

# Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke

## *Stroke Disease Predict Using KNN Algorithm*

Muhammad Naja Maskuri<sup>1</sup>, Harliana<sup>\*2</sup>, Kadek Sukerti<sup>3</sup>, R.M. Herdian Bhakti<sup>4</sup>

<sup>1,2</sup> Program Studi Ilmu Komputer, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama Blitar

<sup>3</sup> Program Studi Sistem Informasi, Fakultas Informatika dan Komputer, ITB STIKOM Bali

<sup>4</sup> Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi, Brebes

e-mail correspondence Author: \*harliana@unublitar.ac.id

### Abstrak

Stroke merupakan gangguan fungsi otak yang disebabkan oleh terganggunya aliran darah ke otak. Penyakit stroke dapat menyebabkan kecacatan pada penderitanya atau bahkan kematian. Data Organisasi Stroke Dunia menyatakan bahwa setiap tahun terdapat 13,7 juta penderita stroke dan 5,5 juta kasus kematian akibat stroke. Penyakit ini merupakan penyakit mematikan nomor tiga dunia. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk melakukan penerapan algoritma K-Nearest Neighbor dalam memprediksi penyakit stroke serta dapat mengetahui akurasi yang dihasilkan algoritma KNN dalam memprediksi penyakit stroke. Melalui hasil perhitungan klasifikasi-prediksi pada data penyakit stroke dengan data latih sebanyak 80 data dan data uji sebanyak 20 data dapat diketahui bahwa algoritma KNN dapat melakukan prediksi penyakit stroke berdasarkan jenis kelamin, umur, hipertensi, riwayat penyakit jantung, status menikah, tipe pekerjaan, tipe tempat tinggal, rerata kadar glukosa, bmi dan status merokok dengan akurasi yang didapatkan sebesar 95% dengan nilai  $k=9$ .

**Kata kunci**—KNN, Prediksi, Stroke,

### Abstract

Stroke is a disorder of brain function caused by disruption of blood flow to the brain. Stroke can cause disability or even death. Data from the World Stroke Organization states that every year there are 13.7 million stroke sufferers and 5.5 million cases of death due to stroke. Stroke is the third deadliest disease in the world. Based on this, this study aims to apply the K-Nearest Neighbor algorithm in predicting stroke and can determine the accuracy of the KNN algorithm in predicting stroke. Through the results of classification-prediction calculations on stroke data with 80 training data and 20 test data, it was found that the KNN algorithm can predict stroke based on gender, age, hypertension, history of heart disease, marital status, type of work, type of place of residence, mean glucose levels, BMI and smoking status with an accuracy obtained of 95% with a value of  $k=9$ .

**Keywords**—KNN, Prediction, Stroke

## PENDAHULUAN

Stroke merupakan gangguan fungsi pada otak yang terjadi secara lokal maupun menyeluruh yang disebabkan oleh terganggunya aliran darah menuju otak [1]. Penyakit stroke pada umumnya dapat mengakibatkan kecacatan atau bahkan kematian. Berdasarkan data Organisasi Stroke Dunia setiap tahun terdapat 13,7 juta penderita stroke dan 5,5 juta kasus kematian akibat stroke, dimana stroke menjadi penyebab kematian nomor tiga di dunia setelah penyakit jantung coroner dan kanker [2].

Banyaknya kasus stroke, membuat prediksi penyakit stroke dibutuhkan sebagai pencegahan dan pendeteksian resiko penyakit stroke sejak dini. Dengan adanya pendeteksian penyakit stroke, penanganan penyakit stroke dapat dilakukan dengan cepat dan dapat memperbesar peluang penyembuhan. Prediksi penyakit stroke dapat dilakukan dengan

