

SISTEM KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (K-NN)

Fitri Yunita

Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

Email:Fitriyun@gmail.com

Abstrak

Diabetes mellitus merupakan penyakit yang banyak diderita oleh masyarakat pada saat ini, penyakit ini timbul karena pola hidup yang kurang baik. Penyakit tersebut ditandai dengan kenaikan kadar gula dalam darah. Penyakit ini biasanya disebabkan pengaruh destruksi sel beta, pola makan, dan kurangnya olahraga yang teratur. Penyakit ini diklasifikasikan menjadi dua kelas/tipe. Dalam mengklasifikasikan suatu permasalahan dengan jumlah data yang besar membutuhkan metode yang sesuai dan tepat yaitu dengan metode K-Nearest Neighbor (K-NN). Metode K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Penelitian ini untuk mengkaji algoritma K-NN dan kemudian mengimplementasikannya dalam klasifikasi data. Data yang digunakan adalah data penyakit diabetes mellitus yang didapatkan dari RSUD Puri Husada Tembilahan. Dari data yang telah dilakukan pengujian, KNN mampu mengklasifikasikan data penyakit diabetes mellitus berdasarkan variabel yaitu usia, pengaruh pola makan, berat badan, tinggi badan, keturunan, dan gangguan destruksi sel beta yang mempengaruhi penyakit diabetes mellitus. K-NN mengklasifikasikan data dengan mencari nilai terdekat pada suatu target yang akan dicapai. Jumlah data latih/training yang digunakan sangat mempengaruhi hasil klasifikasi, semakin banyak jumlah data latih/training, maka nilai akurasi semakin tinggi. Adapun hasil akurasi K-NN dalam klasifikasi penyakit diabetes mellitus adalah 96 %.

Kata kunci: *Klasifikasi Diabetes Mellitus, K-Nearest Neighbor (K-NN)*

1. PENDAHULUAN

Diabetes mellitus merupakan penyakit yang ditandai dengan kenaikan kadar gula dalam darah. Penyakit ini biasanya disebabkan pengaruh destruksi sel beta, diabetonik, dan olah raga yang kurang. Penyakit ini akan mengakibatkan berbagai macam komplikasi kronis terjadi pada mata, ginjal, syaraf, dan pembuluh darah [1].

Berdasarkan sebab yang mendasari kemunculannya, diabetes menjadi beberapa golongan atau tipe. Di antara tipe-tipe diabetes yang termasuk tipe utama adalah diabetes tipe-1 dan tipe-2 [1].

Diabetes tipe-1 biasanya mengenai anak-anak dan remaja. Diabetes ini dulu pernah disebut sebagai *Juvenile diabetes* (diabetes usia muda). Namun karena diabetes tipe ini juga dapat terjadi pada orang dewasa, maka orang lebih suka memakai istilah diabetes tipe-1. faktor penyebab diabetes tipe-1 adalah infeksi virus atau reaksi outo-imun (rusaknya sistem kekebalan tubuh) yang merusak sel-sel penghasil insulin, yaitu sel beta pada

pankreas, secara menyeluruh. Oleh karena itu, pada tipe ini pankreas sama sekali tidak dapat menghasilkan insulin. Untuk bertahan hidup, insulin harus diberikan dari luar dengan cara disuntikan.

Dari seluruh penderita diabetes, jumlah penderita tipe-2 adalah yang paling banyak, yaitu sekitar 90-99%. Diabetes tipe-2 juga disebut diabetes *life style* karena selain faktor keturunan, juga disebabkan gaya hidup yang tidak sehat. Biasanya, tipe ini mengenai orang dewasa. Dulu diabetes ini disebut *maturity onset diabetes*. Namun karena diabetes ini ternyata juga dapat mengenai mereka yang lebih muda, maka istilah diabetes tipe-2 dianggap lebih cocok[1].

Diabetes tipe-2 berkembang sangat lambat, bisa sampai bertahun-tahun. Oleh karena itu, gejala dan tanda-tandanya seringkali tidak jelas. Penderita diabetes tipe-2 biasanya memiliki riwayat keturunan diabetes. Apabila tidak ada gejala klasik, yang biasanya dikeluhkan adalah cepat lelah, berat badan turun walaupun banyak makan,

dan rasa kesemutan di tungkai. Bahkan terkadang ada penderita yang sama sekali tidak merasa perubahan.

