

NASKAH PUBLIKASI

**PERBANDINGAN METODE *FUZZY K-NEAREST NEIGHBOR* DAN
NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR UNTUK DETEKSI
PENYAKIT STROKE**

Program Studi Informatika



Disusun oleh:

SYAMSUL AJI NUGROHO

5150411038

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN ELEKTRO
UNIVERSITAS TEKNOLOGI YOGYAKARTA
2020**

NASKAH PUBLIKASI

**PERBANDINGAN METODE FUZZY K-NEAREST NEIGHBOR DAN
NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR UNTUK DETEKSI
PENYAKIT STROKE**



Disusun oleh:

SYAMSUL AJI NUGROHO

5150411038

Telah disetujui oleh dosen pembimbing



Dr. Enny Itje Sela, S.Si., M.Kom.

Tanggal : 29-01-2020.....

PERBANDINGAN METODE FUZZY K-NEAREST NEIGHBOR DAN NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR UNTUK DETEKSI PENYAKIT STROKE

Syamsul Aji Nugroho

Program Studi Informatika, Fakultas Teknologi Informasi dan Elektro

Universitas Teknologi Yogyakarta

Jl. Ringroad Utara Jombor Sleman Yogyakarta

E-mail : syamsul.ajinugroho@gmail.com

ABSTRAK

Stroke atau Cerebrovascular Accident (CVA) merupakan gangguan fungsi saraf yang disebabkan oleh gangguan aliran darah dalam otak dan menyebabkan gangguan pada aktivitas fungsional. Stroke merupakan penyebab kematian ketiga tersering di negara maju, setelah penyakit jantung dan kanker. Dengan terus bertambahnya penderita stroke tiap tahunnya, belum terdapat upaya efektif dalam menanggulangi penyakit baik dengan meningkatkan kesadaran masyarakat maupun pengelolaan penyakit stroke yang optimal. Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data besar berupa pengetahuan yang selama ini tidak diketahui secara manual. K-Nearest Neighbor merupakan metode yang paling simple, mudah diimplementasikan hanya dengan mengatur satu parameter k. Namun K-Nearest Neighbor juga memiliki beberapa kelemahan utama. Untuk mengatasi kekurangan tersebut dapat melakukan salah satu cara yaitu melakukan perbaikan dengan estimasi probabilitas kelas. Penelitian ini membandingkan metode Fuzzy K-Nearest Neighbor dan Neighbor Weighted K-Nearest Neighbor untuk melakukan deteksi penyakit stroke. Hasil pengujian dengan perbedaan jumlah data uji yaitu 50, 70, 90, 150 dan 200 dan menggunakan nilai ketetangaan (k) yaitu 17 sampai dengan 21 serta nilai eksponensial 2, maka didapatkan hasil rata-rata akurasi 81.272% dan 81.814% dengan menggunakan data seimbang, sedangkan dengan data tidak seimbang secara berurutan yaitu 82.45% dan 82.75%.

Kata kunci : *Neighbor Weighted K-Nearest Neighbor, Fuzzy K-Nearest Neighbor, Klasifikasi*

1. PENDAHULUAN

1.1 Latar Belakang

Stroke atau *Cerebrovascular Accident (CVA)* merupakan gangguan fungsi saraf yang disebabkan oleh gangguan aliran darah dalam otak dan menyebabkan gangguan pada aktivitas fungsional. Stroke merupakan penyebab kematian ketiga tersering di negara maju, setelah penyakit jantung dan kanker [1].

K-Nearest Neighbor merupakan metode yang paling simple, mudah diimplementasikan hanya dengan mengatur satu parameter k. Namun *K Nearest Neighbor* juga memiliki beberapa kelemahan utama antara lain sensitive terhadap fitur-fitur yang kurang relevan, ukuran ketetangaan k, data berderau maupun data pencilan, kompleksitas waktu untuk mencari tetangga terdekat setiap melakukan klasifikasi, dan kompleksitas memori untuk menyimpan semua data latih [2].

Untuk mengatasi kekurangan tersebut dapat melakukan salah satu cara yaitu melakukan perbaikan dengan estimasi probabilitas kelas. Algoritma KNN menggunakan metode suara terbanyak. Konsep ini rentan terhadap data derau atau pencilan. Untuk mengatasi permasalahan tersebut dapat menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* [2].

Akurasi metode *Fuzzy K-Nearest Neighbor* lebih baik dibandingkan dengan *K-Nearest Neighbor* karena tidak hanya mengambil keputusan berdasarkan jumlah kelas mayoritas tetapi menghitung nilai derajat keanggotaan pada tiap kelas yang berperan pada prediksi kelas [3]. Sedangkan menurut penelitian [4], metode *Neighbor Weighted K-Nearest Neighbor* merupakan metode yang lebih baik daripada *K-Nearest Neighbor* untuk mengatasi data tidak seimbang. Berdasarkan dari penelitian sebelumnya, maka disini penulis menggunakan

metode *Fuzzy KNearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor*, dengan tujuan untuk mengetahui metode yang terbaik untuk mendeteksi penyakit stroke.

1.2 Rumusan Masalah

Berdasarkan latar belakang maka rumusan masalah yang akan dibahas dalam penelitian ini adalah:

- a. Apakah metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* dapat mendeteksi penyakit stroke?
- b. Seberapa tinggi akurasi yang diperoleh sistem dengan menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* dalam mendeteksi penyakit stroke?

1.3 Batasan Masalah

Penelitian perbandingan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* untuk deteksi penyakit stroke, yang mencakup berbagai hal, sebagai berikut:

- a. Data yang digunakan adalah data dari data set situs repository Kaggle (<https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data>).
- b. Atribut yang digunakan pada penelitian ini adalah *gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, bmi (body mass index)*, dan *smoking status*.
- c. Sistem yang dibuat hanya untuk mendeteksi penyakit stroke dengan keluaran terkena stroke atau tidak terkena stroke.

