

Penerapan Algoritma *Decision Tree* C4.5 dan Metode *AdaBoost* Untuk Prediksi Kelulusan Mahasiswa

R. Vincencius Andreas Suyanto¹, Eduard Rusdianto², Ernawati³

Program Studi Informatika, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta
Jl. Babarsari No. 44, Sleman 55281, Daerah Istimewa Yogyakarta, Indonesia

Email: vincensuyanto@gmail.com, eduard.rusdianto@uajy.ac.id, ernawati@uajy.ac.id

Abstract. *A student is someone who carries out the learning process at a university. Atma Jaya University Yogyakarta is one of the private universities in Indonesia. Students can be declared to have graduated from a tertiary institution after fulfilling the graduation requirements for four years of study. Based on student data from 2012 to 2016 in the Informatics Study Program, out of 717 students, 291 students graduated on time, and 426 students graduated not on time, resulting in a graduation percentage of 40.6%. Therefore, it is necessary to carry out research to overcome this problem. This research uses classification data mining techniques with the C4.5 decision tree and AdaBoost algorithms. The collected data will be processed by conducting training and data testing. This research produced 13 rules with an accuracy value of 77.78% and an AUC value of 0.901. The study program can use the results and evaluation of the resulting model to help make decisions.*

Keywords: *Students, Classification, Decision Tree C4.5, AdaBoost*

Abstrak. Mahasiswa adalah seseorang yang melakukan proses belajar di sebuah perguruan tinggi. Universitas Atma Jaya Yogyakarta merupakan salah satu perguruan tinggi swasta yang ada di Indonesia. Mahasiswa dapat dinyatakan lulus dari sebuah perguruan tinggi setelah memenuhi syarat-syarat kelulusan dengan lama studi selama empat tahun. Berdasarkan data mahasiswa tahun 2012 hingga 2016 pada Program Studi Informatika, dari 717 mahasiswa, terdapat 291 mahasiswa yang lulus tepat waktu dan 426 mahasiswa yang lulus tidak tepat waktu, sehingga diperoleh persentase kelulusan sebesar 40,6%. Oleh karena itu, perlu dilakukannya penelitian untuk mengatasi permasalahan tersebut. Penelitian ini menggunakan teknik data mining klasifikasi dengan algoritma *decision tree* C4.5 dan algoritma *AdaBoost*. Data yang dikumpulkan akan diolah dengan melakukan pelatihan dan pengujian data. Penelitian ini menghasilkan 13 rule dengan nilai akurasi sebesar 77,78% dan nilai AUC sebesar 0,901. Analisis hasil dan evaluasi model yang dihasilkan dapat digunakan oleh pihak program studi dalam membantu mengambil keputusan.

Kata Kunci: *Mahasiswa, Klasifikasi, Decision Tree C4.5, AdaBoost*

1. Pendahuluan

Mahasiswa adalah orang yang belajar pada jenjang perguruan tinggi [1]. Mahasiswa adalah sebutan untuk orang yang sedang menempuh pendidikan tinggi di sebuah perguruan tinggi yang terdiri atas sekolah tinggi, akademi, dan yang paling umum adalah universitas [2]. Mahasiswa memiliki peranan yang sangat penting dalam kehidupan suatu negara. Mahasiswa berperan untuk menjamin kelangsungan hidup negara dan bangsa serta mewujudkan cita – cita pembangunan nasional. Pada umumnya mahasiswa akan menjalani proses pendidikan yang diselenggarakan oleh perguruan tinggi, yaitu mengikuti proses belajar mengajar, melakukan penelitian dan melakukan pengabdian pada masyarakat. Mahasiswa dapat dinyatakan lulus dari sebuah perguruan tinggi apabila sudah memenuhi syarat-syarat kelulusan yang ada di perguruan tinggi tersebut.

Universitas Atma Jaya Yogyakarta (UAJY) adalah salah satu perguruan tinggi swasta (PTS) yang ada di Indonesia. Universitas ini didirikan pada tanggal 27 September 1965 oleh kaum awam dan dikelola oleh Yayasan Slamet Rijadi Yogyakarta, dibawah naungan Santo Albertus Magnus. Universitas ini memiliki empat kampus dengan satu kampus terletak di

daerah Mrican, Yogyakarta dan tiga kampus terletak di daerah Babarsari, Yogyakarta. Universitas Atma Jaya Yogyakarta termasuk ke dalam 10 universitas swasta terbaik Indonesia versi QS AUR 2023 [3]. Dengan status tersebut maka tidak heran jika banyak siswa yang telah lulus dari jenjang Sekolah Menengah Atas (SMA), berminat untuk mendaftar dan menempuh pendidikan yang lebih tinggi di Universitas Atma Jaya Yogyakarta.

Untuk mempertahankan statusnya sebagai 10 universitas terbaik atau bahkan lebih baik lagi, maka Universitas Atma Jaya Yogyakarta harus memiliki penilaian akreditasi yang unggul. Akreditasi merupakan kegiatan yang dilakukan oleh organisasi atau badan mandiri di luar perguruan tinggi, dalam melakukan penilaian kelayakan dan mutu dari suatu perguruan tinggi atau program studi [4]. Akreditasi bertujuan untuk menentukan kelayakan dan mutu eksternal dari program studi atau perguruan tinggi, berdasarkan kriteria yang mengacu pada Standard Nasional Perguruan Tinggi. Salah satu dari banyak faktor yang mempengaruhi penilaian akreditasi adalah kelulusan mahasiswa. Jika mahasiswa lulus tepat waktu maka akan sangat membantu dalam penilaian akreditasi perguruan tinggi tersebut. Sebaliknya, jika mahasiswa lulus tidak tepat waktu maka dapat memberikan pengaruh negatif terhadap penilaian akreditasi dari perguruan tinggi tersebut. Kriteria dari mahasiswa yang lulus tepat waktu adalah mahasiswa yang dapat menyelesaikan syarat-syarat kelulusan dalam kurun waktu empat tahun atau kurang.

Universitas Atma Jaya Yogyakarta membuka beberapa program studi salah satunya adalah Program Studi Informatika. Setiap tahunnya Program Studi Informatika menerima mahasiswa baru. Selain itu, Program Studi Informatika juga akan meluluskan mahasiswa yang sudah memenuhi syarat-syarat kelulusan di Universitas Atma Jaya Yogyakarta. Dengan adanya hal tersebut maka program studi ini juga akan menghasilkan data mahasiswa dalam jumlah yang besar.

