

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Stroke merupakan salah satu penyebab utama kematian dan kecacatan di seluruh dunia, sehingga prediksi dan pencegahan dini sangat penting dalam upaya kesehatan masyarakat. Kemajuan *Machine Learning* kini banyak digunakan untuk meningkatkan akurasi dalam memprediksi risiko stroke berdasarkan data medis pasien. Berbagai algoritma telah diteliti untuk meningkatkan prediksi, seperti *Random Forest*, *Naïve Bayes*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), dan *Decision Tree*.

Menurut Penelitian Yang di lakukan M. Putri (2024) menunjukkan bahwa algoritma *Random Forest* sangat efektif untuk prediksi stroke karena kemampuannya menangani data dengan fitur yang kompleks serta menghasilkan akurasi yang tinggi dalam mendeteksi dini stroke [1]. Sementara itu, penelitian K. Akmal et al. (2023) melakukan perbandingan antara metode *Naïve Bayes* dan *KNN* dan menemukan bahwa *KNN* memberikan akurasi yang lebih baik dalam klasifikasi stroke, terutama dalam dataset dengan distribusi data yang beragam [2].

Berdasarkan Penelitian dari E. Draitsas et al. (2022) menyoroti bahwa teknik *Machine Learning* memberikan prediksi yang lebih presisi dibandingkan metode tradisional dalam mendeteksi faktor risiko stroke, khususnya ketika menggunakan teknik ensemble seperti *Random Forest* [3]. Selanjutnya, menurut penelitian S.

Mainali, M. E. et al. (2021) mengonfirmasi bahwa penerapan *Machine Learning* tidak hanya meningkatkan kemampuan dalam prediksi stroke, tetapi juga memberikan hasil yang lebih baik dalam diagnosis dini dan perkiraan hasil perawatan pasien [4].

Selain itu, *Machine Learning* juga telah diaplikasikan dalam rehabilitasi pasca-stroke. Menurut Penelitian dari S. Campagnini et al. (2022) menjelaskan bahwa *Machine Learning* dapat digunakan untuk memprediksi pemulihan fungsional pasien pasca-stroke, sehingga memungkinkan pengembangan rencana rehabilitasi yang lebih tepat sasaran [5]. Kemudian I. O. Herlistiono et al. (2024) mengembangkan model prediksi risiko stroke menggunakan kombinasi algoritma *Machine Learning* yang dirancang untuk memaksimalkan akurasi prediksi dalam sistem kesehatan [6].

Penelitian sebelumnya yang dilakukan M. Fadli et al. (2023) mengevaluasi performa algoritma *Random Forest* dan menemukan bahwa model ini memberikan hasil akurasi prediksi hingga 85% untuk deteksi risiko stroke, menjadikannya salah satu algoritma yang direkomendasikan dalam analisis risiko penyakit stroke [7]. Sementara itu, Patmawati (2023) menggunakan algoritma *SVM* untuk prediksi stroke dan melaporkan bahwa *SVM* berhasil mencapai akurasi prediksi sebesar 82%, yang menunjukkan potensi algoritma ini dalam mendeteksi risiko stroke pada pasien [8].

Penelitian lainnya, Latifah et al. (2022) meneliti kombinasi algoritma *Random Forest*, *KNN*, dan *Decision Tree* untuk prediksi stroke, dan hasil prediksi

mereka menunjukkan akurasi tertinggi sebesar 88% pada model *Random Forest*, yang lebih unggul dibandingkan algoritma lainnya [9]. Mereka menemukan bahwa kombinasi algoritma ini memberikan hasil yang bervariasi tergantung pada fitur dan karakteristik data yang digunakan. Dalam Studi yang dilakukan Y. Azhar, A. K. et al. (2022) membandingkan beberapa algoritma klasifikasi *Data Mining* dan menemukan bahwa model *Random Forest* mencapai akurasi 87%, menjadikannya pilihan yang sesuai untuk data stroke dengan karakteristik yang kompleks [10].

Dengan berbagai penelitian ini, dapat disimpulkan bahwa penerapan *Machine Learning* dalam prediksi stroke berpotensi memberikan kontribusi besar pada kesehatan masyarakat melalui deteksi dini risiko, diagnosis yang lebih akurat, dan dukungan dalam perencanaan rehabilitasi pasien pasca-stroke.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan sistem prediksi stroke menggunakan algoritma *Machine Learning* dengan mempertimbangkan berbagai faktor risiko stroke, Sehingga dapat memberikan kontribusi signifikan dalam upaya pencegahan dan penanganan stroke yang lebih tepat sasaran, serta memberikan pemahaman yang lebih baik tentang interaksi kompleks antara berbagai faktor faktor yang mempengaruhi Stroke. Dengan pendekatan ini peneliti mengambil judul “Prediksi Penyakit Stroke Pada Data Kesehatan Menggunakan Algoritma *Machine Learning*”.

