

Stroke Prediction using Machine Learning

Yusra Erlangga Putra¹, Sheryl Anastasya², Resky Auliyah Kartini Askin³, Ivan Betrandi⁴, Amaliah Diah⁵
Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin, Makassar, Indonesia
¹yusraerlangg@gmail.com, ²jane.doe@email.com

Abstrak

Stroke merupakan penyebab utama kecacatan jangka panjang dan kematian di seluruh dunia, dengan risiko yang meningkat seiring bertambahnya usia serta adanya faktor risiko seperti hipertensi dan diabetes. Tujuan penelitian ini adalah mengembangkan model prediksi dini dan akurat untuk stroke guna memungkinkan intervensi perawatan kesehatan preventif yang efektif. Studi ini menggunakan algoritma Random Forest dan Support Vector Machine (SVM). Karena ketidakseimbangan pada dataset, teknik Synthetic Minority Oversampling Technique (SMOTE) diterapkan untuk meningkatkan representasi data minoritas. Model dioptimalkan melalui penyetelan hiperparameter menggunakan Bayesian Optimization. Evaluasi model dilakukan dengan metrik akurasi, precision, recall, dan F1-score, dengan validasi silang untuk memastikan keandalan pada data yang belum terlihat. Hasil eksperimen menunjukkan bahwa algoritma SVM dengan SMOTE dan optimasi Bayesian mencapai akurasi tertinggi. Temuan ini menunjukkan bahwa model pembelajaran mesin yang dioptimalkan dapat memberikan kontribusi signifikan dalam prediksi dini stroke dan mendukung pengambilan keputusan dalam sistem perawatan kesehatan preventif.

Kata kunci: Prediksi Stroke, pembelajaran mesin, Random Forest, Support Vector Machine, SMOTE, Bayesian Optimization

1. Introduction

1.1 Latar Belakang

Stroke merupakan masalah kesehatan global yang serius, menempati posisi kedua sebagai penyebab kematian tertinggi dan posisi ketiga sebagai penyebab kecacatan di dunia. Menurut data WHO, satu dari empat orang berisiko mengalami stroke dalam masa hidupnya, dengan 70% kasus terjadi di negara-negara berpenghasilan rendah dan menengah[1]. Kompleksitas faktor risiko stroke yang meliputi tekanan darah tinggi, kolesterol, diabetes, obesitas, dan gaya hidup, menjadikan prediksi stroke sebagai tantangan yang signifikan dalam dunia kesehatan.[2]

Prediksi dini stroke menggunakan metode konvensional seringkali terbatas dalam kemampuannya mengintegrasikan berbagai faktor risiko secara simultan. Perkembangan pembelajaran mesin membuka peluang baru dalam meningkatkan akurasi prediksi stroke. Dengan kemampuannya menganalisis data kompleks dalam jumlah besar, teknik pembelajaran mesin dapat mengidentifikasi pola-pola tersembunyi dan korelasi antar berbagai faktor risiko yang sulit dideteksi melalui metode konvensional.

Penelitian ini mengusulkan pendekatan inovatif dengan mengombinasikan algoritma Random Forest dan Support Vector Machine (SVM) untuk prediksi stroke. Kedua algoritma ini dipilih karena kemampuannya dalam menangani data kesehatan yang kompleks dan menghasilkan model prediksi yang akurat. Untuk mengatasi masalah ketidakseimbangan data yang umum dalam kasus medis, diterapkan teknik SMOTE (Synthetic Minority Oversampling Technique). Selanjutnya, optimasi Bayesian digunakan untuk meningkatkan performa model melalui penyetelan hiperparameter yang optimal. Pendekatan komprehensif ini diharapkan dapat menghasilkan sistem prediksi stroke yang lebih andal dan aplikatif dalam praktik klinis.

1.2 Literature Review

1.3 Research Rationale

1.4 Research Questions and Objectives

2. Research Methods

2.1 Formulas

$$E = mc^2 \quad (1)$$

3. Results and Discussion

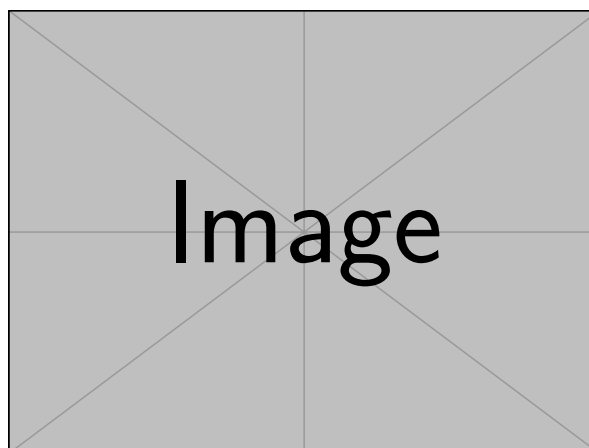


Figure 1: A single-column figure example.

Table 1: Dataset characteristics for stroke prediction

Characteristic	Value	Percentage
Total Samples	5110	100%
Stroke Cases	249	4.87%
Non-stroke Cases	4861	95.13%

4. Conclusion

Acknowledgements

References

- [1] World Health Organization and Dr Poonam Khetrapal Singh, WHO Regional Director for South-East

Asia, “World stroke day,” 2021. Accessed: 27 November 2024.

- [2] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, “Performance analysis of machine learning approaches in stroke prediction,” in *2020 4th international conference on electronics, communication and aerospace technology (ICECA)*, pp. 1464–1469, IEEE, 2020.
-