

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan oleh peneliti pada perbandingan algoritma *naïve bayes classifier* dengan *random forest* untuk klasifikasi penyakit ginjal kronis dapat ditarik kesimpulan, yaitu :

1. Pada penelitian ini data yang digunakan ialah data penyakit ginjal kronis yang didapatkan dari Kaggle dengan jumlah keseluruhan data sebanyak 400 serta 25 atribut yang telah di seleksi menjadi 14 atribut saja yaitu *Age* (umur), *Red Blood Cell* (sel darah merah), *Pus Cell* (sel nanah), *Bacteria*, *Blood Glucose Random* (gula darah), *Blood Urea*, *Sodium*, *Potassium*, *Hypertension*, *Diabetes Mellitus*, *Coranary Artery Disease* (penyakit jantung koroner), *Appetite* (nafsu makan), *Pedal Edema* (Pembengkakan), *Anemia*. Yang bila 400 data itu memiliki Nan (data kosong) yang diisi dengan nilai rata-rata maka data akan full dan jika dibersihkan maka hanya akan menyisakan 158 data saja yang akan menjadi perbandingan. 14 atribut yang telah disebutkan menunjukkan bahwa atribut yang paling berpengaruh terhadap perbandingan algoritma ini ialah Hipertensi.
2. Dari hasil perbandingan 2 (dua) algoritma tersebut didapatkan hasil sebagai berikut. Hasil perbandingan algoritma dimana untuk *naïve bayes classifier* pada dataset 1 (400) memiliki akurasi paling tinggi yaitu 94%

dibandingkan *random forest* yang terpaut 1% lebih rendah dengan 93% saja. Pada dataset

2 (158) *random forest* cenderung memberikan hasil lebih baik dengan memperoleh 87%, serta performa model menurun terlihat dari akurasi 77% maupun nilai r^2 yang negatif -0.06 didapatkan *naïve bayes*. Terakhir menghasilkan nilai *mean absolute error* 0.1 yang berarti *overfitting* dimana model *random forest* menghasilkan data yang sempurna. Dari penjelasan tersebut dinyatakan bahwa *random forest* lebih unggul dalam melakukan kinerjanya dibandingkan *naïve bayes classifier*.

5.2 SARAN

Berdasarkan penelitian yang dilakukan oleh peneliti dapat memberikan saran-saran sebagai berikut :

1. Untuk algoritma *naïve bayes classifier* itu sebaiknya jangan lagi dipilih sebagai model penelitian yang sama atau lebih lanjut serta karakteristik data 158 kurang begitu baik karna ketidakstabilan data sehingga menghasilkan kinerja performa yang kurang baik pula.
2. Serta diharapkan data atau penelitian ini dapat diujikan kembali menggunakan algoritma lainnya.