

BAB I

PENDAHULUAN

1.1 Latar Belakang

Noncommunicable disease (NCDs) atau juga disebut sebagai penyakit kronis, cenderung memiliki jangka waktu lama yang disebabkan oleh hasil kombinasi dari faktor gen, fisiologi, lingkungan dan gaya hidup. Salah satu jenis NCDs adalah penyakit kardiovaskuler, salah satunya adalah stroke[1]. Stroke merupakan penyebab kematian kedua dalam skala global. satu dari empat orang dapat terindikasi bahaya stroke.

Stroke merupakan suatu kondisi terhentinya aliran darah sebab adanya penyumbatan pembuluh darah atau perdarahan secara mendadak yang mengakibatkan gangguan fungsi otak. Pada tahun 2017 stroke merupakan penyakit penyebab kematian tertinggi ketiga dunia[2]. Di *United States America* stroke merupakan penyakit tidak menular yang menyebabkan sekitar 11,0 % dari total kematian dan berada di posisi ke-5 penyebab kematian. Di India stroke menjadi penyebab kematian ke-4[3].

Rumah sakit memiliki data dalam jumlah besar dari pasien yang menderita berbagai jenis penyakit, termasuk stroke, dan kurangnya media atau alat analisis yang digunakan untuk memperoleh informasi tersebut. Dan menjadi tantangan atau masalah bagi paramedis untuk mendapat informasi. Pasien dengan gejala stroke kadangkala menemukan tantangan untuk menjelaskan kondisi mereka dan menyulitkan paramedis dalam pemeriksaan yang tepat[4].

Beberapa penelitian tentang masalah ini telah dilakukan [2] [3] [4] untuk memprediksi dengan menggunakan algoritma dalam menentukan atau mengidentifikasi stroke. Sekaligus menjadi alternatif dalam menemukan akurasi diagnosa stroke pada pasien. Penggunaan algoritma dapat membantu dalam identifikasi dini untuk mendeteksi stroke pada pasien[5].

Menggunakan media atau alat medis dalam mendeteksi stroke pada pasien dapat membantu diagnosa dini. Namun beberapa metode pengklasifikasian memiliki tingkat akurasi, kelemahan, dan kelebihan masing-masing. Tingkat

akurasi juga menjadi hal penting dalam ketepatan proses diagnosa pada pasien. Beberapa metode yang umum digunakan seperti *Naive Bayes*, *KNN*, *Decision Tree*, *Support Vector Machine*, dan *Random Forest* dapat diterapkan pada jenis penelitian seperti ini [2] [3] [4] [6]

Machine learning mempelajari dari pola yang diberikan oleh sampel *training* dan kemudian digunakan untuk mengembangkan prediksi yang berguna. Metode klasifikasi berkembang luas di ranah medis untuk identifikasi dan memprediksi penyakit secara akurat [1]. Penelitian yang ada dengan menggunakan beberapa klasifikasi seperti *Naive Bayes*, *Random Forest*, *KNN* dan *Decision Tree* (*C4.5*) memiliki keunggulan dan kekurangan tersendiri dalam pengolahan pada dataset dalam memprediksi penyakit.[5] Namun tingkat akurasi yang dihasilkan akan berpengaruh terhadap hasil prediksi yang akan didapat.

Sehingga metode yang akan diaplikasikan dalam penelitian ini adalah *Logistic Regresion* untuk mengklasifikasi fitur - fitur yang terdapat di dalam dataset. Selain perhitungan yang tidak begitu kompleks tingkat akurasi yang dihasilkan oleh metode *Logistic Regresion* ini sama baiknya dengan metode lainnya. Seperti : *ANN*, *SVM* , *Random forest* , *KNN* dan *Gradient Boost* [7].

