

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Stroke saat ini menjadi penyebab utama kematian dan kecacatan di Indonesia. Pada tahun 2014, stroke menjadi penyebab kematian (mortalitas) utama dengan pangsa 21,1%. Menurut data WHO, 1 dari 6 orang di dunia mengalami stroke, dengan kasus stroke baru terjadi setiap 2 detik dan 80% kasus stroke mengalami kekambuhan, yang lebih tepatnya harus diantisipasi lebih baik (Rinaldy, dkk, 2021).

Orang yang tidak mengetahui gejala stroke adalah sumber dari banyak masalah yang muncul selama ini. Akibatnya, dokter harus cepat mengenali gejala stroke. Dokter harus mengetahui gejala pasien untuk mengidentifikasi stadium penyakit. Beberapa orang yang terkena stroke akan mengalami kecacatan atau bahkan meninggal dunia (Rinaldy, Soebroto, & Setianto, 2021).

Salah satu cara meminimalisir angka kasus stroke di Indonesia yaitu dengan melakukan deteksi dini penyakit stroke. Namun, stroke adalah sebuah penyakit kompleks yang menunjukkan gejala berbeda antara satu penderita dengan penderita lain sehingga deteksi dini dilakukan dengan teliti. Dokter dapat mengalami kendala dalam melakukan deteksi dini penyakit stroke karena gejala yang berbeda antar pasien, misalnya saja waktu yang dibutuhkan untuk observasi risiko penyakit stroke lebih lama, atau bahkan kesalahan diagnosis karena gejala penyakit stroke yang beragam. Hasil diagnosis antara satu dokter dengan dokter lainnya juga dapat berbeda karena pengetahuan dokter dan perbedaan literatur yang digunakan sehingga hasil diagnosis kurang obyektif (Rinaldy, dkk, 2021).

Berbagai macam permasalahan yang melibatkan pembagian data secara sistematis lebih mudah diselesaikan menggunakan metode klasifikasi. Sebagai contoh dalam bidang medis yaitu, menggunakan klasifikasi yang berfungsi untuk mengklasifikasi luasnya penyakit yang diderita pasien yang dapat

memudahkan para klinisi untuk mendiagnosis penyakit stroke (Meristika, 2013).

Klasifikasi merupakan objek berlandaskan ciri aturan atau *standard* pada objek tersebut. Proses klasifikasi akan membangun suatu model yang berupaya bisa membedakan data ke dalam *atribute* yang berbeda berdasarkan fungsi tersebut. Model tersebut dapat berupa kombinasi dari *k-nearest neighbor*, *fuzzy k-nearest neighbor*, *fuzzy k-nearest neighbor in every class* dan lain-lain (Yunita, 2016).

Dalam ulasan yang dipimpin oleh (Maskuri, Harliana, Sukerti, & Bhakti, 2022) dengan judul “Penerapan *Algoritma K-Nearest Neighbor* (KNN) untuk Memprediksi Penyakit Stroke”. Memiliki nilai K sebesar 9 dan akurasi sebesar 95%, berdasarkan 20 data uji dan 80 data latih.

Pada penelitian berjudul “Perbandingan metode *Fuzzy K-Nearest Neighbor* dan *Neighbor Weighted K-Nearest Neighbor* untuk deteksi penyakit stroke” yang dilakukan oleh (Nugroho, 2020). Hasil pengujian berbagai data antara lain 50, 70, 90, 150, dan 200, serta nilai k yang berkisar antara 17 sampai 21, dan nilai eksponensial 2, mengungkapkan akurasi rata-rata 81,272% dan 81,814% dengan data seimbang, dibandingkan dengan 82,45% dan 82,75% dengan data tidak seimbang. data berurutan..