2. LANDASAN TEORI

2.1 Stroke

Menurut [5], stroke adalah suatu keadaan yang timbul karena gangguan peredaran darah di otak yang menyebabkan terjadinya kematian jaringan otak sehingga mengakibatkan seseorang menderita kelumpuhan atau kematian atau defisit neurologis yang mempunyai serangan mendadak dan berlangsung 24 jam sebagai akibat dari *cardiovascular disease (CVD)*.

Stroke terdiri dari 2 jenis yaitu stroke iskemik dan stroke hemoragik. Stroke iskemik terjadi karena aliran darah di arteri otak terganggu dengan mekanisme yang mirip dengan gangguan aliran darah pada arteri coroner saat serangan jantung sehingga otak menjadi kekurangan oksigen dan nutrisi. Sedangkan stroke hemoragik disebabkan oleh pecahnya pembuluh darah otak. Darah yang keluar akan masuk ke dalam jaringan otak dan menyebabkan terjadinya pembengkakan otak atau

hematom yang akhirnya meningkatkan tekanan di dalam otak [6].

Untuk meminimalisir terkena penyakit stroke dapat dengan menghindari atau melakukan pencegahan faktor risiko yang menyebabkan stroke. Faktor penyebab terjadinya stroke dibagi menjadi dua yaitu faktor resiko yang tidak bisa diubah dan faktor resiko yang dapat diubah meliputi :

- a. Faktor resiko yang tidak dapat diubah antara lain adalah sebagai berikut:
 1. Usia, merupakan salah satu risiko stroke, semakin tua usia maka risiko terkena stroke akan semakin tinggi.
 2. Jenis Kelamin, laki-laki cenderung terkena stroke 3 kali bersiko dibandingkan perempuan namun tingkat kematian akibat stroke lebih banyak terjadi pada perempuan [6].
 3. Ras/bangsa, orang berkulit hitam memiliki risiko terkena stroke lebih tinggi dibandingkan dengan ras kaukasoid [5].
 4. Hereditas, riwayat stroke dalam keluarga, terutama jika dua atau lebih anggota keluarga pernah mengalami stroke pada usia kurang dari 65 tahun maka risiko terkena stroke akan meningkat 2,3 kali lebih besar.
- b. Faktor resiko yang dapat diubah antara lain adalah sebagai berikut:
 1. Hipertensi, semakin tinggi tekanan darah, maka kemungkinan terserang stroke semakin besar 3.8 kali lipat [6].
 2. Hiperkolesterol dan lemak, kadar kolesterol dalam darah yang tinggi pada laki-laki mempunyai risiko 0.80 kali terkena stroke sedangkan pada perempuan mempunyai 0.58 kali terkena stroke [6].
 3. Diabetes Militus.
 4. Penyakit Jantung.
 5. *Transient Ischemic Attack (TIA)*, merupakan tanda-tanda awal terjadinya stroke dan sekitar 10-50% dapat berkembang menjadi stroke komplit.
 6. Merokok, merokok beresiko 2 kali lebih besar terkena stroke karena mamacu peningkatan kekentalan darah, pengerasan dinding pembuluh darah dan penimbunan plak di dinding pembuluh darah [6].
 7. Obesitas, meningkatkan risiko stroke sebesar 15%.
 8. Kehamilan, Wanita yang sedang mengandung dapat mengalami hipertensi terutama tiga bulan sebelum melahirkan..
 9. Penyalahgunaan obat

10. Konsumsi alkohol, dapat mengganggu metabolisme tubuh sehingga terjadi *dyslipidemia*, diabetes mellitus, serta mempengaruhi berat badan dan tekanan darah.

2.2 Fitur

Fitur merupakan elemen setiap vektor yang mempunyai jenis yang beragam, misalnya pada data mengenai informasi fisik manusia ada fitur tinggi badan yang menggunakan nilai yang sifatnya kuantitatif [7]. Umumnya tipe fitur ada dua yaitu kategorikal (kualitatif) dan numerical (kuantitatif). Ada 4 sifat penting yang dimiliki fitur secara umum yaitu:

- Distinctness*, meliputi sama dengan (=) dan tidak sama dengan (\neq).
- Order*, meliputi lebih kecil (<), lebih kecil atau sama dengan (\leq), lebih besar (>) dan lebih besar atau sama dengan (\geq).
- Addition*, meliputi penjumlahan (+) dan pengurangan (-).
- Multiplication*, meliputi perkalian (*) dan pembagian (/).

Dari keempat sifat tersebut maka dapat diturunkan menjadi empat tipe fitur, yaitu nominal, ordinal, interval, dan rasio. Fitur nominal dan ordinal merupakan jenis kategorikal dan nilainya kualitatif. Nilai ini sebenarnya adalah nilai simbolik, tidak mungkin dilakukan operasi aritmatika seperti pada tipe numerik. Sementara fitur interval dan rasio keduanya merupakan jenis numerik dan nilainya kuantitatif. Nilai ini dapat dilakukan operasi aritmatika dan bisa dipresentasikan dengan nilai integer ataupun kontinu.

2.3 Konsep Kedekatan

Menurut [7], metode klasifikasi seperti *K-Nearest Neighbor* menggunakan suatu kuantitas yang disebut kedekatan. Ada dua jenis kedekatan yaitu kemiripan (*similarity*) atau ketidakmiripan (*dissimilarity*). Semakin dekat jarak kedua data maka semakin besar kemiripannya (semakin kecil ketidakmiripannya) dan semakin jauh jarak kedua data maka semakin kecil kemiripannya (semakin besar ketidakmiripannya).