---

### Informasi Artikel:

Submitted: Mei 2022, Accepted: Mei 2022, Published: Mei 2022

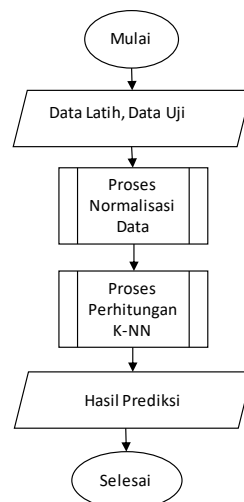
ISSN: 2685-4902 (media online), Website: <http://jurnal.umus.ac.id/index.php/intech>

menggunakan algoritma klasifikasi untuk melakukan pengolahan dan pembelajaran dataset yang ada sebagai penentu hasil prediksi. Algoritma *Analytical Hierarchy Process* (AHP) pada penelitian sebelumnya pernah digunakan untuk memprediksi penyakit stroke, dimana tingkat akurasi yang diperoleh sebesar 73,8% [3]. Berbeda dengan penelitian tersebut, pada penelitian ini peneliti akan mencoba meningkatkan akurasi yang dihasilkan dalam memprediksi penyakit stroke melalui metode *K-Nearest Neighbours*.

Algoritma *K-Nearest Neighbor* disingkat KNN merupakan salah satu metode klasifikasi data mining yang dapat digunakan untuk mengidentifikasi berbagai masalah, seperti penelitian yang telah ada yaitu digunakan untuk memprediksi penyakit diabetes [4], penjualan sepatu [5], hasil produksi [6] dan kelompok UKT mahasiswa [7]. Algoritma KNN pada umumnya digunakan dalam mengklasifikasikan objek berdasarkan pada data pembelajaran yang memiliki nilai selisih kecil dan jarak tetangga terdekat dengan objek. Selanjutnya hasil akhir dari penelitian ini dapat digunakan untuk memprediksi kemungkinan seseorang menderita penyakit stroke. Pada penelitian ini penggunaan algoritma *K-Nearest Neighbours* dipilih karena kemampuannya dalam menentukan kelompok data dengan menghitung jarak tetangga terdekat. Selain itu KNN juga memiliki nilai akurasi yang baik untuk melakukan klasifikasi dengan jumlah data yang besar namun juga memerlukan penentuan nilai parameter jumlah dari tetangga terdekat [8]. Dalam melakukan klasifikasi, algoritma KNN pernah dibandingkan dengan *Naïv Bayes* dan *Neural Network* untuk memprediksi penyakit hepatitis, dimana hasilnya menunjukkan bahwa rata-rata nilai akurasi yang didapatkan oleh KNN menghasilkan nilai akurasi yang lebih tinggi dibandingkan naïv bayes dan Neural Network yaitu 93% ;76,92%; dan 82,97% [9]. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk menerapkan algoritma KNN dalam memprediksi penyakit Stroke

## METODE PENELITIAN

*Dataset Stroke Disease* yang digunakan dalam prediksi penyakit stroke adalah dataset *public* yang berjumlah 100 data, yang diperoleh dari *Dataset Kaggle Repository*. Data yang diperoleh kemudian dilakukan pengolahan data untuk mendapatkan data pembelajaran yang sesuai dengan kebutuhan dalam melakukan prediksi penyakit stroke. Dalam melakukan prediksi penyakit stroke terdapat 12 variabel yang akan digunakan sebagai atribut untuk mengolah dataset, yang terdiri 11 variabel bebas dan 1 variabel terikat. Variabel bebas terdiri dari jenis kelamin, umur, hipertensi, penyakit jantung, pernah menikah, tipe pekerjaan, tipe tempat tinggal, rata-rata kadar glukosa, BMI, dan status merokok, sedangkan variabel terikat yaitu status stroke.