Berbagai kasus yang berkaitan dengan pengelompokan objek dapat diselesaikan lebih mudah dengan menerapkan teknik-teknik klasifikasi. Sebagai contoh pada bidang medis, aplikasi klasifikasi dapat digunakan untuk klasifikasi tingkat penyakit yang diderita oleh seorang pasien sehingga memudahkan dokter dalam memberikan solusi terapi yang tepat. Untuk memecahkan masalah klasifikasi, berbagai macam metode dalam data mining dapat diterapkan. Ketepatan dalam pengklasifikasian objek sangat penting, metode klasifikasi yang baik adalah metode yang menghasilkan kesalahan yang kecil, diantara metode klasifikasi yang telah ada yaitu metode *K-Nearest Neighbor* (K-NN).

Adapun struktur penulisan dalam paper ini terdiri dari pendahuluan yang berisikan penjelasan mengenai penyakit diabetes mellitus dan faktor-faktor yang mempengaruhinya, dan penjelasan penerapan metode *K-Nearest Neighbor* (K-NN). Tinjauan pustaka berisikan teknik-teknik dari metode yang digunakan untuk menyelesaikan kasus penyakit diabetes mellitus. Metodologi penelitian berisikan penjelasan tahapan dari kerangka kerja penelitian yang dilakukan. Bagian selanjutnya yaitu kesimpulan dan saran

2. TINJAUAN PUSTAKA

2.1. Pengertian Data Mining

Data mining sendiri sering disebut sebagai *knowledge discovery in database* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. *Output* dari *data mining* ini dapat digunakan untuk pengambilan keputusan dimasa depan [2].

Data mining memiliki tahapan - tahapan antara lain :Pembersihan data (*data cleaning*), proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Integrasi data (*data integration*), penggabungan data dari berbagai *database* ke dalam satu *database* baru. Seleksi Data (*Data Selection*), Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Transformasi data (*Data Transformation*), data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Proses *mining*, merupakan suatu proses utama saat metode

diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Evaluasi pola (*pattern evaluation*), Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan[3].

2.2. Klasifikasi

Klasifikasi merupakan pengelompokan sebuah objek ke dalam kelas tertentu. Berbagai kasus yang berkaitan dengan pengelompokan objek dapat diselesaikan dengan menerapkan teknik-teknik klasifikasi[4].

Algoritma klasifikasi menggunakan *data training* untuk membuat sebuah model. Model yang sudah dibangun tersebut kemudian digunakan untuk memprediksi label kelas data baru yang belum diketahui. Dalam proses klasifikasi terdapat banyak algoritma yang telah dikembangkan oleh para peneliti seperti *K-Nearest Neighbor*, *Artificial Neural Network*, *Support Vector Machine*, *Decision Tree*, *Naïve Bayes Classifier* dan lain sebagainya. Setiap algoritma dalam klasifikasi *data mining* tersebut memiliki kelebihan dan kekurangan. Tetapi prinsip dari masing-masing algoritma tersebut sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan model dapat memprediksi setiap vektor masukan ke label kelas *output* dengan tepat[5].

Salah satu pengukur kinerja klasifikasi adalah tingkat akurasi. Sebuah sistem dalam melakukan klasifikasi diharapkan dapat mengklasifikasi semua set data dengan benar, tetapi tidak dipungkiri bahwa kinerja suatu sistem tidak bisa 100% akurat.

Untuk mengukur tingkat akurasi dapat digunakan rumus sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data diprediksikan secara benar}}{\text{Jumlah Prediksi yang dilakukan}} \quad (1)$$

2.3. K-Nearest Neighbor (K-NN)

Algoritma K-NN adalah algoritma yang menentukan nilai jarak pada pengujian *data testing* dengan *data training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan *training samples* [6].

K-Nearest Neighbor (K-NN) metode yang melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data lain. Nilai K pada *K-Nearest Neighbor* (K-NN) berarti K-data terdekat dari data uji [3].

Adapun tahapan dari algoritma *K-Nearest Neighbor* (K-NN) dijelaskan sebagai berikut :

1. Persiapan data training dan data testing
2. Menentukan nilai k
3. Menghitung jarak data testing ke setiap data pelatihan.