Kadang ditemukan bahwa fitur yang digunakan dalam data tidak menggunakan tipe yang seragam. Misalnya dari 10 fitur yang digunakan terdapat 9 jenis numerik dan 1 jenis nominal. Cara yang cukup naif adalah dengan menggunakan jarak Manhattan. Menurut [7], untuk mengukur kedekatan vektor dengan fitur campuran tanpa melakukan konversi informasi didefinisikan sebagai berikut ini:

$$s(x, y) = \frac{\sum_{i=1}^r Si(x, y)}{\sum_{i=1}^r wi} \quad 1$$

$Si(x, y)$ adalah ukuran kemiripan di antara fitur ke- i dari x dan y , sedangkan w_i adalah faktor bobot yang berkorelasi dengan fitur i . Jika paling tidak satu dari fitur ke- i pada vektor x dan y tidak teridentifikasi, maka $w_i = 0$. Jika fitur ke- i adalah variable biner dan nilai kemiripannya 0 pada kedua vektor, maka $w_i = 0$. Jika selain dua syarat tersebut, maka $w_i = 1$. Menurut [7], jika nilai dua vektor adalah biner, maka:

$$Si(x, y) = \begin{cases} 1, & \text{jika } xi = yi \\ 0, & \text{lainnya} \end{cases} \quad 2$$

Fitur dari dua vektor x dan y yang bertipe nominal atau ordinal maka $si(x, y)$ bernilai 1 jika xi dan yi mempunyai nilai yang sama. Jika tidak sama, maka bernilai 0. Menurut [7], untuk ukuran kemiripan pada fitur dengan tipe numerik digunakan persamaan berikut:

$$Si(x, y) = 1 - \frac{|xi - yi|}{rangei} \quad 3$$

Rangei adalah panjang interval nilai fitur ke- i . Nilai yang diberikan oleh persamaan tersebut dalam jangkauan [0,1], nilai 1 jika kedua nilai fitur sama dan bernilai 0 jika kedua nilai fitur menempati jarak yang paling jauh di antara semua nilai pada fitur tersebut.

2.3 Fuzzy K-Nearest Neighbor

Fuzzy K-NN merupakan salah satu metode klasifikasi dengan menggabungkan teknik Fuzzy dan K-NN. Metode ini tidak seperti metode lain yang mana pada metode ini akan secara tegas memprediksi kelas yang diikuti oleh data uji berdasarkan perbandingan K terdekat. Dasar dari algoritma Fuzzy KNN adalah untuk menetapkan nilai keanggotaan sebagai fungsi jarak vektor dari KNN dan keanggotaan tetangga mereka di kelas kelas yang memungkinkan [3].

Untuk menghitung nilai keanggotaan pada Fuzzy K-NN, terlebih dahulu melakukan inisialisasi [3].

$$uij = \begin{cases} 0.51 + \left(\frac{nj}{k}\right) * 0.49, & \text{jika } j = i \\ \left(\frac{nj}{k}\right) * 0.49, & \text{jika } j \neq i \end{cases} \quad 4$$

Keterangan :

- uij = nilai keanggotaan kelas i pada vektor j
- nj = jumlah anggota kelas j pada suatu dataset K.
- k = banyaknya tetangga terdekat
- j = kelas target

Selanjutnya menghitung nilai keanggotaan masing masing kelas.

$$ui(x) = \frac{\sum_{j=1}^k uij(1/\|x - xj\|^{\frac{2}{m-1}})}{\sum_{j=1}^k (1/\|x - xj\|^{\frac{2}{m-1}})}$$

5

Keterangan :

- ui(x) = nilai keanggotaan
- k = jumlah tetangga terdekat yang digunakan
- x-xi = selisih jarak data uji dan data latih dalam k tetangga terdekat
- m = bobot pangkat (*weight exponent*) yang besarnya m>1

Setelah nilai tiap kelas ditemukan pilih nilai kelas yang tertinggi yang nantinya akan dijadikan kelas dari data uji.

2.4 Neighbor Weighted K-Nearest Neighbor

Metode *Neighbor Weighted K-Nearest Neighbor* merupakan pengembangan metode *K-Nearest Neighbor* dengan memberikan bobot yang kecil pada kelas mayoritas dan bobot besar pada kelas minoritas [8].

Tahapan-tahapan dari metode *Neighbor Weighted K-Nearest Neighbor* adalah sebagai berikut ini :

- a. Menentukan nilai K tetangga dan nilai Eksponen.
- b. Menghitung nilai kedekatan ketetanggan anantara data uji dan data latih (menggunakan ukuran kemiripan) dengan persamaan 1.
- c. Mengurutkan hasil perhitungan kedekatan ketetanggan berdasarkan jarak atau kedekatan dari yang terbesar ke terkecil.
- d. Menghitung jumlah tiap kelas.
- e. Menurut [8], perhitungan bobot

$$weight\ i = \frac{1}{\left(\frac{Num(c_i^d)}{Min\{Num(c_j^d)|n = 1, \dots, K *\}}\right)^{1/exp}}$$

6

Keterangan :

- Num (C_i^d) = banyaknya data latih d pada kelas i.
- Num (C_j^d) = banyaknya data latih d pada kelas j, dimana j terdapat dalam himpunan k tetangga terdekat.
- Exp = nilai ekponensial lebih dari 1.

- f. Setelah mendapatkan nilai bobot, maka menghitung nilai skor dengan cara mengalikan setiap jumlah skor dari tiap kelas dengan masing-masing bobot dari tiap kelas. Menurut [8], perhitungan skor pada metode *Neighbor Weighted K-Nearest Neighbor*.