**Gambar 1.** Langkah Penelitian

Langkah penelitian ditunjukkan pada **Gambar 1**. Diawali dengan input data latih dan data uji yang diperoleh dari analisis kebutuhan pengumpulan *Dataset Stroke Disease*. Data yang diinputkan akan dinormalisasi untuk meminimalkan redundansi data dan mempermudah proses perhitungan. Selanjutnya dilakukan proses perhitungan KNN yang diawali dengan melakukan perhitungan jarak antara data uji dengan data latih, kemudian melakukan perangkingan data berdasarkan nilai jarak terdekat dan menentukan nilai parameter  $k$ . Dari data yang telah dilakukan perangkingan berdasarkan jarak terdekat dan nilai parameter  $k$  dapat menentukan hasil prediksi penyakit stroke.

**Tabel 1.** Deskripsi Atribut

Atribut	Deskripsi	Tipe Data
<b>Id</b>	Kode Pengenal	Numerik
<b>Jenis Kelamin</b>	Wanita, pria	Numerik
<b>Umur</b>	Umur Pasien (Tahun)	Numerik
<b>Hipertensi</b>	Ya, Tidak	Numerik
<b>Penyakit Jantung</b>	Ya, Tidak	Numerik
<b>Pernah Menikah</b>	Ya, Tidak	Numerik
<b>Tipe Pekerjaan</b>	Pekerjaan pemerintahan, tidak Bekerja, Pribadi, Wiraswasta, Anak-anak	Kategori
<b>Tipe Tempat Tinggal</b>	Perkotaan , Perdesaan	Kategori
<b>Rerata Kadar Glukosa</b>	Rata-rata kadar gula dalam darah (ml/g)	Numerik
<b>BMI</b>	Indeks Massa Tubuh	Numerik
<b>Status Merokok</b>	Tidak diketahui, Sebelumnya merokok, Tidak pernah merokok, Merokok	Kategori
<b>Stroke</b>	Tidak Stroke, Stroke	Numerik

\*Sumber: [www.kaggle.com](http://www.kaggle.com)

Berdasarkan atribut dari *Dataset Stroke Disease* pada **Tabel 1**, terdapat dua jenis tipe data yaitu numerik dan kategori. Dalam KNN perhitungan dapat dilakukan jika data bertipe numerik sehingga dilakukan pengolahan data bertipe kategori menjadi data bertipe numerik melalui pemberian bobot pada subkategori. Kemudian melakukan normalisasi data dengan menggunakan rumus normalisasi *Min-Max* yang ditunjukkan pada **persamaan (1)**.

$$X^* = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (1)$$

Dimana  $x$  merupakan nilai sebelum dinormalisasi,  $\max(X)$  untuk nilai maksimum dari parameter,  $\min(X)$  untuk nilai minimum dari parameter dan  $X^*$  adalah nilai hasil yang telah dinormalisasi. Setelah data dinormalisasi maka selanjutnya data dapat dilakukan perhitungan dengan menggunakan algoritma KNN.

K-Nearest Neighbour (KNN) merupakan salah satu klasifikasi data mining yang bisa digunakan dalam mencari pola baru sebuah data dengan menghubungkan pola data sebelumnya dengan data yang baru untuk mengklasifikasikan data dalam beberapa kelas berdasarkan atribut yang ada, algoritma KNN bekerja dengan menghitung jarak data untuk mengklasifikasikan berdasarkan data yang memiliki kemiripan yang paling besar[8]. Perhitungan jarak pada KNN menggunakan rumus *Euclidian Distance* yaitu[10]:

$$d_i = \sqrt{\sum_{i=1}^k (a_i - b_i)^2} \quad (2)$$

Dengan  $a_i$  adalah nilai pada data training,  $b_i$  adalah nilai pada data uji, sedangkan  $k$  adalah batas banyaknya data.

Langkah- langkah dalam algoritma KNN secara umum yaitu[7] :

1. Menentukan nilai  $k$ , dengan rumus  $k = \sqrt{n}$  (3)
2. Menghitung jarak Euclidian distance dengan rumus pada **Persamaan (2)**.
3. Melakukan pengurutan data dari data yang memiliki nilai jarak Euclid terkecil ke terbesar.
4. Melakukan klasifikasi data berdasarkan kelas terbanyak berdasarkan nilai  $k$  yang telah ditentukan.