Data pelatihan di hitung menggunakan rumus penghitung jarak *Euclidean* sebagai berikut :

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (2)$$

Keterangan :

X_{1i} = Data Training

X_{2i} = Data Testing

4. Tentukan nilai k data training yang memiliki jarak terdekat dengan data testing
5. Periksa label dari k data training terdekat
6. Tentukan label yang frekuensinya paling banyak
7. Masukkan data testing kedalam kelas dengan frekuensi paling banyak
8. Kondisi berhenti.

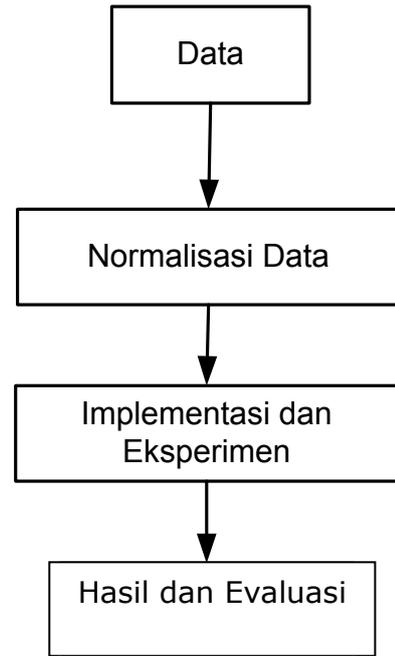
K-Nearest Neighbor (K-NN) memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. *K-Nearest Neighbor* juga memiliki kekurangan yaitu membutuhkan nilai k, jarak dari data percobaan tidak jelas dengan tipe jarak yang digunakan, untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau hanya satu atribut yang telah pasti. Metode *K-Nearest Neighbor* (K-NN) cukup sederhana, tidak ada asumsi mengenai distribusi data dan mudah diaplikasikan. Pemilihan nilai K (jumlah data/tetangga terdekat) ditentukan oleh peneliti. Pemilihan nilai K ini bisa mempengaruhi tingkat akurasi prediksi yang dikerjakan.

Dari keseluruhan nilai untuk setiap hasil klasifikasi memiliki tingkat kesamaan dan perbedaan, maka perhitungan kesamaan data untuk nilai K dapat dilihat pada rumus berikut ini:

$$\begin{aligned} \text{Persentasi Kesamaan} \\ &= \frac{\text{Jumlah kecocokan data}}{\text{jumlah data}} \times 100 \end{aligned} \quad (3)$$

3. METODOLOGI PENELITIAN

Pada metodologi penelitian ini akan dijelaskan bagaimana isi dari kerangka kerja dalam menyelesaikan kasus ini. Kerangka kerja penelitian ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian permasalahan yang akan dibahas dalam kasus ini. Adapun kerangka kerja penelitian yang digunakan dalam penyusunan dan penyelesaian kasus ini dapat dilihat pada gambar berikut.



Gambar 1. Kerangka Kerja Penelitian

3.1. Data

Adapun data penyakit diabetes mellitus bersumber dari RSUD Puri Husada Tembilahan yang menjelaskan faktor-faktor penyebab yang terdapat pada penyakit diabetes mellitus. Terdapat dua tipe/kelas penyakit diabetes mellitus yang merupakan perbedaan indikasi dari penyakit diabetes mellitus dan penyebab yang mempengaruhi timbulnya penyakit diabetes mellitus. Sehingga sistem dapat mengenali kelas/tipe dari penyebab yang mempengaruhi penyakit diabetes mellitus tersebut dengan menggunakan metode *K-Nearest Neighbor* (K-NN).

