

Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta

Tuhfatul Habibah Hasibuan^{1*}, Deni Mahdiana²

¹Fakultas Teknik, Teknik Informatika, Institut Teknologi dan Bisnis Swadharma, Jakarta, Indonesia

²Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail: ^{1*}tuhfatulhabibah@swadharma.ac.id, ²deni.mahdiana@budiluhur.ac.id

(* : corresponding author)

Abstrak

Kelulusan tepat waktu merupakan salah satu ciri keberhasilan mahasiswa dalam meraih gelar sarjana. Namun kenyataannya, mahasiswa tidak selalu menyelesaikannya dalam waktu empat tahun. Kelulusan tepat waktu termasuk salah satu komponen pengukuran mutu lembaga pendidikan tinggi. Karena merupakan salah satu kriteria penilaian dalam proses akreditasi program studi dan institusi oleh Badan Akreditasi Perguruan Tinggi (BANPT). Selama tujuh tahun terakhir, mahasiswa yang lulus tepat waktu belum mencapai 50%. Hal ini dapat menyebabkan devaluasi Akreditasi program studi dan akreditasi lembaga pendidikan tinggi. Pada penelitian ini, teknik penambangan data C4.5 digunakan untuk mendeteksi mahasiswa yang terlambat lulus. Algoritma C4.5 ini adalah algoritma terkenal untuk mengklasifikasikan data dengan fitur dan kategori numerik. Pada penelitian ini algoritma C4.5 memprediksi kelulusan dengan nilai Accuracy 75,52% akurat, Precision 75,50% aku, dan Recall 75,50%.

Kata kunci: Decision Tree, C4.5, Prediksi, Kelulusan Mahasiswa, Data Mining.

Abstract

Graduation on time is one of the characteristics of student success in obtaining a bachelor's degree. However, in reality, students do not always complete it within four years. Graduation on time is one component of measuring the quality of higher education institutions. Because it is one of the assessment criteria in the accreditation process of study programs and institutions by the Higher Education Accreditation Board (BANPT). Over the last seven years, students who graduate on time have not reached 50%. This can lead to a devaluation of accreditation of study programs and accreditation of higher education institutions. In this study, the C4.5 data mining technique was used to detect students who graduated late. This C4.5 algorithm is a well-known algorithm for classifying data by numeric features and categories. In this study the C4.5 algorithm predicts graduation with an accuracy score of 75.52% accurate, Precision 75.50% I, and Recall 75.50%.

Keywords: Decision Tree, C4.5, Prediction, Student Graduation, Data Mining.

1. PENDAHULUAN

Penyelesaian kuliah tepat waktu adalah salah satu karakteristik dari pendidikan sarjana. Namun pada kenyataannya, mahasiswa tidak selalu menyelesaikannya dalam waktu empat tahun. Badan Akreditasi Nasional Perguruan Tinggi (BANPT) dalam hal ini lembaga yang mendapatkan mandat untuk mengukur dan menentukan kualitas dan keberlanjutan program Perguruan Tinggi sebagai penyelenggara pendidikan tinggi, dengan mengakreditasi perguruan tinggi merupakan cara mengevaluasi perguruan tinggi. Pada salah satu standar penilaian Badan Akreditasi Nasional Perguruan Tinggi (BANPT) yakni Standar 3 untuk mahasiswa dan lulusan, penyelesaian kuliah tepat waktu termasuk salah satu komponen untuk menilai mutu terakreditasi. Jika tidak seimbang antara tingkat kelulusan dan kemahasiswaan akan mempengaruhi penilaian akreditasi program sarjana dan Institusi.

Apabila dalam jangka waktu kurang dari atau sama dengan empat tahun mahasiswa dapat menyelesaikan studinya di lembaga pendidikan tinggi maka disebut kelulusan tepat waktu, sedangkan apabila dalam jangka waktu lebih dari empat tahun mahasiswa dapat menyelesaikan studinya di lembaga pendidikan tinggi maka disebut kelulusan mahasiswa tidak tepat waktu [1].

Selama tujuh tahun terakhir, mahasiswa yang lulus tepat waktu belum mencapai 50%. Hal

ini dapat menyebabkan devaluasi Akreditasi program studi dan akreditasi perguruan tinggi.

Dalam penelitian ini, digunakan teknik data mining untuk mendeteksi lebih dini mahasiswa yang lulus terlambat, sehingga Program Studi dan Universitas dapat menyusun rencana strategis untuk penanganan yang lebih cepat dalam meningkatkan kelulusan mahasiswa tepat waktu, sehingga dapat mengurangi penumpukan mahasiswa, dan meningkatkan serta mempertahankan nilai akreditasi Program Studi dan Universitas.

Penelitian-penelitian terdahulu yang terkait dengan prediksi kelulusan mahasiswa tentang “Prediksi Kelulusan Mahasiswa Pada Perguruan Tinggi Kabupaten Majalengka Berbasis *Knowledge Based System*”. Jumlah atribut yang digunakan adalah 7 parameter atribut yang terdiri dari 6 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri Jenis Kelamin, Kota Asal, Agama, IPK, Fakultas, Jurusan Asal Sekolah, dan 1 atribut kelas yaitu dan label yang berisi klasifikasi ketepatan lulus mahasiswa yaitu Tepat Waktu dan Tidak Tepat Waktu. Metode yang digunakan adalah *Back Propagation Neural Network* dan *K-Nearest Neighbor*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Back Propagation Neural Network* dan *K-Nearest Neighbor* dalam memprediksi kelulusan mahasiswa secara berurutan 97,22% dan 97,90% [2].

Penelitian berikutnya tentang “Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree”. Jumlah atribut yang digunakan adalah 9 parameter atribut yang terdiri dari 8 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Nim, Tgl. Lahir, Usia, Jenis Kelamin, IPS-1, IPS-2, IPS-3, IPS-4, dan 1 atribut kelas yaitu label yang berisi klasifikasi ketepatan lulus mahasiswa yaitu lulus tepat waktu dan lulus tidak tepat waktu. Metode yang digunakan adalah C4.5. Dari penelitian ini diperoleh tingkat akurasi algoritma C4.5 dalam memprediksi kelulusan mahasiswa sebesar 91,51% [3].

Penelitian lain yang terkait tentang “Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta”. Jumlah atribut yang digunakan adalah 13 parameter atribut yang terdiri dari 12 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Jenis Kelamin, Jenis seleksi, Pendapatan ayah, Pendidikan ibu, IP semester 1, IP semester 2, IP semester 3, IP semester 4, SKS 1, SKS 2, SKS 3, SKS 4, dan 1 atribut kelas yaitu status yang berisi klasifikasi ketepatan lulus mahasiswa yaitu tepat dan terlambat. Metode yang digunakan adalah *Naïve Bayes*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Naïve Bayes* dalam memprediksi kelulusan mahasiswa sebesar 80,72% [4].