1.2 RUMUSAN MASALAH

Bardasarkan latar belakang diatas, maka dapat dirumuskan bahwa pokok permasalahan yang akan diteliti yaitu :

1. Bagaimana cara mengidentifikasi dan menganalisis faktor-faktor risiko yang berkontribusi terhadap kejadian stroke melalui algoritma *Machine Learning*?
2. Bagaimana akurasi dan kinerja algoritma *Machine Learning*, seperti algoritma *Random Forest*, *Support Vector Machine (SVM)*, dan *Decision Tree*, dalam memprediksi kejadian Stroke berdasarkan data klinis dan faktor risiko yang ada?
3. Apa dampak dari penggunaan *Machine Learning* dalam meningkatkan kesadaran masyarakat mengenai risiko stroke dan mempercepat penanganan medis terhadap kondisi stroke?
4. Bagaimana penerapan *Confusion Matrix* dalam mengevaluasi performa algoritma yang digunakan dalam prediksi stroke?
5. Sejauh mana area di bawah kurva *Receiver Operating Characteristic* dapat menjadi indikator akurasi model dalam prediksi risiko stroke?

1.3 BATASAN MASALAH

Batasan-batasan yang diidentifikasi dalam penelitian ini sebagai berikut :

1. Dataset yang digunakan dalam penelitian ini berasal dari *Kaggle* dengan judul “*Stroke Prediction Dataset*”. Dataset ini memiliki total 5110 data dan 12 fitur serta 2 kelas (0,1).
2. Penelitian ini menggunakan algoritma *Machine Learning* untuk membangun model prediksi stroke berdasarkan Faktor-faktornya.

Algoritma yang akan digunakan dibatasi pada algoritma-algoritma *Machine Learning* seperti *Random Forest*, *SVM* dan *Decision Tree*.

3. Penelitian ini berfokus pada mengembangkan model klasifikasi untuk memperkirakan resiko stroke berdasarkan faktor-faktor tertentu.
4. Penelitian ini menggunakan *Confusion Matrix* untuk mengevaluasi kinerja model, yang mencakup metrik seperti akurasi, presisi, recall, dan F1-score.
5. ROC-AUC (Receiver Operating Characteristic - Area Under Curve) digunakan sebagai metrik tambahan untuk mengevaluasi kemampuan model dalam membedakan antara dua kelas.

1.4 TUJUAN PENELITIAN

Berdasarkan peninjauan yang telah dilakukan penulis, tujuan dari penelitian ini yaitu :

1. Menganalisis faktor risiko utama yang mempengaruhi kemungkinan terjadinya stroke dan membandingkan performa beberapa algoritma *Machine Learning* untuk menentukan algoritma yang paling efektif.
2. Membangun model algoritma *Machine Learning* (*Random Forest*, *SVM*, *Decision Tree*) untuk memprediksi risiko stroke berdasarkan faktor-faktor dalam dataset.
3. Memberikan saran pencegahan berbasis data bagi individu berisiko tinggi, berdasarkan hasil prediksi dari model *Machine Learning*.
4. Menerapkan *Confusion Matrix* untuk mengevaluasi performa prediksi model.

5. Menggunakan *Receiver Operating Characteristic* sebagai indikator akurasi model dalam membedakan risiko stroke.

1.5 MANFAAT PENELITIAN

Berdasarkan uraian yang telah dilakukan penulis, maka diharapkan penelitian ini dapat memberikan manfaat sebagai berikut:

1. Penelitian ini dapat memberikan kontribusi dalam pengembangan teori mengenai penggunaan algoritma *Machine Learning* dalam prediksi risiko stroke, khususnya terkait dengan faktor-faktor resiko stroke.
2. Hasil evaluasi ini dapat memberikan wawasan tentang kelebihan dan kekurangan algoritma *Machine Learning* yang digunakan dalam konteks prediksi stroke.
3. hasil penelitian ini dapat menjadi referensi dalam memilih algoritma yang tepat untuk prediksi penyakit berbasis data.
4. Sistem prediksi ini diharapkan dapat meningkatkan kesadaran akan faktor risiko stroke dan mendorong langkah-langkah pencegahan yang lebih proaktif.
5. Penelitian ini dapat meningkatkan kesadaran masyarakat mengenai pentingnya menjaga kesehatan dan mengelola faktor risiko seperti stroke.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan ini digunakan untuk menjabarkan pembahasan yang ada di setiap bab yang penulis lakukan sebagai berikut :

BAB I : PENDAHULUAN

Bab ini berisikan latar belakang masalah yang dihadapi, perumusan masalah, batasan masalah yang ditetapkan, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI DAN TINJAUAN PUSTAKA

Bab ini mencakup dasar teori yang menjadi acuan dalam melakukan analisis dan juga berisikan teori penunjang dalam memecahkan masalah.

BAB III : METODE PENELITIAN

Bab ini menjabarkan Kerangka kerja penelitian atau alur penelitian, bahan penelitian dan Alat penelitian serta metode yang digunakan pada penelitian penulisan.

BAB IV : PEMBAHASAN DAN HASIL PENELITIAN

Bab ini memaparkan hasil yang diperoleh dari analisis data serta pembahasan yang mendalam tentang hasil tersebut. Hasil yang ditampilkan berupa statistik deskriptif dan hasil evaluasi dari model *Machine Learning* yang digunakan.

BAB V : PENUTUP

Bab ini berisi simpulan dari hasil penelitian yang telah dilakukan serta saran-saran yang dapat diberikan berdasarkan hasil penelitian

ini. Bab ini juga menyarankan arah penelitian lebih lanjut yang dapat dilakukan di masa depan.