Skor(X, Ci)

= Weight_i

$$* \left(\sum djNWKNN \left(\sqrt{\sum_{i=1}^n (X_{2i} - X_{1i}) * \delta(dj, Ci)} \right) \right)$$

7

Keterangan:

- Weight_i = bobot kelas i.
- djNWKNN(x) = data latih dj pada kumpulan tetangga terdekat dari data uji x.
- δ(dj,Ci) = akan bernilai 1 jika jarak ∈ Ci dan bernilai 0 jika jarak ∉ Ci.
- Ci = jenis atau kelas i.

- g. Setelah nilai skor kelas ditemukan pilih nilai skor kelas yang tertinggi yang nantinya akan dijadikan kelas dari data uji.

2.5 Matrik Konfusi

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar. Akan tetapi, tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa bekerja 100% benar. Oleh karena itu, sistem klasifikasi harus di ukur tingkat akurasi. Bisaanya cara mengukur tingkat akurasi klasifikasi menggunakan matriks konfusi.

Menurut [9], matriks konfusi adalah metode yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi.

Tabel 1: Matriks konfusi untuk klasifikasi 2 kelas

f_{ij}		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	f11	f10
	Kelas = 0	f01	f00

Berdasarkan isi matriks konfusi maka dapat diketahui jumlah data dari masing masing kelas yang diprediksi secara benar dan data yang didefinisikan salah. Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui tingkat akurasi dari hasil prediksi. Menurut [7], untuk mengetahui akurasi digunakan formula sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi benar}}{\text{Jumlah prediksi yang dilakukan}}$$

$$\text{Akurasi} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \quad 8$$

3. METODE PENELITIAN

3.1 Bahan/Data dan Peralatan

Pada penelitian objek yang diteliti berupa data penyakit stroke. Penyakit stroke adalah penyakit yang terjadi ketika pasokan darah menuju otak terganggu atau sama sekali berkurang, sehingga jaringan otak kekurangan oksigen dan nutrisi. Data yang digunakan berasal dari data set situs repository Kaggle (<https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data>).

Fitur-fitur yang digunakan pada penelitian ini adalah *gender*, *age*, *hypertension*, *heart disease*, *ever married*, *work type*, *residence type*, *avg glucose level* dengan satuan mg/dl, *bmi* (*body mass index*), dan *smoking status*. Input dalam proses ini adalah data latih dan data uji yang berasal dari data set Kaggle. Karena fitur yang ada dalam data set tersebut beraneka ragam terlihat pada Tabel 2. Maka untuk tipe fitur yang non-numerik diubah menjadi numerik dengan ketentuan pada Tabel 3. Setelah proses perubahan selesai maka data sudah berada di database sistem.

Tabel 2: Data Fitur

No	Fitur	Tipe
1.	Gender	Nominal
2.	Umur	Numerik
3.	Hipertensi	Nominal
4.	Jantung	Nominal
5.	Menikah	Nominal
6.	Pekerjaan	Nominal

Tabel 2: Data Fitur lanjutan

No	Fitur	Tipe
7.	Tinggal	Nominal
8.	Kadar Gula Darah	Numerik
9.	BMI	Numerik
10.	Merokok	Ordinal

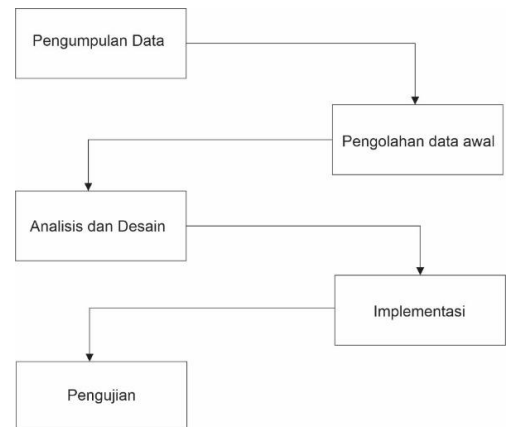
Tabel 3: Matriks konfusi untuk klasifikasi 2 kelas

No	Fitur	Aturan
1	Gender	Female = 0
		Male = 1
2	Menikah	Yes = 1
		No = 0
3	Pekerjaan	Private = 1
		Self-Employed = 2
		Govt_Job = 3
		Children = 4
4	Tinggal	Urban = 1
		Rural = 2
5	Merokok	Never = 1
		Formerly = 2
		Smokes = 3

3.1 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah langkah yang dilakukan oleh penulis dalam rangka mengumpulkan informasi serta melakukan pengolahan data yang telah di dapatkan. Metode penelitian memberikan gambaran mengenai rancangan penelitian yang meliputi antara lain: prosedur dan langkah-langkah yang harus di tempuh, perencanaan pengumpulan data, pengolahan data awal, analisis dan desain, implementasi dan pengujian.

Metode penelitian yang dilakukan penulis dalam penelitian perbandingan metode *fuzzy k-nearest neighbor* (FKNN), *neighbor weighted k-nearest neighbor* (NWKNN) untuk deteksi penyakit stroke adalah sebagai berikut:



Gambar 1: Diagram Proses Penelitian

a. Pengumpulan Data

Teknik pengumpulan data dengan melakukan studi pustaka dan mempelajari materi tentang stroke, faktor risiko stroke, klasifikasi, metode *fuzzy k-nearest neighbor* (FKNN), *neighbor weighted k-nearest neighbor* (NWKNN) dan lain sebagainya. Jumlah data yang didapatkan yang berasal dari Kaggle adalah data test 18.600 dan data train 43.400. Karena data test tidak ada data kelasnya maka yang akan digunakan adalah data train dengan kelas terkena stroke 783 dan tidak terkena stroke 42.617.

