, pp. 760-766, 2020.
- [22] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, dan M. Azmi, “Using K-means *clustering* to *cluster* provinces in Indonesia”, *IOP Publishing In Journal of Physics: Conference Series*, vol. 1028, no. 1, p. 012006, 2018.
- [23] D. J. A. Zahra, A. Triayudi, dan I. D. Solihati, “Application To Classify The Population Census The Total Population By Level Of Education Using K-Means *Clustering* Algorithm”, *Jurnal Mantik*, vol. 3, no. 4, pp. 55-63, 2020.
- [24] Y. P. Sari, A. Primajaya, dan A. S. Y. Irawan, “Implementasi Algoritma K-Means untuk *Clustering* Penyebaran Tuberkulosis di Kabupaten Karawang”, *INOVTEK Polbeng-Seri Informatika*, vol. 5, no. 2, pp. 229-239, 2020.
- [25] M. T. I. Rahmayani, “Analisis *Clustering* Tingkat Keparahan Penyakit Pasien menggunakan Algoritma K-Means (Studi Kasus di Puskesmas Bandar Seikijang)”, *Jurnal Inovasi Teknik Informatika (JITI)*, vol. 1, no. 2, 2018.