

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Seiring dengan perkembangan zaman, khususnya pada era teknologi yang sekarang ini semakin pesat kemajuannya, peningkatan aliran informasi semakin meningkat dalam jumlah besar setiap harinya, peningkatan aliran pada informasi ini dapat menyebabkan terjadinya penumpukan kumpulan data-data berupa teks, baik secara data *online* maupun *offline*. Dokumen teks yang menumpuk setiap harinya menyebabkan terjadinya kesulitan mencari dokumen yang sesuai dengan kebutuhan. Ketersediaan pada data yang berlimpah, kemudian dihasilkan dari penggunaan teknologi informasi yang saat ini hamper disemua bidang kehidupan menimbulkan suatu kebutuhan untuk dapat menggunakan dan memanfaatkan informasi dan pengetahuan yang dimana terdapat banyak manfaat dalam keseluruhan bentuk keseluruhan pada data tersebut, yang kemudian melahirkan *Data Mining*. *Data Mining* merupakan sebuah konsep yang diperuntukkan untuk menemukan pengetahuan atau informasi berharga yang tersembunyi di dalam *database* besar atau data dalam jumlah yang sangat besar. *Data Mining* memiliki beberapa Teknik dalam menghasilkan informasi berharga yang tersembunyi di dalam *database* besar atau data dalam jumlah yang sangat besar.

Penyakit stroke merupakan masalah kesehatan yang utama bagi masyarakat modern saat ini. Stroke menjadi masalah serius yang dihadapi di seluruh dunia. Hal

ini dikarenakan stroke adalah penyebab kematian ketiga terbanyak setelah penyakit jantung koroner dan kanker. Stroke merupakan sebuah gangguan suplai darah pada otak yang biasanya disebabkan karena pecahnya pembuluh darah atau sumbatan oleh gumpalan darah. Hal ini menyebabkan gangguan pasokan oksigen dan nutrisi di otak sehingga terjadi kerusakan pada jaringan otak.

Stroke terbagi atas dua jenis, yaitu stroke isemik dan stroke hemoragik. Stroke isemik merupakan penyakit stroke yang terjadi akibat pembuluh darah terhambat, sehingga menyebabkan aliran darah ke otak terhenti sebagian atau sepenuhnya, stroke jenis ini merupakan kasus yang paling sering terjadi yaitu sekitar 80% dari seluruh kasus stroke. Stroke isemik berdasarkan penyebabnya terbagi menjadi 3 jenis, yaitu trombotik yang disebabkan oleh terbentuknya *thrombus*. *Thrombus* akan menyebabkan pengumpulan darah sehingga aliran darah tidak lancar atau terhenti. Jenis kedua adalah stroke *embolik*, yang disebabkan oleh tertutupnya pembuluh arteri oleh pembekuan darah. Jenis ketiga adalah *hypoperfusion* sistemik yaitu berkurangnya aliran darah keseluruhan bagian tubuh karena adanya gangguan denyut jantung. Sedangkan stroke hemoragik merupakan penyakit stroke yang disebabkan oleh pecahnya pembuluh darah dari otak. Stroke hemoragik Sebagian besar terjadi pada penderita hipertensi. Berdasarkan lokasi pendarahan, stroke hemoragik terbagi menjadi dua yaitu, hemoragik intraserebral pendarahan didalam jaringan otak dan subarahnoid pendarahan pada ruang sempit antara permukaan otak dengan lapisan jaringan yang menutupi otak[1].

Beberapa teknik data mining telah diterapkan pada dunia medis seperti klasifikasi dan data prediktif. Salah satu algoritma klasifikasi yang dapat digunakan

adalah *Naïve Bayes*. *Naïve Bayes* merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistic yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Namun algoritma *Naïve Bayes* memerlukan optimasi untuk meningkatkan tingkat klasifikasi baik berdasarkan akurasi atau AUC menggunakan seleksi fitur *Greedy Forward Selection* sehingga akan mendapatkan fitur atau atribut yang relevan. Kemudian fitur record akan diproses menggunakan algoritma *Naïve Bayes* untuk mendapatkan hasil dari kinerja model yang dibuat.

