

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Kesehatan manusia adalah aset yang sangat berharga. Organisasi Kesehatan Dunia (WHO) mengatakan bahwa kesehatan tidak hanya berarti ketiadaan penyakit, itu juga berarti kesejahteraan fisik, mental, dan sosial. Penyakit tidak menular (*non-communicable diseases*) seperti penyakit jantung, diabetes, dan stroke telah menjadi masalah kesehatan utama dalam beberapa dekade terakhir. Stroke adalah salah satu penyakit yang paling signifikan dalam morbiditas dan mortalitas di seluruh dunia [1].

Penyakit stroke menjadi salah satu penyebab utama kematian dan kecacatan di seluruh dunia. WHO (*World Health Organization*) menyatakan bahwa stroke terjadi ketika aliran darah ke otak terganggu, yang dapat disebabkan oleh pecahnya pembuluh darah (*stroke hemoragik*) atau penyumbatan (*stroke iskemik*). Data menunjukkan bahwa stroke terus meningkat seiring dengan usia populasi dan faktor risiko seperti *hipertensi*, *diabetes*, dan *obesitas* meningkat [2].

Deteksi dini penyakit stroke menjadi aspek penting dalam upaya pencegahan dan pengelolaan penyakit ini, Dengan menggunakan bantuan *machine learning* dapat membantu klasifikasi terhadap risiko terkena penyakit stroke. Kerugian otak dapat dikurangi dan kualitas hidup dapat ditingkatkan dengan penanganan yang cepat dan tepat. Oleh karena itu, membangun sistem deteksi yang efektif dan akurat sangat penting. Teknik pengajaran mesin dapat mendeteksi kondisi medis, termasuk *stroke*, dalam data kesehatan.

Dalam dunia kesehatan, penggunaan kecerdasan buatan seperti *machine learning* telah menjadi subjek penelitian yang berkembang pesat seiring dengan kemajuan teknologi. *Machine learning* memungkinkan pembuatan model prediksi untuk mendeteksi penyakit, termasuk stroke, berdasarkan pola data historis. Karena kesederhanaannya dan kemampuan mereka untuk mengklasifikasikan dataset dengan hasil yang cukup akurat, *k-NN* dan *Naive Bayes* adalah dua algoritma *machine learning* yang paling umum digunakan untuk klasifikasi data kesehatan [3].

Algoritma *K-Nearest Neighbor (KNN)* menghitung jumlah tetangga terdekat dari data yang diuji dan kemudian memprediksi klasifikasi berdasarkan mayoritas tetangga tersebut [4]. Untuk dataset besar, *k-NN* mudah digunakan, tetapi memiliki keterbatasan pada waktu komputasi yang lebih lama. Sedangkan, algoritma *Naive Bayes* yang berbasis probabilitas, sering digunakan untuk memprediksi kategori data yang diuji dengan asumsi bahwa fitur data independen [5]. *Naive Bayes Classifier* bisa dilatih dengan efisien sebagai (*supervised learning*), pada pengaplikasiannya parameter estimasi model *Naive Bayes* menggunakan metode kemungkinan maximum. Dengan kata lain masih bisa bekerja tanpa harus memperdulikan bayesian probabilitas atau bayesian lainnya [6].

Pada penelitian sebelumnya yang dilakukan oleh (Novita Ranti Muntiari, Kharis Hudaiby Hanif 2022) Implementasi klasifikasi menggunakan *machine learning* sudah pernah dilakukan dengan menggunakan perbandingan akurasi yang berjudul "Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning" menyatakan bahwa nilai akurasi dari algoritma

Logistic Regresion, Decision Tree, Naïve bayes dan *k-NN* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00%. Sehingga algoritma logistic regresion, decision tree, naïve bayes dan k-nearest neighbor mempercepat pengambilan keputusan dalam memprediksi dalam klasifikasi penentuan jenis kanker payudara [7].

Penelitian kedua (Suci Anggraini, Muhamad Akbar, Alex Wijaya, Hadi Syaputra, Muhammad Sobri 2021) Mengangkat tema Gejala Penyakit *Coronavirus Disease* dengan judul “*Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Machine Learning*” Nilai akurasi pada dataset dengan 127 data pasien terjangkit covid-19 menggunakan algoritma k-nearest neighbor memperoleh 57,89%, neural network memperoleh 73,68%, random forest memperoleh 68,42%, naïve bayes memperoleh 65,38%. Pada penelitian ini algoritma klasifikasi Neural Network memberikan nilai akurasi yang tertinggi [8].

Selanjutnya penelitian ketiga (Rian, Arief, Donny, 2023) dengan judul “*Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit kanker Payudara menggunakan Algoritma Machine Learning*” Hasil penelitian menunjukkan performa akurasi yang berbeda pada ketiga algoritma tergantung pada metode validasi. *Naïve Bayes* memiliki performa akurasi yang lebih baik saat menggunakan k-fold cross validation dengan akurasi 93.85%. Metode holdout dengan rasio 75:25 terbukti menghasilkan akurasi terbaik untuk klasifikasi data kanker payudara menggunakan *SVM* [9].

