

## BAB V

### PENUTUP

#### 5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan pada *Brain Stroke Dataset*, maka dapat ditarik kesimpulan sebagai berikut :

1. proses menggali informasi stroke otak pada *Brain Stroke Dataset* menggunakan algoritma C5.0 yang merupakan algoritma berbasis *decision tree*, dengan melakukan *prerocessing* yang melibatkan teknik *impute missing value*, *delete data duplication*, *data transformation* dan *data discritization* agar dapat meningkatkan kualitas serta menghindari hasil penambangan yang berkualitas rendah. dengan metode pembuatan model pohon keputusan menggunakan *5-fold cross validation* serta 4 *holdout* yang berbeda yakni 60% : 40%, 70% : 30%, 80% : 20%, dan 90% : 10%.
2. Hasil terbaik didapatkan dari *holdout* 90% : 10% dengan akurasi mencapai 95,18%, *recall* “tidak” 99,37%, *recall* “ya” hanya 8,7%, dan *precision* “tidak” mencapai 95,74% dan *precision* “ya” hanya 40,00%. Hasil ini didapat karena *data training* yang lebih besar dibandingkan dengan *data testing* sehingga model dapat belajar dari *data training* lebih banyak akan tetapi dengan pengujian *data testing* lebih sedikit.
3. Dari beberapa pohon keputusan yang diperoleh didapatlah bahwa atribut yang paling berpengaruh adalah *heart\_desease*, *hypertension* dan *avg\_glucose\_level*. Dimana pada penelitian sebelumnya[12] atribut yang paling berpengaruh adalah *age*, *bmi* dan *avg\_glucose\_level*. Sehingga dapat disimpulkan bahwa

*heart\_disease*, *hypertension* dapat menjadi penyebab stroke otak dan memperkuat kemungkinan *avg\_glucose\_level* merupakan penyebab dari stroke otak.

4. Hasil validasi yang tidak berbeda jauh mengindikasikan bahwa model stabil. Akan tetapi *precision* dan *recall* “ya” pada proses validasi dan testing mengindikasikan adanya imbalance dataset.

## 5.2 SARAN

Adapun saran yang dapat diberikan berdasarkan hasil penelitian yang telah dilakukan yaitu:

1. Diharapkan kedepannya *Brain Stroke Dataset* dapat digunakan untuk penelitian selanjutnya, dimanana penanganan *data imbalance* dapat menjadi perhatian penelitian selanjutnya sebelum membuat model, dengan melakukan *oversampling* ataupun *undersampling* untuk meningkatkan kinerja model yang lebih baik dalam memprediksi stroke otak.
2. Diharapkan pada penelitian selanjutnya untuk menggunakan algoritma *decision tree* dengan penyesuaian parameter agar menghasilkan model yang lebih baik. Seperti nilai *confidence*, *Minimum Impurity Split*, *Minimum Samples Split*, dan *Minimum Samples Leaf*.