

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Stroke adalah penyakit *cerebrovascular* , yaitu disfungsi otak yang berhubungan dengan penyakit pembuluh darah yang mengalirkan darah ke otak. Stroke dapat disebabkan oleh gangguan aliran darah ke otak yang terjadi akibat pembuluh darah pecah, memotong oksigen dan nutrisi serta merusak jaringan otak. Gejala umum stroke termasuk wajah, lengan atau kaki tiba-tiba menjadi kaku atau mati rasa dan lemah. Biasanya terjadi pada salah satu sisi tubuh. Gejala lain termasuk pusing, kesulitan berbicara atau memahami kata-kata, penglihatan kabur pada satu atau kedua mata, kesulitan berjalan, kehilangan koordinasi dan keseimbangan, pingsan atau kehilangan kesadaran, sakit kepala parah yang tidak diketahui penyebabnya[1].

Perkiraan beban stroke *Global Burden of Disease (GBD)* 2019 menunjukkan bahwa stroke tetap menjadi penyebab utama kematian kedua dan penyebab utama ketiga dan kecacatan didunia. Mulai tahun 1990 hingga 2019, meningkat secara substansial yaitu, 70,0% peningkatan stroke insiden, 43,0% kematian akibat stroke, 102,0% stroke yang lazim, dan 143,0% stroke yang tidak lazim. Dengan sebagian besar beban stroke global (86,0% kematian) berada pada negara-negara berpenghasilan rendah dan berpenghasilan menengah ke bawah. Jumlah absolut *disability life adjusted years (DALYs)* stroke pada pria (77,0 juta) melebihi pada wanita (66,0 juta) di tingkat global pada tahun 2019 [2].

Beberapa metode telah digunakan untuk memprediksi penyakit, salah satunya adalah dengan membandingkan kinerja teknik data mining prediktif. Proses pemilihan fitur, algoritma analisis komponen utama, digunakan untuk mengurangi ukuran dan mengimplementasikan algoritma klasifikasi untuk membuat model klasifikasi. Klasifikasi adalah teknik untuk membuat model data yang tidak terklasifikasi untuk meningkatkan kecanggihan yang digunakan dalam pengoptimasian [3].

Klasifikasi data penyakit medis merupakan tugas yang sangat berguna dan penting dalam memprediksi penyakit untuk membantu perawat dan dokter membuat keputusan tentang diagnosis penyakit. Diagnosa dini penting untuk mengurangi jumlah cedera fisik. Aktivitas pasien stroke dalam prediksi menggunakan metode klasifikasi data mining. Berbagai algoritma dapat digunakan untuk klasifikasi, seperti algoritma C4.5, Naïve bayes, Adaboost dan lain-lain.

Berdasarkan penelitian sebelumnya yaitu kelebihan algoritma *naïve bayes* melakukan diagnosis tingkat resiko penyakit stroke dengan 8 atribut dan 3 jenis tingkat resiko rendah, sedang dan tinggi yang kemudian didapatkan hasil klasifikasi dari model-model yang dibangun. Hasil klasifikasi tertinggi yaitu pada proses posterior *Naïve bayes*[4]. Menurut [5] Metode *K-Nearest Neighbor* (KNN) memiliki kelebihan yaitu menggunakan prinsip yang sederhana, bekerja berdasarkan nilai terpendek dari sampel uji ke sampel latih dan tidak memperhitungkan kemungkinan distribusi dari masing-masing kelas.

Dalam melakukan penelitian ini, penulis memilih dua algoritma yang akan digunakan dalam kelanjutan penelitian ini, yaitu *naïve bayes* dan *K-Nearest neighbour*, *naïve bayes* merupakan salah satu klasifikasi dengan menggunakan metode probabilitas yang ditemukan oleh seorang ilmuwan Inggris Thomas Bayes, yaitu masa depan diprediksi berdasarkan pada periode sebelumnya sehingga disebut dengan teorema Bayes. Teorema yang dikombinasikan dengan *Naïve* yaitu diasumsikan kondisi antar petunjuk (atribut) saling bebas[6].