Data mahasiswa yang dihasilkan dapat diolah menggunakan beberapa metode, agar bisa menghasilkan suatu informasi yang dapat dimanfaatkan oleh Program Studi Informatika. Salah satu proses yang bisa digunakan untuk mengolah data adalah dilakukannya proses *data mining*. *Data mining* adalah proses penggalian dan analisis sekumpulan data yang bertujuan untuk menemukan informasi yang belum diketahui sebelumnya [5]. Data yang ada akan diolah sedemikian rupa, sehingga mendapatkan informasi ataupun *insight* yang dapat membantu pembentukan sistem atau kebijakan. Kebijakan yang didapat dari pengolahan data tersebut nantinya dapat menjadi solusi dari permasalahan yang ingin diselesaikan.

Berdasarkan data mahasiswa angkatan 2012 hingga 2016 pada Program Studi Informatika di Universitas Atma Jaya Yogyakarta. Dari 717 mahasiswa, terdapat 291 mahasiswa yang lulus tepat waktu dan 426 mahasiswa yang lulus tidak tepat waktu, sehingga diperoleh persentase kelulusan sebesar 40,6%. Masalah tersebut tentunya akan berdampak negatif pada penilaian akreditasi program studi apabila dibiarkan terus menerus. Selain itu, masalah tersebut juga akan berdampak pada minat calon mahasiswa baru untuk mengambil Program Studi Informatika di Universitas Atma Jaya Yogyakarta.

Guna menyelesaikan permasalahan sesuai dengan uraian di atas maka digunakan proses *data mining* untuk mengolah data mahasiswa yang jumlahnya besar. Penelitian ini bertujuan untuk mendapatkan informasi tersembunyi yang terdapat pada data mahasiswa tersebut. Teknik *data mining* yang akan digunakan adalah klasifikasi. Klasifikasi adalah proses analisis data yang menghasilkan model yang menggambarkan dan membedakan kelas data atau konsep. Model hasil klasifikasi nantinya bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Teknik ini akan membagi mahasiswa menjadi dua yaitu mahasiswa yang lulus tepat waktu dan mahasiswa yang lulus tidak tepat waktu. Untuk algoritma, dalam penelitian ini algoritma yang digunakan adalah algoritma *decision tree* C4.5. Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk pohon keputusan. Algoritma ini memiliki kelebihan seperti dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, dan efisien dalam menangani atribut bertipe kontinyu dan diskrit [6]. Dalam proses pengklasifikasian, akan dilakukan dua buah proses yaitu proses pelatihan dan proses pengujian.

Penelitian ini juga akan menggunakan metode tambahan yaitu metode *ensemble* yang bernama *AdaBoost*. Algoritma *AdaBoost* atau *Adaptive Boosting* merupakan algoritma *machine learning* yang pertama kali diperkenalkan oleh Freund dan Schapire tahun 1995. Algoritma *AdaBoost* merupakan algoritma yang bersifat adaptif, karena hasil pembelajaran yang salah akan diklasifikasikan dan digunakan kembali pada proses klasifikasi selanjutnya. Algoritma *AdaBoost* akan membentuk *strong classification* dengan mengkombinasikan beberapa *weak classification* secara linier [7]. Tujuan dari penggunaan metode tambahan ini adalah diharapkan hasil yang diperoleh dalam penelitian ini, akan semakin baik dan akurat dalam memprediksi kelulusan mahasiswa.

Hasil penelitian diharapkan dapat membantu menemukan informasi ataupun *insight* tersembunyi dari data mahasiswa Program Studi Informatika, Universitas Atma Jaya Yogyakarta. *Insight* yang didapatkan dari penelitian ini diharapkan dapat membantu pihak program studi, dalam membuat sebuah kebijakan atau strategi untuk mengatasi masalah mahasiswa yang lulus tidak tepat waktu.

2. Tinjauan Pustaka

Adapun penelitian lain dilakukan oleh Heryana (2019), dengan judul "*Data Mining Untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naive Bayes*". Pada penelitian tersebut digunakan teknik *data mining* klasifikasi untuk memprediksi kelulusan mahasiswa, dimana prediksi tersebut adalah tepat atau tidak tepat waktunya mahasiswa dalam menyelesaikan masa studinya. Algoritma yang digunakan pada penelitian tersebut adalah algoritma Naive Bayes. Algoritma ini digunakan untuk memprediksi masa depan menggunakan probabilitas yang menerapkan aturan Bayes. Aturan Bayes memprediksikan masa yang akan datang dengan menggunakan pengalaman sebelumnya. Atribut yang digunakan untuk melakukan klasifikasi adalah jenis kelamin, kota kelahiran, tipe sekolah, lokasi sekolah, ekonomi, Indeks Prestasi Kumulatif (IPK), dan Keputusan sebagai label data. Objek yang diteliti pada penelitian tersebut adalah data mahasiswa angkatan 2013 sebagai data latih dan data mahasiswa angkatan 2015 sebagai data uji. Hasil yang diperoleh pada penelitian tersebut menunjukkan nilai akurasi pada data latih sebesar 74,67% dari data yang berjumlah 51. Hasil lain yang diperoleh pada penelitian tersebut menunjukkan bahwa data uji dari 184 mahasiswa angkatan 2015, diprediksi akan lulus tepat waktu sebanyak 42 mahasiswa atau sekitar 22,8%, dan mahasiswa yang lulus tidak tepat waktu sebanyak 142 mahasiswa atau sekitar 77,2% [8].

Penelitian lain dilakukan oleh Rohman dan Rufiyanto, dengan judul "*Implementasi Data Mining Dengan Algoritma Decision Tree C4.5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandanaran*". Pada penelitian tersebut digunakan teknik *data mining* klasifikasi untuk menemukan pola dan memprediksi kelulusan mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *decision tree* C4.5, yang digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola kelulusan mahasiswa. Atribut yang digunakan untuk melakukan klasifikasi adalah jurusan, umur, jenis kelamin, pekerjaan, Indeks Prestasi Kumulatif (IPK) 1 sampai 4, dan atribut label sebagai kelas data. Objek yang diteliti pada penelitian tersebut adalah data kelulusan mahasiswa di Universitas Pandanaran Semarang jenjang pendidikan D3 di Fakultas Teknik. Hasil yang diperoleh pada penelitian tersebut menghasilkan nilai akurasi sebesar 65,98% dan nilai Area Under Curve (AUC) sebesar 0,874, dimana hasil ini dapat dikategorikan sebagai klasifikasi data yang baik [9].

