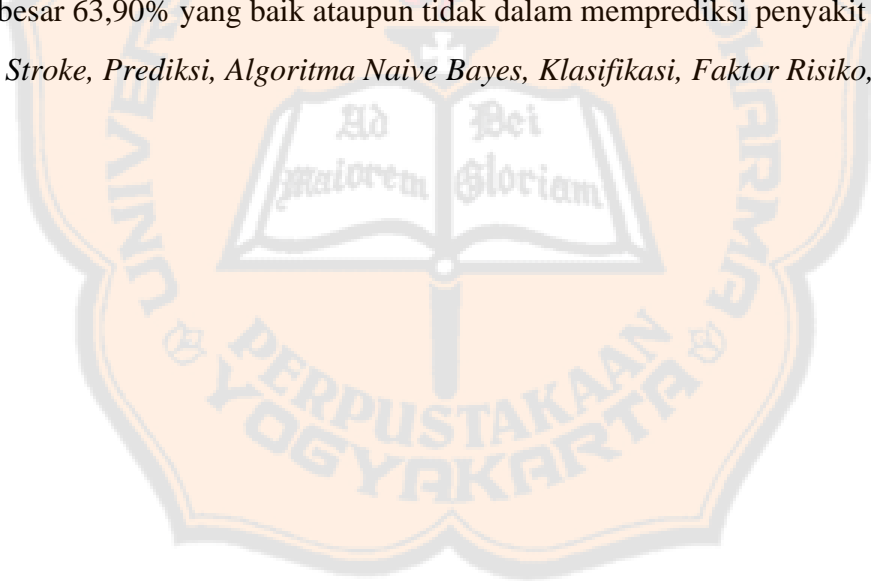


ABSTRAK

Stroke merupakan salah satu penyakit serius yang dapat menyebabkan dampak kesehatan jangka panjang dan bahkan mengancam jiwa. Oleh karena itu, deteksi dini dan prediksi risiko *stroke* menjadi hal penting dalam upaya pencegahan dan penanganan lebih lanjut. Dalam penelitian ini, kami menggunakan algoritma Naive Bayes sebagai metode prediksi untuk mengidentifikasi pasien yang berisiko terkena *stroke*. Data pasien yang terlibat dalam penelitian ini mencakup berbagai variabel seperti usia, jenis kelamin, tekanan darah, indeks massa tubuh (*BMI*), riwayat merokok, dan konsumsi alkohol. Data ini dikumpulkan dari situs *Kaggle* yang merupakan situs penyedia data *open source*.

Metode algoritma Naive Bayes dipilih karena kemampuannya dalam mengelola dan menginterpretasi data dengan berbagai jenis atribut. Algoritma ini mengandalkan asumsi naif bahwa atribut-atribut tersebut bersifat independen, meskipun dalam kenyataannya hal ini mungkin tidak sepenuhnya terpenuhi. Naive Bayes yang akan digunakan ialah model *Gaussian* dimana model ini akan dapat mengatasi data yang bersifat kontinu. Penelitian dilakukan dengan beberapa skenario data diantaranya dengan rasio data latih 0,6 sampai dengan 0,8. Setelah mendapatkan data dengan akurasi optimal, maka melakukan *grid-search* untuk mencari parameter yang menghasilkan akurasi paling baik sebesar 68,06%. Walaupun demikian, penelitian yang telah dilakukan menunjukkan bahwa algoritma *Gaussian* Naive Bayes yang diimplementasikan ke prediksi penyakit *stroke* kurang mendapatkan akurasi yang optimal dan *F1-score* sebesar 63,90% yang baik ataupun tidak dalam memprediksi penyakit *stroke*.

Kata kunci: Stroke, Prediksi, Algoritma Naive Bayes, Klasifikasi, Faktor Risiko, Kesehatan.



ABSTRACT

Stroke is one of the serious diseases that can have long-term health impacts and even threaten life. Therefore, early detection and prediction of the risk of stroke are crucial in preventive efforts and further treatment. In this research, we use the Naive Bayes Algorithm as a predictive method to identify patients at risk of stroke. The patient data involved in this study includes various variables such as age, gender, blood pressure, body mass index (*BMI*), smoking history, and alcohol consumption. This data was collected from the Kaggle site, which is an open-source data provider.

The Naive Bayes algorithm was chosen for its ability to manage and interpret data with various types of attributes. This algorithm relies on the naive assumption that these attributes are independent, although in reality, this may not be entirely fulfilled. The Naive Bayes model to be used is the *Gaussian* model, where this model can handle continuous data. The research is conducted with several data scenarios, including training data ratios ranging from 0.6 to 0.8. After obtaining data with optimal accuracy, a grid-search is then performed to find parameters that result in the best accuracy, reaching 68.06%. However, research that has been carried out shows that the Gaussian Naive Bayes algorithm implemented to predict stroke does not achieve optimal accuracy and the F1-score is 63.90%, which is good or not in predicting stroke.

Keywords: Stroke, Prediction, Naive Bayes Algorithm, Classification, Risk Factors, Health.