b. Pengolahan Data Awal

Pengolahan data awal merupakan tindak lanjut dari pengumpulan data yaitu melakukan preprocessing penghapusan miss data atau data yang tidak lengkap. Hal ini dilakukan untuk mengurangi tingkat noise saat melakukan klasifikasi. Setelah melakukan penghapusan data miss, didapatkan jumlah data train menjadi 29.072 dengan 548 kelas data stroke dan 28.524 data tidak stroke. Setelah itu, mengubah data-data tersebut dari yang non-numerik ke dalam bentuk numerik. Dari data train tersebut diambil secara acak, 600 data digunakan untuk data latih, 25 data uji yang akan dilakukan pengujian data persatu input, 200 data untuk pengujian data seimbang dan 150 data digunakan untuk pengujian data tidak seimbang.

c. Analisis dan Desain

1. Desain Input, untuk memasukkan data dan memprosesnya ke dalam format yang sesuai. Input data yang digunakan dalam sistem ini data yang berformat excel.
2. Desain Proses, tahapan untuk membuat sketsa yang akan terjadi pada setiap modul yang dimiliki sistem. Sketsa tersebut dijadikan acuan untuk membuat algoritma.
3. Desain Output, suatu hasil dari pengolahan yang berupa informasi yang dapat dilihat di sistem.
4. Desain Basis Data, pengembangan basis data yang akan dilakukan oleh sistem dengan menggunakan bahasa pemrograman PHP yang kemudian disimpan di MySQL.
5. Desain Interface.

d. Implementasi

Implementasi merupakan proses pembuatan sistem dari hasil perancangan. Dalam implementasi dilakukan dengan

menggunakan Bahasa pemrograman PHP dan basisdata MySQL.

e. Pengujian

Sistem yang dibuat kemudian dilakukan pengujian untuk mendapatkan akurasi dari setiap metode. Pengujian untuk metode *fuzzy k-nearest neighbor* dan *neighbor weighted k-nearest neighbor* dilakukan pengujian data uji seimbang dengan rasio jumlah kelas 50:50 dan tidak seimbang dengan rasio jumlah kelas, nilai k yang berbeda dan pengujian data uji dengan jumlah data uji yang berbeda serta nilai k sama.

4. ANALISA DAN PERANCANGAN SISTEM

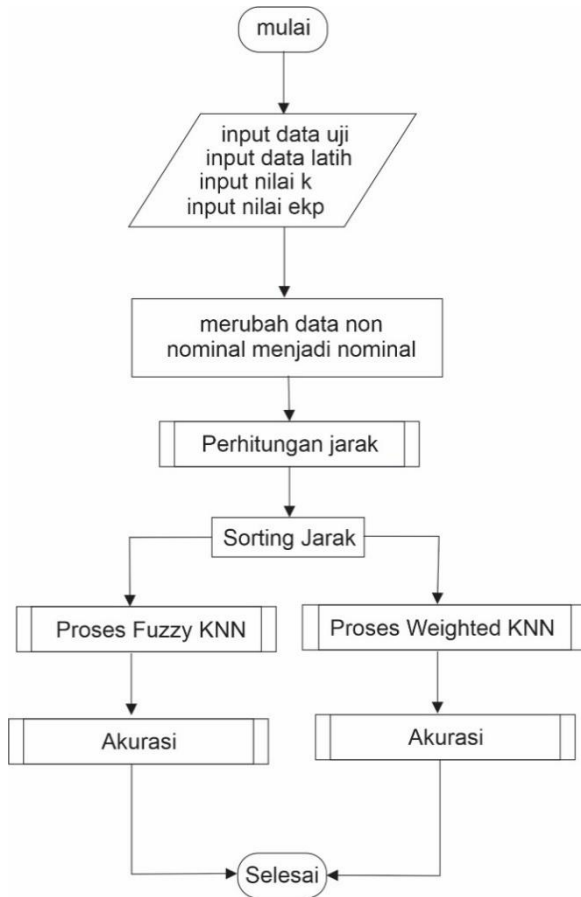
4.1 Analisis

Perancangan sistem deteksi penyakit stroke dengan metode *fuzzy k-nearest neighbor* dan *neighbor weighted k-nearest neighbor* membutuhkan data inputan untuk data latih yaitu *gender, age, hypertension, heart disease, ever married, work type, residence type, avg glucose level, bmi (body mass index), smoking status* dan *class*. Kemudian merubah data non numerik ke numerik dengan cara untuk fitur *gender* akan dibedakan menjadi 1 untuk laki-laki dan 0 untuk perempuan. Untuk fitur *ever married* akan dibedakan menjadi 1 untuk yes dan 0 untuk no. Fitur *work type* dibedakan menjadi 4 yaitu 1 untuk *private*, 2 untuk *self-employed*, 3 untuk *govt job*, dan 4 untuk *children*. Untuk fitur *residence type* dibedakan menjadi 2 yaitu 1 untuk *urban*, 2 untuk *rural*. Untuk fitur *smoking status* di bedakan menjadi 3 yaitu 1 untuk *never smoking*, 2 untuk *formerly smoking*, 3 untuk *smokes*. Setelah data dirubah, melakukan perhitungan jarak dengan persamaan 1 dan urutkan jarak dari yang paling terbesar. Setelah itu memasukkan nilai ketetanggaan (K) dan nilai ekponensial (lebih dari 1) untuk melakukan perhitungan dengan metode *fuzzy k-nearest neighbor* dan *neighbor weighted k-nearest neighbor*. Hitung tiap nilai kelas dan pilih nilai kelas terbesar sebagai hasil keputusan deteksi stroke atau tidak stroke. Setelah itu mencari tingkat akurasi dari deteksi yang di lakukan dengan persamaan 8.

Pada penelitian ini, penulis ingin mencoba mencoba menentukan kelas deteksi stroke menggunakan metode *fuzzy k-nearest neighbor* dan *neighbor weighted k-nearest neighbor*. Setelah itu, penulis akan membandingkan diantara kedua metode tersebut mana yang memiliki tingkat akurasi yang lebih baik.