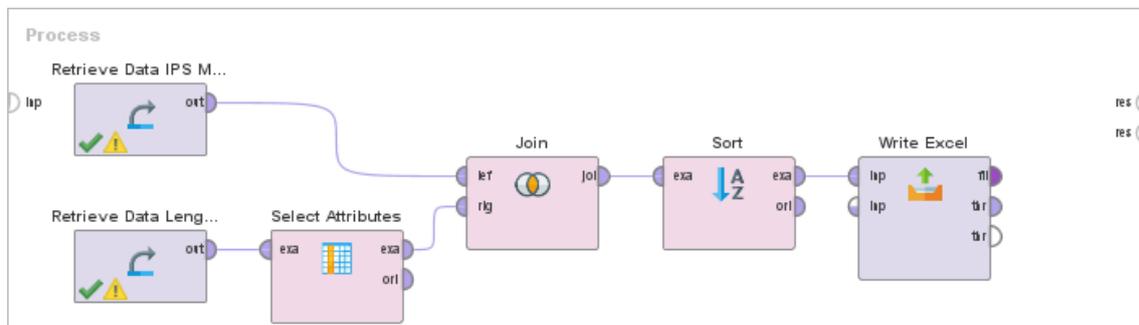
Rohmawan juga melakukan penelitian tentang prediksi kelulusan mahasiswa pada tahun 2018, dengan judul "*Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Artificial Neural Network*". Pada penelitian tersebut teknik *data mining* yang digunakan adalah klasifikasi untuk memprediksi kelulusan mahasiswa. Algoritma yang digunakan pada penelitian tersebut adalah algoritma *decision tree* dan *artificial neural network*, yang akan diimplementasikan untuk memprediksi kelulusan dan menemukan pola kelulusan mahasiswa. Atribut yang digunakan untuk melakukan klasifikasi adalah jenis kelamin, IP semester 1 sampai semester 6, IPK semester 1 sampai semester 6, total sks semester 1 sampai

semester 6, semester 2 sampai semester 6, dan atribut keterangan sebagai label data. Objek yang diteliti pada penelitian ini adalah mahasiswa angkatan 2006 sampai 2010 sebagai data latih pada proses *data mining* dan mahasiswa angkatan 2011 sampai 2012 yang diasumsikan belum lulus sebagai data uji pada proses *data mining*. Hasil yang diperoleh pada penelitian ini yaitu dapat digunakannya algoritma *decision tree* dengan akurasi 74.51% dan *artificial neural network* dengan akurasi 79.74%, untuk melakukan prediksi kelulusan mahasiswa tepat waktu. [10].

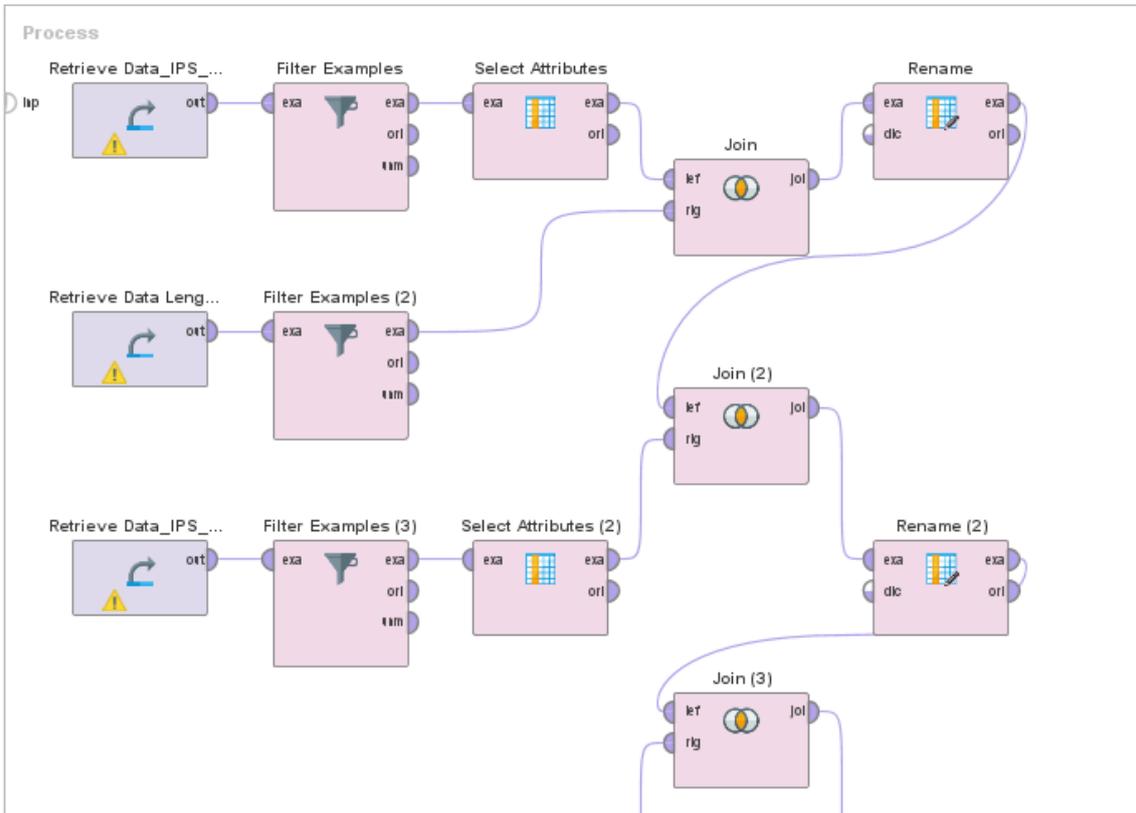
Penelitian lain dilakukan oleh Hakim, dkk., dengan judul "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis *K-Nearens Neighbor* (K-NN)". Pada penelitian ini dibuat sebuah aplikasi yang digunakan untuk melakukan prediksi kelulusan mahasiswa. Teknik *data mining* yang digunakan adalah klasifikasi yang diimplementasikan ke dalam aplikasi. Algoritma yang digunakan pada pembuatan aplikasi prediksi kelulusan mahasiswa adalah algoritma *K-Nearest Neighbor*. Penggunaan algoritma ini bertujuan untuk mengidentifikasi kelulusan mahasiswa pada kasus baru, dengan cara mengadaptasi solusi dari kasus sebelumnya yang memiliki kedekatan dengan kasus baru. Atribut yang digunakan untuk melakukan klasifikasi adalah NIM, Nama Mahasiswa, Jenis Kelamin, Indeks Prestasi Semester 1 sampai 5, dan Keterangan Lulus sebagai label data. Objek yang diteliti pada penelitian ini adalah data – data kelulusan mahasiswa dan data mahasiswa yang masih aktif. Hasil yang diperoleh pada penelitian ini menghasilkan nilai akurasi 98% pada K=1 untuk klasifikasi "Tepat Waktu" [11].

