

BAB V

PENUTUP

5.1 Kesimpulan

Dari penelitian ini ada beberapa kesimpulan yang dapat diambil dan dijabarkan berikut ini.

1. *Dashboard data analytics* dibangun dalam bentuk aplikasi berbasis *website*. *Website* tersebut akan menampilkan visualisasi data pada *dashboard* yang berkaitan dengan penerimaan mahasiswa baru dan memiliki fitur untuk melakukan filter terhadap data yang akan disajikan pada *dashboard* berdasarkan tahun akademik penerimaan mahasiswa baru. Tujuannya, memudahkan pengguna sistem untuk mendapatkan visualisasi data sesuai dengan spesifikasi tahun yang dipilih.
2. Penyajian visualisasi data pada *dashboard data analytics* menggunakan data yang telah melakukan tahap *pre-processing data*. *Dataset* yang digunakan pada tahap *pre-processing* adalah data penerimaan mahasiswa baru di UPN “Veteran” Jakarta dalam kurun waktu 5 (lima) tahun terakhir. Penampilan visualisasi data pada *website* memiliki tujuan utama untuk melakukan *monitoring* terhadap berjalannya sistem penerimaan mahasiswa baru di UPN “Veteran” Jakarta dan membantu pengguna sistem, pihak yang berhak mengambil keputusan, melakukan pertimbangan terhadap keputusan atau kebijakan yang akan dikeluarkan secara objektif berdasarkan data yang ada.
3. Dalam proses melakukan prediksi, *dataset* dilakukan implementasi terhadap 2 (dua) model, yaitu algoritma *Random Forest* dan algoritma *Gradient Boosting*. Keduanya mengimplementasikan dataset yang sama, penerimaan mahasiswa baru dalam kurun waktu 5 (lima) tahun terakhir. Hasil dari setiap prosesnya, didapatkan bahwa penggunaan algoritma *Random Forest* mendapatkan evaluasi akurasi *data training*

sebesar 0.9751 (98%) dan akurasi *data testing* sebesar 0.7996 (80%), sedangkan penerapannya pada algoritma *Gradient Boosting* didapatkan evaluasi akurasi *data training* sebesar 0.8309 (83%) dan akurasi *data testing* sebesar 0.8161 (82%). Dari kedua algoritma yang telah diimplementasikan dapat disimpulkan bahwa algoritma *Gradient Boosting* memiliki selisih nilai akurasi pada *data training* dan *data testing* yang ideal. Walaupun terlihat akurasi pada algoritma *Random Forest* lebih besar daripada *Gradient Boosting*, tetapi algoritma *Random Forest* memiliki selisih nilai akurasi data training dan data testing yang tinggi, yaitu 18%. Tingginya selisih nilai akurasi antara *data training* dan *data testing* membuat model ini dikatakan *overfitting*, model yang dibentuk menyimpang karena mengandung *noise* pada *data training*. Disimpulkan bahwa algoritma Gradient Boosting menjadi algoritma yang diimplementasikan terhadap model untuk memperoleh hasil prediksi terhadap keputusan mahasiswa baru dalam melanjutkan studinya di UPN “Veteran” Jakarta setelah dirinya lolos seleksi penerimaan.

5.2 Saran

Dari hasil penelitian yang telah dijalankan, banyak saran yang diberikan penulis agar penelitian ini dapat dimaksimalkan oleh peneliti lain yang berkesempatan. Saran yang diusulkan oleh peneliti dijabarkan pada poin-poin berikut ini:

1. Penggunaan algoritma pada model sebagai pembanding dapat menggunakan algoritma lain selain dari *Random Forest* dan *Gradient Boosting* untuk melakukan pengecekan ulang mengenai evaluasi akurasi yang paling baik.
2. Pemrosesan data dilakukan tidak hanya pada jenjang sarjana dan diploma, tetapi juga dilakukan pada jenjang lainnya seperti jenjang magister.
3. Cakupan data yang diproses tidak hanya pada jalur penerimaan SNBP, SNBT, dan SEMA UPNVJ saja, tetapi juga pada penerimaan

mahasiswa baru terhadap jalur non-akademik.

4. Sebelum dilakukan pembentukan model terhadap algoritma *machine learning*, *dataset* dilakukan proses keseimbangan data antara kedua label, seperti menerapkan resampling techniques dengan penghapusan sampel dari kelas mayoritas (*undersampling*) atau penambahan lebih banyak contoh dari kelas minoritas (*oversampling*).