Penelitian lainnya tentang “*Prediction of Graduation Delay Based on Student Characteristics and Performance*”. Jumlah atribut yang digunakan adalah 12 parameter atribut yang terdiri dari 11 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri *High School GPA (First and Second Semester GPA)*, *Number of Credit Hours taken by a Student up to Second Semester*, *Enhanced ACT Composite*, *Gender*, *Original Place of Residence*, *HS units of English Composition*, *High School State: ACT Record*, *Ethnicity*, *HS Lab Science units*, *Recent High School Graduate*, dan 1 atribut kelas yaitu *Target Class* yang berisi klasifikasi keterlambatan lulus mahasiswa yaitu *no delay (Class 1)*, *one year delay (Class 2)*, and *two or more years delay (Class 3)*. Metode yang digunakan adalah *Support Vector Machines (SVMs)*, *Gaussian Processes (GPs)*, and *Deep Boltzmann Machines (DBMs)*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Support Vector Machines (SVMs)*, *Gaussian Processes (GPs)*, and *Deep Boltzmann Machines (DBMs)* dalam memprediksi kelulusan mahasiswa secara berurutan 68%, 72% dan 75% [5].

Penelitian berikutnya juga telah dilakukan oleh Testiana tentang “Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah”. Jumlah atribut yang digunakan adalah 8 parameter atribut yang terdiri dari 7 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari NPM, Nama Mahasiswa, Jenjang Pendidikan, Status, Jenis Kelamin, IPK (Indek Prestasi Kumulatif), IP semester, dan 1 atribut kelas yaitu kelulusan tepat waktu yang berisi klasifikasi ketepatan lulus mahasiswa yaitu lulus tepat waktu dan lulus terlambat. Metode yang digunakan adalah *Naïve Bayes*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Naïve Bayes* dalam memprediksi kelulusan mahasiswa sebesar 82.08% [6].

Penelitian berikutnya tentang “Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB”. Jumlah atribut yang digunakan adalah 10 parameter atribut yang terdiri dari 9 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari jenis kelamin, status perkawinan, status penerimaan, status perguruan tinggi S1, sumber biaya pendidikan S2, kelompok instansi, program studi S2, usia masuk S2, IPK S1, dan 1 atribut kelas yaitu Status Mahasiswa yang berisi klasifikasi ketepatan lulus mahasiswa yaitu lulus dan tidak lulus (drop out dan mengundurkan diri). Metode yang digunakan adalah *Rotation Forest*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Rotation Forest* dalam memprediksi kelulusan mahasiswa sebelum SMOTE (*Synthetic Minority Oversampling Technique*) 89%, setelah SMOTE 71,86 % [7].

Penelitian berikutnya tentang “Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C45”. Jumlah atribut yang digunakan adalah 10 parameter atribut yang terdiri dari 9 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Id, IPK, Jumlah_sks, nilai mata kuliah terpilih 1, nilai mata kuliah terpilih 2, nilai mata kuliah terpilih 3, nilai mata kuliah terpilih 4, nilai mata kuliah terpilih 5, nilai mata kuliah terpilih 6, dan 1 atribut kelas yaitu kategori yang berisi klasifikasi ketepatan lulus mahasiswa yaitu tepat dan tidak tepat. Metode yang digunakan adalah C4.5. Dari penelitian ini diperoleh tingkat akurasi algoritma C4.5 dalam memprediksi kelulusan mahasiswa sebesar 73.99% [8].

Penelitian berikutnya tentang “Implementasi Teknik Seleksi Fitur Forward Selection Pada Algoritma Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa Politeknik Indonusa Surakarta”. Jumlah atribut yang digunakan adalah 4 parameter atribut yang terdiri dari 3 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Jenis Kelamin, Kelas, Indeks Prestasi Kumulatif, dan 1 atribut kelas yaitu lama studi yang berisi klasifikasi ketepatan lulus mahasiswa yaitu tepat waktu dan terlambat. Metode yang digunakan adalah *K-Nearest Neighbor*. Dari penelitian ini diperoleh tingkat akurasi algoritma *K-Nearest Neighbor* dalam memprediksi kelulusan mahasiswa 59,52% (tanpa seleksi fitur forward selection) dan 58,19% (dengan seleksi fitur forward selection) [9].

Penelitian berikutnya tentang “Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa”. Jumlah atribut yang digunakan adalah 8 parameter atribut yang terdiri dari 7 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Jurusan, Jenis Kelamin, Umur, IP 1, IP 2, IP 3, IP 4, dan 1 atribut kelas yaitu Status yang berisi klasifikasi kelulusan mahasiswa yaitu tepat dan terlambat. Metode yang digunakan adalah *Neural Network*, *K-Nearest Neighbor* dan *Decision Tree*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Neural Network*, *K-Nearest Neighbor* dan *Decision Tree* dalam memprediksi kelulusan mahasiswa secara berurutan 87,32% ,85.37% dan 83,66% [10].

Penelitian berikutnya tentang “Teknik Bagging Dan Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa”. Jumlah atribut yang digunakan adalah 10 parameter atribut yang terdiri dari 9 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari NIM, IP semester satu sampai empat, jumlah sks yang diambil pada semester satu sampai empat, dan 1 atribut kelas yaitu lama studi yang berisi klasifikasi kelulusan mahasiswa yaitu tepat waktu dan tidak tepat waktu. Metode yang digunakan adalah *Classification and Regression Tree (CART)*. Dari penelitian ini diperoleh tingkat akurasi algoritma *CART* dalam memprediksi kelulusan mahasiswa sebesar 79,592%, Algoritma *CART* dengan teknik *bagging* 81,633%, dan Algoritma *CART* dengan teknik *boosting* 87,755% [11].

Penelitian berikutnya tentang “*Application of Data mining to Prediction of Timeliness Graduation of Students Using C.4. 5 Decision Tree (A Case Study)*”. Jumlah atribut yang digunakan adalah 10 parameter atribut yang terdiri dari 9 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari *Gender*, *Program Studi*, *Type_school*, *Major_SLTA*, *Region*, *IPSmt 1*, *IPSmt 2*, *IPSmt 3*, *IPSmt 4*, dan 1 atribut kelas yaitu *graduation* yang berisi klasifikasi *graduation*/kelulusan mahasiswa yaitu *on time* dan *not on time*. Metode yang digunakan adalah C4.5. Dari penelitian ini diperoleh tingkat akurasi algoritma C4.5 dalam memprediksi kelulusan mahasiswa sebesar 89,82% [12].

