PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE *ADABOOST*

SKRIPSI

UIN SUNAN AMPEL
S U R A B A Y A

DISUSUN OLEH:

MUHAMMAD NAUFAL RABBANI

H76216068

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2020

LEMBAR PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini,

Nama	: Muhammad Naufal Rabbani
NIM	: H76216068
Program Studi	: Sistem Informasi
Angkatan	: 2016

Menyatakan bahwa saya tidak melakukan plagiasi dalam penulisan skripsi saya yang berjudul: "PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE *ADABOOST*", apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 28 Juli 2020

Yang Menyatakan,

METERAI TEMPEL

TGL. - - 20

A1E29AHF548423536

6000

ENAM RIBU RUPIAH

Muhammad Naufal Rabbani

NIM. H76216068

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh:

NAMA	:	MUHAMMAD NAUFAL RABBANI
NIM	:	H76216068
JUDUL	:	PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE ADABOOST

Ini telah diperiksa dan disetujui untuk diujikan

Surabaya, 25 Februari 2020

Dosen Pembimbing 1	Dosen Pembimbing 2
Dwi Rolliawati, MT	Ahmad Yusuf, M. Kom
NIP. 197909272014032001	NIP. 199001202014031003

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh Muhammad Naufal Rabbani ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 19 Juni 2020

Mengesahkan,
Dewan Penguji

Penguji I	Penguji II
Dwi Rolliawati, MT	Ahmad Yusuf, M. Kom
NIP. 197909272014032001	NIP. 199001202014031003
Penguji III	**Penguji IV**
Nita Yalina, S.Kom., M.MT	Khalid, M. Kom
NIP. 198702082014032003	NIP. 197906092014031002

Mengetahui,
PLT. Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. Evi Fatimatur Rusydiyah, M.Ag.
NIP. 197312272005012003

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : MUHAMMAD NAUFAL RABBANI
NIM : H76216068
Fakultas/Jurusan : FAKULTAS SAINS DAN TEKNOLOGI / SISTEM INFORMASI
E-mail address : bosnaufalemail@gmail.com / H76216068@uinsby.ac.id

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
■ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :
PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE *ADABOOST*

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 28 Juli 2020

Penulis

(Muhammad Naufal Rabbani)
nama terang dan tanda tangan

ABSTRAK

PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE *ADABOOST*

Oleh:

Muhammad Naufal Rabbani

Setiap tahunnya perguruan tinggi melakukan penerimaan mahasiswa baru secara rutin untuk membuka awal tahun ajaran baru. Namun tingginya jumlah mahasiswa yang mengundurkan diri menyebabkan banyaknya jumlah kursi kosong yang tersisa. Pengunduran diri yang terjadi bisa diminimalisir apabila seleksi calon mahasiswa baru dilakukan dengan tepat. Salah satu caranya dengan membuat model prediksi berbasis *machine learning* untuk membantu proses seleksi kandidat yang berpotensi menyelesaikan proses penerimaan hingga akhir berdasarkan data yang ada. Agar hal tersebut bisa tercapai, dibuatlah model prediksi menggunakan algoritma *Adaboost* sekaligus membandingkan performanya dengan model alogoritma *Decision Tree.* Untuk memaksimalkan peforma model, maka dilakukan analisa variabel dengan menggunakan *chi square* dalam proses *feature selection*-nya. Hasil akhir menunjukkan bahwa model prediksi *Adaboost* memiliki peforma yang lebih baik daripada model *Decision Tree* dengan skor *f-measure* 90.9%, *precision* 83.7% dan *recall* 99.5%. Selain itu didapatkan juga ciri dari kandidat yang cenderung melanjutkan pendaftaran hingga akhir. Sehingga dengan hasil tersebut bisa membantu pihak perguruan tinggi dalam mengambil keputusan dalam proses seleksi calon mahasiswa baru.

Kata Kunci—Classification, Adaboost, Ensemble Learning, Enrollment

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

ABSTRACT

PENGEMBANGAN SISTEM PREDIKSI DAFTAR ULANG CALON MAHASISWA BARU MENGGUNAKAN METODE *ADABOOST*

Oleh:

Muhammad Naufal Rabbani

Every year the college accepts new students regularly to begin a new school year. However, the high number of students who resigned causes the large number of empty seats left. Resignations that occur can be minimized if the selection of new students is done properly. One of the solution is to create a machine learning based prediction model to help the selection process of candidates who potential to complete the enrollment process untill the end by reading on existing data. In order to achieve it, a prediction model was made by using the Adaboost algorithm and comparing its performance with the Decision Tree algorithm model. To maximize the performance of the model, a variable analysis is performed using chi square analysis for the feature selection process. The final results show that the Adaboost prediction model has a better performance than the Decision Tree model with an f-measure score of 90.9%, precision 83.7% and recall 99.5%. In addition, there were also explanation of the characteristics of candidates who tended to continue the enrollment process until finish. So with these results can help the university to make decisions in the selection process of new students.

Keyword—Classification, Adaboost, Ensemble Learning, Enrollment

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR ISI

HALAMAN JUDUL i
LEMBAR PERSETUJUAN PEMBIMBING ii
PENGESAHAN TIM PENGUJI SKRIPSI iii
LEMBAR PERNYATAAN KEASLIAN iv
LEMBAR PERNYATAAN PUBLIKASI KARYA ILMIAH v
MOTTO vi
KATA PENGANTAR vii
ABSTRAK ix
ABSTRACT x
DAFTAR ISI xi
DAFTAR GAMBAR xii
DAFTAR TABEL xiv
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Perumusan Masalah 3
1.3. Batasan Masalah 4
1.4. Tujuan Penelitian 4
1.5. Manfaat Penelitian 4
1.6. Sistematika Penulisan 5
BAB II TINJAUAN PUSTAKA 7
2.1. Tinjauan Penelitian Terdahulu 7
2.2. Teori-Teori Dasar 10
2.2.1. *Data Mining* 10
2.2.2. Metode Pelatihan Model 11
2.2.3. Klasifikasi 12
2.2.4. Jenis-Jenis Variabel 13
2.2.5. Algoritma *Decision Tree* 14
2.2.6. Algoritma CART 15
2.2.7. *AdaBoost* 16
2.2.8. *Confusion Matrix* 18
2.2.9. *Chi Square* 19

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.10. REST API 20
2.2.11. JSON 21
2.2.12. *Python* 22
2.2.13. *Flask* 22
2.3. Integrasi Keilmuan 22
BAB III METODOLOGI PENELITIAN 24
3.1 Tahapan Penelitian 24
3.1.1. Perumusan Masalah 24
3.1.2. Studi Literatur 25
3.1.3. Pengolahan Data 26
3.1.4. Analisa Variabel 30
3.1.5. Pelatihan Model 30
3.1.6. Pengembangan Sistem 31
3.1.7. Pengujian Sistem 32
3.1.8. Evaluasi Hasil 32
BAB IV HASIL DAN PEMBAHASAN 33
4.1 Hasil 33
4.1.1. Pengolahan data 33
4.1.2. Analisa Variabel 37
4.1.3. Pelatihan Model 41
4.1.4. Pengembangan Sistem 45
4.1.5. Evaluasi Sistem 48
4.2. Pembahasan 49
BAB V KESIMPULAN DAN SARAN 53
5.1. Kesimpulan 53
5.2. Saran 53
DAFTAR PUSTAKA 55

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR GAMBAR

Gambar 2.1: *Training Flow AdaBoost* 17
Gambar 2.2.: Contoh Struktur URL dari REST API 20
Gambar 2.3: Contoh data user dalam format JSON.....22
Gambar 4.1: Perbandingan Jumlah *Class* pada dataset sesudah diolah 37
Gambar 4.2: Cuplikan kode *Adaboost Decision Tree Model*.....42
Gambar 4.3: Desain Arsitektur pada sistem prediksi.....45
Gambar 4.4: *Activity Diagram* pada proses *training model*.....46
Gambar 4.5: *Activity Diagram* pada proses prediksi data.....47
Gambar 4.6: *setting url Postman* untuk uji coba *endpoint /train*.....48
Gambar 4.7: *Response server* hasil *training* model baru 48
Gambar 4.8: *list* nomor pendaftaran untuk uji coba *endpoint /predict* 49
Gambar 4.9: *Response server* hasil prediksi pada data mahasiswa.....49
Gambar 4.10: Perbandingan skenario terbaik dari *Adaboost* dan *Decision Tree*...51

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR TABEL

Tabel 1.1: Data rangkuman pagu pendaftaran tahun 2019 1
Tabel 2.1: Tinjauan Pustaka. 7
Tabel 2.2: Label-label pada Confusion Matri*x* 18
Tabel 2.3: Desain *resource* REST API 21
Tabel 3.1: Atribut-atribut yang ada pada penelitian terdahulu.....24
Tabel 3.2: Atribut yang akan digunakan berdasarkan penelitian terdahulu.....26
Tabel 3.3: Rancangan struktur data untuk hasil akhir pengolahan 27
Tabel 3.4: Struktur Tabel Hasil *Query Database*.....28
Tabel 3.5: Pemilihan Fitur dan Label.....30
Tabel 3.6: Skenario pembagian data 31
Tabel 3.7: Daftar *endpoint* pada *REST API Web Service* 32
Tabel 4.1: Detil tahapan *Pre processing* pada data per atributnya.....33
Tabel 4.2: *Data Validation*.....34
Tabel 4.3: Sampel *dataset* sebelum dilakukan *pre processing* 35
Tabel 4.4: Sampel *dataset* setelah dilakukan *pre processing*.....36
Tabel 4.5: Perbandingan Komposisi data sebelum dan sesudah diolah.....36
Tabel 4.6: Hasil perhitungan *chi square* pada masing-masing atribut.....37
Tabel 4.7: Tabel observasi pada atribut usia.....38
Tabel 4.8: Tabel observasi pada atribut *jenis_sekolah_asal*.....39
Tabel 4.9: Tabel Observasi pada atribut *jalur_masuk*.....39
Tabel 4.10: Tabel Observasi pada atribut *penghasilan_orang_tua* 40
Tabel 4.11: Tabel Observasi pada atribut *prodi_terpilih*.....40
Tabel 4.12: Tabel *Confusion Matrix* model *Adaboost* skenario 1 42
Tabel 4.13: Hasil pengujian *Adaboost Decision Tree*.....43
Tabel 4.14: Hasil pengujian *Decision Tree*. 44

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB I
PENDAHULUAN

1.1. Latar Belakang

Setiap tahunnya perguruan tinggi melakukan penerimaan mahasiswa baru secara rutin untuk membuka awal tahun ajaran baru. Kegiatan ini merupakan salah satu aktivitas penting pada sebuah perguruan tinggi. Jika dilihat dari perspektif *value chain* perguruan tinggi, kegiatan ini adalah masukan logistik pada aktivitas utama yang terjadi di dalam perguruan tinggi. Oleh karena itu, untuk mencapai keberhasilan perguruan tinggi perlu adanya manajemen proses penerimaan mahasiswa baru yang efektif (Niswatin, 2016).

Salah satu pengambilan keputusan yang sulit pada proses penerimaan mahasiswa baru adalah penentuan pagu (jumlah kursi yang tersedia) untuk mahasiswa baru. Hal ini bisa dilihat pada Tabel 1.1 yang menampilkan jumlah sisa kursi yang tidak terisi dari total pagu yang tersedia pada setiap fakultas di Universitas Islam Negeri Sunan Ampel Surabaya dengan total kursi kosong hingga 324 kursi. Hal ini menandakan kurang tepatnya seleksi mahasiswa baru dan banyaknya calon mahasiswa baru yang mengundurkan diri (tidak melanjutkan proses pendaftaran).

Tabel 1.1: Data rangkuman pagu pendaftaran tahun 2019 yang bersumber dari rekam historis proses pendaftaran yang ada pada *database*

Fakultas	Total Pagu	Daftar Ulang	Tidak Daftar Ulang	Sisa pagu
Adab dan Humaniora	460	434	69	26
Dakwah dan Komunikasi	665	645	175	20
Ekonomi dan Bisnis Islam	630	591	116	39
Ilmu Sosial dan Ilmu Politik	333	310	31	23
Psikologi dan Kesehatan	140	134	11	6
Sains dan Teknologi	420	362	27	58
Syariah dan Hukum	760	707	146	53
Tarbiyah dan Keguruan	770	729	131	41

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Fakultas	Total Pagu	Daftar Ulang	Tidak Daftar Ulang	Sisa pagu
Ushuluddin dan Filsafat	595	537	131	58

Apabila seleksi mahasiswa baru dilakukan dengan tepat, maka pemasukan dari lembaga perguruan tinggi dapat dioptimalkan. Hal ini dikarenakan UKT merupakan sumber pemasukkan perguruan tinggi. Sesuai dengan Undang-undang Nomor 12 Tahun 2012 tentang pendidikan tinggi, pemerintah mengatur bahwa pembiayaan pendidikan tinggi bersumber dari Anggaran Pendapatan dan Belanja Negara (APBN) dan mahasiswa yang berupa Uang Kuliah Tunggal (UKT) (Hasanuddin, 2019).

Pemasukan dari uang kuliah tunggal bisa dimaksimalkan bila jumlah kursi kosong diminimalisir. Jika dihitung, rata-rata UKT mahasiswa Universitas Islam Negeri Sunan Ampel Surabaya pada periode tersebut sebesar Rp. 5,600,000.00 dikalikan dengan jumlah kursi yang kosong, maka didapatkan nominal hingga Rp. 1,814,400,000.00. Untuk mengurangi kursi kosong dan mencapai perencanaan yang efektif dalam proses penerimaan mahasiswa baru dapat dilakukan dengan pendekatan *data mining* (Melati dkk., 2018).

Data mining digunakan untuk melakukan prediksi (*Forecasting)* berdasarkan data kuantitatif yang telah tercatat pada beberapa periode penerimaan mahasiswa baru sebelumnya. Sehingga dengan adanya prediksi ini bisa digunakan sebagai pembantu pengambilan keputusan dan menentukan kebutuhan sumber daya di masa yang akan datang (Handoko, 2019).

Data mahasiswa dari proses penerimaan yang berupa data diri, keputusan mahasiswa untuk mendaftar ulang, data orang tua, riwayat pendidikan, dan data demografis dapat dimanfaatkan untuk memprediksi calon mahasiswa yang kemungkinan besar melakukan daftar ulang ataupun tidak dengan menggunakan teknik klasifikasi (Wanjau & Muketha, 2018). Dengan adanya upaya klasifikasi menggunakan *data mining* akan meningkatkan performa dalam proses penerimaan mahasiswa baru (Saini & Kumar Jain, 2013). Terdapat banyak metode klasifikasi yang dapat digunakan sebagai solusi, namun salah satu metode klasifikasi yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sering digunakan pada *Educational Data Mining* (EDM) adalah *Decision Tree* (Saa dkk., 2018).