Selanjutnya untuk melihat nilai akurasi dari algoritma KNN menggunakan rumus *Confusion matrix*, dengan **Persamaan (3)**.

$$\text{Akurasi} = \frac{\text{jumlah pengujian benar}}{\text{jumlah data testing}} \times 100\% \quad (4)$$

## HASIL DAN PEMBAHASAN

### Pengumpulan dan Pengolahan Data

*Dataset Stroke Disease* yaitu data pasien stroke yang diperoleh dari *Kaggle Repository*. *Dataset* yang diambil merupakan data dengan label stroke dan tidak stroke dengan jumlah 100 data. Data dengan label stroke berjumlah 50 data, dan data dengan label tidak stroke berjumlah 50 data. Sehingga perbandingan data yang digunakan adalah 1:1. *Dataset* dapat dilihat sebagaimana pada Tabel 1.

**Tabel 1.** Dataset Stroke Disease

No	Id	Jenis Kelamin	Umur	.	.	BMI	Status Merokok	Status
1	20463	1	81	.	.	28	Tidak diketahui	STROKE
2	44033	1	56	.	.	36	Merokok	STROKE
3	40460	0	68	.	.	41	Sebelumnya merokok	STROKE
4	19824	1	76	.	.	34	Merokok	STROKE
5	58267	1	70	.	.	46	Sebelumnya merokok	STROKE
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
98	61715	1	55	.	.	2	Merokok	TIDAK STROKE
99	44531	1	36	.	.	0	Merokok	TIDAK STROKE
100	40878	1	78	.	.	0	Sebelumnya merokok	TIDAK STROKE

Atribut yang digunakan berjumlah sepuluh atribut yang terdiri dari sub-atribut yang telah diberi nilai pembobotan sebagaimana pada tabel 2.

**Tabel 2.** Atribut dan Sub-Atribut

Kode Atribut	Nama Atribut	Sub-Atribut	Bobot
K1	Jenis Kelamin	Wanita	0
		Pria	1

*Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke*  
(Muhammad Naja Maskuri, Harliana, Kadek Sukerti, R. M. Herdian Bhakti)

<b>K2</b>	Umur	<15	0
		15-24	1
		25-34	2
		35-44	3
		45-54	4
		55-64	5
		65-74	6
		>75	7
<b>K3</b>	Hipertensi	Tidak	0
		Ya	1
<b>K4</b>	Penyakit Jantung	Tidak	0
		Ya	1
<b>K5</b>	Pernah Menikah	Tidak	0
		Ya	1
<b>K6</b>	Tipe Pekerjaan	Pekerja Pemerintahan	0
		Privat	1
		Wiraswata	2
		Tidak Pernah Bekerja	3
		Anak-anak	4
<b>K7</b>	Tipe Tempat Tinggal	Perkotaan	0
		Perdesaan	1
<b>K8</b>	Rerata Kadar Glukosa	<180mg/dl	0
		>=180mg/dl	1
<b>K9</b>	BMI	<30 kg/m	0
		30 kg/m	1
		>30kg/m	2
<b>K10</b>	Status Merokok	Tidak Pernah Merokok	0
		Sebelumnya Merokok	1
		Merokok	2
		Tidak Diketahui	3

*Dataset Stroke Disease* pada Tabel 1 selanjutnya dilakukan pembobotan sesuai dengan nilai bobot untuk setiap sub-atribut yang telah ditentukan sebelumnya (nilai pembobotan terdapat pada tabel 2). Data yang telah dilakukan pembobotan ditunjukkan pada Tabel 3.