Penelitian berikutnya tentang “Analisis C4.5 Pada Klasifikasi CART Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka”. Jumlah atribut yang digunakan adalah 8 parameter atribut yang terdiri dari 7 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari jenis kelamin, jurusan/prodi, IPK, pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua, status mahasiswa, dan 1 atribut kelas yaitu status kelulusan yang berisi klasifikasi status kelulusan mahasiswa yaitu tepat waktu dan tidak tepat waktu. Metode yang digunakan adalah *Classification and Regression Tree* (CART) dan C4.5. Dari penelitian ini diperoleh C4.5 mampu meningkatkan akurasi klasifikasi ketidaktepatan waktu kelulusan mahasiswa yang mencapai konvergen dengan prediksi klasifikasi mencapai 93.23% [13].

Penelitian berikutnya tentang “Prediksi Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes (Studi Kasus 5 PTS di Banda Aceh)”. Jumlah atribut yang digunakan adalah 5 parameter atribut yang terdiri dari 4 atribut prediksi dan 1 atribut kelas. Atribut prediksi terdiri dari Angkatan, IPK, lama studi, PTS dan 1 atribut kelas yaitu tepat waktu yang berisi klasifikasi ketepatan lulus mahasiswa yaitu “ya” tepat waktu dan “tidak” tidak tepat waktu. Metode yang digunakan adalah *Naive Bayes*. Dari penelitian ini diperoleh tingkat akurasi algoritma *Naive Bayes* dalam memprediksi kelulusan mahasiswa sebesar 84% [14].

Penelitian pendahuluan dengan menggunakan software WEKA versi 3.8, sebanyak 1442 data diuji dengan persentase split 70:30 dengan berbagai algoritma klasifikasi, yaitu *K-Nearest Neighbor*, C4.5, dan C4.5. diperoleh akurasi terbaik oleh C4.5. Penelitian pendahuluan dengan tools WEKA versi 3.8 tersaji pada tabel 1

Tabel 1. Penelitian Terdahulu dengan WEKA Versi 3.8

No.	Algoritma	Akurasi		Precision	Recall	Confusion Matrix			
		Correct	Incorrect			TP	FP	FN	TN
1	IBk (KNN)	62.36%	37.64%	62.40%	62.40%	117	93	70	153
2	J48 (C4.5)	73.67%	26.33%	73.70%	73.70%	146	64	50	173
3	Random Forest	70.44%	29.56%	70.70%	70.40%	131	79	49	174

Berdasarkan penelitian terdahulu tersebut, maka dipilih C4.5 untuk memprediksi kelulusan mahasiswa tepat waktu.

Diharapkan dengan pemanfaatan algoritma C4.5 dan penggunaan 10 atribut prediksi dan 1 atribut kelas untuk prediksi kelulusan mahasiswa dapat menghasilkan akurasi yang lebih tinggi dari penelitian sebelumnya.

2. METODE PENELITIAN

2.1 DATA MINING

Konsep data mining adalah menggunakan perangkat keras yang paling kuat, sistem pemrograman yang paling handal dan algoritma yang paling efisien untuk menyelesaikan masalah dalam ilmiah, perdagangan, perawatan kesehatan, pemerintah, humaniora, dan banyak bidang usaha manusia lainnya [5].

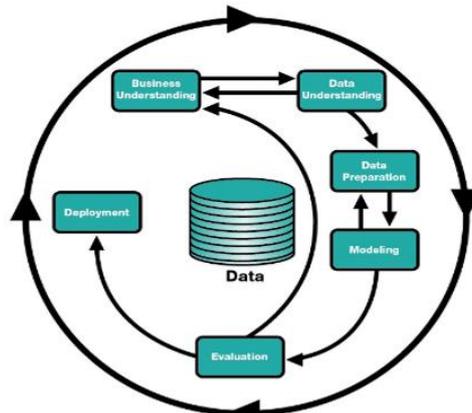
Penambahan data mengacu pada metode analisis bisnis yang melampaui jumlah, teknik deskriptif, pelaporan, dan metode yang didasarkan pada aturan bisnis. Metode penambahan data, dengan kekuatan dan otomatisasinya, memiliki kemampuan untuk mengatasi sejumlah besar data dan mengekstraksi nilai.” [6].

2.2 CRISP DM

Metodologi CRISP-DM dijelaskan dalam hal model proses hirarkis, yang terdiri dari serangkaian tugas yang dijelaskan pada empat tingkat abstraksi (dari umum ke spesifik [15]).

Urutan langkah-langkahnya tidak kaku. Itu selalu diperlukan untuk beralih bolak-balik di antara tahapan yang berbeda. Hasil dari setiap langkah menentukan langkah atau tugas spesifik mana dalam satu langkah yang harus diselesaikan selanjutnya. Panah menunjukkan ketergantungan yang paling penting dan paling sering di antara langkah-langkah tersebut [15].

Model proses penambangan data saat ini memberikan gambaran tentang siklus hidup proyek penambangan data. Siklus hidup proyek data mining terdiri dari beberapa *fase*, yang ditunjukkan pada Gambar 1 [15].



Gambar 1. CRISP DM

Dari gambar 1 menunjukkan 6 tahap siklus hidup data mining yaitu; *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *development*.

2.3 Feature Selection

Seleksi fitur adalah salah satu teknik penambangan data yang umum digunakan pada tahap *pre-processing*. Metode ini digunakan untuk mereduksi kompleksitas atribut yang harus dikelola dalam pemrosesan dan analisis. Metode ini digunakan untuk menemukan subset fitur yang paling menonjol dari kumpulan data nilai mahasiswa. Seleksi fitur sering digunakan untuk mereduksi dimensi model. Pemilihan fitur membantu mereduksi fitur domain dan mengeliminasi fitur yang berlebihan [16].

Metode *Correlation Feature Selection* (CFS) adalah metode berbasis korelasi yang populer untuk menyeleksi fitur terpenting dari sebuah dataset. Metode ini menghitung korelasi antara setiap atribut dengan variabel hasil, kemudian memilih atribut dengan nilai korelasi sedang hingga tinggi (mendekati 1) dan mengeliminasi atribut yang nilai korelasi rendah (mendekati 0). CFS memanfaatkan kekuatan prediktif dan korelasi silang fitur untuk mencari serangkaian fitur yang baik. Percobaan dengan kumpulan data diskrit dan kontinu memperlihatkan bahwa CFS mampu secara dramatis mereduksi ukuran kumpulan data sembari mempertahankan atau meningkatkan kinerja algoritma pembelajaran [16].

