

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stroke adalah masalah kesehatan yang lazim dan serius, termasuk di antara kondisi yang paling mengancam jiwa di seluruh dunia [1]. Stroke adalah penyakit medis yang ditandai dengan gangguan aliran darah ke otak, yang menyebabkan berbagai komplikasi [2]. Menurut statistik kesehatan global, stroke berkontribusi pada sejumlah besar kematian, dengan 92.033 kematian terkait stroke yang dilaporkan di seluruh dunia. Kemajuan dalam penelitian medis dan analisis data memberi para profesional kesehatan alat yang berharga untuk mengidentifikasi dan mengatasi risiko stroke pada individu, menawarkan peluang untuk deteksi dini dan intervensi untuk mengurangi dampak dari kondisi yang melemahkan ini.

Prevalensi stroke di berbagai daerah di Indonesia bervariasi. Beberapa faktor diduga berperan dalam variasi tersebut, diantaranya faktor risiko hipertensi, prevalensi hipertensi dan prevalensi diabetes melitus [3]. Penelitian ini bertujuan untuk menganalisis hubungan antara faktor risiko hipertensi (merokok, aktivitas fisik dan konsumsi garam), prevalensi hipertensi dan prevalensi diabetes melitus dengan perbedaan prevalensi stroke di berbagai provinsi di Indonesia. Penelitian ini menggunakan desain observasional analitik dengan pendekatan studi ekologi. Data diperoleh dari 440 kabupaten dan 33 provinsi di Indonesia yang diambil secara proporsional terhadap ukuran [4]. Sampel penelitian ini adalah laki-laki dan perempuan berusia 18 tahun ke atas. Variabel independen adalah prevalensi stroke, sedangkan variabel dependen adalah faktor risiko hipertensi (merokok, aktivitas fisik dan konsumsi sayur dan buah), prevalensi hipertensi dan prevalensi diabetes melitus. Diagnosis hipertensi, diabetes mellitus dan stroke ditegakkan oleh dokter. Data yang diperoleh dianalisis dengan menggunakan analisis regresi linier berganda

Riset Kesehatan Dasar (Riskesdas) 2007 di 33 provinsi menunjukkan bahwa prevalensi stroke adalah 8,2 per 1000 penduduk, dan prevalensi tertinggi berasal dari provinsi Aceh (16,6%) [5]. Stroke juga merupakan pembunuh nomor 1 di Indonesia (15,4%), Metode: Penelitian observasional prospektif dilakukan dari bulan Oktober 2012 hingga April 2013 dengan menggunakan Formulir Laporan Kasus Stroke yang telah terstandarisasi. Terdapat 11 rumah sakit yang terlibat dalam penelitian epidemiologi ini. Hasil: 1807 pasien stroke yang dikumpulkan dari Oktober 2012 - April 2013, stroke iskemik merupakan kasus terbanyak (67,1%) dan hemoragik 32,9%, dan hipertensi merupakan faktor risiko yang paling banyak ditemukan baik untuk perdarahan (71,2%) maupun stroke iskemik (63,4%), disusul diabetes melitus dan dislipidemia. Kematian tercatat 20,3% kematian setelah 48 jam, 18,3% \leq 48 jam pada stroke perdarahan, dibandingkan dengan 8,3% kematian pada stroke iskemik setelah 48 jam, dan 3,5% \leq 48 jam.

Teknik machine learning dalam dunia medis memiliki peran penting dalam pemrosesan gambar medis, analisis catatan kesehatan elektronik, memprediksi risiko stroke, dan pemantauan pasien secara terus menerus [6]. Algoritma ini dapat secara efektif memproses kumpulan data yang beragam, memungkinkan identifikasi dini risiko terkena penyakit stroke. Prediksi dini dapat meminimalkan keterlambatan penanganan penyakit stroke karena kurangnya pengetahuan tentang gejala awal penyakit stroke. Penerapan pembelajaran machine learning melibatkan analisis data yang dikumpulkan dari individu yang pernah mengalami stroke, yang mencakup berbagai karakteristik dan kebiasaan gaya hidup. Dengan memanfaatkan data ini, algoritma dapat membangun model prediksi untuk mengklasifikasikan pasien memiliki risiko penyakit stroke atau tidak.

Integrasi algoritma machine learning dalam dunia medis memiliki harapan yang signifikan untuk meningkatkan deteksi dini penyakit stroke [7]. Deteksi dini stroke bisa untuk mengurangi risiko komplikasi yang parah dan berkontribusi pada peningkatan kualitas hidup secara keseluruhan bagi individu yang terkena dampak. Seiring dengan kemajuan teknologi, penyempurnaan dan

pengembangan berkelanjutan dalam algoritma machine learning menghasilkan sistem yang belajar dari data historis, yang secara terus menerus meningkatkan akurasi dan efektivitasnya dari waktu ke waktu. Praktisi kesehatan semakin banyak memanfaatkan teknologi komputer untuk mendukung pengambilan keputusan, dengan pembelajaran yang muncul sebagai solusi penting untuk mendiagnosis pasien. Kemampuan analitisnya yang tangguh sangat cocok untuk menangani tugas-tugas besar dan kompleks, termasuk analisis rekam medis, memprediksi terjadinya stroke, dan memeriksa data medis yang rumit.