**Tabel 3.** Dataset Hasil Pembobotan

No	Id	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Status
1	20463	1	7	1	1	1	2	0	1	0	3	STROKE
2	44033	1	5	1	0	1	2	1	1	2	2	STROKE
3	40460	0	6	1	1	1	2	0	1	2	1	STROKE
4	19824	1	7	1	0	1	2	1	1	2	2	STROKE
5	58267	1	6	1	0	1	2	1	1	2	1	STROKE
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
98	61715	1	5	0	0	1	2	1	0	2	2	TIDAK STROKE
99	44531	1	3	0	0	1	2	0	0	0	2	TIDAK STROKE
100	40878	1	6	0	0	1	3	1	0	0	1	TIDAK STROKE

Tahap selanjutnya adalah melakukan normalisasi terhadap *dataset*, dengan menggunakan metode normalisasi *Min-Max* sebagaimana pada persamaan 1. Satu data diambil sebagai contoh

perhitungan normalisasi yaitu data dengan id “44035” (dapat dilihat pada tabel 3 No 1), pada kolom “K2” hasil normalisasi dengan persamaan 1 didapatkan  $X^* = \frac{5-0}{7-0} = 0,714286$

Untuk hasil normalisasi seluruh *dataset* ditunjukkan pada Tabel 4. Didapatkan nilai hasil normalisasi memiliki rentang nilai yang seimbang dibandingkan dengan sebelum dilakukan normalisasi.

**Tabel 4.** Dataset Hasil Normalisasi

No	Id	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Status
1	20463	1	1	1	1	1	0.5	0	1	0	1	STROKE
2	44033	1	0.71	1	0	1	0.5	1	1	1	0.67	STROKE
3	40460	0	0.86	1	1	1	0.5	0	1	1	0.33	STROKE
4	19824	1	1	1	0	1	0.5	1	1	1	0.67	STROKE
5	58267	1	0.86	1	0	1	0.5	1	1	1	0.33	STROKE
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
98	61715	1	0.71	0	0	1	0.5	1	0	1	0.67	TIDAK STROKE
99	44531	1	0.43	0	0	1	0.5	0	0	0	0.67	TIDAK STROKE
100	40878	1	0.86	0	0	1	0.75	1	0	0	0.33	TIDAK STROKE

Pembagian data latih dan data uji menggunakan metode *split validation* dengan perbandingan 80% : 20%. Dengan perbandingan tersebut maka jumlah data latih adalah 80 data sedangkan data uji adalah 20 data. Hasil pembagian data latih metode *split validation* ditunjukkan pada Tabel 5. Sedangkan data uji ditunjukkan pada Tabel 6. Hasil dari pembagian data latih dan data uji ini selanjutnya akan digunakan dalam proses klasifikasi dengan menggunakan KNN.

**Tabel 5.** Data Latih

No	Id	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Status
1	20463	1	1	1	1	1	0.5	0	1	0	1	STROKE
2	44033	1	0.714286	1	0	1	0.5	1	1	1	0.666667	STROKE
3	40460	0	0.857143	1	1	1	0.5	0	1	1	0.333333	STROKE
4	19824	1	1	1	0	1	0.5	1	1	1	0.666667	STROKE
5	58267	1	0.857143	1	0	1	0.5	1	1	1	0.333333	STROKE
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
78	61715	1	0.714286	0	0	1	0.5	1	0	1	0.666667	TIDAK STROKE
79	44531	1	0.428571	0	0	1	0.5	0	0	0	0.666667	TIDAK STROKE
80	40878	1	0.857143	0	0	1	0.75	1	0	0	0.333333	TIDAK STROKE

**Tabel 6.** Data Uji

No	Id	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Status
1	43028	1	0.857143	0	0	1	0.75	1	0	0	0	TIDAK STROKE
2	15313	0	0.857143	1	0	1	0	0	1	1	0.333333	TIDAK STROKE
3	61300	1	0.142857	0	0	0	0.5	0	0	0	0.666667	TIDAK STROKE

*Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke  
(Muhammad Naja Maskuri, Harliana, Kadek Sukerti, R. M. Herdian Bhakti)*