Decision Tree mampu melakukan klasifikasi pada data mahasiswa dengan akurasi yang tinggi (Hoiriyah, 2018). Metode *Decision Tree* juga dinilai lebih baik dari beberapa metode lainnya dalam *Educational Data Mining* (Ab Ghani dkk., 2019; Yahya & Jananto, 2019). Selain itu *Decision Tree* memiliki beberapa algoritma pilihan yang digunakan untuk membentuk pohon keputusan seperti *ID3*, *C4.5*, dan *CART* (Patel & Prajapati, 2018). Namun algoritma *CART* memiliki performa yang lebih baik daripada 2 algoritma lainnya (Lakshmi dkk., 2013). Untuk menambah akurasi dari *Decision Tree,* bisa dilakukan dengan pendekatan *Boosting* yaitu dengan algoritma *Adaboost*. *Adaboost* dinilai mampu meningkatkan skor akurasi dari model *Decision Tree* menjadi lebih baik (Khan dkk., 2020; Rais & Subekti, 2019).

Melalui *Boosting* menggunakan *Adaboost* pada model *Decision Tree (CART)* diharapkan mampu memprediksi jumlah mahasiswa yang melanjutkan proses daftar ulang. Angka dari jumlah mahasiswa yang memiliki probabilitas tinggi melakukan daftar ulang bisa dijadikan dasar untuk membantu pengambilan keputusan dalam seleksi mahasiswa baru. Hal ini menjadi sebab mengapa mengambil pola data penerimaan mahasiswa yang telah lalu menjadi hal penting yang seharusnya dilakukan oleh perguruan tinggi (Aradea dkk., 2011).

1.2. Perumusan Masalah

Dari latar belakang diatas, maka dapat dirumuskan beberapa poin masalah antara lain:

1. Variabel-variabel apa saja yang memiliki hubungan terhadap status daftar ulang mahasiswa?
2. Bagaimana peforma metode *Adaboost* pada *decision tree* untuk prediksi daftar ulang calon mahasiswa baru?

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

1.3. Batasan Masalah

Untuk menjaga penelitian dari masalah-masalah yang ada di luar cakupan maka dibuatlah beberapa batasan masalah, sebagai berikut:

1. Data yang digunakan adalah data calon mahasiswa tahun 2018 dan tahun 2019 yang bersumber dari *database* siakad dan *database* aplikasi UKT Universitas Islam Negeri Sunan Ampel Surabaya.
2. Hasil akhir penelitian berupa *REST API web service*. Sehingga bisa dimanfaatkan untuk integrasi dengan aplikasi lain.

1.4. Tujuan Penelitian

Adapun penelitian ini bertujuan untuk mencapai beberapa hal sebagai berikut:

1. Mencari variabel-variabel yang memilliki hubungan dengan status daftar ulang calon mahasiswa berdasarkan data yang telah diseleksi. Sehingga dapat dijadikan bahan pertimbangan tambahan untuk proses penerimaan mahasiswa baru pada periode berikutnya.
2. Mengetahui peforma adaboost dalam prediksi kemungkinan daftar ulang calon mahasiswa baru.

1.5. Manfaat Penelitian

Diharapkannya agar hasil penelitian ini dapat memberikan manfaat demi memenuhi kebutuhan segala pihak dan lingkungan yang memiliki keterkaitan dalam penelitian,diantaranya:

1. Secara Akademis

a. Menggunakan algoritma *Adaboost* untuk melakukan *Boosting* pada model *Decision Tree* untuk menghasilkan prediksi yang lebih baik terhadap kemungkinan daftar ulang calon mahasiswa baru sebagai dasar pengambilan keputusan oleh perguruan tinggi untuk menentukan mahasiswa yang terpilih.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

b. Mengantisipasi adanya *imbalance data* pada data latih dengan menggunakan algoritma *Adaboost*.

c. Mencari tahu ada atau tidaknya hubungan variabel-variabel terpilih dengan status daftar ulang mahasiswa.

2. Secara Aplikatif

a. Membantu pihak perguruan tinggi dalam melakukan prediksi terhadap kemungkinan daftar ulang oleh calon mahasiswa baru. Sehingga setelah dilakukan prediksi dapat ditentukan hasil seleksi yang ideal.

b. Membantu pihak perguruan tinggi mengolah data dalam proses penerimaan mahasiswa baru sehingga nantinya akan memberikan suatu informasi tambahan tentang pola data calon mahasiswa baru baik yang melanjutkan maupun yang mengundurkan diri.

1.6. Sistematika Penulisan

Adapun penulisan skripsi Program Studi Sistem Informasi Fakultas Sains dan Teknologi memiliki aturan dan sistematika yang dibagi atas beberapa bagian, yakni sebagai berikut:

a. BAB I PENDAHULUAN

Pendahuluan menjelaskan latar belakang permasalahan yang diangkat dalam penelitian yaitu banyaknya kursi kosong karena sulit melakukan seleksi mahasiswa baru yang benar-benar melanjutkan pendaftaran (tidak mengundurkan diri), perumusan masalah, tujuan penelitian, dan manfaat yang diperoleh dari penelitian yang dilakukan.

b. BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang penjelasan penelitian sebelumnya yang relevan dengan penelitian ini, dan penjelasan singkat tentang teori–teori yang terkait dengan penelitian ini seperti konsep sistem informasi, metode yang terkait dengan penelitian ini seperti konsep sistem informasi, metode pengembangan sistem informasi, Bahasa pemrograman, metode *Adaboost* dan sebagainya.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

c. BAB III METODOLOGI PENELITIAN

Bab ini berisikan tentang penjelasan atas alur sistematika daripada cara penelitian yang terdiri dari beberapa tahap, antara lain melakukan identifikasi terhadap kebutuhan, melakukan perencanaan untuk tahap perencanaan dan tahap pengembangan model *prototype* dan tahap implementasi.

d. BAB IV HASIL DAN PEMBAHSAN

Bab ini berisikan paparan hasil dan pembahasan tentang bagaimana alur sebuah data apabila diolah dengan berdasarkan algoritma *Adaboost* dan bagaimana suatu sistem dikembangkan berdasarkan algoritma tersebut.

e. BAB V PENUTUP

Bab ini berisi daripada kesimpulan dari hasil analisa masalah serta beberapa saran dari penulis, sehingga apa yang menjadi tujuan dari penelitian ini dapat terwujud.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB II
TINJAUAN PUSTAKA

2.1. Tinjauan Penelitian Terdahulu

Agar mendapat pemahaman dan wawasan untuk penelitian ini maka dilakukan tinjauan dari penelitian terdahulu yang mempunyai relevansi dengan penelitian yang nantinya dilaksanakan. Pada Tabel 2.1 dijelaskan tinjauan pustaka yang digunakan berserta korelasinya.

Tabel 2.1: Tinjauan Pustaka

No.	Judul	Hasil	Korelasi dengan Penelitian
1.	"*Knowledge Discovery* Data Akademik Untuk Prediksi Pengunduran Diri Calon Mahasiswa" (Melati dkk., 2018)	*data mining* sebagai solusi untuk memperkecil resiko kekurangan mahasiswa baru dan untuk mengetahui pola data calon mahasiswa yang berpotensi untuk melakukan pengunduran diri dalam program pendaftaran.	Menggunakan *data mining* untuk membantu proses penerimaan mahasiswa baru
2.	*"Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions"* (Saini & Kumar Jain, 2013)	Dengan menggunakan metode *data mining* pada proses penerimaan mahasiswa mampu meningkatkan keseluruhan performa proses penerimaan pada institusi perguruan tinggi.	

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

No.	Judul	Hasil	Korelasi dengan Penelitian
3.	*"Mining Student Information System Records to Predict Students' Academic Performance"* (Saa dkk., 2018)	Hasil ekstraksi 141 metode / algoritma data mining dari 34 *papers* dan menginformasikan bahwa metode data mining yang paling banyak digunakan pada penelitian *Educational Data Mining* (EDM) adalah *Decision Tree* dengan frekuensi hingga 24.8%.	Menggunakan metode *Decision Tree* pada penelitian berjenis *Educational Data Mining* (EDM)
4.	"Penerapan Decision Tree Untuk Penentuan Pola Data Penerimaan Mahasiswa Baru" (Aradea dkk., 2011)	*Decision Tree* dapat digunakan untuk mencari variabel-variabel yang berpengaruh pada proses PMB. variabel-variable yang tidak mempengaruhi dapat tereliminasi secara otomatis dalam proses pembentukkan pohon keputusan.	Menggunakan *Decision Tree* untuk membantu proses penerimaan mahasiswa baru

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

No.	Judul	Hasil	Korelasi dengan Penelitian
7.	*"Student Enrolment Prediction Model in Higher Education Institution: A Data Mining Approach"* (Ab Ghani dkk., 2019)	Memprediksi keputusan penerimaan calon mahasiswa menggunakan 3 algoritma sekaligus yaitu: *Logistic Regression, Decision Tree* dan *Naive Bayes* didapatkan hasil akurasi dan *recall* tertinggi dimiliki oleh algoritma *Decision Tree* dengan skor 71% dan 91% daripada 2 algoritma lainnya setelah diuji menggunakan metode *10-fold cross validation*.	Menggunakan *Decision Tree* dengan algoritma *CART* sebagai pilihan metode terbaik dari beberapa algoritma alternatif lainnya
8.	*"An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data"* (Lakshmi dkk., 2013)	Komparasi pada 3 algoritma pembentukan *Decision Tree* (*ID3, C4.5,* dan *CART*) didapatkan hasil algoritma *CART* sebagai algoritma terbaik dibandingkan dengan algoritma *Decision Tree* lainnya.	
9.	*"Study and Analysis of Decision Tree Based Classification Algorithms"* (Patel & Prajapati, 2018)	algoritma *CART* memiliki skor akurasi dan skor presisi tertinggi daripada algoritma *decision tree* lainnya seperti C4.5 dan ID3.	
10.	*"Detecting malicious URLs using binary classification through ada boost algorithm"* (Khan dkk., 2020)	Algoritma *AdaBoost* mampu menambah akurasi pada model.	

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

No.	Judul	Hasil	Korelasi dengan Penelitian

Berdasarkan referensi penelitian terdahulu, metode *Decision Tree* dengan pendekatan *Boosting* menggunakan algoritma *Adaboost* layak digunakan untuk mengembangan sistem prediksi daftar ulang calon mahasiswa baru pada penelitian ini. Tentunya dengan mengimplementasikan algoritma *CART* untuk pembentukan pohonya yang memiliki akurasi lebih baik daripada *C4.5* dan *ID3*.

2.2. Teori-Teori Dasar

2.2.1. *Data Mining*

Salah satu bidang ilmu komputer yang sedang berkembang pesat saat ini adalah *data mining*. Hal ini disebabkan karena data mining menghasilkan suatu *output* berupa model yang bisa memperkirakan kelompok dari suatu objek. Model bisa berbentuk aturan "jika-maka" yang dihasilkan dari proses latihan atau biasa disebut *Model Training* dengan data yang telah ada sebelumnya (Aradea dkk., 2011).

Data Mining juga digunakan untuk mencari suatu pola dari sekumpulan data yang berjumlah cukup banyak. Sehingga dengan data mining, data diam yang tersimpan di *database* bisa dimanfaatkan untuk diolah lebih lanjut hingga menghasilkan pengetahuan atau informasi yang berharga. Data diam ini bisa dijadikan data latih untuk membuat suatu model yang nantinya bisa menjadi bantuan dalam pengambilan keputusan (Amalia dkk., 2019).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tahap pembentukan model pada data mining terbagi menjadi 2 tahap, yaitu: tahap *training* dan tahap *testing*. Pada tahap *training*, data yang telah diberi label atau telah diketahui kelompoknya digunakan untuk melatih model dengan menggunakan suatu algoritma tertentu. Setelah model terlatih telah terbentuk, maka dilanjutkan dengan tahap uji coba atau *testing* yang berguna untuk mengetahui akurasi dari model tersebut. Model yang memiliki akurasi baik bisa digunakan untuk melakukan prediksi kelompok dari data yang belum diketahui kelompoknya (Aradea dkk., 2011).

Perbedaan antara proses pemrograman konvensional dengan pemrograman *machine learning* melalui *data mining* adalah jika pada *machine learning*, model dilatih dari data yang sudah ada sebelumnya (*Learning by example*) sehingga menghasilkan suatu aturan baru. Sedangkan pemrograman konvensional, aturannya diprogram secara manual tanpa melalui proses pelatihan / pembelajaran terlebih dahulu (Rajkomar dkk., 2019).

2.2.2. Metode Pelatihan Model

Terdapat 2 macam pendekatan yang dilakukan dalam proses pelatihan model pada *data mining*, yaitu:

1. Supervised Learning

Pelatihan dengan pendekatan *supervised learning*, prosesnya memerlukan sejumlah data yang memiliki input yang berupa *features* dan *output* yang biasa disebut label. Dengan menggunakan suatu algoritma tertentu, mesin akan belajar melalui observasi data yang telah disediakan. Sehingga nantinya mesin dapat memetakan dari *features* menjadi sebuah *output* (label) dari data baru atau data yang belum pernah dipelajari oleh mesin sebelumnya (Rajkomar dkk., 2019). Secara garis besar *supervised learning* adalah pelatihan model dengan data yang telah ditentukan kelas atau kelompoknya sebelumnya.

2. Unsupervised Learning

Sedangkan pada pelatihan model dengan pendekatan *unsupervised learning* mampu mencari pola dari kumpulan data yang diberikan secara otomatis. *Unsupervised learning* mengelompokkan data berdasarkan suatu pola tertentu yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

belum ditentukan sebelumnya (Amershi & Conati, 2007). Perbedaan antara *supervised learning* dan *unsupervised learning* adalah data yang digunakan untuk training tidak perlu diberi label, mesin yang akan mencari polanya sendiri.

2.2.3. Klasifikasi

Klasifikasi adalah salah satu teknik yang ada pada data mining yang mampu mengelompokkan data berdasarkan pelatihan yang telah dilakukan sebelumnya menggunakan data yang telah tersedia hingga terbentuk aturan baru pada model (Samponu & Kusrini, 2018). Proses training yang dilakukan tentunya menggunakan data yang berlabel karena jika tidak, maka disebut *clustering*.

Untuk melakukan klasifikasi tentu memerlukan algoritma yang menjadi acuan untuk menghitungnya. Klasifikasi memiliki banyak algoritma dan yang umum digunakan adalah algoritma (Wanjau & Muketha, 2018):

1. *Decision Tree*

   klasifikasi dengan membuat pohon keputusan dari *dataset*.

2. *Logistics Regression*

   Klasifikasi hasil dari pengembangan linear regression untuk memprediksi data dengan *output* berupa *variables* yang memiliki kategori.

3. *Nearest Neighbor*

   Melakukan klasifikasi data dengan melihat tetangga terdekatnya dan mengkategorikan objek tersebut pada kategori yang menjadi mayoritas pada tetangga terdekatnya.

4. *Naïve Bayes*

   Klasifikasi dengan metode probabilistik berdasarkan teorema *bayes*.