3. Metodologi Penelitian

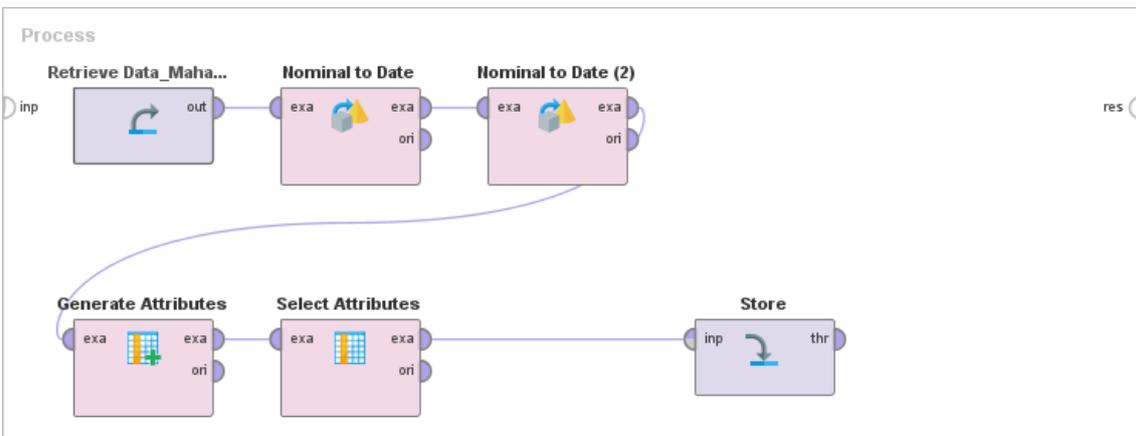
Metodologi penelitian yang digunakan pada penelitian ini adalah sebagai berikut: (1) Studi *Literature* (2) Identifikasi Masalah (3) Pengumpulan Data (4) *Preprocessing Data* (5) Pelatihan dan Pengujian Model (6) Analisis dan Evaluasi Hasil. Setelah melakukan studi *literature* dan identifikasi masalah, dilakukan proses pengumpulan data sekunder yang didapatkan dari pihak Kantor Sistem Informasi, Universitas Atma Jaya Yogyakarta. Data yang dikumpulkan adalah Data Lengkap Mahasiswa dan Data IPS Mahasiswa. Setelah melakukan pengumpulan data, dilakukan *preprocessing data* terhadap data yang dikumpulkan untuk mengubah data mentah menjadi data yang sesuai dan siap untuk dilakukan pemodelan data. Tahapan *preprocessing data* dapat dilihat pada Gambar 1, Gambar 2, dan Gambar 3.



Gambar 1. *Building Process Join Atribut TAHUN_MASUK*



Gambar 2. Building Process Join Dataset



Gambar 3. Building Process Generate Attributes KETERANGAN_LULUS

Proses pada Gambar 1 dilakukan dengan tujuan, untuk menggabungkan atribut TAHUN_MASUK kedalam Data IPS Mahasiswa agar mempermudah penyaringan data pada proses selanjutnya. Proses pada Gambar 2 dilakukan dengan tujuan, untuk mendapatkan nilai Indeks Prestasi Semester (IPS) 1 hingga 8 dari tiap mahasiswa yang digabungkan ke dalam Data Lengkap Mahasiswa. Proses pada Gambar 3 dilakukan dengan tujuan, untuk memilih atribut yang digunakan dalam proses klasifikasi dan mengklasifikasikan mahasiswa menjadi LTW (Lulus Tepat Waktu) dan LTTW (Lulus Tidak Tepat Waktu).

4. Hasil dan Diskusi

4.1. Dataset

Dataset yang digunakan pada penelitian ini adalah Data_Lengkap_Mahasiswa_Final_2012 sebagai data latih dan

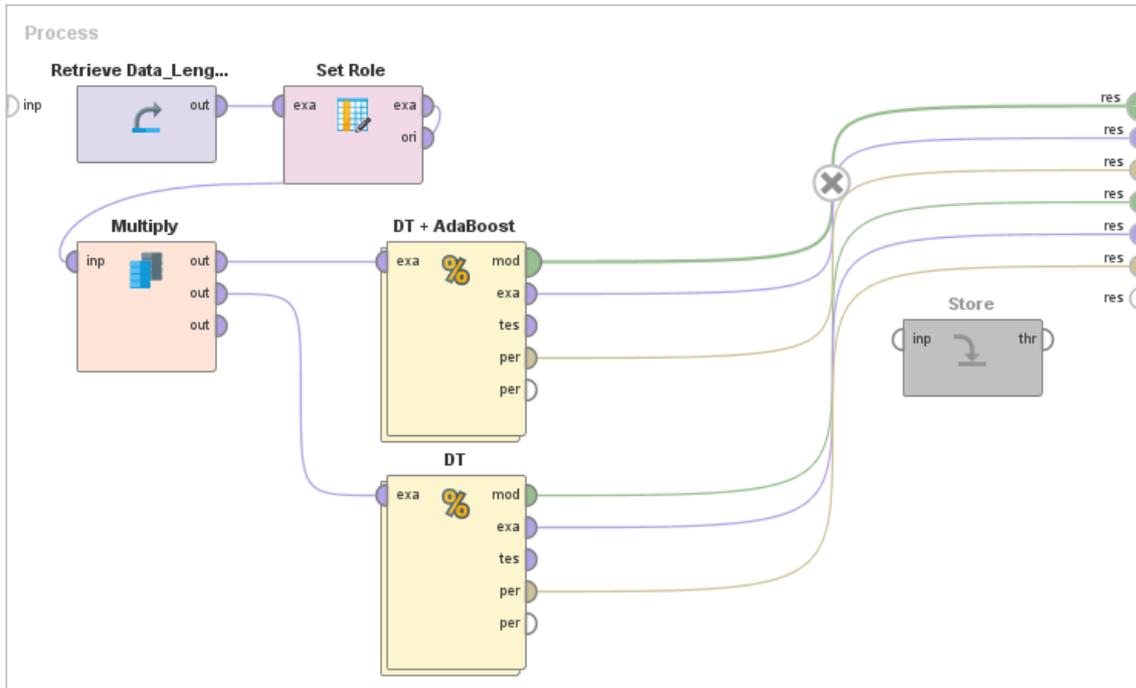
Data_Lengkap_Mahasiswa_Final_2016 sebagai data uji. Data tersebut merupakan hasil dari penggabungan Data Lengkap Mahasiswa dan Data IPS Mahasiswa. Keterangan dari masing-masing atribut pada dataset yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Keterangan Dataset