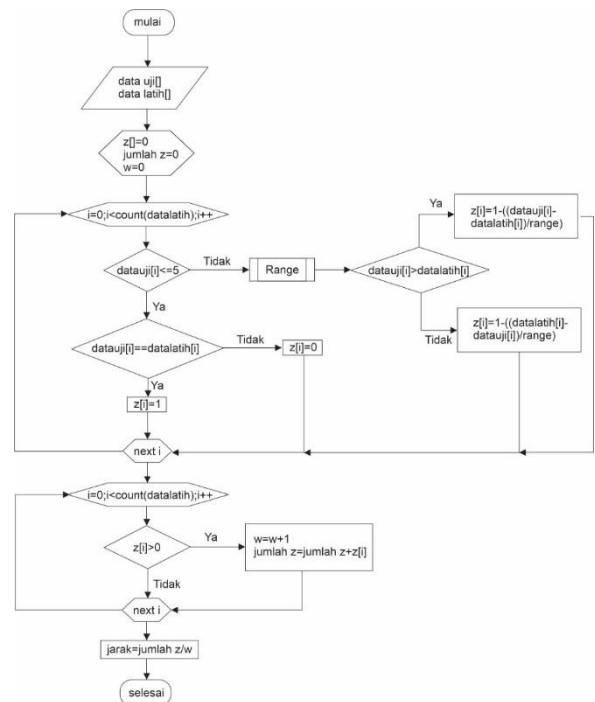
4.2 Rancangan Sistem

Rancangan sistem merupakan proses membuat gambaran sistem yang akan dibuat. Gambaran sistem ditampilkan melalui *flowchart*.



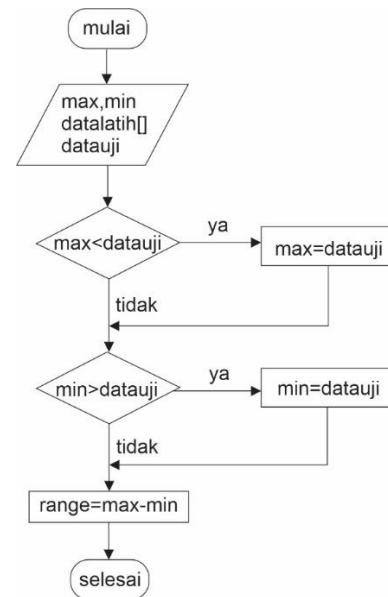
Gambar 2: Flowchart Deteksi Umum

Berikut ini adalah gambaran dari proses perhitungan jarak.



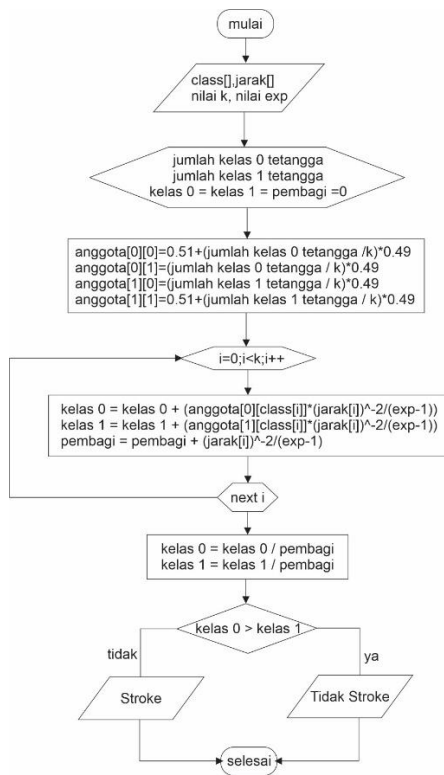
Gambar 3: Flowchart Sub Program Hitung Jarak

Berikut ini adalah gambaran dari proses perhitungan range.



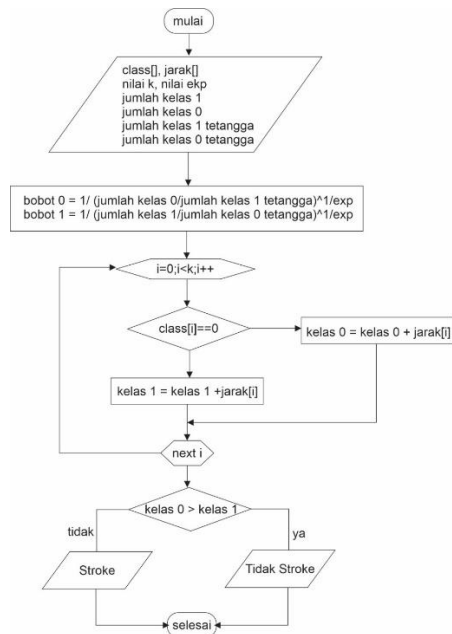
Gambar 4: Flowchart Sub-Proses Perhitungan Range

Berikut ini adalah gambaran dari proses perhitungan fuzzy KNN.



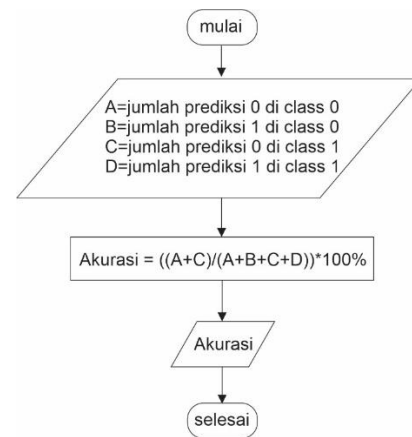
Gambar 5: Flowchart Sub-Proses Perhitungan Fuzzy KNN

Berikut ini adalah gambaran dari proses perhitungan NWKNN.