Data yang diperoleh terdiri atas enam variable input dan target sebanyak dua kelas. Adapun variabel input dalam penelitian ini dapat dilihat pada tabel 1 dibawah ini :

Tabel 1. Variabel Masukan

Variabel	Satuan Nilai	Keterangan
X1	Nilai Usia	Usia (Tahun)
X2	Nilai Berat Badan	Berat Badan (Kg)
X3	Nilai Tinggi Badan	Tinggi Badan (Cm)
X4	0 = Tidak ada	Riwayat Keluarga
	1 = ada	
X5	1 = Ringan	Terdapat Gangguan Destruksi Sel Beta
	2 = Berat	
X6	1 = Rendah	Pengaruh Pola Makan
	2 = Sedang	
	3 = Tinggi	

Sedangkan sebagai unit output atau target yang akan dihasilkan dapat dilihat pada tabel 2 dibawah ini :

Tabel 2. Variabel Target

No	Variabel	Keterangan
1	1	Tipe-1
2	2	Tipe-2

Berikut informasi awal yang digunakan dalam klasifikasi data penyakit diabetes mellitus menggunakan metode *K-Nearest Neighbor* (K-NN) adalah :

1. Jumlah data = 25
2. Jumlah variabel input = 6 buah. (Umur, Berat Badan, Tinggi Badan, Riwayat Keluarga, Gangguan Destruksi Sel Beta, Pola Makan).
3. Jumlah kelas yang diinginkan = 2 kelas.

Adapun data yang diperoleh dari RSUD Puri Husada Tembilahan sebanyak 339 data. Adapun data yang di jadikan sampel yaitu sebanyak 25 data. Adapun datapenyakit diabetes mellitus dapat dilihat pada tabel 3 dibawah ini :

Tabel 3. Data Sampel Penyakit Diabetes Mellitus

Input Data Ke	Variabel						Target
	X1	X2	X3	X4	X5	X6	
1	67	85	167	1	1	1	2
2	60	67	150	1	1	2	2
3	53	75	156	1	2	2	2
4	48	80	174	1	1	2	2
5	56	67	156	0	1	2	1
6	47	65	153	1	1	2	2
7	68	70	164	1	2	2	2
8	70	80	173	0	1	2	2
9	57	59	168	0	1	2	2
10	54	56	170	0	2	2	2
11	40	75	165	0	1	3	2
12	65	88	175	1	1	3	2
13	60	68	155	1	2	2	2
14	63	67	168	1	1	3	2
15	59	86	173	1	1	3	2
16	43	56	150	1	1	3	2
17	50	80	155	1	2	2	2
18	45	59	149	1	1	2	2
19	42	64	169	0	1	2	2
20	61	75	157	1	2	2	2
21	65	73	172	0	1	1	2
22	51	59	173	1	1	1	2
23	50	54	169	1	1	2	2
24	55	76	157	0	1	2	2
25	43	68	150	1	1	1	2

3.2. Normalisasi Data

Suatu teknik untuk mengorganisasikan data kedalam tabel-tabel untuk memenuhi kebutuhan pemakai di dalam suatu

organisasi. Data-data yang dilakukan normalisasi dengan membagi nilai data tersebut dengan nilai *range* data (nilai data maksimum – nilai data minimum).

$$x_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Keterangan

Xn = Nilai data normal
 X0 = Nilai data aktual
 Xmin = Nilai Minimum data aktual keseluruhan
 Xmax = Nilai maksimum data aktual keseluruhan

Normalisasi data input bertujuan untuk menyesuaikan nilai *range* data dengan fungsi aktivasi dalam sistem ELM. Ini berarti nilai kuadrat input harus berada pada range 0 sampai 1. Sehingga *range* input yang memenuhi syarat adalah nilai data input dari 0 sampai 1. Oleh karena itu output yang dihasilkan pun akan berada pada range 0 sampai 1. Kemudian untuk mendapatkan nilai sebenarnya dari output perlu dilakukan normalisasi.

Setelah data sampel diketahui, berikutnya dilakukan normalisasi data pada masing-masing variabel (X) dengan tujuan agar menyamakan data dalam bentuk rata-rata sehingga antara data yang satu dengan data yang lain memiliki *range* yang berdekatan (tidak terlalu jauh).