2.4 Penanganan Missing Data

Nilai yang hilang (*missing value*) memengaruhi pelatihan dan prediksi. Data pelatihan: Berisi nilai "tidak dikenal", Prediksi: Input pada waktu prediksi berisi nilai "tidak dikenal"[17]. Strategi penanganan *missing data* sebagai berikut:

- 1) Pemurnian dengan melewati (*Purification by skipping*)
 - a. Pemurnian dengan melewati / menghapus,
 - b. Abaikan titik data dengan nilai yang hilang,
 - c. Abaikan fitur dengan nilai yang hilang.

- 2) Pemurnian dengan imputasi (*Purification by imputing*)
 - a. Fitur kategorikal menggunakan mode: Nilai paling populer (mode) x_i yang tidak hilang,
 - b. Fitur numerik menggunakan rata-rata atau median: Nilai rata-rata atau median dari x_i yang tidak hilang.
- 3) Sesuaikan algoritma pembelajaran agar kuat dengan nilai yang hilang (*Adapt learning algorithm to be robust to missing values*) [17].

2.5 ALGORITMA C.45

Algoritma C4.5 merupakan algoritma yang terkenal digunakan untuk mengelompokkan data dengan ciri khas numerik dan kategori. Proses pengelompokkan menghasilkan rule yang dapat dimanfaatkan untuk meramal nilai atribut khas diskrit dari record baru. Algoritma C4.5 juga merupakan algoritma ID3 yang dikembangkan, yang dirancang untuk mengatasi data yang hilang, untuk mengatasi data yang terus menerus dan terpotong [18].

Secara umum, algoritma C4.5 untuk membangun decision tree terlihat seperti ini:

- 1) Pilih atribut sebagai root.
- 2) Buat cabang untuk setiap nilai.
- 3) Distribusikan kasus pada cabang.
- 4) Ulangi proses buat setiap cabang hingga seluruh instance pada cabang mempunyai kelas yg sama.

Nilai gain tertinggi dari atribut yang ada, digunakan untuk pemilihan root atribut sesuai dengan Rumus menurut Persamaan [17] digunakan untuk menghitung gain.

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_i \text{Entropy (S}_i) \quad (1)$$

Dimana :

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

2.6 Confusion Matrix

Dalam penerapannya, pada umumnya pengukuran akurasi bersumber dari kerangka kebingungan, juga disebut matriks klasifikasi. Kerangka ini menjelaskan kelompok benar dan salah yang dibuat oleh pengklasifikasi untuk kumpulan data tertentu. Baris dan kolom dari matriks kebingungan masing-masing sesuai dengan kelas yang diprediksi dan benar (aktual). Arti setiap sel dari matriks kebingungan ditunjukkan pada Gambar 2 [19].

		Actual Class	
		C ₁	C ₂
Predicted Class	C ₁	$n_{1,1}$ = number of C ₁ records classified correctly	$n_{2,1}$ = number of C ₂ records classified incorrectly as C ₁
	C ₂	$n_{1,2}$ = number of C ₁ records classified incorrectly as C ₂	$n_{2,2}$ = number of C ₂ records classified correctly

Gambar 2. Matriks Kebingungan : Makna Setiap Sel

Dari gambar 2 terlihat kerangka dari confusion matriks yang menjelaskan kelompok benar dan salah tergambar dalam baris dan kolom sesuai dengan kelas yang di prediksi dan yang actual.

Metodologi penelitian berkaitan dengan tahapan CRISP DM yaitu Wawasan bisnis (*Business Understanding*), wawasan data (*Data Understanding*), persiapan data (*Data Preparation*), pemodelan (*Modelling*), evaluasi (*Evaluation*), dan implementasi (*Deployment*).

Metode pengambilan sampel adalah *Stratified Random Sampling* dan memakai data akademik utama berdasarkan mahasiswa tahun 2012-2014. Pengumpulan data dilakukan pada tahun 2018. Terdapat 4205 baris data, langkah-langkah penelitian ini adalah:

1) Memahami bisnis (*Business Understanding*)

Memahami tujuan dan persyaratan proyek dari perspektif bisnis, dan kemudian mengubah informasi ini menjadi definisi masalah data mining dan rencana awal untuk mencapai tujuan tersebut merupakan fokus utama pada tahap awal ini.

2) Memahami data (*Data Understanding*)

Untuk memahami data, mulailah dengan pengumpulan data awal dan kemudian beralih ke aktivitas yang memproses data, mengidentifikasi masalah kualitas data, mendapatkan wawasan awal dari data, dan/atau mengidentifikasi subkumpulan data yang menjadi minat tertentu. Membuat asumsi atau hipotesis tentang informasi yang tidak tampak.

3) Persiapan data (*Data Preparation*)

Semua operasi yang diperlukan untuk membuat kumpulan data akhir (data untuk alat pemodelan) dari data mentah asli dilakukan pada fase persiapan data. Mencakup tugas pemrosesan data sering dilakukan berulang kali dan tanpa urutan tertentu. Tujuan mencakup tabel, kumpulan data, dan pemilihan atribut serta transformasi dan pembersihan data untuk alat pemodelan.

Secara singkat, terdapat dua fungsi utama yang dilakukan dalam fase ini yaitu:

a) Persiapan Dataset

Menyiapkan data akademik tahun 2012-2014 untuk mahasiswa tahun 2018 yang didapatkan dari Universitas Islam Negeri Syarif Hidayatullah Jakarta melalui Pusat Informasi dan Data (PUSTIPANDA) yang berisi informasi: jurusan, jenis_seleksi, tahun_angkatan, kelamin, propinsi, ipsmt_1, ipsmt_2, ipsmt_3, ipsmt_4, ipsmt_5, ipsmt_6, ipsmt_7, ipsmt_8, ipsmt_9, ipsmt_10, ipsmt_11, ipsmt_12, ipsmt_13, ipsmt_14, status, jenis_sekolah, jurusan_sekolah

b) Melakukan *feature selection*.

Setelah menghasilkan kumpulan data Pemilihan atribut dilakukan dengan menggunakan teknik *Correlation Attribute Eval* Menggunakan metode penelitian Ranker. Atribut yang dipilih digunakan Ramalan/prediksi.

4) Pemodelan (*Modelling*)

Metode pemodelan yang berbeda dipilih dan diterapkan serta parameter dikalibrasi ke nilai optimal dilakukan pada tahapan ini.