5. *Support Vector Machines*

   Klasifikasi yang menggunakan konsep kernel dan menggunakan *hyper-plane* untuk memperjelas perbedaan pada masing-masing kelas (Wulandari dkk., 2019).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.4. Jenis-Jenis Variabel

Pada sebuah data kuantitatif yang diolah terdapat berbagai macam jenis variabel. Sebelum mengolahnya melalui pendekatan data mining dengan menggunakan suatu metode tertentu, baiknya telah dikenali jenis-jenis variabel yang ada di dalamnya. Hal ini bertujuan agar memudahkan dalam proses pemilihan algoritma yang tepat. Adapun beberapa jenis-jenis variabel yaitu (Sugianto, 2016):

1. Variabel Diskrit

Merupakan variabel yang berupa data yang sifatnya memiliki kategori atau dapat dikelompokkan pada suatu kelas tertentu. Kategori atau kelas yang terbentuk memiliki nilai yang setara. Contohnya atribut yang memiliki jenis variabel diskrit adalah jenis kelamin. Data jenis kelamin memiliki 2 kategori yaitu pria dan wanita. Dua kategori tersebut memiliki nilai yang setara secara data.

2. Variabel Kontinyu

Variabel kontinyu adalah data yang berbentuk angka, bisa berbentuk bilangan pecahan maupun bilangan bulat. Sehingga variabel kontinyu dapat digunakan untuk proses operasi hitung. Contoh atribut yang berjenis variabel kontinyu adalah umur. Karena umur bisa digunakan untuk proses operasi hitung seperti dihitung rata-ratanya.

3. Variabel Ordinal

Variabel Ordinal merupakan data yang berbentuk angka yang memiliki peringkat. Dengan adanya peringkat, memungkinkan variabel ordinal tidak setara satu sama lain. Contoh variabel ordinal adalah pada atribut *ranking* kelas. Data rangking kelas 1 tidak sama dengan data rangking kelas 2, kedua angka tersebut memiliki peringkat yang tidak setara.

4. Variabel Interval

Data angka juga bisa dikatakan sebagai variabel interval apabila datanya memiliki rentang dengan jarak yang jelas. Contoh atribut yang memliki jenis variabel interval adalah rentang gaji ataupun rentang usia. Dimisalkan terdapat data rentang gaji dengan nilai 1 juta - 3 juta, 5 juta - 7 juta, dan lain sebagainya. Data-data tersebut memiliki rentang dan jarak yang jelas yaitu 2 juta.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

5. Variabel Rasio

Variabel rasio merupakan data angka yang telah diproses melalui operasi hitung yang kompleks. Angka pada variabel rasio bukanlah sebuah simbol ataupun kategori melainkan merupakan angka yang sebenarnya. Contoh variabel rasio adalah data tinggi badan. Dimisalkan data tinggi badan A bernilai 100 cm dan data tinggi badan B bernilai 200 cm, maka skala rasio tinggi badan A adalah setengah dari tinggi badan B.

2.2.5. Algoritma *Decision Tree*

Metode klasifikasi yang umum digunakan salah satunya adalah *Decision Tree* atau pohon keputusan. Tak hanya umum digunakan, metode *decision tree* juga dikenal sebagai metode yang efektif untuk melakukan klasifikasi dan prediksi pada suatu data (Darmawan, 2018). Proses klasifikasi dilakukan dengan membuat *node* (cabang) dan *leafs* (daun) dari *dataset* yang ada (Wanjau & Muketha, 2018). Aturan-aturan yang dihasilkan disajikan dalam bentuk pohon keputusan yang besar yang bersumber dari fakta yang ditemukan pada *dataset* (Darmawan, 2018).

Terdapat 2 jenis *Decision Tree* yaitu: *Classification Tree*, dan *Regression Tree* bahkan ada gabungan dari keduanya, biasa disebut *Classification and Regression Trees* (CART) (Patel & Prajapati, 2018). *Classification Tree* digunakan untuk memprediksi data diskrit (Dou dkk., 2019). Data diskrit adalah data yang sifatnya memiliki kategori atau dapat dikelompokkan pada suatu kelas tertentu (Sugianto, 2016). Sedangkan *Regression Tree* adalah pohon keputusan yang digunakan untuk memprediksi *continuous variable* atau data yang sifatnya kontinu tanpa kategori (Dou dkk., 2019). Data kontinyu adalah data yang berbentuk angka, bisa berbentuk bilangan bulat maupun bilangan pecahan. Sehingga data kontinyu biasanya didapat dari proses operasi hitung (Sugianto, 2016).

Metode *Decision Tree* memiliki beberapa algoritma yang digunakan untuk perhitungan dalam proses pembentukan pohon keputusan. Algoritma-algoritma ini membuat percabangan pada tiap atributnya kemudian menguji setiap node untuk mengetahui apakah pembagiannya sudah paling baik atau apakah kriteria pembagiannya sudah identik pada tiap kelasnya. Adapun beberapa algoritma yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

biasa digunakan, yaitu: ID3 (*Iterative Dichotomiser 3*), C4.5, CART (*Classification and Regression Trees*) (Patel & Prajapati, 2018).

2.2.6. Algoritma CART

CART adalah singkatan dari (*Classification And Regression Tree*) yang merupakan salah satu algoritma pembentuk pohon keputusan pada metode *Decision Tree*. Algoritma CART diperkenalkan oleh Brieman (Lakshmi dkk., 2013). Algoritma CART biasa digunakan untuk proses klasifikasi. Dibandingkan dengan algoritma decision tree lainnya seperti C4.5 maupun ID3, CART memiliki performa yang lebih unggul (Patel & Prajapati, 2018; Sugianto, 2016).

Pohon keputusan CART hanya menanyakan *yes/no questions*. Tiap-tiap cabang pertanyaan akan membagi data menjadi 2 bagian yang memiliki tingkat *homogeneity* maksimal. Proses pembagian data ini akan terus dilakukan secara berulang hingga hanya tersisa kelompok data yang memiliki 1 kelas saja atau tidak bisa dibagi lagi (Timofeev, 2004).

Pada tahap pemilihan atribut pemisah yang hendak digunakan untuk membentuk pohon keputusan, CART menggunakan *gini index* sebagai perhitungan untuk menentukan atribut mana yang layak dijadikan *root* dari percabangan (Lakshmi dkk., 2013). Variabel pemisah terbaik bisa didapatkan dari memilih atribut dengan nilai *gini index* terendah. Atribut tersebut akan dijadikan sebagai *node root*. *Gini index* dihitung dengan menggunakan Persamaan 1 (Zulfa dkk., 2019):

$$Gini(T) = 1 - \sum_{j=1}^{N} P_j^{\,2} \qquad (1)$$

Keterangan:

P_j : *probability* kelas j, j adalah kelas positif maupun negatif

N : Penomoran label kelas

Jika data D hendak dibagi menjadi 2 bagian yaitu D_1 dan D_2 berdasarkan atribut A, maka

$$Gini_{Split(A)}(D) = \frac{D_1}{D}\, Gini(D_1) + \frac{D_2}{D}\, Gini(D_2) \qquad (2)$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Keterangan:

$Gini(D_1)$dan $Gini(D_2)$ dihitung dengan persamaan (1)

Setelah menghitung *gini index* pada tiap-tiap kemungkinan cabang yang ada, maka dipilih percabangan yang memiliki nilai *gini index* terendah. Percabangan baru akan menghasilkan 2 populasi data baru yang terbentuk karena pemisahan. Populasi data cabang kiri dan populasi data cabang kanan. Data yang baru dijadikan acuan untuk proses *splitting* berikutnya. Langkah ini terus berulang hingga data tak bisa dibagi lagi atau hanya memiliki 1 kelas saja (Timofeev, 2004).

2.2.7. *AdaBoost*

AdaBoost (*Adaptive Boosting*) adalah salah satu algoritma *boosting* yang ditemukan oleh Freund and Schapire pada tahun 1997. *Boosting* sendiri merupakan konsep pembelajaran mesin dengan mengkombinasikan beberapa *classifier* yang lemah untuk membentuk *classifier* yang kuat. Sedangkan *Adaptive boosting* merupakan algoritma *boosting* yang mampu menyesuaikan secara adaptif nilai *error* yang dihasilkan oleh *classifier* lemah untuk dijadikan acuan pada proses pelatihan *classifier* berikutnya (Wu dkk., 2020).

Classifier lemah yang umum digunakan pada *AdaBoost* adalah *Decision Tree* dengan kedalaman 1 level atau biasa disebut *decision stump.* Semua *classifier* (model) lemah dilatih secara bergantian (tidak bersamaan) menggunakan *training set* yang sama. Setiap model yang terbentuk akan memperbaiki bobot *dataset* untuk pelatihan model berikutnya. Proses iterasi terus berlanjut hingga tercapai jumlah model yang diinginkan atau dataset pelatihan tidak dapat lagi diperbaiki (Khan dkk., 2020). Persamaan 3 hingga 11 merupakan penjelasan *pseudocode* dari algoritma *Adaboost* (Schapire, 2013):

Given: $(x_1, y_1), ., (x_m, y_m)$ where $y_i \in \{-1, +1\}$.

(3)

Inisiasi: $D_1 = 1/m$ for $i = 1, ., m$. (4)

For $t = 1, ..., T$: (5)

Latih model lemah (*weak learner*) dengan distribusi D_t

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Ambil hipotesis lemah $h_t : X \rightarrow \{-1, +1\}$ (6)

Pilih hipotesis h_t dengan bobot *error* paling rendah

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]. \quad (7)$$

Hitung bobot dari model:

$$\alpha_t = \frac{1}{2} ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (8)$$

Lakukan *update* pada distribusi *dataset*, for $i = 1, ., m$: (9)

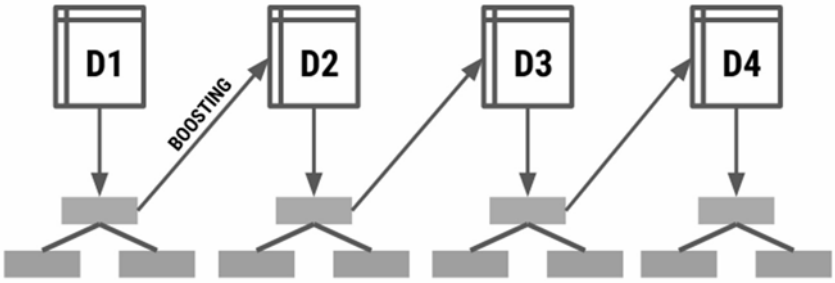
$$D_{t+1}(i) = \frac{D_t(i) \ exp(-\alpha_t y_i h_t(x_i))}{Z_t} \quad (10)$$

Z_t adalah faktor normalisasi untuk memastikan bobot data pada distribusi selanjutnya.

Output hasil prediksi akhir adalah:

$$H(x) = sign \left(\sum_{t=1}^{T} \alpha_i h_t(x) \right) \quad (11)$$

Untuk memudahkan pemahaman tentang alur algoritma *AdaBoost*, maka digambarkan *training flow* dari *AdaBoost* yang menggunakan *Decision Tree* sebagai *weak learner* seperti pada pada Gambar 2.1. Gambar tersebut menjelaskan bahwa pada tiap akhir proses *training* 1 *weak learner* maka dilakukan *boosting* atau distribusi ulang pada *dataset* berdasarkan perhitungan *error* atau *missclassified* dari *weak learner* yang baru saja dilatih. Sehingga distribusi data berikutnya (D2, D3, D4) akan berbeda dengan *dataset* awal (D1). perlu digarisbawahi bahwa yang berbeda hanyalah bobot atau distribusi datanya saja.



Gambar 2.1: *Training Flow AdaBoost*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

2.2.8. *Confusion Matrix*

Confusion Matrix adalah tabel matriks yang dapat digunakan untuk mengevaluasi model klasifikasi yang telah dilatih atau dibuat (Amornsamankul dkk., 2019). Terdapat 2 kelas pada tabel *confusion matrix* yaitu kelas positif dan negatif. Pada tabel *confusion matrix* juga terdapat 4 *cell* yang memiliki label masing-masing: *True Positive* (TP), *False Positive* (FP), *False Negative* (FN) dan *True Negative* (TN) seperti yang ada pada Tabel 2.2.

Tabel 2.2: Label-label pada *Confusion Matrix*

Confusion Matrix		**Kelas hasil prediksi**	
		Positif	**Negatif**
Kelas Sebenarnya	**Positif**	TP	FP
	Negatif	FN	TN

Label-label pada tabel matriks memiliki maksud untuk membedakan hasil prediksi dari model. Label *True Positive* (TP) digunakan untuk mempresentasikan jumlah data berkelas positif yang diklasifikasikan sebagai kelas positif juga. Sedangkan label *False Positive* (FP) adalah jumlah data berkelas positif yang diklasifikasikan sebagai kelas negatif. Adapun label *False Negative* (FN) yang menampilkan jumlah data berkelas negatif yang diklasifikasikan sebagai kelas positif. Yang terakhir adalah label *True Negative* (TN) adalah data berkelas negatif yang diklasifikasikan sebagai kelas negatif juga.

Dari informasi pada tabel matriks yang dihasilkan, maka model dapat dievaluasi dengan menghitung skor *precision* , dan *recall*. Masing-masing skor memiliki rumus yang berbeda. Berikut adalah rumus perhitungan dari akurasi, presisi dan *recall* secara berturut-turut:

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

F-measure merupakan hasil dari perhitungan *harmonic mean* dari *precision* dan *recall*. *Recall* biasa disebut *Sensitivity* yaitu proporsi dari kasus *real positive*

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

yang diprediksi secara benar (*True Positive*) (David M. W ., 2011) . Sedangkan *Precision* biasa disebut *positive predictive value (PPV)*. Penggunaan *harmonic mean* ini lebih proporsional daripada menggunakan *arithmetic mean* biasa. Berikut adalah rumus perhitungan *F-measure.* (Sasaki, 2007).

$$F = 2\frac{PR}{P+R} \tag{14}$$

Dimisalkan ada kasus dimana kita memiliki sistem deteksi wajah (*face recognition)* yang memiliki *precision* 1 sedangkan skor *recall* hanya 0.3. Jika dihitung performanya, maka seharusnya didapatkan skor yang rendah, karena sistem tersebut hanya mampu melakukan *cover* 30% dari wajah yang telah didaftarkan. Sayangnya, jika dihitung dengan arithmetic mean biasa akan didapatkan skor 0.65. Sedangkan jika dihitung menggunakan *harmonic means* maka didapatkan hasil 0.46. Contoh ini yang membuat perhitungan *F-measure* lebih proporsional (*reasonable)* (Sasaki, 2007). Dengan mengetahui skor presisi, *recall*, dan *F-measure* maka dapat dinilai apakah model berjalan dengan baik atau kurang baik.

2.2.9. *Chi Square*

Chi square adalah salah satu perhitungan statistik yang digunakan untuk mencari tahu ada atau tidaknya ketergantungan maupun hubungan antara dua variabel. *Chi Square* dapat digunakan untuk pemilihan fitur (*feature selection*) untuk proses pelatihan model *machine learning*. Adapun manfaat dari melakukan proses pemilihan fitur yaitu: mengurangi ukuran data, dan mempercepat proses pelatihan (karena adanya pengurangan data). Dengan melalui proses *feature selection* diharapkan mampu meningkatkan performa dari model yang dibuat. Rumus perhitungan *chi square* tertulis pada persamaan 16 (Nisa & Darwiyanto, 2019).