No	Atribut	Deskripsi	Spesifikasi
1	ID	Berisikan id dari masing – masing mahasiswa (<i>primary key</i>)	152 unique value (2012), 135 unique value (2016), <i>No missing value</i>
2	JALUR_MASUK	Berisikan keterangan jalur masuk yang ditempuh mahasiswa untuk masuk ke Universitas Atma Jaya Yogyakarta	20 nominal value, <i>No missing value</i>
3	TAHUN_MASUK	Berisikan tahun awal masuknya mahasiswa menjadi mahasiswa UAJY	<i>No missing value</i>
4	SKS_KUMULATIF	Berisikan jumlah total sks yang sudah ditempuh oleh mahasiswa hingga lulus	<i>No missing value</i>
5	IPS_1	Berisikan nilai IP dari mahasiswa pada semester 1	<i>No missing value</i>
6	IPS_2	Berisikan nilai IP dari mahasiswa pada semester 2	<i>No missing value</i>
7	IPS_3	Berisikan nilai IP dari mahasiswa pada semester 3	<i>No missing value</i>
8	IPS_4	Berisikan nilai IP dari mahasiswa pada semester 4	<i>No missing value</i>
9	IPS_5	Berisikan nilai IP dari mahasiswa pada semester 5	<i>No missing value</i>
10	IPS_6	Berisikan nilai IP dari mahasiswa pada semester 6	1 missing value (2012) 1 missing value (2016)
11	IPS_7	Berisikan nilai IP dari mahasiswa pada semester 7	1 missing value (2012)
12	IPS_8	Berisikan nilai IP dari mahasiswa pada semester 8	3 missing value (2012) 4 missing value (2016)
13	IPK	Berisikan nilai IPK (Indeks Prestasi Kumulatif) mahasiswa	<i>No missing value</i>
14	PREDICATE	Berisikan predikat kelulusan mahasiswa	6 nominal value, <i>No missing value</i>
15	KETERANGAN_LULUS	Berisikan keterangan tepat atau tidak tepat waktu kelulusan mahasiswa	2 nominal value, <i>No missing value</i>

4.2. Pelatihan Data

Pada tahap ini, dilakukan proses menerapkan algoritma *decision tree* C4.5 dan algoritma *AdaBoost* yang menggunakan algoritma C4.5 sebagai dasar pemodelan data terhadap data latih. Data latih yang digunakan adalah Data_Lengkap_Mahasiswa_Final_2012. Pada tahap ini dilakukan juga proses perbandingan performa antara algoritma *AdaBoost* dengan algoritma *decision tree* C4.5. Dari hasil perbandingan performa tersebut, model dari algoritma dengan performa terbaik akan digunakan pada proses pengujian model. *Building Process* dari penerapan algoritma pada tahap ini dapat dilihat pada Gambar 4.



Gambar 4. Building Process Penerapan Algoritma

Pada awal proses pada Gambar 4, dilakukan *set role* untuk mengubah *role* atribut KETERANGAN_LULUS menjadi label. Langkah selanjutnya adalah melakukan pengujian performa algoritma menggunakan operator *cross validation*. Algoritma yang diuji performanya adalah algoritma *AdaBoost* dan algoritma *decision tree* C4.5. Proses iterasi pada algoritma *AdaBoost* dilakukan sebanyak delapan iterasi. Hal ini dikarenakan nilai akurasi yang dihasilkan sebelum dan setelah iterasi kedelapan selalu di bawah nilai akurasi yang dihasilkan pada iterasi kedelapan. Setelah itu dilakukan perbandingan hasil performa dari algoritma yang digunakan. Nilai akurasi yang dihasilkan pada percobaan iterasi algoritma *AdaBoost* dapat dilihat pada Tabel 2. Hasil perbandingan performa dari kedua algoritma yang diuji dapat dilihat pada Tabel 3.

Tabel 2. Nilai Akurasi Percobaan Iterasi Algoritma *AdaBoost*

Jumlah Iterasi	Akurasi
1	84.83%
2	75.83%
3	84.96%
4	83.54%
5	84.83%
6	86.12%
7	86.79%
8	87.46%
9	85.56%
10	86.79%

Tabel 3. Perbandingan Performa Algoritma

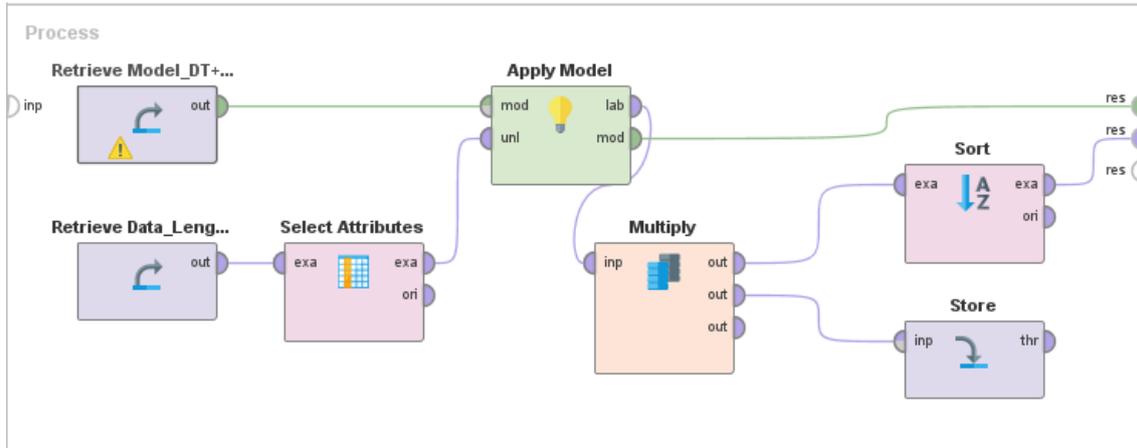
	<i>AdaBoost</i>	<i>Decision Tree</i> C4.5
Akurasi	87.46%	81.58%
<i>Area Under Curve</i>	0.920	0.804

Dari tabel perbandingan performa algoritma pada Tabel 3, dapat dilihat bahwa hasil dari algoritma *AdaBoost* lebih baik dari algoritma *decision tree* C4.5. Oleh karena itu, algoritma yang digunakan pada proses pengujian adalah algoritma *AdaBoost* dengan algoritma C4.5 sebagai dasar pemodelan data. Setelah melakukan perbandingan performa, model algoritma

dengan performa terbaik disimpan ke dalam *local repository* RapidMiner menggunakan operator *store*.

