

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Stroke telah dikenal sebagai salah satu penyakit paling berbahaya yang menjadi penyebab utama kecacatan dan kematian di seluruh dunia. Menurut data yang dihimpun selama tiga dekade terakhir, stroke tidak hanya merupakan penyebab kematian urutan kedua yang paling umum namun juga merupakan faktor penting yang berkontribusi pada *disability-adjusted life years (DALYs)* – suatu ukuran yang merefleksikan tahun-tahun kehidupan yang hilang akibat kecacatan[1]. Studi yang dilakukan oleh Murphy dan Werring pada tahun 2020 menunjukkan adanya peningkatan signifikan dalam insiden stroke sepanjang dekade terakhir, termasuk stroke iskemik yang disebabkan oleh penyumbatan pembuluh darah serta stroke hemoragik yang diakibatkan oleh pendarahan[2]. Dalam hal kejadian per kapita, angka ini bervariasi antara 85 hingga 94 kasus per 100.000 orang secara internasional, dengan prevalensi yang meningkat tajam di kalangan individu yang berusia di atas 75 tahun, dimana angka kejadiannya mencapai antara 1151 hingga 1216 kasus per 100.000 orang. Studi yang sama juga menekankan disparitas yang signifikan berdasarkan status ekonomi sebuah negara, dengan negara-negara berpenghasilan rendah melaporkan sekitar 85% dari total kematian akibat stroke, serta menyumbang hingga 87% dari total beban kecacatan yang berkaitan dengan kondisi ini. Dalam konteks nasional, prevalensi stroke di Indonesia tidak kalah mengkhawatirkan, dimana penyakit ini menempati posisi teratas sebagai penyebab kematian, dengan data yang menunjukkan adanya satu kasus kematian akibat stroke setiap sepuluh detik[3]. Dampak yang ditimbulkan oleh stroke bukan hanya mempengaruhi pasien tetapi juga keluarga dan sistem kesehatan secara keseluruhan [4]. 67% orang yang pernah mengalami stroke dan masih hidup berusia di bawah 70 tahun [5]. Terdapat 5 faktor yang menjadi risiko utama yang menyebabkan stroke yaitu tekanan darah sistolik tinggi (55%), *body mass index* tinggi (24,3%), glukosa darah tinggi puasa (20,2%), polusi partikel sekitar (20,1%), dan merokok (17,6%) [5]. Beban ekonomi yang diakibatkan oleh stroke sangat besar, termasuk biaya kesehatan langsung dan biaya tidak langsung seperti hilangnya produktivitas dan perawatan jangka panjang untuk disabilitas. Dengan prediksi dan pencegahan yang efektif, dampak ekonomi ini dapat berkurang

secara signifikan [6]. Mengingat tingginya angka prevalensi dan konsekuensi fatal yang dapat ditimbulkan oleh stroke, pentingnya deteksi dini menjadi sangat krusial. Dengan deteksi yang tepat waktu, tidak hanya dapat meningkatkan peluang pemulihan pasien, tetapi juga secara substansial mengurangi risiko komplikasi lebih lanjut.

Berbagai metode telah dikembangkan untuk mengidentifikasi penyakit stroke pada tahap awal, termasuk pendekatan tradisional seperti penggunaan pencitraan resonansi magnetik (MRI) dan pemindaian tomografi terkomputasi (CT), yang meskipun efektif, diketahui memerlukan biaya yang signifikan dan bersifat invasif[7][8]. Alternatif yang sedang berkembang adalah penerapan teknik pembelajaran mesin[7], yang menawarkan potensi deteksi yang lebih cepat dan biaya yang lebih rendah. Studi terkait menunjukkan bahwa faktor waktu memegang peranan krusial dalam penanganan stroke, di mana respons yang cepat dapat menjadi faktor determinan dalam mengurangi risiko kematian dan kerusakan otak yang irreversibel[7][9][10][11]. Selanjutnya, pendeteksian dini yang akurat juga dapat mencegah berkembangnya komplikasi sekunder yang dapat ditimbulkan oleh stroke[12].

Perkembangan di bidang kecerdasan buatan dan teknik pembelajaran mesin membuka potensi untuk menciptakan sistem prediksi yang lebih terjangkau dan efisien. Pembelajaran mesin diakui sebagai instrumen yang berpotensi dalam meramalkan terjadinya stroke dengan memanfaatkan faktor-faktor risiko yang dikaitkan dengan pasien[7]. Banyak algoritma pembelajaran mesin telah diaplikasikan dalam konteks prediksi medis[13]. Diantara berbagai algoritma tersebut, termasuk Naïve Bayes, Support Vector Machine, Random Forest, K-Nearest Neighbors, Decision Tree, Stacking, Majority Voting, Linear Discrimination Analysis, Gaussian Naïve Bayes, dan Logistic Regression, studi-studi sebelumnya telah mengindikasikan bahwa Gaussian Naïve Bayes khususnya menawarkan tingkat akurasi yang memadai serta kecepatan dalam proses prediksi[7][11].