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{16}$$

Keterangan:

E_i = nilai ekspektasi ke-i

O_i = nilai observasi ke-i

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Adapun beberapa syarat yang harus dipenuhi agar dapat menghitung nilai *chi square* dari suatu data. Berikut syarat uji *Chi Square* (Nisa & Darwiyanto, 2019):

1. Tidak boleh ada data dengan frekuensi aktual (F_0) yang bernilai 0 (nol).
2. Jika tabel kontingensi data memiliki bentuk 2x2, maka harus tidak ada data yang memiliki nilai frekuensi harapan (F_h) kurang dari 5.
3. Jika bentuk tabel kontingensi lebih dari 2x2 (contoh: 2x3), maka proporsi jumlah sel dengan nilai frekuensi harapan (F_h) kurang dari 5 tidak lebih dari 20%.

2.2.10. REST API

REST API merupakan suatu kumpulan *resource* dan operasi yang dapat dipanggil melalui *resource* yang lain. Pemanggilan REST API dapat memicu jalannya suatu operasi di dalamnya untuk memproses suatu data ataupun hanya sekedar menyampaikan data sesuai permintaan dari *client*. *Client* melakukan pemanggilan REST API melalui protokol HTTP, oleh sebab itu REST API dapat dipanggil melalui *request browser* (Surwase, 2016).

Karena diakses melalui protokol HTTP, tentu REST API memiliki struktur url tertentu yang disediakan untuk para *consumer* dari *resource*. Struktur url dipresentasikan pada Gambar 2.2. REST API memiliki *base path* yaitu alamat *path url* utama yang digunakan untuk melakukan isolasi antara satu versi REST API dengan yang lainnya. Contoh *base path* adalah '/api/v1' dan '/api/v2' dimana kedua *base path* ini dibuat untuk membedakan versi *resource* yang diakses dan tentunya berbeda pula aturan penggunaan *resource* pada masing-masing versinya (Surwase, 2016).

Gambar 2.2.: Contoh Struktur URL dari REST API

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Selain *base path*, REST API juga memiliki *relative path* yang digunakan untuk mempresentasikan *resource* atau operasi yang bisa diakses (Surwase, 2016). *Relative path* bersifat relatif terhadap *base path*. Maksudnya adalah dimisalkan terdapat *resource* yang memiliki *relative path* '/buku' dan memiliki *base path* '/api/v1', itu artinya *resource* tersebut bisa diakses melalui url '/api/v1/buku'. Contoh desain daftar *resource* dari *REST API* seperti pada Tabel 2.3.

Tabel 2.3: Desain resource *REST API*

Resource	**Deskripsi**
/buku	Mengakses daftar semua daftar buku yang ada di *database*
/buku/{id}	Mengakses *detail* satuan buku sesuai parameter id yang diminta

Tidak hanya mampu diakses melalui protokol HTTP dengan metode *GET* saja, REST API juga dapat diakses melalui metode yang lain seperti *POST, PUT, PATCH* maupun *DELETE*. Metode-metode ini memudahkan *REST API* untuk membedakan operasi apa yang hendak dilakukan walaupun *client* mengakses dengan url yang sama. Tiap metode *request* yang dilakukan *client* mempresentasikan jenis operasi yang akan dilakukan oleh REST API. Misalkan mengakses '/buku/{id}' dengan metode *DELETE*, maka *client* mengharapkan REST API melakukan operasi penghapusan buku sesuai id yang dipilih.

2.2.11. JSON

JSON merupakan format data yang sering digunakan untuk pertukuran *resource* maupun data. Format data JSON mudah dibaca oleh manusia (*Human Readable*) dan mudah untuk diolah oleh komputer. JSON memiliki performa yang jauh lebih baik daripada XML yaitu salah satu alternatif format data selain JSON. JSON lebih cepat hingga 100 kali lipat daripada XML (Nurseitov dkk., 2019).

JSON dan REST API memiliki keterkaitan. REST API dapat memberikan *response* dalam berbagai format data, tentu JSON bisa dijadikan pilihan sebagai

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

format data yang nantinya disampaikan kepada *consumer*. Pada Gambar 2.3 adalah contoh JSON *syntax* yang mempresentasikan data *user*.

```
{
  "nama": "Naufal Rabbani",
  "jenis_kelamin": "laki-laki"
}
```

Gambar 2.3: Contoh data *user* dalam format JSON

2.2.12. *Python*

Salah satu bahasa pemrograman yang paling populer untuk mengimplementasikan layanan *backend* pada suatu aplikasi *web* adalah *python*. Hal ini bisa dilihat terdapat 500 ribu lebih proyek yang menggunakan bahasa *python* pada Github dan *index* pada Tiobe menunjukkan bahwa *python* menduduki *ranking* 1 sebagai 4 bahasa terpopuler pada tahun 2016 lalu (Vogel dkk., 2017).

2.2.13. *Flask*

Flask merupakan salah satu framework andalan komunitas *python*. *Framework* ini unggul karena kesederhanaannya dan fleksibilitasnya dengan menyediakan fitur paling minimum untuk membuat sebuah *web server* sehingga membuatnya ringan dan tidak berlebihan pada *codebase.* Hal ini yang membuat flask dikategorikan sebagai salah satu *micro-framework*. Untuk membuat *"Hello World*" cukup dengan 5 baris kode saja (Vogel dkk., 2017).

2.3. Integrasi Keilmuan

Model klasifikasi yang dibuat adalah model yang dilatih berdasarkan histori data yang sudah ada sebelumnya. Dengan *data mining*, mesin atau sistem mampu mempelajari data calon mahasiswa pada tahun-tahun sebelumnya untuk dijadikan acuan dalam melakukan prediksi kedepannya. Konsep ini bersinggungan dengan ayat pada Al-Quran yaitu surah Al-Hasyr ayat 18 tentang pentingnya introspeksi:

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوا۟ ٱتَّقُوا۟ ٱللَّهَ وَلْتَنظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ ۖ وَٱتَّقُوا۟ ٱللَّهَ ۚ إِنَّ ٱللَّهَ خَبِيرٌۢ بِمَا تَعْمَلُونَ

Artinya:

"Wahai orang-orang yang beriman! Bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat), dan bertakwalah kepada Allah. Sungguh, Allah Maha Teliti terhadap apa yang kamu kerjakan. (QS Al-Hasyr : 18)"

Berdasarkan wawancara dengan bapak Husnul Muttaqin, S. Ag. S.Sos, M.S.I salah satu dosen sosiologi fakultas Ilmu Sosial dan Politik UIN Sunan Ampel Surabaya, narasumber setuju dengan pendapat bahwa pada ayat tersebut Allah menyuruh orang-orang yang beriman untuk memperhatikan apa yang telah diperbuatnya agar mendapatkan kebaikkan di hari esok (akhirat). Kata "telah" mengindikasikan sesuatu yang ada di masa lampau atau sudah terjadi. Oleh sebab itu, dengan melakukan pembelajaran terhadap pola data yang telah ada sebelumnya diharapkan mampu memperbaiki keputusan pemilihan mahasiswa baru yang diloloskan pada proses penerimaan mahasiswa baru agar tidak terlalu banyak kursi kosong.

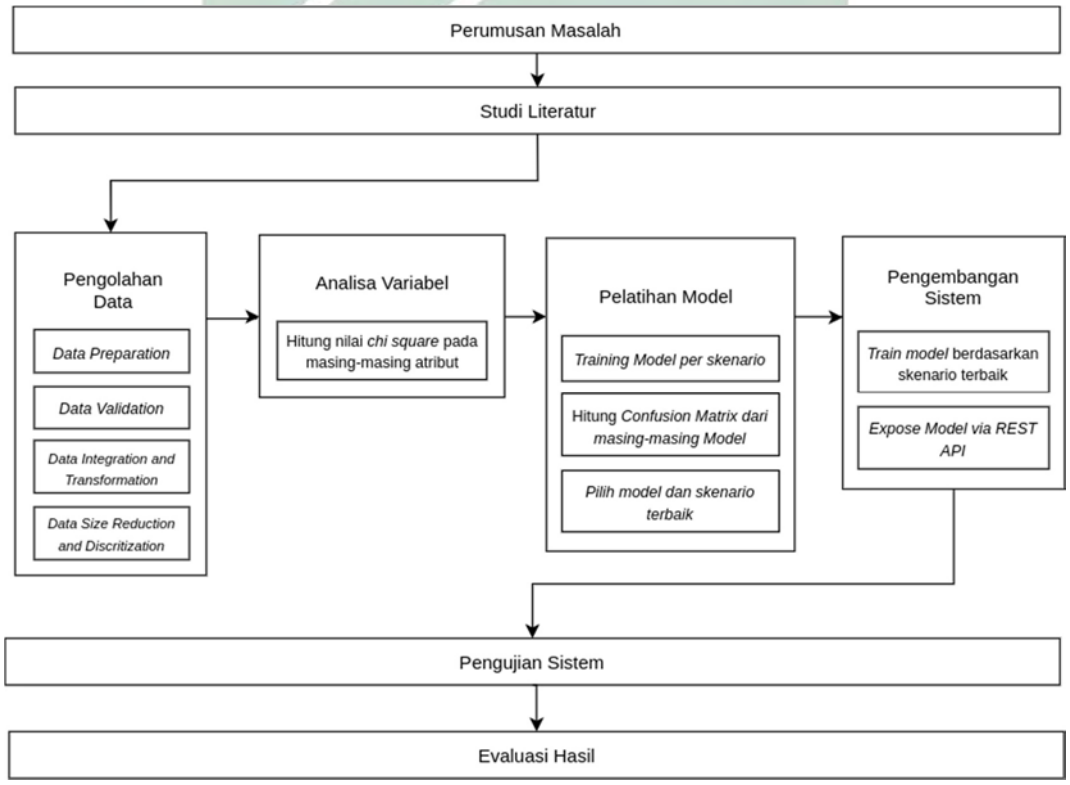
Penelitian ini bertujuan untuk mengurangi resiko kursi kosong pada proses penerimaan mahasiswa baru. Dengan mempelajari *pattern* data calon mahasiswa pada tahun sebelumnya bisa ditemukan ciri mahasiswa yang mengundurkan diri. Sehingga pihak perguruan tinggi mampu mempertimbangkan ulang calon mahasiswa baru yang diprediksi oleh sistem akan mengundurkan diri. Hasil prediksi ini dijadikan bahan tambahan untuk pengambilan keputusan untuk meminimalisir jumlah kursi kosong. Dengan begitu, tidak ada sumber daya yang terbuang secara sia-sia.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB III
METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Proses penelitian dipresentasikan ke dalam bentuk diagram alur. Metode yang digunakan adalah metode kuantitatif yang disesuaikan. Diagram alur penelitian dibuat agar mempermudah penyampaian informasi langkah-langkah yang hendak dilakukan. Diagram alur dari penelitian yang disajikan pada Gambar 3.1.



Gambar 3.1: Diagram Alur Penelitian

Pembahasan dari masing-masing proses pada alur penelitian yang telah digambarkan akan dijelaskan pada sub bab 3.

3.1.1. Perumusan Masalah

Pada tahap ini, masalah yang dirumuskan telah menjadi latar belakang penelitian. Masalah yang diangkat seperti halnya tertera dalam latar belakang yakni mengenai pembuatan sistem prediksi daftar ulang calon mahasiswa baru. Sistem

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

prediksi daftar ulang calon mahasiswa baru merupakan salah satu upaya dari pihak perguruan tinggi untuk melakukan seleksi yang lebih baik lagi dalam proses penerimaan mahasiswa baru. Dengan melakukan metode klasifikasi *data mining* dari data calon mahasiswa yang sudah ada, memungkinkan sistem untuk memprediksi hasil apakah calon mahasiswa tersebut memiliki kemungkinan besar untuk melanjutkan proses pendaftaran atau mengundurkan diri. Nantinya hasil tersebut akan dijadikan referensi tambahan dalam proses pengambilan keputusan untuk seleksi calon mahasiswa baru yang seharusnya diterima.

3.1.2. Studi Literatur

Untuk melakukan suatu klasifikasi harus ditentukan terlebih dahulu fitur atau atribut pada data yang hendak dijadikan acuan. Berdasarkan studi literatur yang telah dilakukan, didapatkan pemetaan atribut dari penelitian sebelumnya yang ditampilkan pada Tabel 3.1.

Tabel 3.1: Atribut-atribut yang ada pada penelitian terdahulu.

Penelitian Terdahulu					
(Wanjau & Muketha, 2018)	***(Melati dkk., 2018)***	***(Yahya & Jananto, 2019)***	***(Aradea dkk., 2011)***	***(Rozi, 2015)***	***(Ab Ghani dkk., 2019)***
Career Flexiblity	Jenis Sekolah SMA	Kota Asal	Jurusan Pilihan	Agama	Gender
High School Final Grade	Gelombang	Jenis Kelamin	Jurusan SMA	Penghasilan Orang Tua	Age
Math Grade	Pekerjaan Orang tua	Asal Pendidikan	Daerah Asal	Usia	Religion
Pre - University awareness	Jurusan SMA	Prodi	Rata-Rata UN	Jumlah Saudara Kandung	Location
Teacher Inspiration		Agama	Diterima di Pilihan ke	Waktu Tempuh	Family income
Financial Aid		Kelas Kuliah	skor ujian masuk		Tuition Cost
Extracurricular					Scholarship
Societal Expectation					Ranking
Parent Career					Fasilitas
Self Efficacy					Course Structure
Career Earning					Job Opportunity

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Penelitian Terdahulu					
Gender					Pendidikan Orang Tua
Age					Jumlah anak pada keluarga
Family Income					Departement
					Recidency

Dari pemetaan yang ada pada Tabel 3.1, maka dipilih beberapa fitur yang hendak digunakan untuk proses *training model* berdasarkan ketersediaan data yang ada saat ini. Fitur terpilih yang dituliskan pada Tabel 3.2.

Tabel 3.2: Atribut yang akan digunakan berdasarkan penelitian terdahulu

No.	Atribut	Penelitian Terdahulu					
		(Wanjau & Muketha, 2018)	*(Melati dkk., 2018)*	*(Yahya & Jananto, 2019)*	*(Aradea dkk., 2011)*	*(Rozi, 2015)*	*(Ab Ghani dkk., 2019)*
1.	Jenis Kelamin	✔		✔			
2.	Usia	✔				✔	
3.	Kota Asal			✔	✔		✔
4.	Jenis Sekolah Asal		✔	✔			
5.	Pekerjaan Orang Tua	✔	✔				
6.	Penghasilan Orang Tua	✔				✔	
7.	Prodi			✔	✔		✔

3.1.3. Pengolahan Data

Data yang digunakan untuk pengembangan sistem didapat dari *database* sistem informasi akademik Universitas Islam Negeri Sunan Ampel Surabaya. Data yang diambil adalah data mahasiswa pendaftar yang melanjutkan pendaftaran maupun yang mengundurkan diri selama 1 tahun terakhir yaitu data pendaftaran periode 2019 dan 2018.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Adapun data yang telah didapatkan dari *database* sistem informasi akademik Universitas Islam Negeri Sunan Ampel Surabaya akan diolah hingga berbentuk seperti yang dideskripsikan pada Tabel 3.3.