4.3. Pengujian Model

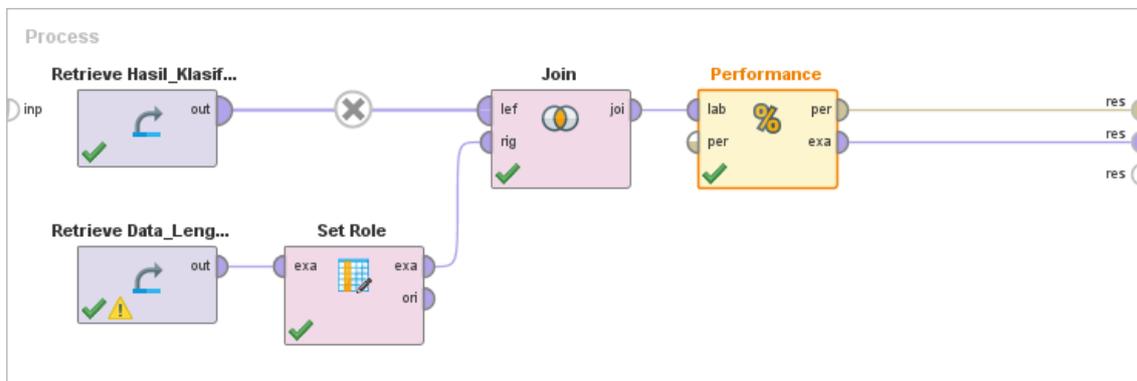
Setelah mendapatkan model algoritma, penelitian dilanjutkan dengan melakukan pengujian terhadap *Data_Lengkap_Mahasiswa_Final_2016*. Pengujian dilakukan untuk mendapatkan prediksi kelulusan mahasiswa pada mahasiswa angkatan 2016. Pada penelitian ini, diasumsikan bahwa mahasiswa angkatan 2016 merupakan mahasiswa yang belum lulus. *Building process* pada tahap ini dapat dilihat pada Gambar 5.



Gambar 5. *Building Process* Pengujian Model

Pada awal proses yang terlihat pada Gambar 5, dilakukan dua pemanggilan data yaitu *Model_DT+AdaBoost_Klasifikasi_KelulusanMahasiswa* dan *Data_Lengkap_Mahasiswa_Final_2016*. Setelah itu, dilakukan pemilihan atribut pada *Data_Lengkap_Mahasiswa_Final_2016* untuk membuang atribut *KETERANGAN_LULUS*. Langkah selanjutnya adalah melakukan *apply model* untuk memprediksi kelulusan mahasiswa pada mahasiswa angkatan 2016. Setelah mendapatkan hasil prediksi, dilakukan penyimpanan data prediksi ke dalam *local repository*. *File* prediksi kelulusan mahasiswa akan disimpan dengan nama *Hasil_Klasifikasi_ModelKelulusanMahasiswa_2016*.

Setelah mendapatkan hasil prediksi kelulusan mahasiswa, langkah selanjutnya melakukan pengukuran performa dari hasil prediksi tersebut. Pengukuran performa dilakukan dengan menggabungkan data hasil prediksi dengan data asli yang terjadi. *Building process* dari pengukuran performa hasil prediksi dapat dilihat pada Gambar 6.



Gambar 6. *Building Process* Pengukuran Performa Hasil Prediksi

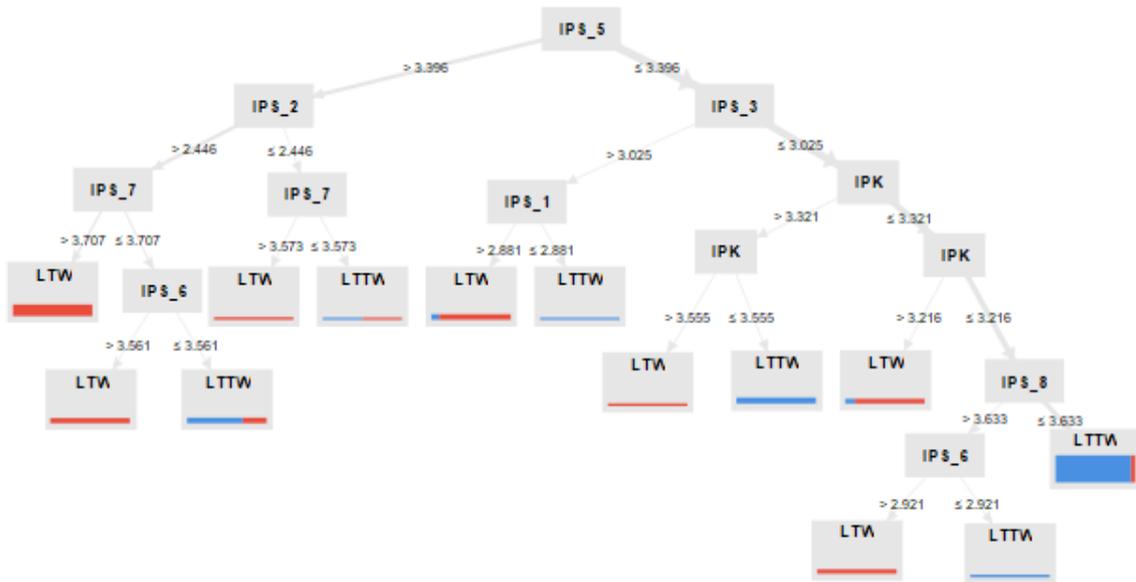
Pada awal proses pada Gambar 6, dilakukan dua pemanggilan data yaitu Hasil_Klasifikasi_ModelKelulusanMahasiswa_2016 dan Data_Lengkap_Mahasiswa_Final_2016. Selanjutnya, mengubah *role* atribut KETERANGAN_LULUS pada Data_Lengkap_Mahasiswa_Final_2016 menjadi label. Setelah itu, melakukan penggabungan data menggunakan operator *join* dengan tipe *right join* dengan atribut ID sebagai atribut kunci. Setelah melakukan penggabungan data, langkah selanjutnya adalah melakukan pengukuran performa dari hasil prediksi menggunakan operator *performance*. Pengukuran performa dilakukan untuk membandingkan kelulusan mahasiswa dari hasil prediksi dengan data yang sebenarnya terjadi. Akurasi dan *Area Under Curve* (AUC) dari prediksi kelulusan mahasiswa dapat dilihat pada Tabel 4.