Permasalahan klasifikasi sering muncul dalam berbagai bidang, salah satunya adalah diagnosis penyakit. Namun, tantangan utama dalam klasifikasi

adalah menangani data yang tidak terstruktur, ketidakseimbangan data, dan tingkat akurasi model yang dipengaruhi oleh kualitas dataset dan pemilihan algoritma. Penelitian terdahulu telah banyak membahas pengembangan model klasifikasi dengan berbagai algoritma seperti *Decision Tree*, *Support Vector Machine (SVM)*, *k-NN* dan *Naïve bayes*, namun masih terdapat keterbatasan pada aspek akurasi, efisiensi, dan kemampuan model dalam menangani data dengan dimensi tinggi atau data yang mengandung noise. Oleh karena itu, penelitian lanjutan diperlukan untuk mengatasi kendala tersebut dengan pendekatan yang lebih inovatif dan adaptif, sehingga dapat menghasilkan model yang lebih robust dan relevan untuk aplikasi dunia nyata.

Oleh karena itu pada penelitian ini akan di ujikan perbandingan klasifikasi algoritma *k-NN* dan *Naïve bayes* dengan parameter terdapat pada dataset.

1.2 RUMUSAN MASALAH

Berdasarkan dari uraian latar belakan diatas, maka didapatkan rumusan masalah pada penelitian ini sebagai berikut :

1. Bagaimana *Klasifikasi machine learning* dapat mendeteksi penyakit *stroke* dengan algoritma *k-NN* dan *Naïve bayes*?
2. Bagaimana menganalisis perbandingan performa algoritma *K-Nearest Neighbors* dan *Naïve Bayes* dalam mendeteksi penyakit *stroke* berdasarkan matrix evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score*.

1.3 BATASAN MASALAH

Untuk mempersempit fokus penulis terhadap penelitian ini, berikut adalah batasan masalah yang terdapat pada penelitian :

1. Algoritma yang digunakan pada machine learning ini adalah *K-Nearest Neighbors* dan *Naïve Bayes*
2. Nilai evaluasi program ini digunakan dari *Accuracy, Precision dan Recall*
3. Data set diperoleh dari website kaggle yang terdapat berbagai parameter
4. Parameter data stroke terdiri dari *gender, age, hypertension, heart disease, ever_married, work type, residence type, avg glucose, bmi, smoking status*
5. Menggunakan *Google colab* sebagai platform untuk menguji model *machine learning*

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang didapat, maka tujuan penelitian ini adalah sebagai berikut :

1. Untuk menganalisis klasifikasi *machine learning* terhadap penyakit stroke dengan algoritma *k-nearest neighbors* dan *naïve bayes*
2. Untuk mengetahui algoritma mana yang lebih optimal dalam mendeteksi penyakit stroke

1.4.2 Manfaat Penelitian

Adapun penelitian ini memiliki manfaat sebagai berikut :

1. Penelitian ini diharapkan dapat menjadi salah satu cara penyelesaian dalam menentukan prediksi seseorang terkena penyakit stroke dengan menggunakan machine learning
2. Menjadi bahan referensi untuk penelitian selanjutnya mengenai deteksi penyakit stroke
3. Menambah pengetahuan bagi mahasiswa dan pembaca mengenai machine learning dan penggunaannya

1.5 SISTEMATIKA PENULISAN

Untuk memberikan gambaran umum mengenai keseluruhan bab yang akan dibahas dapat dilihat dalam sistematika penulisan berikut :

BAB I : PENDAHULUAN

Bab ini menjelaskan mengenai latar Belakang Masalah, Perumusan Masalah, Batasan masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan

BAB II : LANDASAN TEORI

Pada bab ini berisikan tentang teori-teori terkait yang berhubungan dengan masalah yang diteliti dari pokok permasalahan yang diangkat, yakni klasifikasi machine learning untuk mendeteksi penyakit stroke menggunakan algoritma k-NN dan naïve bayes. Dengan poin-poin yang tertera pada penelitian yaitu stroke, machine learning, algoritma k-NN dan naïve bayes. bahasa pemrograman yang digunakan. Serta penelitian sejenis yang dapat mendukung proyek penelitian ini.

BAB III : METODE PENELITIAN

Bab ini berisi langkah penelitian, waktu penelitian dan tempat penelitian, alat yang digunakan hingga prosedur penelitian dari awal sampai akhir penelitian

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini berisi mengenai hasil yang digunakan dan pembahasan mengenai hasil tersebut

BAB V : PENUTUP

Bab ini menjelaskan kesimpulan dan saran untuk penelitian selanjutnya