

INFORMASI INTERAKTIF

JURNAL INFORMATIKA DAN TEKNOLOGI INFORMASI

PROGRAM STUDI INFORMATIKA – FAKULTAS TEKNIK -UNIVERSITAS JANABADRA

METODE KLASIFIKASI DATA MINING ALGORITMA C4.5 DAN PART UNTUK PREDIKSI WAKTU KELULUSAN MAHASISWA DI UNIVERSITAS DARWAN ALI

Selviana Yunita, Nurahman

PENERAPAN SISTEM PAKAR UNTUK IDENTIFIKASI ANAK BERKEBUTUHAN KHUSUS MENGGUNAKAN METODE RULE BASED SYSTEM

Yumarlin MZ, Hanang Indrianta

IMPLEMENTASI SMOTE UNTUK MENGATASI *IMBALANCED DATA* PADA SENTIMEN ANALISIS SENTIMEN HOTEL DI NUSA TENGGARA BARAT DENGAN MENGGUNAKAN ALGORITMA SVM

Erry Maricha Oki Nur Haryanto, Adhien Kenya Anima Estetikha, Rahmad Arif Setiawan

IMPLEMENTASI DASHBOARD MICROSOFT POWER BI UNTUK VISUALISASI DATA COVID 19 INDONESIA

Jemmy Edwin Bororing

RANCANG BANGUN APLIKASI KASIR USAHA MIKRO KECIL MENENGAH MENGGUNAKAN METODE *BLOCK PROGRAMMING* (STUDI KASUS : PELANGI STORE)

Agit Amrullah, Fata Aulia

PERANCANGAN APLIKASI PENGENALAN LITERASI COVID-19 MENGGUNAKAN *ACTIONSCRIPT 3.0* PADA *MACROMEDIA FLASH*

Jeffry Andhika Putra, Erry Maricha Oki Nur Haryanto

PENGARUH SMOTE DAN *FORWARD SELECTION* DALAM MENANGANI KETIDAKSEIMBANGAN KELAS PADA ALGORITMA KLASIFIKASI

Ika Nur Fajri, Femi Dwi Astuti

MEDIA KOMUNIKASI KESEHATAN UNTUK TUNA RUNGU DAN TUNA WICARA BERBASIS ANDROID

Ryan Ari Setyawan, Rizqi Mirza Fadilla

IMPLEMENTASI *USER EXPERIENCE DESIGN* PADA PERANCANGAN APLIKASI PEMBELAJARAN PRAKTIKUM *ONLINE* BERBASIS *MOBILE*

Eri Haryanto, Agustin Setiyorini

PROTOTYPE PENGENALAN CANDI DI YOGYAKARTA BERBASIS *AUGMENTED REALITY*

Fatsyahrina Fitriastut, Ryan Ari Setyawan, Helio Rofino Correia



DEWAN EDITORIAL

- Penerbit** : Program Studi Informatika Fakultas Teknik Universitas Janabadra
- Ketua Penyunting
(Editor in Chief)** : Fatsyahrina Fitriastuti, S.Si., M.T. (Universitas Janabadra)
- Penyunting (Editor)** : 1. Jemmy Edwin B, S.Kom., M.Eng. (Universitas Janabadra)
2. Ryan Ari Setyawan, S.Kom., M.Eng. (Universitas Janabadra)
3. Yumarlin MZ, S.Kom., M.Pd., M.Kom. (Universitas Janabadra)
- Alamat Redaksi** : Program Studi Informatika Fakultas Teknik
Universitas Janabadra
Jl. Tentara Rakyat Mataram No. 55-57
Yogyakarta 55231
Telp./Fax : (0274) 543676
E-mail: informasi.interaktif@janabadra.ac.id
Website : <http://e-journal.janabadra.ac.id/>
- Frekuensi Terbit** : 3 kali setahun

JURNAL INFORMASI INTERAKTIF merupakan media komunikasi hasil penelitian, studi kasus, dan ulasan ilmiah bagi ilmuwan dan praktisi dibidang Informatika. Diterbitkan oleh Program Studi Informatika Fakultas Teknik Universitas Janabadra di Yogyakarta, tiga kali setahun pada bulan Januari, Mei dan September.

