

ABSTRAK

Meningkatnya prevalensi penyakit stroke mengakibatkan kenaikan tingkat kematian dan kecatatan. Dengan demikian, maka perlu dilakukan pencegahan dini pada penyakit stroke. Salah satu solusi yang dapat digunakan yaitu dengan melakukan klasifikasi penyakit stroke berdasarkan atribut yang memiliki nilai *information gain* tidak sama dengan 0. Adapun atribut kriterianya yaitu umur, hipertensi, tekanan darah, tinggi, penyakit jantung, diabetes, berat badan (BMI), status merokok, pekerjaan, status pernikahan, dan tempat tinggal. Klasifikasi dengan *data mining* harus memperhatikan keseimbangan kelas pada data yang digunakan. Penelitian ini menggunakan data sebanyak 4981 dan setelah melalui tahapan *preprocessing* berkurang menjadi 4940 dengan kategori kelas stroke (1) sebanyak 248 dan kelas tidak stroke (0) sebanyak 4692. Terdapat ketidakseimbangan kelas pada data yang digunakan, sehingga dilaksanakan metode *balancing* untuk menyeimbangkan data. Adapun metode *balancing* yang digunakan dalam penelitian ini yaitu *Syntetic Over-sampling Technique (SMOTE)*, *Adaptive Synthetic (ADASYN)*, *Random Oversampling (ROS)*, dan *SMOTE Tomek Link*. Hasil akurasi dari data yang tidak dilakukan *balancing* sebesar 86%. Hasil tersebut lebih tinggi dibandingkan dengan hasil akurasi dengan metode *balancing* menggunakan *SMOTE Tomek Link* sebesar 78%. Meskipun nilai akurasinya lebih rendah dibandingkan dengan data yang tidak dilakukan *balancing*, tetapi nilai presisi, *recall*, dan *f1 score* yang dihasilkan pada metode *balancing* ini lebih tinggi. Berdasarkan hasil presisi, *recall*, dan *f1 score* yang dihasilkan oleh setiap metode *balancing* menunjukkan bahwa *SMOTE Tomek Link* merupakan metode yang memberikan hasil presisi, *recall*, dan *f1 score* yang paling optimal. Pengujian dengan *SMOTE Tomek Link* menghasilkan akurasi tertinggi sebesar 78%, presisi sebesar 76%, *recall* sebesar 83%, dan *f1 score* sebesar 79%. Berdasarkan hasil pengujian dari setiap perbandingan kelas pada data diperoleh bahwa semakin seimbang kelas maka nilai presisi, *recall*, dan *f1 score* semakin meningkat untuk semua metode *balancing* yang digunakan yaitu *SMOTE*, *ADASYN*, *ROS*, dan *SMOTE Tomek Link*.

Kata kunci: *Klasifikasi Penyakit Stroke, Metode Balancing, Akurasi, Recall, Presisi, F1 Score*

ABSTRACT

The increasing prevalence of stroke has resulted in an increase in death and disability rates. Thus, it is necessary to do early prevention of stroke. One solution that can be used is to classify stroke based on attributes that have an information gain value not equal to zero. The attribute criteria are age, hypertension, blood pressure, height, heart disease, diabetes, weight (BMI), smoking status, occupation, marital status, and place of residence. Classification with data mining must pay attention to class balance in the data used. This study used 4981 data and after going through the preprocessing stage, the amount of data used was reduced to 4940 with 248 stroke class categories (1) and 4692 non-stroke class categories (0). There was a class imbalance in the data used, so balancing method was implemented to balance data. The balancing methods used in this study are Synthetic Over-sampling Technique (SMOTE), Adaptive Synthetic (ADASYN), Random Oversampling (ROS), and SMOTE Tomek Link. The accuracy of the data that is not balanced is 86%. These results are higher than the results of the accuracy of the balancing method using SMOTE Tomek Link of 78%. Even though the accuracy value is lower than the data that is not balanced, the value of precision, recall, and f1 score produced in this balancing method is higher. Based on the precision, recall, and f1 score results produced by each balancing method, it shows that SMOTE Tomek Link is the method that provides the most optimal precision, recall, and f1 score results. Testing with SMOTE Tomek Link resulted in the highest accuracy of 78%, precision of 76%, recall of 83%, and F1 score of 79%. Based on the test results of each class comparison on the data, it was found that the more balanced the classes, the value of precision, recall, and f1 scores increased for all balancing methods used, namely SMOTE, ADASYN, ROS, and SMOTE Tomek Link.

Keywords: Classification of Stroke, Balancing Method, Accuracy, Recall, Precision, F1 Score