“Sistem Pendukung Keputusan Diagnosis Penyakit Stroke menggunakan Metode *Fuzzy K-Nearest Neighbor* (FK-NN) (Studi Kasus Puskesmas Kendal Kerep Kota Malang) demikian judul penelitian yang dilakukan oleh (Rinaldy, Soebroto, & Setianto, 2021). sistem tersebut telah disusun dengan memanfaatkan metode FK-NN menjadi risiko stroke rendah, bahaya stroke sedang dan risiko stroke tinggi. Studi ini menunjukkan bahwa metode FK-NN secara akurat mendeteksi stroke dini dengan akurasi masing-masing 61,1% pada k = 4 dan 50% data latih.

Pada penelitian yang dilakukan oleh (Prasetyo, 2012) dengan judul “*Fuzzy K-Nearest Neighbor in every Class* untuk Klasifikasi Data”. Pada penelitian ini dilakukan perhitungan nilai keanggotaan data uji pada setiap kelas dengan basis akumulasi jarak K tetangga terdekat yang ditemukan. Kelas dengan nilai

keanggotaan terbesar akan dipilih sebagai kelas hasil prediksi. Akurasi yang didapatkan dari pengujian yang dilakukan dalam penelitian ini menunjukkan bahwa akurasi yang dihasilkan *Fuzzy K-Nearest Neighbor in Every Class* relatif lebih tinggi dari pada *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor*, yaitu 97%. Keakuratan klasifikasi objek sangat penting, dan metode klasifikasi yang baik adalah metode yang dapat membantu meningkatkan nilai akurasi, yaitu metode *Fuzzy K-Nearest Neighbor in every Class* (Rosyid, Prasetyo, & Agustin, 2016).

*Fuzzy K-Nearest Neighbor in every class (FK-NNC)* merupakan pengembangan dari *K-Nearest Neighbor (K-NN)* dan *Fuzzy K-Nearest Neighbor (FK-NN)*. Metode ini sedikit memodifikasi *Fuzzy K-Nearest Neighbor (FK-NN)* dengan memberikan beberapa  $k$  tetangga terdekat untuk setiap label kelas dari data uji (setiap label kelas memiliki jumlah tetangga yang sama sebanyak  $k$ ). Label dalam metode ini sangat berpengaruh, yang bertujuan untuk mengurangi kelas yang tidak seimbang (Dhya'attulhaq, dkk, 2021)

Berdasarkan hal tersebut maka penulis melakukan suatu penelitian mengenai “Klasifikasi Penyakit Stroke Menggunakan *Fuzzy K-Nearest Neighbor in every Class*”. Dengan terbuatnya penelitian ini, diharapkan terciptanya suatu analisis yang dapat digunakan untuk mendiagnosis penyakit Stroke dengan tingkat akurasi yang baik.

## 1.2 Rumusan Masalah

Berapa tingkat akurasi, *recall* dan *precision* yang dipengaruhi sejumlah nilai data  $k$  pada *dataset* Stroke menggunakan algoritma *Fuzzy K-Neares Neighbor in every Class*.

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah mengukur tingkat akurasi, *recall* dan *precision* penyakit stroke menggunakan metode *Fuzzy K-Nearest Neighbor in every Class*.

## 1.4 Manfaat Penelitian

Dalam penelitian kali ini diharapkan menghasilkan manfaat, yaitu:

1. Dapat mengetahui performa *Fuzzy K-Nearest Neighbor in every Class* dalam mengklasifikasi penyakit Stroke.
2. Dapat dijadikan sarana rujukan penelitian selanjutnya.

### 1.5 Batasan Penelitian

Adapun hal-hal yang digunakan dalam penelitian dibatasi pada masalah yang dibahas, yaitu:

- a. Data Stroke yang digunakan untuk penelitian ini diperoleh dari <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset> tahun 2020.
- b. Jumlah data yang digunakan sebanyak 500 data.
- c. Terdapat 10 atribut yang digunakan yaitu, *gender, age, hypertension, heart disease, ever married, residence type, avg glucose level, bmi (body mass index), dan smoking status, stroke*.
- d. Output dari diagnosa Stroke yaitu *positive* Stroke (1) dan *negative* Stroke (0).
- e. Pembangunan model menggunakan *platform Google Colab*.