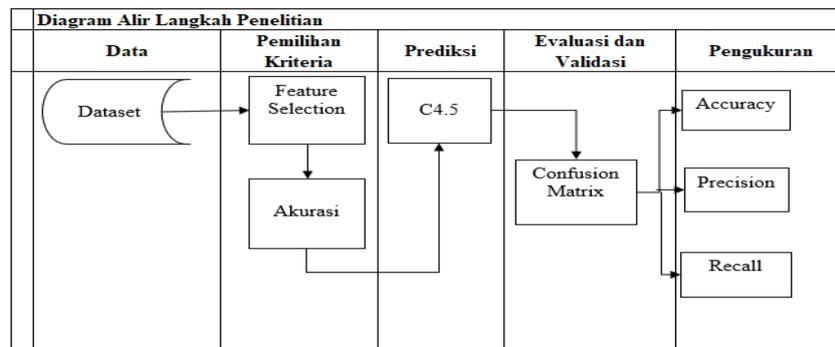
5) Evaluasi (*Evaluation*)

Model (atau model) yang tampak berkualitas tinggi dari perspektif analisis data dibangun di tahapan ini. Mengevaluasi dan mempelajari secara detail tahapan pembuatannya adalah sangat penting untuk dapat mengetahui apakah model tersebut benar-benar mencapai tujuan bisnis, sebelum akhirnya menggunakan model. Mengetahui ada tidaknya masalah bisnis penting yang tidak ditangani dengan cukup serius di akhir fase merupakan tujuan utama pada fase ini. Pada fase ini keputusan harus dibuat tentang penggunaan data mining.

6) Implementasi (*Deployment*)

Di tahap ini dilakukan pengujian kepada user untuk memastikan yang hasilnya sesuai dengan kebutuhan dan harapan user.

Untuk flowchart langkah penelitian dapat dilihat pada gambar 3.



Gambar 3. Flowchart Langkah Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Terdapat data yang tidak signifikan antara mahasiswa yang masuk dan lulus. Hal ini terlihat pada data akreditasi yang tersedia di data.uinjkt.ac.id, diperoleh jumlah mahasiswa S1 yang lulus tepat waktu dalam 7 tahun terakhir tidak mencapai 50 persen. Penelitian ini memprediksi tingkat kelulusan tepat waktu mahasiswa UIN Syarif Hidayatullah Jakarta dengan menggunakan metode C4.5.

3.2 Data Understanding

Untuk prediksi atau ramalan kelulusan, peneliti menggunakan data akademik mahasiswa 2012-2014 didapatkan dari Universitas Islam Negeri Syarif Hidayatullah Jakarta melalui Pusat Informasi dan Data (PUSTIPANDA).

Informasi yang diperoleh berupa dua dokumen Microsoft Excel Dokumen pertama berisi informasi mahasiswa tahun 2012-2016 terdiri dari 4205 baris data dan 23 kolom yaitu jurusan, jenis_seleksi, tahun_angkatan, nim, nama, kelamin, propinsi, ipsmt_1 sampai ipsmt_14, ipk, dan status. Dokumen kedua berisi data mahasiswa angkatan 2012-2016 yang terdiri dari 4205 baris data dan 9 kolom yaitu:tahun_angkatan,jurusan,jenis_seleksi, nim, nama, kelamin, status, jenis_sekolah dan jurusan_sekolah.

Kemudian kedua dokumen Format Microsoft Excel digabungkan menjadi satu spreadsheet komposit 4205 catatan dan 32 kolom. Berikut data awal mahasiswa:

	A	B	C	D	E	F	G	H	I
1		jenis_seleksi	tahun_angkatan	nim	nama	kelamin	propinsi	jenis_sekolah	jurusan_sekolah
2	Agribisnis	SNMPTN	2012	1.11209E+12	AGUS HERLAMBAH	Laki-laki	DKI Jakarta	SMAN	SMAN IPA
3	Agribisnis	SNMPTN	2012	1.11209E+12	INDRI SULISTIANGINGSH	Perempuan	Jawa Barat	SMAN	SMAN IPA
4	Agribisnis	SNMPTN	2012	1.11209E+12	RANI SEPTIANI	Perempuan			
5	Akuntansi	SNMPTN	2012	1.11208E+12	LIDIYNA KHOIRUL FATHI	Perempuan	Banten	Pesantren	IPS
6	Akuntansi	SNMPTN	2012	1.11208E+12	LAILA RAMADIANA	Perempuan	DKI Jakarta	SMAS	SMAS IPS
7	Akuntansi	SNMPTN	2012	1.11208E+12	NOVA YULIANTI	Perempuan	Banten	MAN	MAN IPS
8	Bahasa dan Sastra Arab	SNMPTN	2012	1.11202E+12	KHOIRUM MILLATIN	Perempuan	Banten	Pesantren	IPS
9	Biologi	SNMPTN	2012	1.1121E+12	ANNISA AMALIA	Perempuan	DKI Jakarta	SMAS	SMAS IPA
10	Ekonomi Pembangunan	SNMPTN	2012	1.11208E+12	SANDRA DESTIAWATI	Perempuan	Jawa Barat	MAS	MAS IPS
11	Ekonomi Pembangunan	SNMPTN	2012	1.11208E+12	RAFIDA ZAHRUDDIN	Perempuan			
12	Farmasi	SNMPTN	2012	1.1121E+12	AMELIA GUSTIN	Perempuan	Bengkulu	SMAN	SMAN IPA
13	Farmasi	SNMPTN	2012	1.1121E+12	SITI WINDI HARIANI	Perempuan	Banten	SMAN	SMAN IPA
14	Farmasi	SNMPTN	2012	1.1121E+12	MOETHIA	Perempuan	Kalimantan Barat	SMAS	SMAS IPA
15	Farmasi	SNMPTN	2012	1.1121E+12	CHALILA DELI GAYO	Perempuan	Sumatera Utara	MAN	MAN IPA
16	Fisika	SNMPTN	2012	1.1121E+12	NURUL FADILLAH	Perempuan			
17	Hukum Keluarga (Ahwal Syakhshiyah)	SNMPTN	2012	1.11204E+12	SODIKIN FEBRIANTO	Laki-laki	Jawa Tengah	MAS	MAS Agama
18	Hukum Keluarga (Ahwal Syakhshiyah)	SNMPTN	2012	1.11204E+12	NAILINNAFIS	Perempuan	Jawa Timur	MAN	MAN Agama
19	Ilmu Hubungan Internasional	SNMPTN	2012	1.11211E+12	HIKA DAYAMA SITORUS	Perempuan	Sumatera Utara	Pesantren	IPS
20	Ilmu Hubungan Internasional	SNMPTN	2012	1.11211E+12	DEDE ABDURACHMAN	Laki-laki	DKI Jakarta	SMAN	SMAN IPS
21	Ilmu Hubungan Internasional	SNMPTN	2012	1.11211E+12	REFDYNNAL JUMAR	Laki-laki	Jawa Barat	SMAS	SMAS IPS