4	72547	1	0.714286	0	0	1	0.5	1	0	1	0	TIDAK STROKE
5	41033	0	0.285714	0	0	1	0	1	0	1	0.333333	TIDAK STROKE
6	44001	0	0.428571	0	0	1	0.5	0	0	1	0	TIDAK STROKE
7	50491	1	1	0	0	1	0.75	0	0	0	1	TIDAK STROKE
8	56831	0	0.714286	0	0	1	0.5	0	0	0	1	TIDAK STROKE
9	50238	1	0	0	0	0	1	0	0	0	0	TIDAK STROKE
10	67921	0	0	0	0	0	1	0	0	0	0	TIDAK STROKE
11	5317	0	1	0	1	1	0.5	0	1	0	0.666667	STROKE
12	65105	1	1	0	0	1	0.5	0	1	0	0	STROKE
13	7937	1	0.714286	1	0	1	0	0	1	0	1	STROKE
14	43717	1	0.714286	1	0	1	0.5	0	1	1	1	STROKE
15	1210	0	0.857143	0	0	1	0.5	1	1	1	0	STROKE
16	29552	0	0.714286	1	1	1	0.5	0	1	1	1	STROKE
17	64778	1	1	0	1	1	0.5	1	1	1	0	STROKE
18	11762	0	1	0	0	1	0.5	0	1	1	0	STROKE
19	20387	0	0.857143	1	0	1	0.75	1	1	0	0.666667	STROKE
20	25974	1	1	0	0	1	0.75	0	1	0	0	STROKE

### Klasifikasi

Setelah dilakukan pengumpulan dan pengolahan data, selanjutnya adalah melakukan proses perhitungan klasifikasi data dengan menggunakan Algoritma K-Nearest Neighbor dengan langkah-langkah perhitungan sebagai berikut :

- 1) Menentukan nilai  $k$  dengan menggunakan persamaan 3, diperoleh  $k = \sqrt{80} = 8,94 = 9$ .
- 2) Melakukan perhitungan jarak (*Euclidean Distance*) setiap data uji dengan seluruh data latih. Berikut kasus perhitungan jarak *Euclid* data uji satu (data uji pada Tabel 6, No 1, dengan data latih pada Tabel 5, No 1) dengan menggunakan rumus *Euclidean Distance* sebagaimana pada persamaan 1 diperoleh :

$$d = \sqrt{(1-1)^2 + (1-0,857143)^2 + (1-0)^2 + (1-0)^2 + (1-1)^2 + (0,5-0,75)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{5,082908} = 2,2545306$$

Lanjutkan perhitungan jarak data uji satu dengan seluruh data pada data latih, sehingga diperoleh sebagaimana pada Tabel 7.

**Tabel 7.** Jarak Euclid Data Uji Satu Terhadap Data Latih

No	Id	Jarak Euclidean
1	20463	2.2545306
2	44033	1.8781248
3	40460	2.4846753
4	19824	1.8781248
5	58267	1.7814632
.	.	.
.	.	.
.	.	.
79	44531	1.3002376
80	40878	0.3333333

3) Hasil dari perhitungan jarak data uji satu dengan data latih sebagaimana Tabel 7, dilakukan pengurutan dari nilai jarak terkecil ke nilai jarak terbesar, kemudian dilakukan pemberian label kelas sesuai dengan data asli (data latih). Data yang telah diurutkan dan diberi label kelas ditunjukkan pada Tabel 8.

**Tabel 8.** Pengurutan jarak dan Pemberian Label Data Uji Satu

No	Id	Jarak Euclidean	Status
1	40878	0.333333	TIDAK STROKE
2	39927	0.831034	TIDAK STROKE
3	58422	0.863813	TIDAK STROKE
4	55051	0.912949	TIDAK STROKE
5	43872	1.040629	TIDAK STROKE
.	.	.	.
.	.	.	.
.	.	.	.
79	53401	2.450907	STROKE
80	40460	2.484675	STROKE