Adapun data sampel penyakit diabetes mellitus yang datanya telah dinormalisasi dapat dilihat pada tabel 3 dibawah ini :

Tabel 4. Data Penyakit Diabetes yang Telah Dinormalisasi

Input Data Ke	X1	X2	X3	X4	X5	X6	Tipe
1	0,900	0,912	0,692	1,000	0,000	0,000	2
2	0,667	0,382	0,038	1,000	0,000	0,500	2
3	0,433	0,618	0,269	1,000	1,000	0,500	2
4	0,267	0,765	0,962	1,000	0,000	0,500	2
5	0,533	0,382	0,269	0,000	0,000	0,500	1
6	0,233	0,324	0,154	1,000	0,000	0,500	2
7	0,933	0,471	0,577	1,000	1,000	0,500	2
8	1,000	0,765	0,923	0,000	0,000	0,500	2
9	0,567	0,147	0,731	0,000	0,000	0,500	2
10	0,467	0,059	0,808	0,000	1,000	0,500	2
11	0,000	0,618	0,615	0,000	0,000	1,000	2
12	0,833	1,000	1,000	1,000	0,000	1,000	2
13	0,667	0,412	0,231	1,000	1,000	0,500	2
14	0,767	0,382	0,731	1,000	0,000	1,000	2
15	0,633	0,941	0,923	1,000	0,000	1,000	2
16	0,100	0,059	0,038	1,000	0,000	1,000	2
17	0,333	0,765	0,231	1,000	1,000	0,500	2
18	0,167	0,147	0,000	1,000	0,000	0,500	2
19	0,067	0,294	0,769	0,000	0,000	0,500	2
20	0,700	0,618	0,308	1,000	1,000	0,500	2
21	0,833	0,559	0,885	0,000	0,000	0,000	2
22	0,367	0,147	0,923	1,000	0,000	0,000	2
23	0,333	0,000	0,769	1,000	0,000	0,500	2
24	0,500	0,647	0,308	0,000	0,000	0,500	2
25	0,100	0,412	0,038	1,000	0,000	0,000	2

3.3 Implementasi dan Eksperimen

Proses *K-Nearest Neighbor* (K-NN) dalam sistem ini merupakan proses yang akan dilakukan pertama kali, metode ini akan melakukan pemilihan nilai K (jumlah data/tetangga terdekat) ditentukan oleh peneliti. Pemilihan nilai K ini bisa mempengaruhi tingkat akurasi prediksi yang dikerjakan.

Berdasarkan algoritma K-NN, dapat dilakukan perhitungan sebagai berikut :

1. Tentukan nilai K, misalkan K = 1

2. Hitung jarak setiap training data dengan data testing.

Adapun data yang digunakan diambil dari data nomor 1 sebagai data training, 2, 3 dan seterusnya sampa data 5 sebagai data testing. Adapun contoh perhitungan nilai jaraknya dapat dilakukan dengan cara sebagai berikut:

$$d_1 = \frac{\sqrt{(0,900 - 0,667)^2 + (0,912 - 0,382)^2}}{\sqrt{(0,692 - 0,038)^2 + (1,000 - 1,000)^2} + \sqrt{(0,000 - 0,000)^2 + (0,000 - 0,500)^2}} = 1,01$$

$$d_2 = \frac{\sqrt{(0,900 - 0,433)^2 + (0,912 - 0,618)^2}}{\sqrt{(0,692 - 0,269)^2 + (1,000 - 1,000)^2} + \sqrt{(0,000 - 1,000)^2 + (0,000 - 0,500)^2}} = 1,32$$

Adapun hasil perhitungan nilai jarak penyakit diabetes mellitus dapat dilihat pada tabel 5 berikut ini :

Tabel 5. Hasil Perhitungan Nilai Jarak Penyakit Diabetes

Data Ke-i	Nilai Jarak	Tipe/Kelas
2	1,0061	2
3	1,3165	2
4	0,8633	2
5	1,3578	2
6	1,1534	2
7	1,2079	2
8	1,1554	2
9	1,3955	1
10	1,7829	2
11	1,7036	2
12	1,0521	2
13	1,3295	2
14	1,14	2
15	1,0608	2
16	1,6718	2
17	1,3438	2
18	1,3608	2
19	1,527	2
20	1,2347	2

Data Ke-i	Nilai Jarak	Tipe/Kelas
21	1,0798	2
22	0,9605	2
23	1,1867	2
24	1,2759	1
25	1,1478	2

Hasil perhitungan nilai jarak yang digambarkan pada tabel 5 didapatkan nilai jarak terkecil terletak pada data input ke empat, dengan nilai jarak 0,8633.