Gambar 4. Data Awal Mahasiswa

Dari gambar 4 terlihat ada beberapa data yang masih belum lengkap atau masih ada yang kosong Seperti propinsi, jenis_sekolah, dan jurusan_sekolah

3.3 Data Preparation

Informasi akademik mahasiswa tahun ajaran 2012-2014 yang terdapat pada 4205 baris dan 32 kolom dari Pusat Informasi dan Data (PUSTIPANDA) Universitas Islam Negeri Syarif Hidayatullah Jakarta tidak dapat langsung dijadikan data dalam penelitian ini. Selanjutnya dilakukan teknik pengolahan data sebagai berikut:

1) Data Integration

Pada tahap awal ini, menggabungkan dua dokumen Microsoft Excel menjadi satu tabel yang terdiri dari 4205 record dan 32 kolom yang berisi: jurusan, jenis_seleksi, tahun_angkatan, nim, nama, kelamin, propinsi, ipsmt_1, ipsmt_2, ipsmt_3, ipsmt_4, ipsmt_5, ipsmt_6, ipsmt_7, ipsmt_8, ipsmt_9, ipsmt_10, ipsmt_11, ipsmt_12, ipsmt_13, ipsmt_14, ipk, status, tahun_angkatan, jurusan, jenis_seleksi, nim, nama, kelamin, status, jenis_sekolah, jurusan_sekolah.

2) Data Reduction

Pada tahap ini dilakukan *data preprocessing* dengan menghilangkan duplikasi kolom (atribut) dan data serta menghilangkan data atau *record* yang tidak diperlukan dalam penelitian ini, yaitu : Menghapus 1 kolom jurusan; Menghapus 1 kolom jenis_seleksi.; Menghapus 1 kolom tahun_angkatan; Menghapus kolom nim; Menghapus kolom nama; Menghapus kolom kelamin; Menghapus kolom ipsmt_8; Menghapus kolom ipsmt_9; Menghapus kolom ipsmt_10; Menghapus kolom ipsmt_11; Menghapus kolom ipsmt_12; Menghapus kolom ipsmt_13; Menghapus kolom ipsmt_14; Menghapus kolom ipk; Menghapus kolom status; Menghapus record mahasiswa yang berstatus Drop Out sebanyak 483 record.; Menghapus record mahasiswa yang berstatus mengundurkan diri sebanyak 3 record; Menghapus record data mahasiswa angkatan 2015 sebanyak 1106 record dan angkatan 2016 sebanyak 1171 record.

3) Data cleaning

Kemudian untuk tahap ini dilakukan data preprocessing dengan melakukan penanganan *missing value* dengan imputasi, yaitu

- a) Mengisi data jenis_sekolah yang kosong sebanyak 59 record
- b) Mengisi data jurusan_sekolah yang kosong sebanyak 60 record.

Karena datanya bersifat kategorikal, maka pengisian dilakukan dengan data yang populer.

Untuk pengisian data jenis_sekolah mempertimbangkan data Angkatan, Jenis Seleksi, Jenis Kelamin, Propinsi, IPS, dan Status

Sedangkan untuk pengisian data jurusan_sekolah mempertimbangkan Angkatan, Jenis Seleksi, Jenis Kelamin, Propinsi, Jenis Sekolah, IPS, Status.

4) Data Transformation

Pada tahap ini dilakukan *Attribute/Feature Construction* yang diperlukan yaitu dengan membuat 1 atribut baru yaitu prediksi yang merupakan label kelas yang berisi kategori "Tepat Waktu" dan "Terlambat".

Setelah dilakukan transformasi data terbentuk dataset dengan 15 atribut yang terdiri dari 1442. Dengan perbandingan jumlah tepat waktu dengan terlambat adalah 671:771. Atribut-atribut tersebut adalah : jurusan; jenis_seleksi; tahun_angkatan; kelamin; propinsi; jenis_sekolah; jurusan_sekolah; ipsmt_1; ipsmt_2; ipsmt_3; ipsmt_4; ipsmt_5; ipsmt_6; ipsmt_7; Prediksi.

5) Feature Selection

Setelah terbentuk dataset dengan 15 atribut, kemudian dilakukan seleksi atribut dengan menggunakan teknik *CorrelationAttributeEval* dengan *Ranker Research Method*. Yang menghasilkan atribut terpilih yang menghasilkan nilai optimal untuk memprediksi kelulusan

mahasiswa. Atribut tersebut adalah : jurusan; jenis_seleksi; tahun_angkatan;kelamin; propinsi; jenis_sekolah; jurusan_sekolah; ipsmt_1; ipsmt_2; ipsmt_3

Setelah terbentuk dataset sejumlah 1442 record dengan 11 atribut terpilih, kemudian dilakukan kembali uji coba dan transformasi data untuk melihat performa akurasi terbaik. Semua uji coba dilakukan dengan persentase split 70 % data training (1009 record) dan 30 % data testing (433 record). Uji coba dilakukan menggunakan tools WEKA versi 3.8. dengan 11 atribut.

Beberapa uji coba dan transformasi data yang dilakukan adalah :

- 1). Membuat kategori baru untuk atribut jurusan.

Dari 51 kategori menjadi 2 kategori yaitu agama dan umum.

Hasil uji coba dari transformasi atribut jurusan menjadi 2 kategori dapat dilihat pada tabel 2.

Tabel 2. Hasil Uji Coba 11 Atribut dengan Transformasi jurusan

No.	Algoritma	Akurasi		Precision	Recall	Confusion Matrix			
		Correct	Incorrect			TP	FP	FN	TN
1	IBk (KNN)	60.74%	39.26%	60.70%	60.70%	115	95	75	148
2	J48 (C4.5)	73.90%	26.10%	74.00%	73.90%	145	65	48	175
3	Random Forest	68.82%	31.18%	69.00%	68.80%	129	81	54	169

Dari tabel 2 hasil uji coba dengan transformasi data jurusan diperoleh akurasi algoritma C4.5 tertinggi yaitu 73,90%, Precision 74%, dan Recall 73,90 %.

- 2). Membuat kategori baru untuk atribut jurusan, jenis_sekolah, dan jurusan_sekolah.

Untuk atribut jurusan tetap 2 kategori sebagaimana transformasi pada poin 1, yaitu agama dan jurusan. Sedangkan untuk atribut jenis sekolah, dari 7 kategori menjadi 6 kategori yaitu MAN, MAS, Pesantren, SMAN, SMAS, dan Lainnya.