Kemajuan dalam machine learning telah memicu peningkatan minat dalam mengidentifikasi indikator potensial stroke pada pasien [8]. Algoritma machine learning memainkan peran penting dalam melatih dan mengevaluasi model yang dirancang untuk memprediksi kemungkinan stroke berdasarkan berbagai faktor risiko. Keakuratan prediksi ini bergantung pada kualitas data, kinerja model, dan pemilihan parameter pelatihan yang tepat. Dataset yang dilatih menghasilkan prediksi dengan tingkat akurasi yang berbeda-beda, yang dipengaruhi oleh faktor-faktor seperti kualitas data, kinerja model, dan parameter pelatihan. Dalam bidang medis, akurasi komparatif dari beragam algoritma menjadi sangat penting dalam menentukan model yang paling efektif dan tepat, yang menawarkan referensi berharga untuk pengambilan keputusan yang tepat dalam deteksi dan manajemen stroke.

Dalam penelitian ini, peneliti mengusulkan prediksi penyakit stroke dengan menggunakan tiga algoritma klasifikasi: Naïve Bayes, Random Forest, dan Logistic Regression. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi metode klasifikasi yang paling akurat untuk mendiagnosis stroke. Penerapan algoritma ini dalam prediksi stroke bertujuan untuk meningkatkan akurasi proses diagnostik, memberikan wawasan yang berharga tentang metode yang efektif untuk mengidentifikasi dan mengkategorikan individu yang berisiko mengalami stroke.

1.2 Rumusan Masalah

Berdasarkan uraian yang dijelaskan didalam latar belakang, maka

disusunlah sebuah rumusan masalah yaitu “Apa metode terbaik untuk melakukan prediksi terhadap penyakit stroke menggunakan penerapan machine learning?”

1.3 Batasan Masalah

Untuk mencegah penyimpangan dari focus penelitian, maka perlu ditentukan Batasan masalah yang akan dibahas dan diteliti. Berikut ini adalah batasan masalah yang digunakan dalam penelitian ini:

1. Penelitian ini terbatas pada evaluasi performa model dalam melakukan klasifikasi.
2. Klasifikasi terbatas pada dataset penyakit stroke.
3. Metrik evaluasi terbatas pada beberapa penjabaran turunan dari confusion metric seperti akurasi, presisi, recall, f1-score dan Area Under Curve (AUC).
4. Penerapan machine learning pada penelitian ini terbatas pada enam algoritma yaitu Naive Bayes, Random Forest, Logistic Regression, K-Nearest Neighbors, Gradient Boosting, Decision Tree

1.4 Tujuan Penelitian

Berdasarkan pernyataan penelitian kali ini, hasil yang akan dicapai adalah:

1. Menerapkan ilmu machine learning agar memberikan manfaat bagi ilmu pengetahuan dan pengembangan penelitian.
2. Berperan dalam kontribusi teknologi di bidang medis, terutama dalam Tindakan preventif melalui prediksi awal risiko penyakit stroke.
3. Membantu masyarakat dalam memahami prediksi dini risiko penyakit stroke, sehingga mereka dapat mengambil langkah pencegahan yang tepat.

1.5 Manfaat Penelitian

Adapun manfaat penelitian kali ini yaitu :

1. Menghadirkan referensi bagi pengembangan ilmu pengetahuan dan kelanjutan penelitian, terutama di bidang machine learning.

2. Berpotensi diterapkan dalam konteks medis sebagai Tindakan preventif, khususnya dalam deteksi dini risiko penyakit stroke. Dengan demikian, keterbatasan jumlah tenaga medis tidak akan menghalangi peran medis dalam memberikan pelayanan yang efektif.
3. Meningkatkan kesadaran masyarakat tentang pentingnya gaya hidup sehat dan pencegahan dini terhadap risiko penyakit stroke.

1.6 Sistematika Penulisan

Pada penelitian kali ini penulis memuat uraian secara garis besar isi Skripsi untuk tiap-tiap bab. Pada bagian ini mendeskripsikan apa saja isi masing-masing bab yang akan disusun dan akan dijelaskan secara singkat dari Bab I, Bab II, Bab III, Bab IV Bab V, sebagai berikut :

Bab 1 PENDAHULUAN: Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Sistematika Penulisan

Bab 2 TINJAUAN PUSTAKA : Studi Literature, Tabel Keaslian Penelitian, Machine Learning, Algoritma Naïve Bayes, Algoritma Random Forest, Algoritma Logistic Regression

Bab 3 METODE PENELITIAN : Alur Penelitian, Exploratory Data Analysis (EDA), Data Preprocessing, Anlysis (Train The Model), Evaluasi Kinerja.

Bab 4 HASIL DAN PEMBAHASAN : Klasifikasi Algoritma Naive Bayes, Klasifikasi Algoritma Random Forest, Klasifikasi Algoritma Logistic Regression, Kurva ROC.

Bab 5 PENUTUP : Kesimpulan, Saran, Referensi