Adapun data hasil klasifikasi penyakit diabetes mellitus, didapatkan dari perhitungan nilai jarak terkecil masing-masing data yang dilakukan dapat dilihat pada tabel berikut ini :

Tabel 6. Hasil Perhitungan Nilai Jarak terkecil Penyakit Diabetes

Data Ke-1	Nilai Jarak terkecil	Tipe/Kelas
1	0,863	2
2	0,452	2
3	0,182	2
4	0,646	2
5	0,27	2
6	0,243	2
7	0,385	2
8	0,567	2
9	0,519	1
10	1,012	2
11	0,619	2
12	0,222	2
13	0,222	2
14	0,606	2
15	0,222	2
16	0,514	2
17	0,182	2
18	0,243	2
19	0,623	2
20	0,222	2
21	0,567	2
22	0,544	2

Data Ke-1	Nilai Jarak	Tipe/Kelas
23	0,544	2
24	0,27	1
25	0,537	2

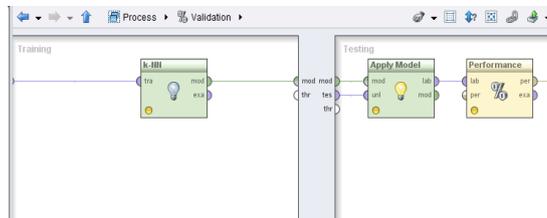
Dari hasil klasifikasi berdasarkan jarak data terkecil yang digambarkan pada tabel 6 di atas, didapatkan hasil kelas tipe-1 sebanyak 2 dan kelas tipe-2 sebanyak 23.

4. HASIL DAN EVALUASI

4.1. Pengolahan Data Rapid Miner

Rapid Miner merupakan salah satu software untuk pengolahan data mining, dan berbasis open-source. Metode yang digunakan untuk mengklasifikasikan data penelitian ini adalah *K-Nearest Neighbor* (K-NN).

Setelah dilakukan perhitungan manual, maka data diuji dengan software *Rapid Miner* dengan memasukkan semua data klasifikasi penyakit diabetes mellitus dan data tersebut terdiri dari data training dan data testing seperti dilihat pada gambar 2 berikut:



Gambar 2. Data training dan data testing

Adapun hasil penentuan kesamaan data penyakit diabetes mellitus dengan metode *K-Nearest Neighbor* menggunakan software *Rapid Miner* dengan nilai $K=1$ dapat dilihat pada tabel 7. Sebagai berikut :

Tabel 7. Validasi Klasifikasi KNN

Data Ke-i	Klasifikasi Rapid Miner	Klasifikasi KNN	Validasi KNN
1	2	2	Sama
2	2	2	Sama
3	2	2	Sama
4	2	2	Sama
5	1	2	Tidak Sama
6	2	2	Sama
7	2	2	Sama
8	2	2	Sama
9	2	1	Tidak

Data Ke-i	Klasifikasi Rapid Miner	Klasifikasi KNN	Validasi KNN
			Sama
10	2	2	Sama
11	2	2	Sama
12	2	2	Sama
13	2	2	Sama
14	2	2	Sama
15	2	2	Sama
16	2	2	Sama
17	2	2	Sama
18	2	2	Sama
19	2	2	Sama
20	2	2	Sama
21	2	2	Sama
22	2	2	Sama
23	2	2	Sama
24	2	1	Tidak Sama
25	2	2	Sama

Keterangan :

- Sama = Terjadi apabila hasil perhitungan K-NN sama dengan perhitungan Ketetapan
- Tidak Sama = Terjadi apabila hasil perhitungan K-NN berbeda dengan hasil ketetapan

Dari perhitungan jarak validasi ketetapan dengan metode *K-Nearest Neighbor* didapatkan hasil nilai yang sama 22 dan yang tidak sama 3. Maka, perhitungan persentase kesamaan data untuk nilai $K = 1$ sebagai berikut :

persentaseKesamaan($k = 1$)

$$= \frac{\text{Jumlahkecocokandata}}{\text{jumlahdata}} \times 100$$

$$= \frac{22}{25} \times 100\% = 88\%$$

Dari tabel 7. dapat disimpulkan bahwa persentase kesamaan hasil klasifikasi antara *rapid miner* yang telah diperoleh dengan metode *K-Nearest Neighbor* (K-NN) yang tertinggi yaitu pada saat pembobot bernilai $K = 1$ dengan tingkat kesamaan persentase keberhasilan sebesar 88%.