Penerapan metode *Logistic Regression* dalam klasifikasi penyakit stroke ini diharapkan mampu memperoleh tingkat akurasi yang baik dan tingkat pemrosesan yang lebih cepat karena metode *Logistic Regresion* tidak memerlukan optimasi hyperparameter apapun serta metode ini dapat dengan mudah diimplementasikan. Penerapan ini diharapkan dapat membantu kinerja rekan dibidang medis dalam prediksi apakah pasien tersebut berkemungkinan besar terserang penyakit stroke ini.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Pemrosesan data dalam jumlah besar tentang pasien yang terindikasi penyakit stroke memerlukan waktu dan perhitungan yang cukup kompleks apabila dilakukan secara manual hal ini akan banyak menguras banyak waktu dan tenaga para pakar medis dalam mengidentifikasi pasien yang berkemungkinan terserang penyakit stroke karena itu untuk memanfaatkan perkembangan teknologi dan ilmu matematis yang ada dapat diterapkan sebuah metode untuk mengidentifikasi fitur-

fitur yang ada di dalam data yang didapat menggunakan metode pengklasifikasian sehingga menghasilkan nilai akurasi yang baik.

1.2.2 Usulan Solusi

Usulan solusi pada penelitian ini yaitu membuat sistem untuk identifikasi penyakit stroke pada dataset stroke dengan fitur-fitur yang ada menggunakan metode *Logistic Regression*.

1.2.3 Pertanyaan Penelitian

Pertanyaan pada penelitian ini yaitu berapa tingkat akurasi yang dihasilkan dari penggunaan metode *Logistic Regression* dalam mengidentifikasi dataset stroke?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan

Tujuan dari penelitian ini yaitu mengukur tingkat akurasi dalam mengidentifikasi dataset stroke menggunakan metode *Logistic Regression*.

1.3.2 Manfaat

Dari penelitian ini diharapkan mampu memberikan manfaat dalam membangun dan mengembangkan ataupun sebagai acuan untuk membangun sistem yang dapat mengidentifikasi data penyakit stroke dengan tingkat akurasi yang baik, sehingga dapat membantu rekan tenaga medis dalam mengambil tindakan dini untuk pasien.

1.4 Batasan Masalah

Batasan-batasan masalah pada penelitian ini, sebagai berikut :

1. *Raw Dataset* yang digunakan berasal dari database public yang disediakan oleh Kaggle di <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>.
2. *HealthCare Stroke Dataset* berjumlah 5110 data.
3. Didalam dataset *HealthCare Stroke Dataset* mempunyai 10 fitur dan 1 kelas target.

1.5 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dipaparkan mulai dari latar belakang penelitian, rumusan masalah yang akan dijawab, tujuan penelitian dan manfaat dari penelitian serta batasan masalah serta sistematika penulisan.

BAB II STUDI PUSTAKA

Pada bab ini akan dipaparkan dasar teori yang akan menjadi pijakan dari penelitian ini, yang mana di atasnya penelitian ini akan ditegakkan. Mulai dari teori dan materi secara umum yang menunjang penelitian ini sampai dengan hal-hal yang secara detil membahas tentang alur penelitian dalam bab ini. Juga penelitian terkait sebelumnya tentang klasifikasi penyakit.

BAB III METODE USULAN

Pada bab ini membahas tentang metode yang diusulkan yakni *Logistic Regression*. selain itu, bab ini juga menjelaskan arsitektur dari rancangan sistem, dataset yang digunakan, perhitungan dan uji coba manual menggunakan *Microsoft Excel* serta tahapan – tahapan yang dilakukan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang hasil dan implementasi pembuatan sistem berdasarkan rancangan arsitektur yang dibuat sebelumnya dan penerapan metode yang diusulkan ke dalam sistem.

BAB V PENUTUP

Dalam bab ini, berisikan tentang kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian, dan saran mengenai metode atau hal-hal apa yang perlu dikembangkan kedepannya.

REFERENSI

Berisi tentang daftar jurnal atau hasil penelitian sebelumnya yang dijadikan acuan dalam penelitian ini.