4) Langkah terakhir adalah melakukan klasifikasi kelas untuk data uji satu berdasarkan nilai jarak terdekat dan nilai  $k$ . Nilai  $k$  seperti yang telah ditentukan sebelumnya adalah 8, sehingga dari seluruh data pengurutan nilai jarak pada Tabel 8, akan diambil 9 data teratas, sebagaimana pada Tabel 9. Hasil prediksi penyakit stroke dapat dilihat dengan melihat perbandingan kelas yang ada. Jika jumlah data dengan kelas stroke lebih banyak dibandingkan dengan kelas tidak stroke maka hasil prediksinya adalah stroke. Dan jika jumlah data dengan kelas stroke lebih sedikit dari kelas tidak stroke maka hasil prediksinya adalah tidak stroke. Dari Tabel 9 didapatkan bahwa jumlah data dengan kelas tidak stroke berjumlah 9 sedangkan data dengan kelas stroke berjumlah 0, sehingga hasil prediksi data uji satu, “43028” adalah tidak stroke. Hasil prediksi dari keseluruhan data uji ditunjukkan pada Tabel 10.

**Tabel 9.** Hasil Prediksi Data Uji Satu dengan  $k = 9$

No	Id	Jarak Euclidean	Status
1	40878	0.333333	TIDAK STROKE
2	39927	0.831034	TIDAK STROKE
3	58422	0.863813	TIDAK STROKE
4	55051	0.912949	TIDAK STROKE
5	43872	1.040629	TIDAK STROKE
6	38131	1.210311	TIDAK STROKE
7	60810	1.235345	TIDAK STROKE
8	63561	1.235861	TIDAK STROKE
9	61715	1.235861	TIDAK



## STROKE

**Tabel 10.** Hasil Prediksi Seluruh Data Uji

No	Id	Status
1	43028	TIDAK STROK
2	37404	TIDAK STROK
3	61300	TIDAK STROK
4	72547	TIDAK STROK
5	41033	TIDAK STROK
6	44001	TIDAK STROK
7	50491	TIDAK STROK
8	56831	TIDAK STROK
9	50238	TIDAK STROK
10	67921	TIDAK STROK
11	5317	STROK
12	65105	TIDAK STROK
13	7937	STROK
14	43717	STROK
15	1210	STROK
16	29552	STROK
17	64778	STROK
18	11762	STROK
19	20387	STROK
20	25974	STROK

**Tabel 11.** Perbandingan Kelas Data Uji Prediksi Dan Data Uji Asli

No	Id	Data Asli	Data Prediksi	Keterangan
1	43028	TIDAK STROK	TIDAK STROK	Sesuai
2	37404	TIDAK STROK	TIDAK STROK	Sesuai
3	61300	TIDAK STROK	TIDAK STROK	Sesuai
4	72547	TIDAK STROK	TIDAK STROK	Sesuai
5	41033	TIDAK STROK	TIDAK STROK	Sesuai
6	44001	TIDAK STROK	TIDAK STROK	Sesuai
7	50491	TIDAK STROK	TIDAK STROK	Sesuai
8	56831	TIDAK STROK	TIDAK STROK	Sesuai
9	50238	TIDAK STROK	TIDAK STROK	Sesuai
10	67921	TIDAK STROK	TIDAK STROK	Sesuai
11	5317	STROK	STROK	Sesuai

12	65105	STROK	TIDAK STROK	Tidak Sesuai
13	7937	STROK	STROK	Sesuai
14	43717	STROK	STROK	Sesuai
15	1210	STROK	STROK	Sesuai
16	29552	STROK	STROK	Sesuai
17	64778	STROK	STROK	Sesuai
18	11762	STROK	STROK	Sesuai
19	20387	STROK	STROK	Sesuai
20	25974	STROK	STROK	Sesuai

Selanjutnya melakukan perhitungan akurasi dengan membandingkan kelas pada setiap data uji asli (dapat dilihat pada Tabel 6) dengan kelas pada setiap data uji hasil prediksi (dapat dilihat pada Tabel 10). Perbandingan label kelas antara data uji asli dengan label kelas data uji prediksi dapat dilihat pada Tabel 11.