Tabel 3.3: Rancangan struktur data untuk hasil akhir pengolahan

No.	Nama Data	Jenis Variabel	Deskripsi
1.	Usia	Kontinu	Digunakan untuk klasifikasi dalam perihal segmentasi usia. Contoh: “20 tahun”, “19 tahun”
2.	Jenis Kelamin	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi jenis kelamin. Contoh: “Laki-laki”, “Perempuan”
3.	Kabupaten Asal	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi domisili pendaftar. Contoh: “Lamongan”, “Surabaya”
4.	Jenis Sekolah Asal	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi asal sekolah pendaftar. Contoh: “SMA”, “MA”, “SMK”
5.	Pekerjaan Orang Tua	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi pekerjaan orang tua pendaftar. Contoh: “swasta”, “PNS”, “lain-lain”
6.	Penghasilan Orang Tua	Interval	Digunakan untuk klasifikasi dalam perihal segmentasi penghasilan orang tua pendaftar. Contoh: “< 1 juta”, “1 juta s.d. 2 juta”

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

No.	Nama Data	Jenis Variabel	Deskripsi
7.	Prodi Terpilih	Nominal	Digunakan untuk klasifikasi dalam perihal segmentasi prodi terpilih untuk pendaftar. Contoh: “Sistem Informasi”, “Ilmu Komunikasi”, “Sastra Inggris”

Agar data yang diperoleh *valid* dan berkualitas, maka perlu dilakukan beberapa tahap pengolahan, yaitu: *Data Validation, Data Integration and Transformation, Data Size Reduction and discritization* (Hoiriyah, 2018). Berikut beberapa tahapan yang akan dilakukan:

1. Data Preparation

Tahap paling pertama adalah melakukan ekstraksi data dari *database*. Hasil datanya akan dijadikan sebagai input. Data yang didapat adalah data mentah dengan bentuk awal seperti yang dituliskan pada Tabel 3.4. Data tersebut yang nantinya akan diolah lebih lanjut hingga menghasilkan data dengan bentuk seperti yang dipresentasikan pada Tabel 3.3.

Tabel 3.4: Struktur Tabel Hasil *Query Database*

No.	Nama	Tipe Data	Contoh Data
1.	sex	string	“P” atau “L”
2.	tgllahir	date	“1997-04-25”, “1998-01-23”
3.	namapendapatan	string	“< 1.000.000”, “1.000.000 s.d. 2.000.000”
4.	pekerjaan_ayah	string	“Swasta”, “Tani”, “Wiraswasta”, “PNS”, “Lain-lain”

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

No.	Nama	Tipe Data	Contoh Data
5.	namakota	string	“KOTA SURABAYA”, “KAB. GRESIK”, “KOTA LAMONGAN”
6.	namaprodi	string	“SISTEM INFORMASI”, “MATEMATIKA”, “SASTRA INGGRIS”
7.	jenissekolah	string	“SMAN”, “MAS”, “SMAS”, “MAN”
9.	lanjut_daftar	string	0 atau 1

2. Data Validation

Tidak semua data yang didapat akan digunakan dalam proses *training* model. Pada tahap ini data akan diseleksi, akan dihilangkan *outliers* dan *noise*-nya. Sehingga tidak ada lagi data yang tidak konsisten maupun data yang tidak lengkap atau *missing value* (Hoiriyah, 2018). Proses Data validation dilakukan menggunakan *script python.*

3. Data Integration and Transformation

Setelah dihilangkan *outliers* dan *noise* pada datanya. Data akan ditransformasi melalui *script python* hingga memiliki struktur seperti pada Tabel 3.3. Hasilnya akan diekspor dalam bentuk CSV agar tidak perlu lagi di ekstrak ulang dari *database* sehingga datanya bisa digunakan berulang kali tanpa harus mengulangi proses dari awal.

4. Data Size Reduction and Discritization

Data yang valid dan bersih selanjutnya akan dioptimasi lebih lanjut untuk mempersingkat proses eksekusi pelatihan pada model. Untuk itu perlu dilakukan pemilihan fitur yang penting yang hendak digunakan untuk proses *training*. Sehingga tidak semua fitur digunakan. Berikut pemilihan fitur dan label yang akan digunakan seperti pada Tabel 3.5.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 3.5: Pemilihan Fitur dan Label

No.	Nama Data	Deskripsi
1.	Usia	Fitur *X1*
2.	Jenis Kelamin	Fitur *X2*
3.	Alamat Asal	Fitur *X3*
4.	Asal Sekolah	Fitur *X4*
5.	Pekerjaan Orang Tua	Fitur *X5*
6.	Penghasilan Orang Tua	Fitur *X6*
7.	Prodi Terpilih	Fitur *X7*
9.	Melanjutkan Pendaftaran	Label *Y*

3.1.4. Analisa Variabel

Tahap analisa variabel dilakukan guna mencari tahu apakah tiap-tiap atribut memiliki hubungan terhadap pada status daftar ulang calon mahasiswa baru. Analisa dapat dilakukan dengan menghitung korelasi. Hanya saja, karena jenis variabel pada dataset pada penelitian ini berjenis kategorial yang sifatnya setara, maka digunakan perhitungan *chi square.* Hasil dari analisa variabel adalah mengetahui ada atau tidaknya hubungan antara variabel terpilih dengan status daftar ulang calon mahasiswa.

Dengan mengetahui apa saja variabel yang memiliki hubungan dengan status daftar ulang mahasiswa, maka dapat dijadikan informasi tambahan untuk membuat program atau inovasi-inovasi baru yang mampu mengurangi kemungkinan calon mahasiswa mengundurkan diri.

3.1.5. Pelatihan Model

Proses *training model* dilakukan menggunakan *python script* dengan bantuan *library scikit-learn.* Pada tahap ini data akan digunakan untuk membentuk sebuah model dengan metode *Adaboost* yang telah dijelaskan *pseucode*-nya pada

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

rumus (3-11). Sedangkan untuk *weak learner*-nya akan menggunakan metode *decision tree* CART dengan perhitungan sesuai dengan rumus (1) dan (2).

Proses pelatihan dilakukan dengan beberapa skenario pembagian data. Dari beberapa skenario yang ada, akan dipilih model dan skenario terbaik berdasarkan perhitungan *F Measure* dan *confusion matrix* dengan perhitungan menggunakan rumus (12-14). Perhitungan ini akan dilakukan pada masing-masing modelnya. Adapun beberapa skenario pembagian data *training* dan data *testing* dituliskan pada Tabel 3.6.

Tabel 3.6: Skenario pembagian data

Skenario nomor ke-i	Data Training	Data Testing
1	70%	30%
2	50%	50%
3	60%	40%

Sedangkan pengujian atau *testing* yang digunakan untuk memvalidasi model yang telah dibuat pada tahap *training* dilakukan dengan cara memprediksi data yang sudah mempunyai label sebelumnya (bukan data baru). Jumlah data yang harus dites sesuai dengan skenario yang tertulis pada Tabel 3.6. Metode yang digunakan untuk melakukan *testing model* adalah dengan menghitung *confusion matrix* pada model yang telah dibuat untuk didapatkan skor *precision*, *recall,* dan *f-measure*.

3.1.6. Pengembangan Sistem

Model yang terlatih akan disimpan dalam bentuk file sehingga proses klasifikasi berikutnya tidak perlu membuat pohon keputusan atau model baru. Dengan cara ini proses klasifikasi bisa lebih cepat. Proses konversi model yang awalnya berada pada memori sementara, diubah menjadi *file*, dilakukan dengan menggunakan *cPickle* pada *python*.

Sistem yang diekspos berbentuk REST API *Web Service*. Hal ini bertujuan agar *classifier model* yang telah dibuat bisa diintegrasikan dengan sistem atau aplikasi lain sehingga bisa digunakan untuk menambah fitur pada aplikasi yang

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

sudah ada sebelumnya misal: SIAKAD UINSA Surabaya. Bentuk data yang dihasilkan dari REST API berupa JSON.

REST API *Web Service* dibuat menggunakan *framework Flask Python* dan terdiri dari 2 *endpoint* sesuai Tabel 3.7

Tabel 3.7: Daftar *endpoint* pada REST API *Web Service*

No.	***Endpoint***	***Method***	**Parameter**	**Keterangan**
1.	*/predict*	POST	list nomor pendaftaran (opsional)	Untuk mengetahui hasil prediksi pada data mahasiswa
2.	*/train*	POST	-	Untuk membuat model baru dan melakukan proses *training* mulai awal

3.1.7. Pengujian Sistem

Selain menguji model yang telah dibuat, juga dilakukan pengujian sistem yang berupa REST API. Hasil dari REST API akan diuji untuk dipastikan bisa mempresentasikan data respon sesuai dengan perhitungan atau prediksi dari model yang dibuat.

3.1.8. Evaluasi Hasil

Tahap paling akhir adalah evaluasi hasil. Pada tahap ini diambil kesimpulan berdasarkan hasil pengujian sistem yang dilakukan. Mulai dari proses analisa variabel yang akan menjabarkan variabel apa saja yang memiliki hubungan terhadap status daftar ulang mahasiswa, mencari tahu ciri mahasiswa yang cenderung melanjutkan daftar ulang, mengevaluasi pengaruh *feature selection* pada model, mencari tahu peforma *Adaboost,* dan apakah sistem mampu melakukan prediksi dengan baik. Dan memberikan pemaparan skenario terbaik, dan model terbaik yang telah dibuat.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil

Pada bagian ini akan dipaparkan hasil penelitian berdasarkan tahapan yang telah dijelaskan di metode penelitian pada bab sebelumnya. Tahapan tersebut mulai dari proses pengolahan data, analisa variabel hingga proses pelatihan model dengan beberapa skenario yang telah ditentukan. Dan juga akan menjelaskan rancangan dan alur sistem prediksi yang telah dibuat beserta uji sistem yang dilakukan.

4.1.1. Pengolahan data

Sebelum melakukan traning model dengan beberapa skenario yang telah disiapkan, diperlukan data yang bersih dari *noise* dan valid. *Data training* tersebut haruslah sesuai dengan rancangan pada Tabel 3.3. Untuk mendapatkan data yang valid dan sesuai dengan rancangan, maka dilakukan 4 tahapan *pre processing* secara umum yaitu: *Data Preparation, Data Validation, Data Integration and Transformation*, dan *Data Size Reduction*. Tetapi pada setiap tahapnya, atribut-atribut terpilih perlu perlakuan yang berbeda agar mendapatkan data yang valid dan bersih. Untuk itu dijelaskan beberapa detil tahapan per atributnya seperti pada Tabel 4.1.

Tabel 4.1: Detil tahapan *Pre processing* pada data per atributnya.

Atribut	Perlakuan	Ekspektasi Hasil
tgl_lhr (usia)	– Menghapus data yang tidak valid – Mengubah tanggal lahir menjadi usia berdasarkan tahun masuk	17, 18, 19, 20, 21, ...
jenis_kelamin	– Menghapus data yang tidak valid	P dan L
kode_kabupaten (kabupaten_asal)	– Menghapus data yang tidak valid	3515, 3525, 3524, 3578, 3522, ...

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Atribut	Perlakuan	Ekspektasi Hasil

Tahapan yang telah dijelaskan pada Tabel 4.1 tersusun setelah melakukan pengamatan keberagaman nilai dari setiap atribut dalam *dataset*. Hasil pengamatan *dataset* menunjukkan bahwa terdapat nilai atribut yang bersifat valid dan tidak valid yang dilustrasikan pada Tabel 4.2

Tabel 4.2: Data Validation

Atribut	Contoh nilai yang valid	Contoh nilai yang tidak valid
tgl_lhr	11-02-2001, 14-03-2000, 14-03-2000, ...	None, 0, 00-00-0000
jenis_kelamin	P, L	None, 1
kode_kabupaten	3515, 3525, 3524, 3578, 3522, ...	0, nan

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Atribut	Contoh nilai yang valid	Contoh nilai yang tidak valid

Jika diambil beberapa sampel dari irisan *dataset* sebelum dilakukan *pre processing*, maka didapatkan data seperti yang dipresentasikan pada Tabel 4.3.

Tabel 4.3: Sampel *dataset* sebelum dilakukan *pre processing*.

tgl_lhr	jenis kelamin	kode_kabupaten	asal_sekolah	pekerjaan_ayah	penghasilan_total	prodi	jalur masuk	status
14-03-2000	L	3525.0	MAS MAARIF NU SIDOMUKTI	Pegawai Swasta	2250000	AKUN	1	U
14-04-2000	P	3515.0	SMAS KEMALA BHAYANGKARI 3	Pegawai Swasta	5250000	AKUN	1	A
08-06-2000	L	3524.0	MAS MA. ISLAMIYAH ATTANWIR	Pegawai Swasta	1500000	AKUN	1	T
30-03-2000	P	3515.0	SMAS HANG TUAH 2 GEDANGAN	Pegawai Swasta	5250000	AKUN	1	A

Proses pengolahan data dilakukan menggunakan bahasa *python*. Tahapan prosesnya sesuai dengan yang telah dijelaskan pada Tabel 4.1. Sehingga setelah selesai melakukan proses *pre processing data*, didapatkan irisan *dataset* seperti pada tabel 4.4.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.4: Sampel *dataset* setelah dilakukan *pre processing*.

usia	jenis kelamin	kabupaten asal	jenis asal sekolah	pekerjaan orang tua	penghasilan orang tua	prodi terpilih	jalur masuk	melanjutkan pendaftaran
17	P	3515.0	smas	swasta	5.000.001 - 7.500.000	Ekonomi dan Bisnis Islam	SNMPTN	daftar ulang
18	L	3525.0	mas	swasta	2.000.001 - 2.500.000	Ekonomi dan Bisnis Islam	SNMPTN	tidak daftar ulang
18	P	3515.0	smas	swasta	5.000.001 - 7.500.000	Ekonomi dan Bisnis Islam	SNMPTN	daftar ulang
18	L	3524.0	mas	swasta	1.000.001 - 1.500.000	Ekonomi dan Bisnis Islam	SNMPTN	tidak daftar ulang
18	P	3515.0	smas	swasta	5.000.001 - 7.500.000	Ekonomi dan Bisnis Islam	SNMPTN	daftar ulang

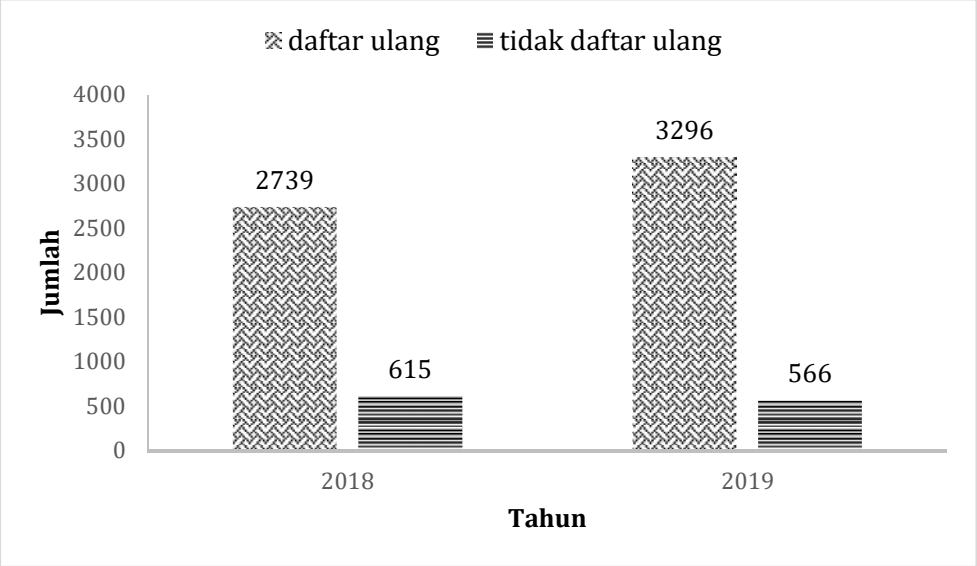
Data mentah awal berjumlah **11735** data. Setelah diolah berkurang hingga menjadi **7216** data. Pada Tabel 4.5 menjelaskan komposisi data yang telah diolah. Dari tabel tersebut bisa dilihat perbandingan jumlah data yang tidak seimbang (*imbalance*) antara data mahasiswa yang melanjutkan daftar ulang dan yang tidak melanjutkan daftar ulang. Perbandingan jumlah data yang tak seimbang ini dapat dilihat pada Gambar 4.1

