

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Stroke adalah kondisi bahaya yang perlu ditangani secepatnya, karena sel otak dapat mati dalam hitungan menit. Menurut data Riskedas, prevalensi stroke di Indonesia di tahun 2013 meningkat dibandingkan dengan data 2007 dengan nilai 8,3%, naik hingga 12,1% per 1.000 orang (RI, 2013). Gejala yang biasanya terjadi pada penyakit stroke antara lain pada bagian wajah, lengan yang sering merasah lemas. Efek dari stroke bermacam – macam tergantung pada bagian otak mana yang mengalami gangguan dan seberapa parah gangguan tersebut. Kematian secara mendadak mungkin terjadi apabila seorang pasien mengalami yang sangat parah (Woto, 2014).

Penanganan yang cepat dapat mengurangi tingkat kerusakan pada otak dan kemungkina timbulnya komplikasi. Maka perlu dilakukan prediksi orang tersebut terkena penyakit stroke atau tidak. Salah satu cara untuk memprediksi penyakit stroke yaitu menggunakan klasifikasi. Penyakit stroke perlu diklasifikasi agar dapat memprediksi penyakit dengan akurat. Hasil prediksi yang akurat membantu praktisi kesehatan dalam mengambil keputusan dengan tepat.

Klasifikasi merupakan proses penemuan model deskripsi dan pembeda sebuah kelas data atau konsep yang dirancang untuk bisa digunakan memprediksi kategori objek yang labelnya belum diketahui (Kamber, 2006). Banyak algoritma yang dapat diaplikasikan untuk mengklasifikasi sebuah data. Beberapa diantaranya yaitu K-Nearest Neighbor dan Gaussian Naive Bayes. K-Nearest Neighbor yang memiliki konsisten yang kuat, dengan mencari kasus berdasarkan pencocokan bobot antara kasus lama dengan kasus baru (Kusrini dan Lutfhi, 2009). K-Nearest Neighbor merupakan teknik yang sederhana, efisiensi dan efektif dalam bidang pengenalan pola, kategori teks, pengolahan objek dan lain lain, karena sederhana dalam pengolahannya dan dapat melakukan training data dalam jumlah yang besar. Penerapan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* sebagai proses

klasifikasi yang dirancang untuk mendiagnosa penyakit dengan mengklasifikasi gejala yang dialami pasien dengan memperoleh nilai akurasi tertinggi. Maka dari itu, penulis memutuskan menggunakan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes* pada proses klasifikasi dalam melakukan perbandingan dan menentukan algoritma yang memiliki akurasi lebih tinggi dalam proses klasifikasi.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Wiria Nugraha Deni dkk (2017) membuat sistem untuk klasifikasi penyakit stroke menggunakan metode *Naive Bayes Classifier* menggunakan data Rumah Sakit Umum daerah Undita Palu. Berdasarkan pengujian yang telah dilakukan memperoleh hasil akurasi 89,65%. Sedangkan penelitian yang dilakukan oleh Y A Gerhana dkk (2019) dengan judul perbandingan algoritma *Naive Bayes* dan C.45 dalam memprediksi masa belajar siswa, pada penelitian ini mendapatkan hasil akurasi dari algoritma *Naive Bayes* lebih baik yaitu sebesar 88%, sedangkan dengan menggunakan algoritma C.45 mendapatkan hasil akurasi sebesar 87%. Pada penelitian yang dilakukan oleh Devi Yunita (2017) yang berjudul perbandingan algoritma *Decision Tree* untuk penentuan resiko kredit kepemilikan mobil, pada penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* mendapatkan hasil akurasi yang akurat dalam penentuan kelayakan konsumen dengan nilai akurasi sebesar 98,18%. Pada penelitian yang dilakukan oleh Arma Adzhana dkk (2020) yang berjudul penerapan *Naive Bayes Gaussian* pada klasifikasi jenis jamur berdasarkan ciri statistik orde pertama, pada penelitian pengujian menggunakan *cross validation* dengan nilai  $K=4$  didapatkan nilai akurasi sebesar 98,75%. Melihat dari penelitian-penelitian sebelumnya perbedaan dengan penelitian sebelumnya yaitu dapat dilihat dari data yang digunakan, atribut yang digunakan dan pada tujuan penelitian ini.

Berdasarkan penjelasan diatas, maka penulis tertarik untuk mengklasifikasi penyakit stroke dengan judul “Perbandingan Metode *K-Nearest Neighbor* Dan *Gaussian Naive Bayes* Untuk Klasifikasi Penyakit Stroke”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang didapat yaitu berapa hasil perbandingan akurasi, presisi, *recall* pada klasifikasi penyakit stroke dengan menggunakan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes*?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini adalah mengetahui nilai perbandingan akurasi, presisi, *recall* pada klasifikasi penyakit stroke dengan menggunakan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes*.

## 1.4 Batasan Masalah

1. Menggunakan data penyakit stroke yang dipublish pada tanggal 27 Januari 2021 yang bersumber dari situs Kaggle.com.
2. Jumlah data yang digunakan sebanyak 360.
3. Pada penelitian ini menggunakan 10 atribut yaitu *Gender, Age, Hypertension, Heartdisease, Ever married, Work type, Residence type, Avg glucose level, Bmi, Smoking status*.
4. *K-fold* yang digunakan pada penelitian ini yaitu *k-fold* 2 sampai 10.
5. Pada penelitian ini menggunakan tools *Jupyter Notebook*.

## 1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu :

1. Untuk mengetahui perbandingan algoritma yang terbaik.
2. Bisa digunakan sebagai deteksi awal mengenai penyakit stroke.
3. Bisa dijadikan sebagai refrensi pada penelitian selanjutnya.