Gambar 6: Flowchart Sub-Proses Perhitungan NWKNN

Berikut ini adalah gambaran dari proses perhitungan akurasi.



Gambar 7: Flowchart Sub-Proses Perhitungan Akurasi

5. HASIL DAN PEMBAHASAN

5.1 Hasil

Hasil pengujian sistem dilakukan untuk memperoleh nilai akurasi dari proses deteksi penyakit stroke menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor*. Berikut ini skenario pengujian yang dilakukan:

a. Pengujian dengan data seimbang

Proses pengujian dilakukan dengan menggunakan jumlah data uji yang berbeda yaitu 50, 70, 90, 150 dan 200. Data yang digunakan adalah data seimbang dengan persentasi tiap kelas 50:50. Nilai keanggotaan yang digunakan yaitu 17 sampai dengan 21 serta nilai eksponensial yang digunakan adalah 2. Hasil akurasi dapat dilihat pada Tabel 4.

Tabel 4: Akurasi pengujian dengan data seimbang

Data	K	FKNN	NWKNN
50	17	78%	78%
	18	82%	78%
	19	82%	80%
	20	82%	80%
	21	82%	80%
70	17	78.57%	78.85%
	18	77.14%	81.43%
	19	81.43%	81.43%
	20	75.71%	81.43%
	21	78.57%	78.57%
90	17	82.22%	82.22%
	18	83.33%	85.56%
	19	84.44%	84.44%
	20	81.11%	84.44%
	21	84.44%	84.44%

Tabel 4: Akurasi pengujian dengan data seimbang lanjutan

Data	K	FKNN	NWKNN
150	17	83.33%	83.33%
	18	82%	83.33%
	19	83.33%	83.33%
	20	82.67%	83.33%
	21	82%	82%
200	17	82.5%	82.5%
	18	80%	83.5%
	19	81.5%	81.5%
	20	80.5%	83%
	21	81%	81%
Rata-rata		81.272%	81.814%

b. Pengujian dengan data tidak seimbang

Proses pengujian dilakukan dengan menggunakan jumlah data uji yang sama yaitu 100. Data yang digunakan adalah data tidak seimbang dengan persentasi 40 data stroke : 60 data tidak stroke, 60 data stroke : 40 data tidak stroke, 70 data stroke : 30 data tidak stroke dan 30 data stroke data : 70 data tidak stroke. Nilai keanggotaan yang digunakan yaitu 17 sampai dengan 21 serta nilai eksponensial yang digunakan adalah 2. Hasil akurasi dapat dilihat Tabel 5.

Tabel 5: Akurasi pengujian dengan data tidak seimbang

Perbandingan	K	100	
		FKNN	NWKNN
40 data stroke, 60 data tidak stroke	17	83%	84%
	18	84%	84%
	19	83%	83%
	20	84%	83%
	21	83%	82%
60 data stroke, 40 data tidak stroke	17	82%	83%
	18	81%	86%
	19	83%	84%
	20	81%	85%
	21	84%	83%
30 data stroke, 70 data tidak stroke	17	85%	81%
	18	83%	80%
	19	82%	80%
	20	83%	80%
	21	83%	80%
70 data stroke, 30 data tidak stroke	17	83%	84%
	18	80%	83%
	19	81%	83%
	20	80%	83%
	21	81%	84%
Rata-rata		82.45%	82.75%

c. Pengujian dengan per satu data uji

Proses ini dilakukan menggunakan 600 data latih dan 25 data uji yang diambil secara acak dari data set. Proses pengujian dilakukan dengan cara satu-satu secara bergantian. Nilai keanggotaan yang digunakan yaitu 7 serta

nilai eksponensial yang digunakan adalah 2. Hasil akurasi dapat dilihat pada Tabel 6.

Tabel 6: Hasil pengujian dengan 1 data uji

No	Asal	FKNN	NWKNN
1	Stroke	Stroke	Stroke
2	Stroke	Stroke	Stroke
3	Stroke	Tidak Stroke	Tidak Stroke
4	Stroke	Tidak Stroke	Tidak Stroke
5	Stroke	Stroke	Stroke
6	Stroke	Tidak Stroke	Tidak Stroke
7	Stroke	Stroke	Stroke
8	Stroke	Stroke	Stroke
9	Stroke	Stroke	Stroke
10	Tidak Stroke	Tidak Stroke	Tidak Stroke
11	Tidak Stroke	Tidak Stroke	Tidak Stroke
12	Tidak Stroke	Tidak Stroke	Tidak Stroke
13	Tidak Stroke	Tidak Stroke	Tidak Stroke
14	Tidak Stroke	Tidak Stroke	Tidak Stroke
15	Tidak Stroke	Stroke	Stroke
16	Tidak Stroke	Tidak Stroke	Tidak Stroke
17	Tidak Stroke	Tidak Stroke	Tidak Stroke
18	Tidak Stroke	Tidak Stroke	Tidak Stroke
19	Tidak Stroke	Tidak Stroke	Tidak Stroke
20	Tidak Stroke	Tidak Stroke	Tidak Stroke
21	Tidak Stroke	Stroke	Stroke
22	Tidak Stroke	Tidak Stroke	Tidak Stroke
23	Tidak Stroke	Tidak Stroke	Tidak Stroke
24	Tidak Stroke	Tidak Stroke	Tidak Stroke
25	Stroke	Tidak Stroke	Tidak Stroke

Perhitungan akurasi dari metode Fuzzy K-Nearest Neighbor dan Neighbor Weighted K-Nearest Neighbor dengan persamaan 8 sebagai berikut:

Tabel 7: Matriks Konfusi FKNN

f _{ij}	Kelas hasil prediksi (j)		
	Kelas = 1	Kelas = 0	
Kelas asli (i)	Kelas = 1	8	1
	Kelas = 0	4	12

$$\begin{aligned}
 \text{Akurasi FKNN} &= \left(\frac{6 + 13}{6 + 4 + 2 + 13} \right) * 100\% \\
 &= 76\%
 \end{aligned}$$

Tabel 8: Matriks Konfusi NWKNN

f _{ij}		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	8	1
	Kelas = 0	4	12

$$\text{Akurasi FKNN} = \left(\frac{6 + 13}{6 + 4 + 2 + 13} \right) * 100\% = 76\%$$

5.2 Pembahasan

Penelitian ini memfokuskan pada perbandingan metode *Fuzzy K-Nearest Neighbor* dengan *Neighbor Weighted K-Nearest Neighbor* untuk mendeteksi penyakit stroke. Data yang digunakan yakni dataset penyakit stroke situs repository Kaggle (<https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data>).