Tabel 4.5: Perbandingan Komposisi data sebelum dan sesudah diolah (berdasarkan tahun)

Status	**Sebelum**			**Sesudah**		
	2018	***2019***	***Total***	***2018***	***2019***	***Total***
Daftar Ulang	3921	4444	8365	2739	3296	6035
Tidak Daftar Ulang	2102	1266	3370	615	566	1181
Total Data			11735			7216

Gambar 4.1 merupakan grafik perbandingan jumlah kelas positif dan negatif yang ada pada *dataset*. Grafik ini dipresentasikan guna memperlihatkan ketidakseimbangan kelas yang signifikan antara kelas Daftar Ulang dan kelas Tidak Daftar Ulang.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.1: Perbandingan Jumlah *Class* pada dataset sesudah diolah

4.1.2. Analisa Variabel

Untuk mengetahui apakah suatu fitur berpengaruh pada status daftar ulang calon mahasiswa baru dapat dilakukan dengan menghitung korelasi. Hanya saja, karena jenis variabel pada *dataset* pada penelitian ini berjenis kategorial yang sifatnya setara, maka digunakan perhitungan *chi square*. Hal ini bertujuan untuk mengetahui ada atau tidaknya hubungan suatu atribut tertentu terhadap status daftar ulang mahasiswa. Tabel 4.6 menjelaskan rangkuman dari hasil perhitungan *chi square*.

Tabel 4.6: Hasil perhitungan *chi square* pada masing-masing atribut

Atribut	lambda	chi2	dof	p	cramer	power	memiliki hubungan
usia	1	12.92	4	0.012	0.042	0.834	Ya
jenis_kelamin	1	3.638	1	0.056	0.022	0.479	Tidak
kabupaten_asal	1	33.659	20	0.029	0.073	0.969	Ya
jenis_asal_sekolah	1	70.622	10	0	0.1	1	Ya
pekerjaan_orang_tua	1	18.492	14	0.185	0.051	0.804	Tidak
penghasilan_orang_tua	1	35.875	11	0	0.071	0.994	Ya
prodi_terpilih	1	37.895	12	0	0.073	0.995	Ya
jalur_masuk	1	158.72	4	0	0.148	1	Ya

Kolom Hasil pada Tabel 4.6. diambil berdasarkan perbandingan nilai p dan nilai a (alpha = 0.05). Jika nilai p kurang dari a ($p < a$) maka atribut tersebut

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

memiliki hubungan dengan status daftar ulang mahasiswa. Setelah melakukan perhitungan, dapat disimpulkan jika ada 2 variabel yang tidak memiliki hubungan terhadap status daftar ulang calon mahasiswa baru yaitu atribut *jenis_kelamin* dan *pekerjaan_orang_tua*.

Adapun analisa data potensial yang dapat dipaparkan agar bermanfaat untuk mengetahui ciri mahasiswa yang berpotensi untuk melanjutkan proses penerimaan mahasiswa baru hingga daftar ulang dan menjadi mahasiswa aktif. Analisa ini dilakukan dengan cara sederhana dengan menampilkan hasil observasi data dari frekuensi munculnya data pada *class* daftar ulang dan *class* tidak daftar ulang. Data dikelompokkan berdasarkan *value* unik (*distinct*) dari masing-masing atribut. Presentasi tabel observasi tertera pada Tabel 4.7 hingga Tabel 4.11.

Tabel 4.7: Tabel observasi pada atribut usia

usia	daftar ulang	tidak daftar ulang	probabilitas daftar ulang
26	1	0	100
25	1	0	100
24	1	0	100
22	11	1	91.67
39	9	1	90
21	69	8	89.61
20	375	63	85.62
18	3336	638	83.95
19	2023	406	83.29
17	205	62	76.78
23	4	2	66.67

Tabel 4.7 menunjukkan jika pada calon mahasiswa baru S1 dengan rentang usia 19 - 21 makin tua usianya makin tinggi probabilitas untuk melanjutkan daftar ulang.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.8: Tabel observasi pada atribut *jenis_sekolah_asal*

jenis_asal_sekolah	daftar ulang	tidak daftar ulang	probabilitas daftar ulang
pondok	52	3	94.55
smks	174	19	90.16
pkbm	15	2	88.24
lain-lain	158	22	87.78
smas	824	117	87.57
smk	117	18	86.67
smta	13	2	86.67
ma	216	35	86.06
smkn	165	29	85.05
man	1297	231	84.88
madrasah	77	14	84.62
mas	939	172	84.52
sma	631	124	83.58
sman	1357	393	77.54

Sedangkan Tabel 4.8 menunjukkan jika calon mahasiswa baru yang berasal dari pondok memiliki probabilitas tertinggi untuk melanjutkan daftar ulang, sedangkan yang berasal dari SMAN (sekolah Menengah Atas Negeri) memiliki probabilitas terendah untuk melanjutkan daftar ulang.

Tabel 4.9: Tabel Observasi pada atribut *jalur_masuk*

jalur_masuk	daftar ulang	tidak daftar ulang	probabilitas daftar ulang
SBMPTN	1186	149	88.84
SNMPTN	455	70	86.67
UMPTKIN	2195	386	85.04
MANDIRI	1460	272	84.3
SPANPTKIN	739	304	70.85

Berikutnya Tabel 4.9 menunjukkan bahwa mahasiswa yang melakukan pendaftaran melalui jalur SBMPTN lebih cenderung untuk melakukan daftar ulang hingga akhir, sedangkan yang mendaftar melalui jalur SPANPTKIN memiliki probabilitas rendah untuk melanjutkan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.10: Tabel Observasi pada atribut *penghasilan_orang_tua*

penghasilan_orang_tua	daftar ulang	tidak daftar ulang	probabilitas daftar ulang
3.000.001 - 4.000.000	909	144	86.32
4.000.001 - 5.000.000	605	102	85.57
2.500.001 - 3.000.000	557	100	84.78
1.500.001 - 2.000.000	582	106	84.59
2.000.001 - 2.500.000	508	94	84.39
500.001 - 1.000.000	630	121	83.89
1.000.001 - 1.500.000	766	148	83.81
5.000.001 - 7.500.000	582	125	82.32
7.500.001 - 10.000.000	299	65	82.14
10.000.001 - 15.000.000	112	26	81.16
500.000 atau kurang	367	110	76.94
15.000.000 lebih	118	40	74.68

Selanjutnya Tabel 4.10 menunjukkan bahwa calon mahasiswa yang penghasilan orang tuanya berada pada rentang 1,500,000 - 4,000,000 memiliki probabilitas tinggi untuk melanjutkan. sedangkan yang penghasilannya 15,000,000 hingga lebih cenderung kecil untuk melanjutkan daftar ulang.

Tabel 4.11: Tabel Observasi pada atribut *prodi_terpilih*

prodi_terpilih	daftar ulang	tidak daftar ulang	probabilitas daftar ulang
Psikologi dan Kesehatan	173	22	88.72
Adab dan Humaniora	515	81	86.41
Ilmu Sosial dan Politik	402	67	85.71
Ekonomi dan Bisnis Islam	960	171	84.88
Tarbiyah dan Keguruan	904	161	84.88
Dakwah dan Komunikasi	922	199	82.25
Syari'ah dan Hukum	603	136	81.6
Sains dan Teknologi	470	111	80.9
program_magister	182	46	79.82
Ushuluddin dan Filsafat	537	140	79.32

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Terakhir pada Tabel 4.11 menunjukkan jika prodi dengan probabilitas paling tinggi untuk melanjutkan daftar ulang adalah prodi-prodi dari fakultas Psikologi dan Kesehatan, sedangkan yang paling rendah adalah dari fakultas Ushuluddin dan Filsafat. Di Universitas Islam Negeri Surabaya Sunan Ampel tentu didominasi oleh prodi-prodi islam namun ada beberapa fakultas yang memliki jurusan umum seperti fakultas Piskologi dan Kesehatan dan fakultas Sains dan Teknologi. Sayangnya probabilitas daftar ulang di fakultas Sains Teknologi cenderung rendah berada pada peringkat 3 dari bawah.

Dari Tabel 4.7 hingga Tabel 4.11 yang merupakan tabel observasi berisikan probabilitas (dalam bentuk persen) kemungkinan daftar ulang pada tiap jenis *value*, maka didapatkan kesimpulan jika ciri calon mahasiswa baru yang cenderung melanjutkan proses pendaftaran adalah sebagai berikut:

1. Memiliki rentang usia 19 hingga 21 tahun, semakin tua usianya semangkin tinggi potensi daftar ulangnya
2. Merupakan tamatan atau alumni dari pondok
3. Mendaftar Melalui jalur SBMPTN
4. Memiliki orang tua dengan penghasilan berada pada rentang Rp. 1,500,000 hingga Rp. 4,000,000.

4.1.3. Pelatihan Model

Berdasarkan Tabel 3.6 terdapat 3 skenario pelatihan model. Variasi skenario ini bertujuan agar dapat dipilih model terbaik dan skenario terbaik. Sehingga pada proses pelatihan model berikutnya, cukup dilakukan 1 pelatihan saja menggunakan skenario terbaik.

Pelatihan model dilakukan menggunakan bahasa *python* dengan bantuan *library sci-kit learn.* Pelatihan model menggunakan 2 algoritma, yaitu algoritma *Adaboost Decision Tree* dan *Decision Tree* (saja). Hal ini dilakukan guna mengetahui perbandingan peforma kedua algoritma. Gambar 4.2 merupakan cuplikan kode dari model *Adaboost Decision Tree*. Potongan kode tersebut menampilkan parameter apa saja yang digunakan saat proses pelatihan.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

```
model = AdaBoostClassifier(
  DecisionTreeClassifier(max_depth=2),
  algorithm="SAMME.R",
  n_estimators=25
)
```

Gambar 4.2: Cuplikan kode *Adaboost Decision Tree Model*

Tiap-tiap model yang tercipta akan dihitung skor *confusion matrix* masing-masing untuk mengetahui mana model terbaik. Sebelum itu, agar mendapatkan pemahaman yang lebih, maka diberikan contoh perhitungan *confusion matrix* pada pelatihan model *adaboost* skenario 1 dengan atribut kabupaten. Model yang tercipta telah melakukan prediksi terhadap *data test* (data yang sudah memiliki label sebelumnya) dan didapatkan tabel *confusion matrix* seperti pada Tabel 4.12.

Tabel 4.12: Tabel *Confusion Matrix* model *Adaboost* skenario 1

Confusion Matrix		**Kelas hasil prediksi**	
		Positif	***Negatif***
Kelas Sebenarnya	***Positif***	1782	352
	Negatif	19	12

Diketahui:

TP = 1782

TN = 12

FP = 352

FN = 19

Maka dapat dihitung skor *precision, recall,* dan *f-measure* menggunakan persamaan 12, 13, dan 14 seperti di bawah ini:

$$\square\square\square\square\square\square\square\square\square\square\square = \frac{\square\square}{\square\square + \square\square}$$

$$= \frac{1782}{1782 + 352}$$

$$= 0.835051$$

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

$$\square\square\square\square\square\square = \frac{\square\square}{\square\square + \square\square}$$

$$= \frac{17882}{1782 + 19}$$

$$= 0.9894503$$

$$\square = 2\frac{\square\square}{\square + \square}$$

$$= 2\ \frac{0.83\ \square\ 0.99}{0.83 + 0.99}$$

$$= 2\ \frac{0.8217}{1.82}$$

$$= 0.90296703$$

Pada sub bab analisa variabel, juga dijelaskan jika atribut *pekerjaan_orang_tua* dan *jenis_kelamin* tidak memiliki hubungan terhadap status daftar ulang calon mahasiswa baru. Oleh karena itu, akan ditambahkan skenario tambahan dengan melatih model menggunakan data yang telah diolah dengan menghapus atribut *pekerjaan_orang_tua, jenis_kelamin* dan keduanya. Proses menghapus atau menyeleksi atribut tersebut biasa disebut *feature selection* (FS). Dari total 9 skenario tersebut didapatkan 2 hasil pengujian yaitu pada pada Tabel 4.13 dan Tabel 4.14.

Tabel 4.13: Hasil pengujian *Adaboost Decision Tree*.

Scenario	***Train Size***	***Test Size***	***F Measure***	***Precision***	***Recall***
Menggunakan semua atribut					
1	70%	30%	90.571792	83.505155	98.945031
2	60%	40%	90.573224	83.321654	99.207343
3	50%	50%	90.366059	83.593086	98.333333
Tanpa atribut ***jenis_kelamin*** **dan** ***pekerjaan_orang_tua***					
1	70%	30%	90.756729	83.6219	99.222654
2	60%	40%	90.63214	83.362522	99.29078
3	50%	50%	90.23569	83.418223	98.266667
Tanpa atribut ***pekerjaan_orang_tua***					
1	70%	30%	90.895257	83.660131	99.500278
2	60%	40%	90.621426	83.432783	99.165624
3	50%	50%	90.180483	83.323912	98.266667
Tanpa atribut ***jenis_kelamin***					
1	70%	30%	90.651127	83.364399	99.333703
2	60%	40%	90.412529	83.374428	98.748436
3	50%	50%	90.398412	83.384962	98.7

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Tabel 4.14: Hasil pengujian *Decision Tree*.

