

BAB V

KESIMPULAN

5.1 KESIMPULAN

Untuk kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan pada bab-bab sebelumnya, yaitu :

1. Penelitian ini menggunakan *dataset online* yang bersumber dari www.kaggle.com (<https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset>) yang berjumlah 615 data. *Dataset* ini memiliki 13 atribut termasuk *class* yang akan menjadi hasil keputusan yang berbentuk sebuah pohon (*decision tree*). Setelah di-*cleaning* dataset yang berjumlah 615 data berubah menjadi 589 data.
2. Berikut merupakan atribut yang ada pada dataset untuk klasifikasi penyakit *Hepatitis C* : *Category, Age, Sex, Albumin Blood Test, Alkaline phosphatase, Alanine Transaminase, Aspartate Transaminase, Bilirubin, Acetylcholinesterase, Cholesterol, Creatinine, Gamma-Glutamyl Transferase, Proteins*.
3. Dari hasil klasifikasi dengan menggunakan metode *decision tree* dapat dilihat dari atribut yang paling banyak yaitu "*Blood Donor*" dalam dataset dapat dilihat penyakit *Hepatitis C* paling banyak tertular atau terinfeksi melalui donor darah atau yang bersentuhan langsung dengan darah.

4. Hasil klasifikasi menggunakan *Data Training* sebagai model pohon keputusan menghasilkan bentuk pohon dan jumlah *node* yang berbeda, seperti berikut :
 - a.) Menggunakan *Data Training* 60% Pada WEKA = 3 node.
 - b.) Menggunakan *Data Training* 70% Pada WEKA = 9 node.
 - c.) Menggunakan *Data Training* 80% Pada WEKA = 7 node.
 - d.) Menggunakan *Data Training* 60% Pada RapidMiner = 11 node.
 - e.) Menggunakan *Data Training* 70% Pada RapidMiner = 10 node.
 - f.) Menggunakan *Data Training* 80% Pada RapidMiner = 12 node.
5. Hasil klasifikasi dengan pengujian *Data Testing* untuk membandingkan akurasi tertinggi yang diperoleh dari 2 *tools* dan 2 model pengujian :
 - a.) WEKA menggunakan *Use Training Set* = 91.94% (*data testing* 40%)
 - b.) WEKA menggunakan *10 Fold-Cross Validation* = 84.74% (*data testing* 40%)
 - c.) *RapidMiner* menggunakan *Use Training Set* = 92.37% (*data testing* 40%)
 - d.) *RapidMiner* menggunakan *10 Fold-Cross Validation* = 83.90% (*data testing* 40%)

Dapat disimpulkan bahwa *tools* yang dipakai hasilnya sama baik, hasil akurasi tertinggi dengan menggunakan *Use Training Set* adalah *RapidMiner* sedangkan dengan menggunakan *10 Fold-Cross Validation* adalah WEKA.

5.2 SARAN

Adapun saran yang dapat penulis berikan dari hasil penelitian ini adalah sebagai berikut :

1. Diharapkan untuk penelitian selanjutnya dapat digunakan untuk meningkatkan performa klasifikasi *Hepatitis C*.
2. Diharapkan penelitian ini dapat diujikan menggunakan metode dan algoritma *data mining* lainnya, khususnya yang memiliki performa yang lebih baik dan klasifikasi lainnya.
3. Diharapkan kedepannya adanya penelitian yang melakukan perbandingan dengan metode dan algoritma lainnya.