Algoritma klasifikasi Naive Bayes terbukti efektif dalam berbagai situasi, menunjukkan tingkat akurasi yang signifikan yang seringkali melampaui model-model klasifikasi lain, meskipun hanya menggunakan volume data pelatihan yang relatif rendah untuk mencapai performa yang optimal[14]. Penelitian-penelitian yang telah dilakukan untuk mengevaluasi kemampuan algoritma ini dalam skenario yang berbeda-beda termasuk satu studi yang membandingkan performa algoritma

Naive Bayes dengan lima algoritma lain—Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, dan Support Vector Machine—dalam pengolahan data teks[15]. Hasil penelitian tersebut menunjukkan bahwa Naive Bayes berhasil mencapai akurasi tertinggi, dengan angka sebesar 82%. Lebih lanjut, varian Gaussian Naïve Bayes juga telah menunjukkan kinerja yang superior dalam klasifikasi Diabetes Melitus dibandingkan dengan algoritma K-Nearest Neighbor, dengan menghasilkan akurasi yang lebih tinggi[12]. Penelitian lainnya dalam prediksi penyakit stroke melaporkan terjadi peningkatan dalam akurasi, presisi, dan nilai recall yang diperoleh melalui penerapan algoritma Gaussian Naïve Bayes[16]. Sebuah penelitian yang berfokus pada deteksi kanker paru-paru melaporkan bahwa algoritma ini mencatat akurasi hingga 96%[17]. Analisis serupa pada dua set data yang berbeda dalam konteks deteksi kanker lainnya juga mengkonfirmasi efektivitas algoritma Naive Bayes dalam menghasilkan akurasi yang tinggi[18]. Di lain pihak, sebuah penelitian yang dilakukan dalam lingkup klasifikasi berbagai kondisi kesehatan pada pengguna narkoba menemukan bahwa prediksi yang dibuat oleh algoritma ini lebih akurat daripada metode alternatif yang dipertimbangkan[19]. Keefektifan algoritma Naive Bayes ini juga dibuktikan dalam aplikasi yang memanfaatkan pendekatan berbasis ontologi, di mana nilai F-measure yang dicapai adalah 87%[20].

Dari pembahasan sebelumnya, diketahui bahwa algoritma Gaussian Naïve Bayes menunjukkan performa yang memuaskan dengan tingkat akurasi yang memadai, yang mendukung pemilihannya sebagai instrumen prediksi dalam penelitian ini. Meskipun demikian, masih terdapat tantangan dalam memperluas aksesibilitas solusi prediksi yang akurat ke masyarakat luas. Untuk mengatasi hal ini, penelitian ini bertujuan untuk mengembangkan sebuah aplikasi web yang dapat memudahkan masyarakat dalam mengakses dan menggunakan model prediksi stroke yang berbasis algoritma Gaussian Naïve Bayes.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dijabarkan, berikut adalah rumusan masalah dalam penelitian ini:

1. Bagaimana proses rancang bangun aplikasi prediksi penyakit stroke yang mengintegrasikan algoritma Gaussian Naïve Bayes?
2. Bagaimana kinerja algoritma Gaussian Naïve Bayes dalam memprediksi

kejadian penyakit stroke berdasarkan variabel-variabel prediktif yang tersedia?

### **1.3 Batasan Permasalahan**

Berikut adalah batasan-batasan yang akan diterapkan dalam penelitian ini: Berikut merupakan batasan-batasan yang digunakan dalam penelitian ini:

1. Penelitian ini hanya akan menggunakan atribut-atribut tertentu sebagai variabel prediktif dalam pengembangan model prediksi, yakni jenis kelamin, umur, punya hipertensi, pernah terkena penyakit jantung, pernah menikah, tipe pekerjaan, tipe tempat tinggal, rata-rata tingkat glukosa dalam darah, Body Mass Index (BMI), status merokok, dan pernah stroke.
2. Penelitian ini dibatasi pada evaluasi kinerja algoritma Gaussian Naïve Bayes dalam memprediksi penyakit stroke, dengan fokus pada metrik-metrik evaluasi seperti akurasi, presisi, recall, dan f1-score.
3. Penelitian ini menggunakan dataset yang bersumber dari Kaggle yaitu dataset yang disediakan oleh pengguna dengan nama fedesoriano, yang diberi judul "Stroke Prediction Dataset".

### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah, berikut adalah penjabaran tujuan penelitian ini:

1. Merancang dan membangun aplikasi prediksi penyakit stroke yang mengintegrasikan algoritma Gaussian Naïve Bayes.
2. Mengevaluasi kinerja algoritma Gaussian Naïve Bayes dalam memprediksi kejadian penyakit stroke berdasarkan variabel-variabel prediktif yang tersedia.

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Menyediakan suatu alat bantu yang dapat digunakan oleh praktisi kesehatan serta individu untuk mendeteksi dini penyakit stroke, dengan harapan dapat memfasilitasi pencegahan yang lebih baik dan intervensi medis yang tepat waktu
2. Memberikan kontribusi terhadap literatur akademis dalam bidang pembelajaran mesin dan kesehatan masyarakat dengan menyajikan studi kasus aplikasi algoritma Gaussian Naïve Bayes dalam konteks medis yang praktis dan dapat diakses oleh pengguna.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN  
Berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- Bab 2 LANDASAN TEORI  
Berisi landasan teori terkait algoritma Gaussian Naïve Bayes.
- Bab 3 METODOLOGI PENELITIAN  
Berisi metodologi penelitian yang digunakan, proses perancangan sistem, dan rancangan antarmuka sistem.
- Bab 4 HASIL DAN DISKUSI  
Berisi hasil penelitian mulai dari spesifikasi perangkat lunak dan perangkat keras yang digunakan, rancang bangun sistem, pengujian aplikasi, serta evaluasi akhir dari sistem aplikasi yang dibuat.
- Bab 5 KESIMPULAN DAN SARAN  
Berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian berikutnya.