Scenario	Train Size	Test Size	F Measure	Precision	Recall
1	70%	30%	81.756757	82.924043	80.621877
2	60%	40%	82.325383	84.080035	80.64247
3	50%	50%	82.356937	83.939079	80.833333
Tanpa atribut *jenis_kelamin* dan *pekerjaan_orang_tua*					
1	70%	30%	81.895093	83.209169	80.621877
2	60%	40%	81.717687	83.31166	80.183563
3	50%	50%	82.944162	84.226804	81.7
Tanpa atribut pekerjaan_orang_tua					
1	70%	30%	82.51551	83.839542	81.232649
2	60%	40%	82.283047	83.721934	80.892783
3	50%	50%	82.793694	84.235943	81.4
Tanpa atribut jenis_kelamin					
1	70%	30%	81.598199	82.715345	80.510827
2	60%	40%	82.312925	83.918509	80.767626
3	50%	50%	82.105263	83.66782	80.6

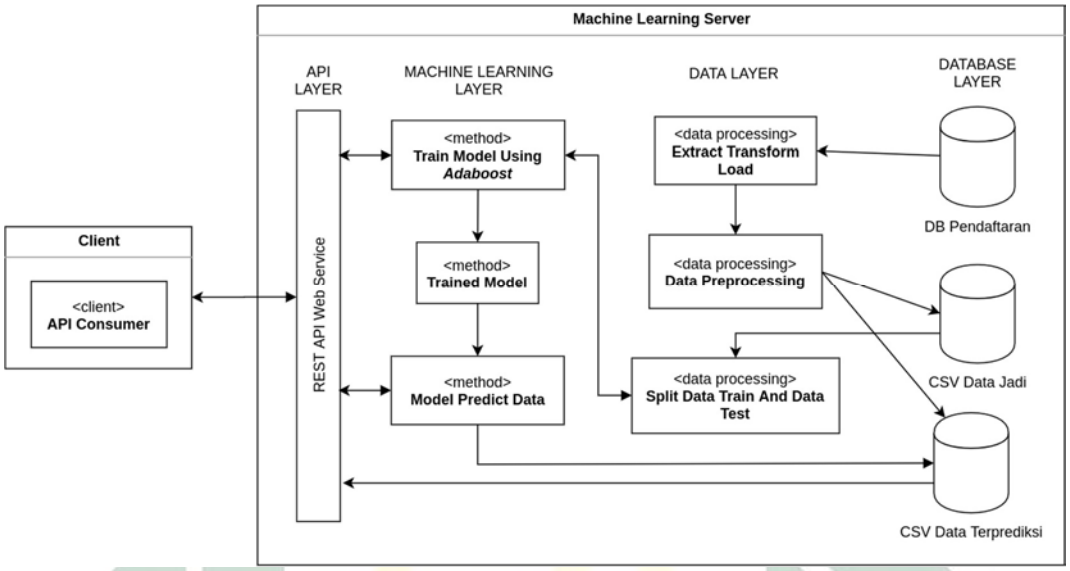
Berdasarkan Tabel 4.13 dan Tabel 4.14 maka dapat disimpulkan jika perubahan jumlah data *training* tidak begitu signifikan terhadap skor *confusion matrix*. Yang mempengaruhi skor dari masing-masing model adalah ada atau tidaknya *feature selection*. Selain itu berdasarkan hasil pengujian pada Tabel 4.13 dan Tabel 4.14, didapatkan bahwa model *Adaboost* terbaik dihasilkan oleh skenario 1 tanpa *pekerjaan_orang_tua* dengan nilai *F Measure* 90.9, *Precision* 83.7 dan *recall* 99.5. Model terbaik ini nantinya akan dijadikan acuan dalam proses prediksi melalui *Rest API* yang dibuat.

Semua perbandingan yang tertera pada Tabel 4.13 dan Tabel 4.14 membuktikan bahwa peforma *Adaboost Decision Tree* lebih baik daripada hanya *Decision Tree* dalam menangani data yang *imbalance* pada semua skenario. Selain itu, dari keseluruhan proses pelatihan model, dapat disimpulkan bahwa jumlah data *train* dan data *test* tidak mempengaruhi peforma model, justru yang mempengaruhi adalah ada atau tidaknya atribut *pekerjaan_orang_tua* maupun *jenis_kelamin*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.1.4. Pengembangan Sistem

Sistem prediksi daftar ulang calon mahasiswa baru dikembangkan berdasarkan desain arsitektur sistem, seperti yang ditunjukan pada Gambar 4.3.



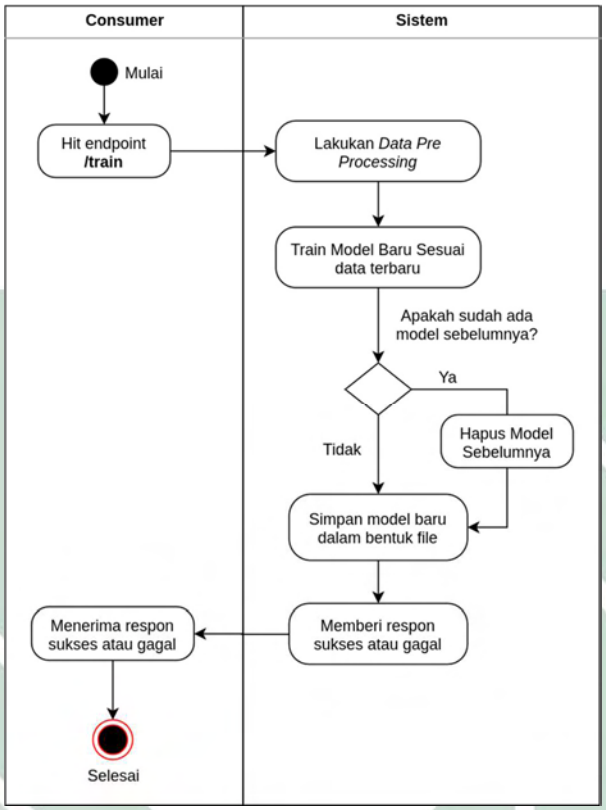
Gambar 4.3: Desain Arsitektur pada sistem prediksi

Pada *Machine Learning server* terdapat 4 *layer*, yaitu *Database Layer* untuk mengambil data dari database dan membuat file CSV sebagai tempat untuk melakukan *warehousing data* yang telah diolah. *Layer* berikutnya adalah *Data Layer* yang dimana bertugas untuk melakukan proses *ETL* dan *Data Preprocessing* sekaligus *split* data *train* dan *data test* untuk dilanjutkan menuju *layer* berikutnya yaitu *Machine Learning Layer*. Pada *layer* ini mesin dilatih menggunakan *data output* dari *Data Layer* sehingga menghasilkan sebuah model baru. Dan layer terakhir pada *machine learning server* adalah API *layer* yang bertugas mengekspos REST API *Web Service* untuk memungkinkan *consumer* atau *client* melakukan transaksi menggunakan JSON.

Selanjutnya *Activity Diagram,* diagram ini bertugas untuk menjelaskan alur-alur aktivitas yang dilakukan oleh *user* pada sistem sekaligus memudahkan *user* dalam memahami alur sistem yang dikembangkan. *User* utama dari sistem ini adalah *consumer* API *Web Service.* Berikut adalah alur-alur dari *activity diagram* pada sistem prediksi kemungkinan daftar ulang calon mahasiswa baru.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Gambar 4.4 merupakan *activity diagram* yang menunjukkan aktivitas untuk *training model*.

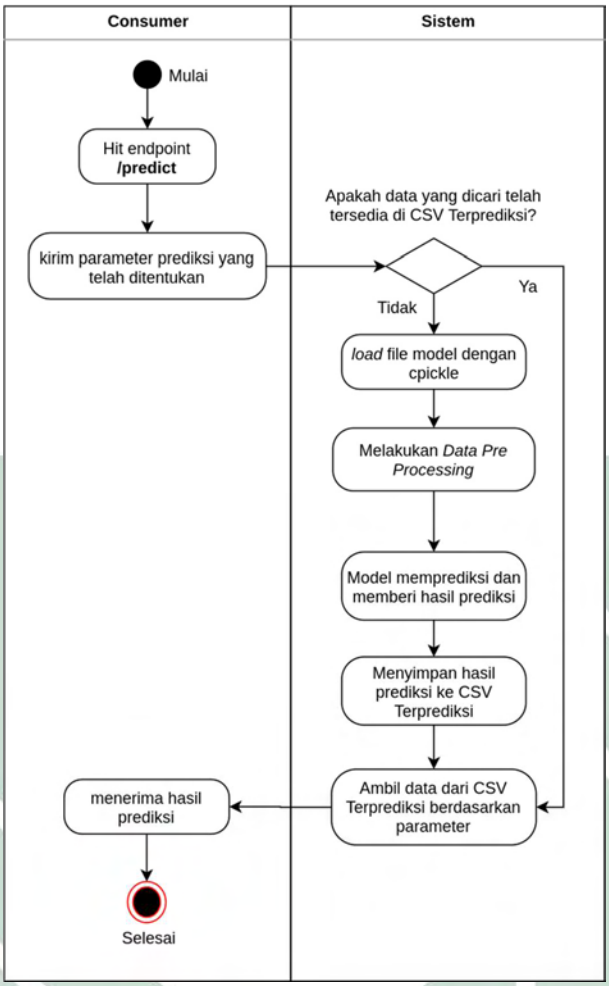


Gambar 4.4: *Activity Diagram* pada proses *training model*

Activity diagram pada Gambar 4.4 menjelaskan alur untuk melakukan *training model* baru. *Consumer* melakukan *hit* pada *endpoint /train* tanpa mengirimkan parameter apapun, sistem akan menerima *request* dari *consumer* lalu menindaklanjuti *request* tersebut dengan melakukan pelatihan untuk membuat model baru. Apabila tidak ada *error* yang terjadi, maka *user* akan menerima respon berupa *success*.

Berikutnya pada Gambar 4.5 merupakan *activity diagram* yang menunjukkan aktivitas untuk prediksi data calon mahasiswa baru. Berikut adalah *activity diagram* dari prediksi data mahasiswa.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id



Gambar 4.5: *Activity Diagram* pada proses prediksi data

Diagram pada Gambar 4.5 menjelaskan alur porses perdiksi data mahasiswa secara akumulatif yaitu data keseluruhan data pendaftar pada suatu periode tertentu yang belum memiliki hasil prediksi dari model. *Consumer* melakukan *hit* pada *endpoint /predict* dengan mengirimkan data *payload* berupa *list* nomor pendaftaran mahasiswa yang ingin dilihat hasil prediksinya. Sistem akan menerima *list* tersebut lalu mencarinya pada file CSV Terprediksi. Jika ada maka sistem akan langsung memberikan hasilnya. Tapi jika belum, sistem akan melakukan *loading* file model hasil *training* sebelumnya, melakukan *pre processing data*. Lalu model akan melakukan perhitungan pada data sehingga menghasilkan respon berupa prediksi apakah mahasiswa tersebut cenderung melanjutkan pendaftaran atau mengundurkan diri yang hasilnya nanti akan dikirim kepada *consumer* sebagai nilai balikan dari REST API.

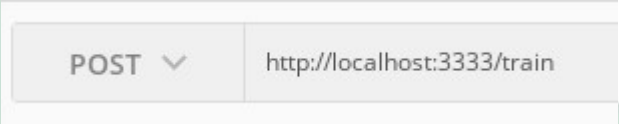
digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

4.1.5. Evaluasi Sistem

Sistem yang dievaluasi pada penelitian ini adalah REST API *Web Service* yang dibuat menggunakan *Flask* dengan bahasa python dengan mengimplentasikan *library scikit-learn*. Pengujian sistem dilakukan dengan bantuan *software Postman*. Terdapat 3 *endpoint* dan berikut masing-masing hasil uji coba yang dilakukan pada REST API *Web Service* pada sistem.

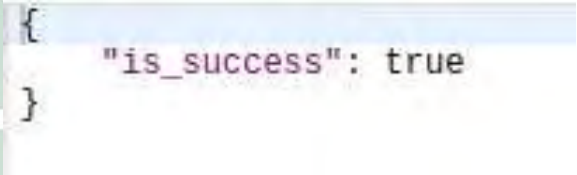
1. *Endpoint /train*

Uji coba menggunakan method request *POST* pada *endpoint /train* tanpa *data payload* sehingga untuk melakukan *training* harus dijembatani oleh sebuah aplikasi yang bisa melakukan *POST request* seperti yang tertera pada Gambar 4.6.



Gambar 4.6: setting url *Postman* untuk uji coba *endpoint /train*

Setelah melakukan *request*, maka *server* merespon dengan hasil prediksi dengan bentuk data seperti Gambar 4.7.



Gambar 4.7: Response server hasil training model baru

Nilai pada *field is_success* memiliki tipe data *boolean* yang akan bernilai *true* jika model berhasil dilatih dan akan bernilai *false* apabila proses *training* gagal.

2. *Endpoint /predict*

Uji coba menggunakan method request *POST* pada *endpoint /predict* dengan *data payload* yang dikirim seperti pada Gambar 4.8.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

```
{
    "nomor_pendaftaran": [
        "4180124568",
        "4180242319",
        "4180429834",
        "4180464002",
        "4180289919",
        "4180164300"
    ]
}
```


Gambar 4.8: *list* nomor pendaftaran untuk uji coba *endpoint /predict*

Setelah melakukan *request*, maka *server* merespon dengan hasil prediksi dengan bentuk data seperti pada Gambar 4.9:


```
{
    "nomor_pendaftaran": {
        "4180124568": true,
        "4180242319": true,
        "4180429834": false,
        "4180464002": true,
        "4180289919": false,
        "4180164300": true
    }
}
```


Gambar 4.9: *Response server* hasil prediksi pada data mahasiswa

Tiap *field* memiliki *key index* berupa nomor pendaftaran yang dikirimkan ke *server*, sedangkan nilai *field* memiliki tipe data *boolean* yang akan bernilai *true* jika data mahasiswa tersebut diprediksi akan melanjutkan daftar ulang, dan akan bernilai *false* apabila diprediksi sistem akan mengundurkan diri.

4.2. Pembahasan

Bagian ini akan membahas tentang hasil penelitian yang telah dipaparkan pada bab sebelumnya untuk menjawab beberapa permasalahan yang telah maupun sedang diteliti. Sehingga akan ditemukan *insight* baru yang bisa dijadikan referensi untuk penelitian berikutnya.

Hasil analisa variabel menunjukkan bahwa tidak adanya hubungan atribut *pekerjaan_orang_tua* dan *jenis_kelamin* terhadap status daftar ulang calon mahasiswa baru. Ini tidak sesuai dengan penelitian yang dilakukan oleh (Aradea

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

dkk., 2011) yang menyatakan bahwa variabel daerah asal tidak memiliki hubungan pada penerimaan mahasiswa baru dan pada penelitian (Melati dkk., 2018) yang menyatakan bahwa variabel pekerjaan orang tua memiliki hubungan terhadap penerimaan mahasiswa baru. Namun variabel lainnya memiliki hubungan terhadap status daftar ulang calon mahasiswa seperti: jalur masuk / gelombang, dan jenis asal sekolah yang juga disinggung pada penelitian (Melati dkk., 2018). Sedangkan penelitian-penelitian lainnya yang menjadi referensi dalam pemilihan atribut tidak secara spesifik mengatakan bahwa atribut tertentu memiliki hubungan. Namun demikian penelitian ini menyatakan bahwa variabel usia, penghasilan total, dan prodi terpilih memiliki hubungan terhadap status daftar ulang calon mahasiswa baru.