DAFTAR ISI

	<i>halaman</i>
Metode Klasifikasi Data Mining Algoritma C4.5 Dan Part Untuk Prediksi Waktu Kelulusan Mahasiswa Di Universitas Darwan Ali <i>Selviana Yunita, Nurahman</i>	1 - 7
Penerapan Sistem Pakar Untuk Identifikasi Anak Berkebutuhan Khusus Menggunakan Metode <i>Rule Based System</i> <i>Yumarlin MZ, Hanang Indrianta</i>	8 - 15
Implementasi SMOTE Untuk Mengatasi <i>Imbalanced Data</i> Pada Sentimen Analisis Sentimen Hotel Di Nusa Tenggara Barat Dengan Menggunakan Algoritma SVM <i>Erry Maricha Oki Nur Haryanto, Adhien Kenya Anima Estetikha, Rahmad Arif Setiawan</i>	16 - 20
Implementasi <i>Dashboard</i> Microsoft Power BI Untuk Visualisasi Data Covid 19 Indonesia <i>Jemmy Edwin Bororing</i>	21 - 29
Rancang Bangun Aplikasi Kasir Usaha Mikro Kecil Menengah Menggunakan Metode <i>Block Programming</i> (Studi Kasus : Pelangi Store) <i>Agit Amrullah, Fata Aulia</i>	30 - 37
Perancangan Aplikasi Pengenalan Literasi Covid-19 Menggunakan <i>Actionscript</i> 3.0 Pada <i>Macromedia Flash</i> <i>Jeffry Andhika Putra, Erry Maricha Oki Nur Haryanto</i>	38 - 44
Pengaruh SMOTE Dan <i>Forward Selection</i> Dalam Menangani Ketidakseimbangan Kelas Pada Algoritma Klasifikasi <i>Ika Nur Fajri, Femi Dwi Astuti</i>	45 - 49
Media Komunikasi Kesehatan Untuk Tuna Rungu Dan Tuna Wicara Berbasis Android <i>Ryan Ari Setyawan, Rizqi Mirza Fadilla</i>	50 - 59
Implementasi <i>User Experience Design</i> Pada Perancangan Aplikasi Pembelajaran Praktikum <i>Online</i> Berbasis <i>Mobile</i> <i>Eri Haryanto, Agustin Setiyorini</i>	60 - 69
Prototype Pengenalan Candi Di Yogyakarta Berbasis <i>Augmented Reality</i> <i>Fatsyahrina Fitriastut, Ryan Ari Setyawan, Helio Rofino Correia</i>	70 - 78

PENGANTAR REDAKSI

Puji syukur kami panjatkan kehadirat Allah Tuhan Yang Maha Kuasa atas terbitnya JURNAL INFORMASI INTERAKTIF Volume 7, Nomor 1, Edisi Januari 2022. Pada edisi kali ini memuat 10 (sepuluh) tulisan hasil penelitian dalam bidang informatika.

Harapan kami semoga naskah yang tersaji dalam JURNAL INFORMASI INTERAKTIF edisi Januari tahun 2022 dapat menambah pengetahuan dan wawasan di bidangnya masing-masing dan bagi penulis, jurnal ini diharapkan menjadi salah satu wadah untuk berbagi hasil-hasil penelitian yang telah dilakukan kepada seluruh akademisi maupun masyarakat pada umumnya.

Redaksi

METODE KLASIFIKASI DATA MINING ALGORITMA C4.5 DAN PART UNTUK PREDIKSI WAKTU KELULUSAN MAHASISWA DI UNIVERSITAS DARWAN ALI

Selviana Yunita¹, Nurahman²

^{1,2} Universitas Darwan Ali
Jl. Batu Berlian No. 10 Sampit, Kotawaringin Timur, Kalimantan Tengah, 74321

Email: ¹selviana.yunita.ax@gmail.com, ²nurrahman.ikhtiar@gmail.com

ABSTRACT

College is one of the most influential education aspect for a nation's improvement. The quality of a college is important to support and explore student potential. The quality of a college helps student to prepare themselves in working world. One of the qualities of a college can be seen from the punctuality of graduating time for students. It is become important for a college to find out the factors that influence the punctuality of graduating time for students. Darwan Ali University is one of university located in Sampit, Central Kalimantan. Based on their Information System Management, in 2011, there are 707 new students. In 2015 only 290 students passed. It shows that only 41% of students graduate on time. The source of this research data comes from Management Information System of Darwan Ali University. The purpose of this research is to find the rules which affects the accuracy of student graduation. The data used in this study include department of study programs, the GPA from first to fourth semester, and gender of students. This study uses two algorithms, namely the C4.5 and PART algorithms. The researcher also found that the C4.5 algorithm has better accuracy than PART, with an accuracy level of 83.004 %.