Penelitian ini bukanlah penelitian yang pertama dilakukan, penelitian terkait dengan menggunakan dataset yang berbeda, “Analyzing the Performance of Stroke Prediction using ML Classification Algorithms” dengan tingkat akurasi menggunakan metode *Logistic regression* 78%, *Decision tree classification* 66%, *Random forest classification* 73%, *K-nearest neighbors classification* 80%, *Support vector machine* 80%, *Naïve Bayes classification* 82%[2]. Dari penelitian tersebut bahwa metode *Naïve Bayes classification* memiliki akurasi yang lebih tinggi dari pada metode yang lain, tingkat akurasinya mencapai 82%, namun didalam penelitian ini akan dilakukan peningkatan akurasi dari *Naïve Bayes* dengan menggunakan seleksi fitur *Greedy Forward Selection*.

Berdasarkan latar belakang permasalahan yang telah dibahas, maka penulis mengangkat kasus stroke ini dikarenakan melihat dari penderita stroke yang mengalami peningkatan setiap tahunnya dan tergolong penyakit yang beresiko tinggi. Dengan itu, penulis memutuskan untuk menerapkan algoritma *Naïve Bayes*

dan *Greedy Forward Selection* untuk prediksi penyakit stroke. Yang penulis tuangkan dalam judul “**PREDIKSI STROKE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN GREEDY FORWARD SELECTION**”.

1.2 RUMUS MASALAH

Berdasarkan latar belakang yang sudah dijelaskan, maka dapat diambil suatu rumusan masalah yaitu sebagai berikut :

1. Bagaimana cara menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection*?
2. Bagaimana cara menghitung keakurasian menggunakan algoritma *Naïve Bayes* dan *Greedy Forward Selection*?

1.3 BATASAN MASALAH

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi :

1. Data set yang digunakan adalah *Stroke Prediction Dataset* dari situs web [Kaggle.com](https://www.kaggle.com)
2. Algoritma yang digunakan untuk penelitian ini adalah *Algoritma Naïve Bayes* dan *Greedy Forward Selection*.
3. Alat bantu (*tools*) yang digunakan pada penelitian kali ini adalah *RapidMiner*.

4. Atribut yang digunakan yaitu : id, gender, age, hypertension, heart disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, smoking_status, stroke.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Berdasarkan permasalahan yang tertera diatas, maka dapat disimpulkan bahwa tujuan dari penelitian ini adalah :

1. Menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi stroke.
2. Untuk mengetahui tingkat keakurasian dari penerapan algoritma *Naïve Bayes* dan *Greedy Forward Selection* untuk prediksi stroke.

1.4.2 Manfaat Penelitian

Serta manfaat yang akan didapat dari penelitian, yaitu:

1. Dapat melihat bagaimana pengaruh dari atribut yang memiliki banyak data dalam memprediksi stroke.
2. Penulis dapat menambah pengetahuan mengenai menerapkan algoritma *Naïve Bayes* dan *Greedy Forward Selection*.
3. Dapat digunakan sebagai acuan untuk melakukan penelitian selanjutnya.

1.5 SISTEMATIKA PENULISAN

Adapun sistematika penulisan tugas akhir ini, penulis menguraikan dalam beberapa bab yaitu :

BAB I : PENDAHULUAN

Pada Bab ini akan diuraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penulisan, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada Bab landasan teori ini membahas tentang teori-teori dan pendapat para ahli yang berhubungan dengan permasalahan yang dianalisis. Teori-teori yang digunakan antara lain mengenai data mining, klasifikasi, prediksi, stroke, *Naïve Bayes*, *Greedy Forward Selection* dan alat bantu aplikasi *RapidMiner*.

BAB III : METODOLOGI PENELITIAN

Pada Bab ini menjelaskan tentang penjelasan dari metode-metode penelitian yang digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian.

BAB IV : ANALISIS

Pada Bab ini menjelaskan tentang gambaran umum objek penelitian, serta menganalisis dataset yang didapat dari website kaggle.com dengan melakukan persiapan data, kemudian penulis akan melakukan *Exploratory Data Analysis* untuk mengetahui prediksi

tingkat akurasi data menggunakan metode klasifikasi algoritma *Naïve Bayes* dan menggunakan seleksi fitur *Greedy Forward Selection* serta menggunakan alat bantu atau *tools* yaitu *RapidMiner*. Hasil dari analisis yang telah dilakukan diinterpretasikan menjadi informasi mengenai status prediksi penyakit stroke serta menggunakan alat bantu atau *tools* yaitu *RapidMiner*.

BAB V : KESIMPULAN

Bab ini berisikan kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.