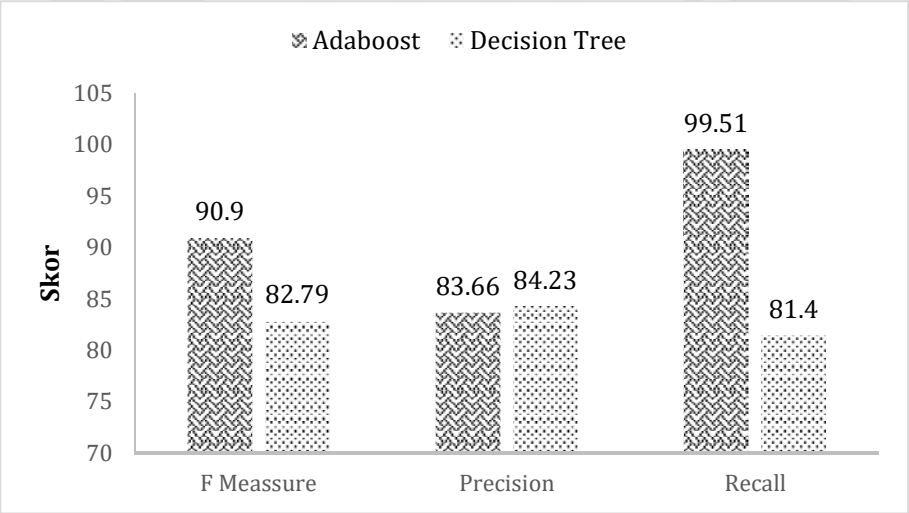
Analisa variabel terhadap status daftar ulang mahasiswa dihitung menggunakan metode statistik *chi square* yang hasilnya dijadikan acuan untuk melakukan proses *feature selection.* Juga ditambahkan skenario tanpa atribut *pekerjaan_orang_tua, jenis_kelamin,* dan keduanya. Hasil riset memperlihatkan jika proses *feature selection* atau pemilihan fitur yang tepat menggunakan *chi-square* mampu menambah skor metriks dari model yang dilatih. Penambahan skor metrik sebesar 0.32% dihitung berdasarkan skor *F-measure* tertinggi dari skenario semua atribut dan skenario *feature selection*. Selisih skor *F-measure* tidak terlalu signifikan, namun karena adanya *feature selection*, maka jumlah atribut yang diproses jadi lebih sedikit. Hasil ini selaras dengan penelitian (Hoiriyah, 2018) dan bertentangan dengan hasil penelitian (Nisa & Darwiyanto, 2019) yang menyatakan jika proses *feature selection* menggunakan *chi square* tidak mempengaruhi peforma model.

Algoritma *Adaboost* memiliki peran penting dalam proses pelatihan model. Ini disebabkan karena pada proses pengolahan data didapatkan hasil perbandingan data kelas positif (melanjutkan daftar ulang) dan data kelas negatif (tidak melanjutkan daftar ulang) yang tidak seimbang (*imbalance*) atau memiliki perbedaan jumlah yang sangat signifikan. Hasil ini sesuai dengan penelitian (Rais & Subekti, 2019) yang menyatakan bahwa algoritma *Adaboost* mampu mengatasi

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

imbalance class pada data. Ketidakseimbangan data ini sangat mencolok dan telah dibahas pada bab pengolahan data.

Untuk membuktikan bahwa *Adaboost* mampu menangani data yang tak seimbang melalui proses *feature selection* maupun tidak, dapat dilihat pada perbandingan skor *F-Measure, Precision*, dan *Recall* di setiap skenario pelatihannya. Perbandingan telah dibahas pada bab pelatihan model. Untuk menampilkan perbandingan yang lebih kongkrit lagi, maka dipilih skenario terbaik dari masing-masing model. Skenario 1 tanpa menggunakan *pekerjaan_orang_tua* merupakan skenario terbaik pada proses pelatihan model *Adaboost Decision Tree*. Sedangkan skenario 3 tanpa menggunakan *pekerjaan_orang_tua* merupakan skenario terbaik dari proses pelatihan model *Decision Tree*. Jika kedua skenario terbaik tersebut dibandingkan, maka didapatkan hasil seperti pada Gambar 4.10.



Gambar 4.10: Perbandingan skenario terbaik dari *Adaboost* dan *Decision Tree*

Dari Gambar 4.10, dapat disimpulkan jika dalam kasus ini, peforma *Adaboost Decision Tree* lebih baik daripada peforma *Decision Tree*. Memiliki selisih skor *F Measure* sebesar 8.11% dan selisih nilai *Recall* sebesar 18.11%.

Srivastava dan Saiprasath G. dkk. juga melakukan komparasi beberapa algoritma *machine learning* diantaranya adalah *Decision Tree* dan *Adaboost*. Hasil menunjukkan bahwa algoritma *Adaboost* mengungguli alogoritma *decision tree*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Walau *adaboost* bukanlah yang terbaik pada kedua penelitian tersebut, namun mampu memprediksi lebih baik daripada *Decision Tree* (Srivastava, 2019; Saiprasath G. dkk., 2019). Hal ini sesuai dengan hasil pelatihan model yang telah dilakukan pada penelitian ini. Sekaligus membuktikan bahwa *Adaboost* memliki peformansi yang lebih baik daripada *Decision Tree*, terlebih lagi dalam pengolahan data yang *imbalance* (Rais & Subekti, 2019).

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

BAB V
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari serangkaian proses penelitian yang dilakukan dan telah dijelaskan hasil pada masing-masing prosesnya pada bab sebelumnya. Maka dapat disimpulkan:

1. Berdasarkan perhitungan *chi-square,* variabel-variabel (beserta nilai *chi-square)* yang memiliki hubungan terhadap status daftar ulang mahasiswa berturut-turut adalah jalur masuk (158.72), jenis asal sekolah (70.622), prodi terpilih (37.895), penghasilan orang tua (35.875), kabupaten asal (33.659), usia (12.92). Kesimpulan ini didapat setelah melakukan tahap analisa variabel menggunakan *chi-square*.
2. Peforma metode *Adaboost* pada *decision tree* untuk prediksi daftar ulang calon mahasiswa baru memiliki skor metriks *precision* 83.66%*, recall* 99.50%, dan *f-measure* 90.89%. Berbeda jauh dengan metode *decision tree* tanpa *Adaboost* yang memiliki rata-rata skor metriks 82%. Hal ini membuktikan bahwa *Adaboost* pada *decision tree* mampu mengatasi *imbalance data* dan memilik peforma yang lebih baik daripada *decision tree* biasa (tanpa *Adaboost*).

5.2. Saran

Dalam penelitian Pengembangan Sistem Prediksi Daftar Ulang Calon Mahasiswa Baru Menggunakan Metode *Adaboost* yang telah dilakukan tentu masih terdapat banyak kekurangan. Oleh karena itu penulis memberikan rekomendasi dan saran untuk perbaikan pada penelitian berikutnya. Yaitu:

1. Memperbanyak *feature* untuk proses *training* seperti 'nilai rapor sekolah tingkat atas', 'jurusan sekolah tingkat atas', 'semua jurusan yang terpilih', dan 'urutan prioritias jurusan yang terpilih', dan lain sebagainya. Hal ini bertujuan agar model memiliki variasi data yang lebih luas lagi.
2. Membandingkan antara metode *Adaboost* dengan metode *ensemble learning* yang lainnya seperti *random forest* menggunakan data yang tak

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

seimbang. Karena pada penelitian (Srivastava, 2019) dan (Saiprasath G. dkk., 2019) *random forest* mendapatkan skor tertinggi namun tidak dijelaskan menggunakan data yang seimbang atau tak seimbang.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

DAFTAR PUSTAKA

Ab Ghani, N. L., Che Cob, Z., Mohd Drus, S., & Sulaiman, H. (2019). Student Enrolment Prediction Model in Higher Education Institution: A Data Mining Approach. Dalam M. A. Othman, M. Z. A. Abd Aziz, M. S. Md Saat, & M. H. Misran (Ed.), *Proceedings of the 3rd International Symposium of Information and Internet Technology (SYMINTECH 2018)* (Vol. 565, hlm. 43–52). Springer International Publishing. https://doi.org/10.1007/978-3-030-20717-5_6

Amalia, H., Pohan, A. B., & Masripah, S. (2019). PENERAPAN FEATURE WEIGHTING OPTIMIZED PADA NAÏVE BAYES UNTUK PREDIKSI PROSES PERSALINAN. *Jurnal Pilar Nusa Mandiri*, *15*(1), 15–20. https://doi.org/10.33480/pilar.v15i1.3

Amershi, S., & Conati, C. (2007). Unsupervised and supervised machine learning in user modeling for intelligent learning environments. *Proceedings of the 12th International Conference on Intelligent User Interfaces - IUI '07*, 72. https://doi.org/10.1145/1216295.1216315

Amornsamankul, S., Pimpunchat, B., Triampo, W., Charoenpong, J., & Nuttavut, N. (2019). A Comparison of Machine Learning Algorithms and Their Applications. *International Journal of Simulation: Systems, Science & Technology*. https://doi.org/10.5013/IJSSST.a.20.04.08

Aradea, Satriyo A., Ariyan Z., & Yuliana A. (2011). PENERAPAN DECISION TREE UNTUK PENENTUAN POLA DATA PENERIMAAN MAHASISWA BARU. *Jurnal Penelitian Sitrotika*.

Darmawan, E. (2018). C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education. *Jurnal Online Informatika*, *3*(1), 22. https://doi.org/10.15575/join.v3i1.171

David M. W . (2011). EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION. *Journal of Machine Learning Technologies*.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C.-W., Khosravi, K., Yang, Y., & Pham, B. T. (2019). Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of The Total Environment*, *662*, 332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Handoko, W. (2019). PREDIKSI JUMLAH PENERIMAAN MAHASISWA BARU DENGAN METODE SINGLE EXPONENTIAL SMOOTHING (STUDI KASUS: AMIK ROYAL KISARAN). *Jurnal Teknologi dan Sistem Informasi*, *2*, 8.

Hasanuddin. (2019). *PERSEPSI MAHASISWA TERHADAP IMPLEMENTASI UANG KULIAH TUNGGAL DI FAKULTAS ILMU SOSIAL UNIVERSITAS NEGERI MAKASSAR*.

Hoiriyah. (2018). *ALGORITMA C4.5 BERBASIS SELEKSI ATRIBUT MENENTUKAN KEMUNGKINAN PENGUNDURAN DIRI MAHASISWA*. *9*, 9.

Khan, F., Ahamed, J., Kadry, S., & Ramasamy, L. K. (2020). Detecting malicious URLs using binary classification through adaboost algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(1), 997. https://doi.org/10.11591/ijece.v10i1.pp997-1005

Lakshmi, T. M., Martin, A., Begum, R. M., & Venkatesan, V. P. (2013). An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data. *International Journal of Modern Education and Computer Science*, *5*(5), 18–27. https://doi.org/10.5815/ijmecs.2013.05.03

Melati, I. S., Linawati, L., & Giriantari, I. A. D. (2018). Knowledge Discovery Data Akademik Untuk Prediksi Pengunduran Diri Calon Mahasiswa. *Majalah Ilmiah Teknologi Elektro*, *17*(3), 325. https://doi.org/10.24843/MITE.2018.v17i03.P04

Nisa, A., & Darwiyanto, E. (2019). *Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi*. 10.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Niswatin, R. K. (2016). *SISTEM SELEKSI PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE WEIGHTED PRODUCT (WP)*. 6.

Nurseitov, N., Paulson, M., Reynolds, R., & Izurieta, C. (2019). *Comparison of JSON and XML Data Interchange Formats: A Case Study*. 6.

Patel, H. H., & Prajapati, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74–78. https://doi.org/10.26438/ijcse/v6i10.7478

Rais, A. N., & Subekti, A. (2019). Integrasi SMOTE Dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing. *Jurnal Informatika*, *6*(2), 278–285. https://doi.org/10.31311/ji.v6i2.6186

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Rozi, A. F. (2015). Sistem Pendukung Keputusan Seleksi Penerimaan Calon Siswa/i Baru Menggunakan Algoritma C4.5 (Studi Kasus: SDIT An-Najah Jatinom Klaten). *Teknoin*, *21*(1). https://doi.org/10.20885/teknoin.vol21.iss1.art2

Saa, A. A., Al-Emran, M., & Shaalan, K. (2018). Mining Student Information System Records to Predict Students' Academic Performance. Dalam A. E. Hassanien, A. T. Azar, T. Gaber, R. Bhatnagar, & M. F. Tolba (Ed.), *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)* (Vol. 921, hlm. 229–239). Springer International Publishing. https://doi.org/10.1007/978-3-030-14118-9_23

Saiprasath G., Naren Babu R., ArunPriyan J., Vinayakumar R., Sowmya V., & Soman K. P. (2019). PERFORMANCE COMPARISON OF MACHINE LEARNING ALGORITHMS FOR MALARIA DETECTION USING MICROSCOPIC IMAGES. *IJRAR*.

Saini, P., & Kumar Jain, A. (2013). Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions. *International Journal of Computer Applications*, *84*(14), 37–41. https://doi.org/10.5120/14646-2966

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Samponu, Y. B., & Kusrini, K. (2018). Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu. *Jurnal ELTIKOM*, *1*(2), 56–63. https://doi.org/10.31961/eltikom.v1i2.29

Sasaki, Y. (2007). *The truth of the F-measure*. 5.

Schapire, R. E. (2013). Explaining AdaBoost. Dalam B. Schölkopf, Z. Luo, & V. Vovk (Ed.), *Empirical Inference* (hlm. 37–52). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_5

Srivastava, A. K. (2019). *Comparison Analysis of Machine Learning algorithms for Steel Plate Fault Detection*. *06*(05), 4.

Sugianto, A. (2016). *JENIS-JENIS DATA VARIABEL (VARIABEL DISKRIT DAN VARIABEL KONTINYU)*. 6.

Surwase, V. (2016). *REST API Modeling Languages—A Developer's Perspective*. *2*(10), 4.

Timofeev, R. (2004). *Declaration of Authorship*. 40.

Vogel, P., Klooster, T., Andrikopoulos, V., & Lungu, M. (2017). A Low-Effort Analytics Platform for Visualizing Evolving Flask-Based Python Web Services. *2017 IEEE Working Conference on Software Visualization (VISSOFT)*, 109–113. https://doi.org/10.1109/VISSOFT.2017.13

Wanjau, S. K., & Muketha, G. M. (2018). Improving Student Enrollment Prediction Using Ensemble Classifiers. *International Journal of Computer Applications Technology and Research*, *07*(03), 122–128. https://doi.org/10.7753/IJCATR0703.1003

Wu, Y., Ke, Y., Chen, Z., Liang, S., Zhao, H., & Hong, H. (2020). Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *CATENA*, *187*, 104396. https://doi.org/10.1016/j.catena.2019.104396

Wulandari, Y., Wiranto, & Wiharto. (2019). *KLASIFIKASI KELUHAN PELANGGAN BERBASIS TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) (STUDI KASUS POS INDONESIA)*. 9.

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id

Yahya, N., & Jananto, A. (2019). *KOMPARASI KINERJA ALGORITMA C.45 DAN NAIVE BAYES UNTUK PREDIKSI KEGIATAN PENERIMAANMAHASISWA BARU (STUDI KASUS : UNIVERSITAS STIKUBANK SEMARANG)*. 8.

Yan, M., Sun, H., Wang, X., & Liu, X. (2012). WS-TaaS: A Testing as a Service Platform for Web Service Load Testing. *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, 456–463. https://doi.org/10.1109/ICPADS.2012.69

Zulfa, H., Nhita, F., & Saepudin, D. (2019). Hibridisasi Algoritma Classification and Regression Tree (CART) dan Artificial Neural Network (ANN) untuk Prediksi Harga Bawang Merah di Kabupaten Bandung. 9

digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id