Adapun hasil pengujian banyak data penyakit diabetes mellitus yang terdiri 339 data dapat dilihat pada tabel 8 berikut ini :

Tabel 8. Hasil Pengujian Data

No	Jumlah data	Jumlah data yang benar	Jumlah data yang salah	Akurasi
1	25	22	3	88%
2	60	53	7	88%
3	95	87	8	91%
4	150	137	13	91%
5	339	318	21	93%

Rata-rata

Akurasi :

$$\frac{88\% + 88\% + 91\% + 91\% + 93\%}{5} = 90\%$$

Dari tabel 8. dapat disimpulkan bahwa persentase kesamaan hasil klasifikasi antara data kelas yang telah diperoleh dengan metode *K-Nearest Neighbor* (KNN) yang tertinggi yaitu pada saat pembobot bernilai $K = 1$ dengan tingkat kesamaan persentase keberhasilan sebesar 90 %.

Adapun pengujian berdasarkan jumlah data training yang lebih banyak, jumlah data yang di testing adalah 25. Hasil pengujian berdasarkan jumlah data training pada *K-Nearest Neighbor* (K-NN) dapat dilihat pada Tabel 9 berikut ini :

Tabel 9. Hasil Pengujian Data

No	Jumlah training	Jumlah data benar	Jumlah data salah	Akurasi
1	1	22	3	88%
2	5	23	2	92%
3	10	25	0	100%
4	20	25	0	100%
5	25	25	0	100%

Rata-rata

Akurasi :

$$\frac{88\% + 92\% + 100\% + 100\% + 100\%}{5} = 96\%$$

Dari tabel 9. dapat disimpulkan bahwa persentase akurasi data penyakit diabetes mellitus dengan jumlah data training lebih banyak sebesar 96 %.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. K-NN mampu mengklasifikasikan data penyakit diabetes mellitus berdasarkan usia, pengaruh pola makan dan beberapa faktor lain yang

mempengaruhi penyakit diabetes melitus.

2. Jumlah data latih/training yang digunakan sangat mempengaruhi hasil klasifikasi, semakin banyak jumlah data latih, maka nilai akurasi semakin tinggi.
3. Berdasarkan hasil evaluasi akurasi K-NN dalam klasifikasi penyakit diabetes mellitus adalah 96 %.

5.2 Saran

1. Penelitian dapat dikembangkan dengan menambah variabel faktor yang mempengaruhi penyakit diabetes melitus lainnya seperti, Jenis kelamin dll.
2. Dalam penelitian ini mungkin masih ditemukan beberapa kekurangan, dan masih dapat dikembangkan lagi dengan metode lainnya.

DAFTAR PUSTAKA

- [1] H. Kurniadi, U. Nurrahmani. *Stop Diabetes Hipertensi Kolestrol Tinggi Jantung Koroner*. Yogyakarta: Istana Media, 2014.
- [2] T. Alfina, B. Santosa, and A. R. Barakbah. *Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data*. Jurnal Teknik, Vol. 1, pp. 521-522, 2012.
- [3] B. Merluarini, D. Safitri, and A. Hoyyi. *Perbandingan Analisis Klasifikasi Menggunakan Metode K-Nearest Neighbor (K-NN) dan Multivariate Adaptive Regression Spline (Mars) pada data Akreditasi Sekolah Dasar Negeri di Kota Semarang*. Jurnal Gaussian, Vol. 3, no. 3, pp. 314-317, 2014.
- [4] N. Syafitri. *Perbandingan metode K-NearestNeighbor (KNN) dan Meode Nearest Cluster Classifier (NCC) Dalam Pengklasifikasian kualitas batik tulis*. Teknologi Informasi & pendidikan, Vol 2, no. 1, pp. 45-46, 2010.
- [5] B. A. Mukhtar. *Implementasi dengan Naive Bayes Classifier untuk mendukung Strategi pemasaran di Bagian Humas STMIK AMIKOM Yogyakarta* [Laporan Penelitian], Yogyakarta, 2013.
- [6] N. Krisandi, Helmi, and B. Prihandono, "Algoritma K-Nearest Neighbor dalam Klasifikasi Data Hasil Produksi Kelapa Sawit PT. MINAMAS Kecamatan Paridu," Vol. 2, no.1, pp. 34-35, 2013.