Pada penelitian ini didapatkan bahwa akurasi tertinggi di pengujian dengan jumlah dataset berbeda yaitu 50, 70, 90, 150 dan 200, nilai keanggotaan yang digunakan yaitu 17 sampai dengan 21 serta nilai eksponensial yang digunakan adalah 2. Didapatkan hasil dengan menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* secara berurutan rata-rata akurasi 81.272% dan 81.814% dengan menggunakan data seimbang, sedangkan dengan jumlah data uji yang sama yaitu 100. Data yang digunakan adalah data tidak seimbang dengan persentasi 40 data stroke : 60 data tidak stroke, 60 data stroke : 40 data tidak stroke, 70 data stroke : 30 data tidak stroke dan 30 data stroke data : 70 data tidak stroke secara berurutan yaitu 82.45% dan 82.75%.

Sedangkan untuk pengujian dengan satu data uji disimpulkan bahwa akurasi dengan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* sama, yaitu 76%. Hasil tersebut tidak lebih baik dari penelitian [10], dengan menggunakan metode Dempster Shafer mendapatkan tingkat akurasi 90%. Hal ini dikarenakan berbedanya jenis fitur yang digunakan dan data yang digunakan berbeda sumber.

6. PENUTUP

6.1. Kesimpulan

Setelah melakukan penelitian ini didapatkan beberapa kesimpulan antara lain:

- Sistem dapat mendeteksi penyakit stroke dengan menerapkan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor*.
- Berdasarkan hasil pengujian dengan data uji seimbang dengan menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* secara berurutan rata-rata

akurasi 81.272% dan 81.814%, sedangkan untuk hasil pengujian data uji tidak seimbang dengan beragam rasio jumlah kelas di dapatkan hasil secara berturut-turut 82.45% dan 82.75%. Untuk hasil pengujian dengan jumlah data latih 600 dan data uji 25 di ambil secara acak maka didapatkan hasil tingkat akurasi menggunakan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* sama yaitu 76%.

6.2. Saran

Berikut ini adalah saran yang diberikan untuk proses pengembangan selanjutnya agar sistem menjadi lebih baik:

- Ditambahkan persentase akurasi deteksi pada tampilan deteksi umum untuk lebih meyakinkan pengguna.
- Dapat mencari alternative perhitungan jarak dengan fitur campuran yang lain seperti menggunakan jarak Manhattan.
- Sistem dikembangkan ke perangkat mobile seperti android atau ios.

DAFTAR PUSTAKA

- Maulina, M. and Rahayu, M.S. (2018), "Korelasi Rasio Kolesterol Total Terhadap HDL dengan prediksi outcome stroke iskemik akut," *Qanun Medika*, Vol 2(1), 59–71.
- Suyanto (2018), "Machine Learning Tingkat dasar dan lanjut," Bandung: Informatika.
- Nugraha, S.D., Regasari, R., Putri, M. dan Wihandika, R.C. (2018), "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi," *Jurnal Pengembangan Teknologi dan Ilmu Komputer*, Vol 1(9), 925-932.
- Ridok, A. dan Latifah, R. (2015), "Klasifikasi Teks Bahasa Indonesia Pada Corpus Tak Seimbang Menggunakan NWKNN," *Journal of Environment Engineering & Sustainable Technology*, Vol 2(4), 9–10.
- Agustina, B.D. (2019), "Asuhan Keperawatan Gawat Darurat Pada Pasien Dengan Stroke diinstalasi gawat darurat rumah sakit umum daerah sleman yogyakarta," Skripsi, A.Md.Kep., Akademi Kesehatan Karya Husada, Yogyakarta.
- Jayanti, A.A. (2015), "Hubungan Hipertensi dengan kejadian stroke di Sulawesi Selatan tahun 2013," Skripsi, S.Kes, Universitas Islam Negeri Syarif Hidayatullah, Jakarta.
- Prasetyo, E. (2014), "Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab," A. Sahala, Ed. Yogyakarta: ANDI Publisher..
- Hadi, A.H., Ratnawati, D.E. dan Dewi, C. (2018), "Identifikasi Penyakit Gagal Ginjal Menggunakan Metode *Neighbor Weighted K-Nearest Neighbor*

- (NWKNN),” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol 2(9), 2562–2569.
- [9] Abdillah, S. (2015), “Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santra Maria Pematang,” Jurnal Teknik Informatika, Vol 3(4), 1–12.
- [10] Putri Indraswari, D., Andy Soebroto, A., dan Arisetijono Marhaendraputro, E. (2015). “Sistem Pendukung Keputusan Deteksi Dini Penyakit Stroke Menggunakan Metode Dempster-Shafer,” Journal of Environmental Engineering and Sustainable Technology, Vol 2(2), 97–104.
- [11] Sela E. I. dan Ihsan M., (2017), “Deteksi Kualitas Telur Menggunakan Analisis Tekstur,” IJCCS, Vol 11(2), 199-208.