Berdasarkan Tabel 11, sebanyak 19 kelas data uji dari hasil prediksi sesuai dengan data uji asli. Sehingga nilai akurasi dapat dihitung dengan menggunakan Persamaan 4 didapatkan akurasi =  $\frac{19}{20} \times 100\% = 95\%$ . Dengan demikian kinerja algoritma *K-Nearest Neighbor* dengan perbandingan data latih dan data uji 80% : 20% dengan nilai  $k = 9$ , memiliki nilai akurasi yang tinggi.

## KESIMPULAN

Penelitian ini melakukan pemodelan terhadap dataset penyakit stroke dengan menggunakan algoritma K-Nearest Neighbor. Dengan Jumlah data yang digunakan dalam penelitian sebanyak 100 data, pembagian data latih dan data uji menggunakan metode *split validation* dengan perbandingan 80%:20%. Nilai  $k$  yang digunakan untuk menguji kinerja algoritma KNN dalam memprediksi penyakit stroke yaitu  $k=9$ , dimana didapatkan nilai akurasi sebesar 95%. Berdasarkan hasil tersebut algoritma KNN memiliki tingkat akurasi yang baik dalam melakukan proses prediksi penyakit stroke.

## DAFTAR PUSTAKA

- [1] R. Ramadhan Saepul, S. Rizal Amegia, and F. Dasya Arif, "Komparasi Algoritma C4.5 Berbasis PSO Dan GA Untuk Diagnosa Penyakit Stroke," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 1, p. 155, 2020.
- [2] S. Putri Ayundari, "Diagnosis Dan Tatalaksana," *J. Med. Utama*, vol. 02, no. 01, pp. 402–406, 2020.
- [3] K. Ibrahim, S. Arief Andy, and R. Dian Eka, "Sistem Untuk Deteksi Penyakit Stroke Menggunakan Metode Analytical Hierarchy Process dan Weighted Product," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 6, pp. 2182–2191, 2018.
- [4] M. M Syukri and S. I Wayan, "Implementasi Algoritma K-Nearest Neighbor ( KNN ) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. VIII, no. 1, pp. 1–10, 2019.
- [5] H. Bagus and R. Fahrur, "Prediksi penjualan sepatu menggunakan metode k- nearest neighbor," *JOEICT(Jurnal Educ. Inf. Commun. Technol.*, vol. 04, pp. 13–18, 2020.
- [6] U. Farkhina Dwi, S. Amril Mutoi, and W. Deden, "Implementasi Algoritme K-Nearest Neighbor ( KNN ) untuk Prediksi Hasil Produksi," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 1, pp. 21–25, 2020.
- [7] Sukamto, A. Yanti, and A. Rizka, "Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *JUITA J. Inform.*, vol. 8, no. 1, p. 121, 2020, doi:

- 10.30595/juita.v8i1.6267.
- [8] M. Kukuh Wiliam, S. Yuita Arum, and A. Achmad, “Optimasi K-Nearest Neighbour Menggunakan Particle Swarm Optimization pada Sistem Pakar untuk Monitoring Pengendalian Hama pada Tanaman Jeruk,” vol. 2, no. 9, pp. 3333–3344, 2018.
  - [9] Sulastri, H. Kristophorus, and A. Muchamad Taufiq, “ANALISIS PERBANDINGAN KLASIFIKASI PREDIKSI PENYAKIT HEPATITIS DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR , NAÏVE BAYES DAN NEURAL NETWORK,” vol. 24, no. 2, pp. 82–91, 2019.
  - [10] S. B. Otonog and H. B. Raden Mohamad, “Penentuan Status Stunting pada Anak dengan Menggunakan Algoritma KNN,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 02, pp. 130–137, 2021, doi: 10.46772/intech.v3i02.533.