Keywords: *Classification, data mining, C4.5, PART*

1. PENDAHULUAN

Sebagai salah satu faktor yang berpengaruh dalam kemajuan bangsa, kualitas perguruan tinggi menjadi sangat penting demi mendukung dan menggali potensi mahasiswa dalam terjun ke dunia profesional. Demi menghasilkan lulusan tepat waktu serta memiliki kompetensi, perlu dilakukan evaluasi secara berkala sehingga hasil belajar mahasiswa dapat terus ditingkatkan.

Ketepatan waktu mahasiswa dalam menyelesaikan pendidikannya, merupakan salah satu indikator mutu yang harus diperhatikan dan dievaluasi. Dengan adanya evaluasi terhadap ketepatan waktu lulus ini, akan mampu meningkatkan kualitas proses pembelajaran pada pendidikan perguruan tinggi. Namun, dalam prakteknya, perguruan tinggi tidak melakukan antisipasi kepada mahasiswa yang memiliki potensi untuk mengalami hambatan selama proses pembelajaran [1].

Kurangnya antisipasi serta evaluasi terhadap mahasiswa yang mengalami hambatan dalam proses belajar ini berdampak terhadap rendahnya hasil belajar mahasiswa, yaitu angka indeks prestasi. Selain berdampak terhadap angka indeks prestasi, hambatan yang dialami mahasiswa akan mempengaruhi waktu lulus

mahasiswa, sehingga mahasiswa tidak dapat lulus tepat waktu.

Berdasarkan data yang dirilis oleh PDDIKTI pada 2018 mengenai statistik Pendidikan Tinggi, secara nasional terdapat 1.732.308 mahasiswa baru yang masuk, dan 1.247.116 mahasiswa yang lulus. Dari data tersebut, terdapat penurunan sekitar 28% terhadap jumlah kelulusan mahasiswa jika dibandingkan dengan data mahasiswa masuk [2]. Besarnya angka penurunan data kelulusan juga berdampak terhadap banyaknya peneliti yang mencoba memprediksi kelulusan dengan berbagai metode serta menjadi topik menarik untuk diteliti dan dicari pemecahan masalahnya. Penelitian mengenai prediksi kelulusan telah dilakukan oleh beberapa peneliti sebelumnya [3][4][5][6][7].

Penelitian ini dilakukan di Universitas Darwan Ali, yang berlokasi di Sampit, Kotawaringin Timur. Berdasarkan data, mahasiswa yang masuk pada tahun 2011 adalah 707 orang, dan pada tahun 2015 total mahasiswa yang lulus adalah 290 orang. Dari data tersebut, dapat dilihat jika mahasiswa yang lulus tepat waktu adalah sebesar 41% dari jumlah awal mahasiswa yang masuk. Universitas Darwan Ali sampai saat ini belum memanfaatkan data kondisi akademik mahasiswa secara maksimal. Dengan memanfaatkan data yang ada pada SIM

(Sistem Informasi Manajemen) milik Universitas Darwan Ali, akan dilakukan proses *data mining* terhadap riwayat akademik mahasiswa sehingga diharapkan dapat dilakukan antisipasi dan penanganan sejak dini kemungkinan mahasiswa lulus tidak tepat waktu. Dalam proses pengolahan data digunakan dua algoritma sebagai pembanding, yaitu algoritma C4.5 dan algoritma PART. Kumpulan data akademik mahasiswa yang ada di dalam SIM Universitas Darwan Ali akan diproses dengan kedua algoritma tersebut untuk mendapatkan pola yang berupa *rule* yang akan menjadi landasan dalam melakukan prediksi data hasil belajar mahasiswa.

Berdasarkan permasalahan dan uraian diatas, perlu dilakukan analisis klasifikasi ketepatan lulus mahasiswa. Hal ini bertujuan untuk membantu mengembangkan suatu sistem cerdas serta variabel yang sesuai untuk mendukung sistem tersebut. Selain itu, analisis ini akan membantu menentukan algoritma mana dengan akurasi yang lebih tinggi sehingga sesuai dengan karakteristik data dan analisis yang perlu dilakukan. Penelitian ini menghasilkan *classification rule* yang dapat digunakan sebagai dasar dalam pembangunan sistem cerdas yang akan membantu pihak Universitas dalam memprediksi waktu kelulusan mahasiswa. Dengan melakukan penelitian di Universitas Darwan Ali mengenai ketepatan waktu lulusan bagi mahasiswa, dapat meningkatkan mutu pendidikan Universitas Darwan ali, dan meningkatkan kualitas sumber daya manusia di Sampit