Tabel 4. Hasil Performa Dari Prediksi Kelulusan Mahasiswa

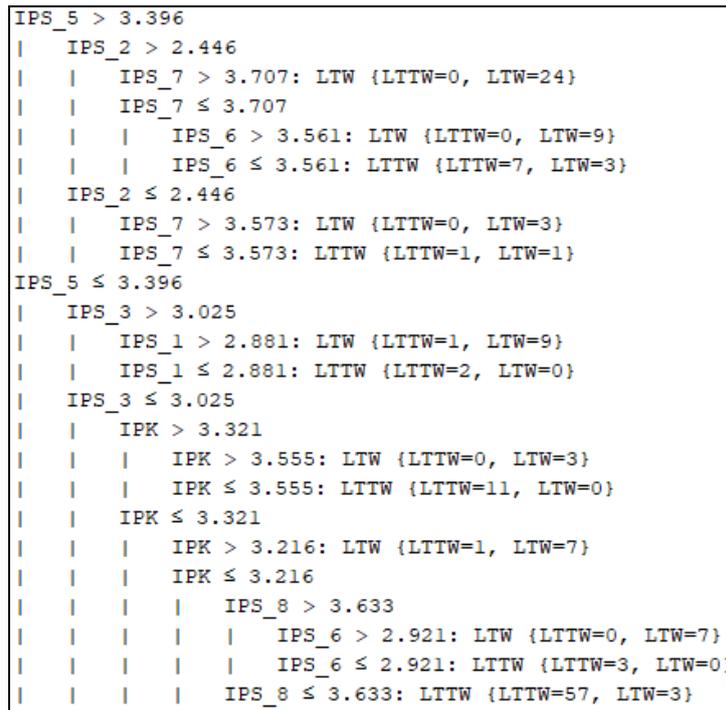
Hasil Performa	
Akurasi	77.78%
Area Under Curve	0.901

4.4. Analisis Hasil dan Evaluasi Model

Setelah mendapatkan performa dari hasil prediksi, selanjutnya dilakukan analisis hasil prediksi dari model yang terbentuk. Analisis hasil dilakukan guna menggali informasi dari model yang terbentuk. Hasil yang diperoleh merupakan *rule* dari model pohon keputusan yang terbentuk. Selain itu, ditampilkan juga persentase dari seluruh data mahasiswa untuk masing-masing *rule* yang terbentuk. Pohon keputusan yang terbentuk dari hasil mengolah data dapat dilihat pada Gambar 7. *Rule* pohon keputusan dapat dilihat pada Gambar 8. Persentase data untuk masing-masing *rule* yang terbentuk dapat dilihat pada Tabel 5.



Gambar 7. Pohon Keputusan



Gambar 8. Rule Pohon Keputusan

Tabel 5. Persentase Data Untuk Masing-masing Rule Yang Terbentuk

Rule	Persentase Data
IPS_5 > 3.396 AND IPS_2 > 2.446 AND IPS_7 > 3.707	15,79%
IPS_5 > 3.396 AND IPS_2 > 2.446 AND IPS_7 ≤ 3.707 AND IPS_6 > 3.561	5,92%
IPS_5 > 3.396 AND IPS_2 > 2.446 AND IPS_7 ≤ 3.707 AND IPS_6 ≤ 3.561	6,58%
IPS_5 > 3.396 AND IPS_2 ≤ 2.446 AND IPS_7 > 3.573	1,97%
IPS_5 > 3.396 AND IPS_2 ≤ 2.446 AND IPS_7 ≤ 3.573	1,32%
IPS_5 ≤ 3.396 AND IPS_3 > 3.025 AND IPS_1 > 2.881	6,58%
IPS_5 ≤ 3.396 AND IPS_3 > 3.025 AND IPS_1 ≤ 2.881	1,32%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK > 3.321 AND IPK > 3.555	1,97%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK > 3.321 AND IPK ≤ 3.555	7,24%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK ≤ 3.321 AND IPK > 3.216	5,26%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK ≤ 3.321 AND IPK ≤ 3.216 AND IPS_8 > 3.633 AND IPS_6 > 2.921	4,61%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK ≤ 3.321 AND IPK ≤ 3.216 AND IPS_8 > 3.633 AND IPS_6 ≤ 2.921	1,97%
IPS_5 ≤ 3.396 AND IPS_3 ≤ 3.025 AND IPK ≤ 3.321 AND IPK ≤ 3.216 AND IPS_8 ≤ 3.633	39,47%

Dari Gambar 8 dapat diketahui bahwa dari 13 atribut yang digunakan untuk melakukan klasifikasi ke dalam label data, ada lima atribut yang tidak mempengaruhi proses klasifikasi. Atribut tersebut adalah TAHUN_MASUK, JALUR_MASUK, SKS_KUMULATIF, IPS_4, dan PREDICATE. Selain itu, berdasarkan rule yang terbentuk dari proses klasifikasi dapat diketahui kriteria atau penyebab dari mahasiswa yang lulus tidak tepat waktu. Salah satu rule menunjukkan bahwa mahasiswa yang lulus tidak tepat waktu adalah mahasiswa yang memiliki IPS_5 > 3,396, IPS_2 > 2,446, IPS_7 ≤ 3,707, dan IPS_6 ≤ 3,561. Artinya mahasiswa yang memiliki dan memenuhi kriteria tersebut akan di prediksi tidak akan lulus tepat waktu. Namun, rule pohon keputusan yang terbentuk tidak sepenuhnya masuk akal dalam memprediksi kelulusan mahasiswa. Hal ini dikarenakan ada beberapa faktor lain yang dapat mempengaruhi tepat atau tidak tepat waktu, kelulusan seorang mahasiswa. Faktor-faktor lain yang mempengaruhi yaitu cuti studi, tugas akhir, dan mengulang mata kuliah. Pada data mahasiswa angkatan 2016 terdapat lima mahasiswa melakukan cuti studi, 61 mahasiswa mengambil tugas

akhir pada semester 9 yang diasumsikan pada semester 8 mereka sudah mengambil tugas akhir, dan 19 mahasiswa mengulang mata kuliah. Data tersebut didapatkan dengan melakukan analisis dan penyaringan data pada Data IPS Mahasiswa.

Hasil perbandingan performa algoritma yang diperoleh pada penelitian ini menunjukkan bahwa algoritma *AdaBoost* memiliki performa yang lebih baik. Algoritma *AdaBoost* menghasilkan nilai akurasi sebesar 87,46% dan nilai AUC sebesar 0,920. Sedangkan algoritma *decision tree* C4.5 menghasilkan nilai akurasi sebesar 81,58% dan nilai AUC sebesar 0,804. Hasil pengolahan data baru dengan menggunakan model yang terbentuk dari algoritma *AdaBoost* juga memiliki performa yang cukup baik. Performa dari prediksi kelulusan mahasiswa menghasilkan nilai akurasi sebesar 77,78% dan nilai AUC sebesar 0,901. Dari hasil performa prediksi yang didapatkan, menunjukkan bahwa algoritma *AdaBoost* dengan algoritma C4.5 sebagai dasar pemodelan data bisa digunakan untuk mengolah *dataset* mahasiswa.