Berdasarkan uraian permasalahan diatas, maka penulis tertarik untuk melakukan pengklasifikasian penyakit stroke dengan menerapkan metode *Naïve bayes* dan *K-nearest neighbour* dengan judul “KOMPARASI METODE *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOUR* UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE”.

1.2 RUMUSAN MASALAH

Berdasarkan uraian diatas, maka dapat dirumuskan masalah yang akan dikaji dalam penelitian ini adalah :

- a. Bagaimana penerapan metode *naïve bayes* dan *K-Nearest neighbour* untuk memprediksi penyakit stroke ?
- b. Bagaimana mengkomparasi penerapan algoritma *Naïve bayes* dan *K-Nearest neighbour* untuk memprediksi penyakit stroke ?

1.3 BATASAN MASALAH

Untuk menghindari permasalahan yang meluas, maka penulis membatasi pembahasan permasalahan sebagai berikut :

- a. Data *healthcare dataset* stroke dengan jumlah 4.844 dataset, bersumber pada www.kaggle.com
- b. Algoritma *naïve bayes* dan *K-Nearest neighbour* untuk mengklasifikasi menggunakan *WEKA*.
- c. *Tools* yang digunakan, *Microsoft Office Word*, *Microsoft Excel* dan *WEKA*
- d. Variabel data yang digunakan berjumlah 13 variabel
- e. Variabel yang digunakan diantaranya yaitu : id, jenis kelamin, usia, hipertensi, penyakit jantung, pernah menikah, tipe kerja, tempat tinggal, genetik, rata-rata glukosa darah, indeks massa tubuh, status merokok, dan stroke.
- f. Data keseluruhan yaitu berjumlah 4.844 data.
- g. Dengan *cross validation folds 5* pada metode *naïve bayes* yaitu didapat akurasi dengan sebesar 91,5566%, dan metode *K-Nearest neighbor* yaitu sebesar 94,3848%.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

- a. Untuk melihat bagaimana penerapan metode *Naïve bayes* dan *K-Nearest neighbor* dalam memprediksi penyakit stroke.
- b. Untuk mengkomparasi penerapan algoritma *naïve bayes* dan *K-Nearest neighbor* dalam memprediksi penyakit stroke.

1.4.2 Manfaat penelitian

Adapun manfaat yang didapat dalam penelitian ini sebagai berikut :

- a. Mengembangkan ide penulis dalam mengklasifikasi suatu data .
- b. Dihasilkan metode prediksi penyakit stroke yang baik dengan akurasi yang tinggi.

1.5 SISTEMATIKA PENULISAN

Sistematika dibuat untuk memberi suatu gambaran yang jelas terhadap penulisan ilmiah agar menghindari terjadinya pembahasan diluar dari tema dan judul penelitian. Secara garis besar penulisan ilmiah terdiri dari 5 bab, dapat dilihat dari sistematika penulisan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini merupakan bab pendahuluan yang membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini memuat tentang landasan teori yang mendasari pembahasan laporan secara khusus dan berisikan definisi – definisi yang melandasi penelitian yang dilakukan dengan melakukan studi *literature* sebagai dasar dalam melakukan analisa dan perancangan.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas secara rinci kerangka kerja penelitian, metode atau pendekatan yang digunakan, dan *tools* (alat bantu) yang digunakan dalam menjawab masalah penelitian untuk mencapai suatu tujuan.

BAB IV : HASIL ANALISIS DAN PEMBAHASAN

Bab ini membahas tentang analisa *dataset* yang akan digunakan atau terhadap *dataset* healthcare dataset stroke meliputi : analisa *dataset*, analisa penyakit, serta berisi hasil analisa dari data yang digunakan atau hasil pengkomparasian prediksi penyakit stroke.

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran-saran dari penelitian guna mengembangkan analisa prediksi penyakit yang telah di uji coba.