Hasil uji coba dari transformasi atribut jurusan, jenis_sekolah, dan jurusan_sekolah dapat dilihat pada tabel 3.

Tabel 3. Hasil Uji Coba 11 Atribut dengan Transformasi jurusan, jenis_sekolah, dan jurusan_sekolah

No.	Algoritma	Akurasi		Precision	Recall	Confusion Matrix			
		Correct	Incorrect			TP	FP	FN	TN
1	IBk (KNN)	61.43%	39.26%	61.40%	61.40%	116	94	73	150
2	J48 (C4.5)	75.52%	24.48%	75.50%	75.50%	156	54	52	171
3	Random Forest	70.67%	29.33%	70.70%	70.70%	138	72	55	168

Dari tabel 3 hasil uji coba dengan transformasi data jurusan, jenis_sekolah, dan jurusan_sekolah diperoleh akurasi algoritma C4.5 tertinggi yaitu 75,52%, Precision 75.50%, dan Recall 75.50%.

- 3). Melakukan uji coba dengan transformasi atribut jenis_sekolah dengan kategori sebagaimana pada tabel 3.

Hasil uji coba dari transformasi atribut jenis_sekolah dapat dilihat pada tabel 4.

Tabel 4. Hasil Uji Coba 11 Atribut dengan Transformasi jenis_sekolah

No.	Algoritma	Akurasi		Precision	Recall	Confusion Matrix			
		Correct	Incorrect			TP	FP	FN	TN
1	IBk (KNN)	62.36%	37.64%	62.40%	62.40%	117	93	70	153

2	J48 (C4.5)	73.67%	26.33%	73.70%	73.70%	146	64	50	173
3	Random Forest	70.67%	29.33%	70.90%	70.70%	132	78	49	174

Dari tabel 4 hasil uji coba dengan transformasi data jenis_sekolah diperoleh akurasi algoritma C4.5 tertinggi yaitu 73,67%, *Precision* 73.70%, dan *Recall* 73.70%.

Dari uji coba yang telah dilakukan, dengan *feature selection* dan transformasi data dapat diketahui bahwa hasil akurasi algoritma C4.5 dapat meningkat dan lebih tinggi dari algoritma KNN dan Random Forest dengan hasil akurasi maksimal 75,52%, *Precision* 75.50%, dan *Recall* 75.50% .Oleh karena itu, maka pada penelitian ini dipilih algoritma C4.5 untuk memprediksi kelulusan mahasiswa dengan menggunakan 10 atribut untuk prediksi dan 1 label kelas.

3.4 Modelling

Setelah melakukan berbagai uji coba maka dipilih metode C4.5 dalam pembuatan model untuk memprediksi kelulusan mahasiswa. Beberapa contoh rule yang diperoleh sebagai berikut:

```
tahun_angkatan <= 2013
| ipsmt_1 <= 2.86
| | propinsi = DKI Jakarta
| | | jurusan_sekolah = IPA: Terlambat (7.0)
| | | jurusan_sekolah = IPS
| | | | ipsmt_2 <= 2.95: Terlambat (2.0)
| | | | ipsmt_2 > 2.95: Tepat Waktu (2.0)
| | | | jurusan_sekolah = Agama: Tepat Waktu (2.0)
| | | | jurusan_sekolah = Lainnya: Terlambat (0.0)
| | propinsi = Jawa Barat
| | | jenis_sekolah = SMAN
| | | | ipsmt_2 <= 3.43: Terlambat (7.0)
Number of Leaves :    139
Size of the tree :    179
```

Pembentukan *tree* dilakukan berdasarkan perhitungan entropy dan gain masing-masing atribut sehingga setiap atribut mempunyai kelas. Berikut ini contoh pembentukan *tree*:

1) Perhitungan *Entropy*

$$\begin{aligned}
 \text{Entropy Total} &= (-47/101 * (\text{LOG}_2(47/101))) + (-54/101 * \text{LOG}_2(54/101)) \\
 &= (-0.46535 * \text{LOG}_2(0.46535)) + (-0.53465 * \text{LOG}_2(0.53465)) \\
 &= (-0.46535 * -1.10362) + (-0.53465 * -0.90332) \\
 &= 0.51357 + 0.48297 \\
 &= 0.99653
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy jurusan (Agama)} &= (-16/29 * (\text{LOG}_2(16/29))) + (-13/29 * \text{LOG}_2(13/29)) \\
 &= (-0.55172 * \text{LOG}_2(0.55172)) + (-0.44828 * \text{LOG}_2(0.44828)) \\
 &= (-0.55172 * -0.85798) + (-0.44828 * -1.15754) \\
 &= 0.47337 + 0.51890 \\
 &= 0.99227
 \end{aligned}$$

2) Perhitungan Gain

$$\begin{aligned}
 \text{Gain jurusan} &= 0.99653 - ((29/101 * 0.99227) + (72/101 * 0.98604)) \\
 &= 0.99653 - (0.28491 + 0.70292) \\
 &= 0.99653 - 0.98783 \\
 &= 0.00870
 \end{aligned}$$

3) Penentuan Node

Karena Gain tertinggi adalah ipsmt_2, maka ipsmt_2 menjadi node, dan menjadi acuan untuk perhitungan entropy dan gain selanjutnya. Proses ini dilakukan secara berulang hingga masing atribut memiliki kelas.

3.5 *Prototipe*

1) Menu Login

Untuk halaman login ditunjukkan pada gambar 5.

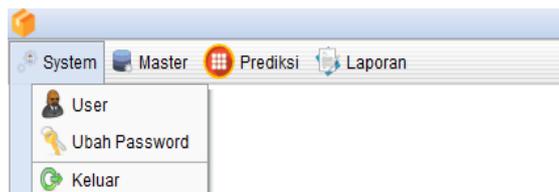


Gambar 5. Halaman Login

Halaman login adalah satu halaman yang muncul saat sistem dimulai untuk pertama kalinya. Di sisi pengguna ini Masukkan username dan password untuk mengakses system

2) Menu *System*

Untuk menu system terdapat pada gambar 6.