Pada penelitian ini, juga dilakukan proses simulasi pengukuran akurasi menggunakan beberapa atribut Indeks Prestasi Semester (IPS) mahasiswa. Pengukuran akurasi dilakukan dengan tujuan, untuk mengetahui pada semester berapa pembuatan keputusan prediksi kelulusan mahasiswa dapat dilakukan. Batas minimum akurasi yang dihasilkan untuk mulai membuat keputusan adalah 70,00%. Hasil pengukuran akurasi dengan menggunakan beberapa atribut IPS dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengukuran Akurasi Menggunakan Atribut IPS

Atribut IPS	Akurasi
IPS_1 hingga IPS_2	67,78%
IPS_1 hingga IPS_3	69,63%
IPS_1 hingga IPS_4	72,59%
IPS_1 hingga IPS_5	71,11%
IPS_1 hingga IPS_6	71,11%
IPS_1 hingga IPS_7	71,85%

Dari Tabel 6 dapat diketahui bahwa, pengambilan keputusan prediksi kelulusan mahasiswa dapat dilakukan setelah mahasiswa mendapatkan nilai IPS 4. Artinya, pihak program studi bisa mulai melakukan prediksi kelulusan mahasiswa setelah mahasiswa memiliki nilai IPS 1 hingga IPS 4. Hasil analisis data dan penggalian informasi terhadap *dataset* mahasiswa yang diperoleh pada penelitian ini, bisa digunakan untuk membantu pihak program studi dalam mengambil keputusan. Misalnya, dari salah satu kriteria *rule* yang memprediksi mahasiswa lulus tidak tepat waktu, dapat digunakan sebagai acuan dalam membuat suatu kebijakan. Kebijakan yang dibuat diharapkan dapat mengatasi masalah kelulusan mahasiswa yang tidak tepat waktu.

5. Kesimpulan dan Saran

Berdasarkan hasil penelitian yang diperoleh, dapat disimpulkan bahwa teknik *data mining* yang diterapkan pada penelitian ini yaitu klasifikasi dan penggunaan algoritma *AdaBoost*, dapat digunakan untuk mengolah *dataset* mahasiswa. Hal ini dibuktikan dengan hasil performa dari algoritma *AdaBoost* yang menggunakan algoritma C4.5 sebagai dasar pemodelan data lebih besar daripada algoritma *decision tree* C4.5. Hasil performa algoritma *AdaBoost* memperoleh nilai akurasi sebesar 87,46% dan nilai AUC sebesar 0,920. Sedangkan performa algoritma *decision tree* C4.5 memperoleh nilai akurasi sebesar 81,58% dan nilai AUC sebesar 0,804. Hasil analisis data dan *rule* dari model pohon keputusan yang terbentuk pada penelitian ini, juga dapat digunakan oleh pihak program studi untuk membantu mengambil keputusan. Hasil keputusan yang terbentuk diharapkan dapat menangani masalah mahasiswa yang lulus tidak tepat waktu.

Saran untuk penelitian yang akan datang guna menyempurnakan penelitian ini yaitu, dapat menggunakan algoritma klasifikasi lainnya seperti *K-Nearest Neighbor* (K-NN), *Support Vector Machine* (SVM), *Naïve Bayes*, dan algoritma lainnya. Penelitian yang akan datang juga dapat menggunakan algoritma metode *ensemble* lainnya seperti *Bagging* dan *Stacking*. Atribut

yang digunakan juga bisa lebih beragam seperti ditambahkannya atribut STATUS_BEKERJA dan STATUS_MENIKAH. Penelitian yang akan datang juga dapat meneliti faktor-faktor kuat yang mempengaruhi ketepatan waktu kelulusan mahasiswa. Hasil dari penelitian nantinya dapat dibandingkan dengan hasil dan kinerja dari algoritma pemodelan dan algoritma *ensemble* yang digunakan pada penelitian ini. Berdasarkan saran yang diberikan, diharapkan penelitian yang akan datang bisa mendapatkan hasil yang lebih maksimal dan bermanfaat.

Referensi

- [1] D. A. N. Wulan and S. M. Abdullah, "Prokrastinasi Akademik Dalam Penyelesaian Skripsi," *Jurnal Sosio-Humaniora*, vol. 5, Mei 2014.
- [2] A. M. Rizki, 7 Jalan Mahasiswa. Sukabumi: CV Jejak, 2018.
- [3] S. D. Caesaria, "18 Kampus Swasta Terbaik Indonesia Versi QS AUR 2023, Mana Kampusmu?," [Online]. Tersedia: <https://www.kompas.com/edu/read/2022/11/09/181900471/18-kampus-swasta-terbaik-indonesia-versi-qs-aur-2023-mana-kampusmu-?page=all>. [Diakses: Sep. 12, 2023]
- [4] BAN-PT, "Akreditasi Perguruan Tinggi Kriteria Dan Prosedur IAPT 3.0," Jakarta, 2019. [Online]. Tersedia: https://www.banpt.or.id/wp-content/uploads/2019/09/Lampiran-02-PerBAN-PT-3-2019-Kriteria-dan-Prosedur-IAPT-3_0.pdf. [Diakses: Sep. 12, 2023]
- [5] L. Muflikhah, D. E. Ratnawati, dan R. R. M. Putri, *Data Mining*. Malang: UB Press, 2018.
- [6] D. Kamagi and S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa", *Ultimatics : Jurnal Teknik Informatika*, vol. 6, no. 1, pp. 15-20, Jun. 2014.
- [7] R. T. Yunardi and N. Z. Dina, *Data Mining dan Machine Learning dengan Orange3 Tutorial dan Aplikasinya*. Surabaya: Airlangga University Press, 2022.
- [8] D. Heryana, "Data Mining Untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika UIN Raden Intan Lampung Menggunakan Naive Bayes," Universitas Islam Negeri Raden Intan, Lampung, 2019.
- [9] A. Rohman and A. Ruffyanto, "Implementasi Data Mining Dengan Algoritma Decision Tree C4.5 Untuk Prediksi Kelulusan Mahasiswa di Universitas Pandanaran," *Proceeding SINTAK 2019*, Semarang, 2019, hal 134-139.
- [10] E. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network", *jurnalmatrik*, vol. 20, no. 1, pp. 21–30, Jan. 2019.
- [11] L. A. Rahman Hakim, A. A. Rizal, and D. Ratnasari, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN)," *Jurnal Teknologi Informasi dan Multimedia*, vol. 1, pp. 30–36, Mei 2019.