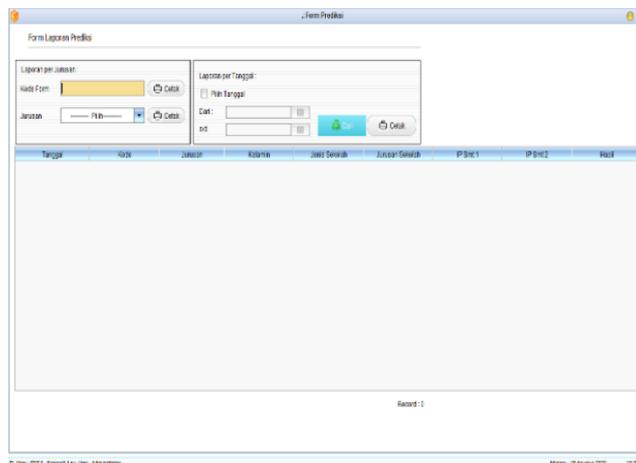


Gambar 6. Tampilan Menu System

Menu sistem adalah menu yang ditampilkan saat setelah masuk, pengguna memilih menu sistem. Di sisi pengguna ini Administrator dapat menambah, menyimpan, memperbarui, dan menghapus pengguna dan Kata sandi.

3) Halaman Laporan

Untuk halaman laporan ditunjukkan pada gambar 7.



Gambar 7. Tampilan Halaman Laporan

Pengguna dapat melihat dan cetak laporan prediksi pada menu Form Laporan Prediksi halaman ini muncul jika pengguna memilih menu tersebut setelah login. Pengguna dapat melihat dan Cetak laporan prediksi berdasarkan kode form, jurusan dan tanggal.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa pengembangan model prediksi dengan menggunakan algoritma C4.5 dan 10 atribut prediktif yaitu :jurusan, jenis_seleksi, tahun_angkatan, kelamin, propinsi, jenis_sekolah, jurusan_sekolah, ipsmt_1, ipsmt_2, ipsmt_3 dan 1 atribut kelas yaitu Prediksi, untuk kelulusan tepat waktu, dapat dikembangkan. Sebagaimana diperoleh *Accuracy* 75.52%, *Precision* 75.52%, dan *Recall* 75.52%.

Untuk pengembangan dan implementasi hasil penelitian ini beberapa hal yang perlu dilakukan sebagai berikut: menguji dengan algoritma yang sama atau berbeda, dengan atribut yang berbeda dan jumlah atribut yang lebih sedikit untuk prediksi kelulusan mahasiswa dan dapat menghasilkan akurasi yang lebih tinggi dari penelitian ini, perlu dilakukan Training terhadap user agar data yang dimasukan dapat terjaga validitasnya, perlu dilakukan maintenance secara berkala untuk menghindari merusakkan sistem.

DAFTAR PUSTAKA

- [1] Y. Al Amin, F.N.; Indahwati, Angraini, "Analisis Ketepatan Waktu Lulus Berdasarkan Karakteristik Mahasiswa Fem Dan Faperta Menggunakan Metode Chart," *Xplore J. Stat.*, vol. 1, no. 2, pp. 1–8, 2013, doi: 10.29244/xplore.v1i2.12411.
- [2] I. S. Prasetyo, T.F., Susandi D., dan Widianingrum, "Prediksi Kelulusan Mahasiswa Pada Perguruan Tinggi Kabupaten Majalengka Berbasis Knowledge Based System," in *Seminar Nasional Telekomunikasi dan Informatika (SELISIK 2016)*, 2016, pp. 32–37.
- [3] H. Romadhona, A., Suprapedi, S. dan Himawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree," *J. Teknol. Inf.*, vol. 13, pp. 69–83, 2017.
- [4] D. Salmu, S. dan A. Solichin, "Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naïve Bayes : A Case Study at Islamic State University Syarif Hidayatullah Jakarta," *Pros. Semin. Nas. Multidisiplin Ilmu*, April, hal. 701–709, 2017.
- [5] T. Ojha, "Prediction Of Graduation Delay Based On Student Performance," 2017.
- [6] G. Testiana, "Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah," *JUSIFO (Jurnal Sist. Informasi)*, vol. 4, pp. 49–62, 2018.
- [7] A. Wijaya, J., Soleh, A.M. dan Rizki, "Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB," *Xplore J. Stat.*, vol. 2, no. 2, pp. 32–40, 2018, doi: 10.29244/xplore.v2i2.99.
- [8] S. Yunianita, S., Setiani, N., dan Mulyati, "Prediksi Ketepatan Masa Studi Mahasiswa dengan Algoritma Pohon Keputusan C45," *SNATI*, pp. 23–29, 2018.
- [9] N. Supriyanti, W. , dan Puspitasari, "Implementasi Teknik Seleksi Fitur Forward Selection Pada Algoritma Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa Politeknik Indonusa Surakarta," *J. Inf. Politek. Indonusa Surakarta*, vol. 4, 2018.
- [10] M. Rohman, A., dan Rochcham, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa," *J. Neo Tek.*, vol. 5, no. 1, 2019.
- [11] dan I. Arrahimi, A.R., Ihsan, M.K., Kartini, D., Faisal, M.R., "Teknik Bagging Dan

- Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa,” *J. Sains dan Inform.*, vol. 5, no. 1, pp. 21, 2019, doi: 10.34128/jsi.v5i1.171.
- [12] Y. H. K. Wirawan, C., Khudzaeva, E., Hasibuan, T.H., Karjono, K., dan Lubis, “Application of Data mining to Prediction of Timeliness Graduation of Students (A Case Study),” *2019 7th Int. Conf. Cyber IT Serv. Manag.*, pp. 1–4, 2019, doi: 10.1109/CITSM47753.2019.8965425.
- [13] D. Suwardika, G. dan I. . K. P. ut. Suniantara, “Analisis Random Forest Pada Klasifikasi Cart Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 13, no. 3, pp. 177–184, 2019, doi: 10.30598/barekengvol13iss3pp177-184ar910.
- [14] T. Munawir, M. dan Iqbal, “Prediksi Kelulusan Mahasiswa menggunakan Algoritma Naive Bayes (Studi Kasus 5 PTS di Banda Aceh),” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 3, no. 2, pp. 1–5, 2019.
- [15] D. Chapman, P., Clinton, J., Kerber, R.Khabaza, T. Reinartz, T.Shearer, C., “CRISP-DM 1.0,” *Cris. Consort.*, hal. 76, 2000.
- [16] I. made B. Adnyana, “Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa,” *J. Sist. Dan Inform.*, vol. 13, pp. 72–76, 2019.
- [17] E. Fox, “Handling missing data,” 2018.
- [18] E. Elisa, “Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti,” *J. Online Inform.*, vol. 2, no. 1, pp. 36, 2017, doi: 10.15575/join.v2i1.71.
- [19] K. C. Shmueli G., Bruce, P.C., Yahav I., Patel N. R., dan Lichtendahl, Jr., *Data Mining For Business Analytics Concepts, Techniques, and Applications in R*, First. John Wiley & Sons, Inc., 2018.