2. METODE PENELITIAN

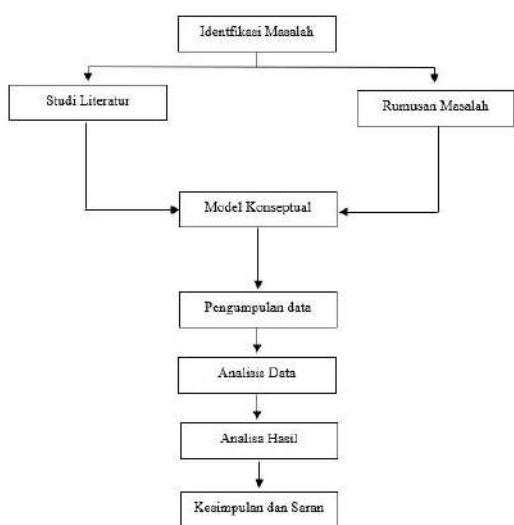
Ada beberapa tahapan untuk menyelesaikan permasalahannya, secara garis besar tahapan penelitian yang dilakukan adalah identifikasi masalah, studi literature, merumuskan masalah yang didapatkan pada proses identifikasi, merancang model konseptual, pengumpulan data, analisis data, analisa hasil, kesimpulan dan saran. Penelitian dilakukan sesuai dengan tahapan yang telah ditentukan dalam penelitian ini. Penentuan tahapan penelitian dilakukan agar penelitian dapat dilakukan dengan terarah sesuai dengan ketentuan ketentuan standar dalam suatu konsep ilmiah. Adapun uraian setiap tahapan adalah

- Identifikasi masalah adalah suatu tahap permulaan dari penguasaan sebuah masalah dimana suatu objek tertentu dan dalam

situasi tertentu dapat kita kenali sebagai suatu masalah.

- Studi literatur dilakukan dengan penelaahan buku, literatur, catatan dan laporan-laporan yang ada hubungannya dengan masalah yang akan dipecahkan, studi literatur dilakukan juga untuk mengumpulkan data penunjang mengenai teori-teori yang mendukung penelitian.
- Setelah melakukan identifikasi masalah dan studi literatur maka didapatkan perumusan masalah yang akan menjadi fokus dalam penelitian ini, tujuan dari perumusan masalah adalah memperjelas ruang lingkup penelitian.
- Dalam tahapan kerangka konseptual diharapkan akan memberi gambaran dan mengarahkan asumsi mengenai variabel-variabel yang akan diteliti, kerangka konseptual memberikan petunjuk kepada peneliti dalam merumuskan masalah penelitian.
- Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan dalam rangka mencapai tujuan penelitian. Pada tahap ini dilakukan pengumpulan data yang riwayat akademik mahasiswa yang terdapat pada SIM (Sistem Informasi Manajemen) Universitas Darwan Ali.
- Analisis dan Pemrosesan data dilakukan untuk mengolah data yang telah berhasil dikumpulkan, pemrosesan data meliputi proses *dataset development*, *data cleansing*, ekstraksi fitur. Setelah fitur di ekstraksi, dilakukan pemrosesan dengan menggunakan algoritma C4.5 dan algoritma PART.
- Analisa Hasil dilakukan dengan data-data yang telah dikumpulkan mengalami proses validasi data dan analisis data akan memberikan suatu hasil dari penelitian yang, dari hasil penelitian itu kita akan mengetahui jawaban dari pertanyaan-pertanyaan penelitian.
- Di dalam tahap kesimpulan terdapat pernyataan yang menunjukkan terjawab tidaknya pertanyaan penelitian dan dalam kesimpulan tidak boleh terlepas dari pertanyaan penelitian yang diajukan.

Secara singkat, tahapan penelitian digambarkan melalui diagram alir seperti pada gambar 1.



Gambar 1. Diagram Alir

3. TINJAUAN PUSTAKA

3.1.1 Data Mining

Data Mining merupakan proses ekstraksi pengetahuan dari berbagai database besar dan identifikasi informasi yang bermanfaat dengan menggunakan berbagai teknik seperti statistika, matematika, kecerdasan buatan, dan *machine learning*. Beberapa jenis tugas yang dapat diselesaikan dengan *data mining* adalah seperti: *description*, *estimation*, *prediction*, *classification*, *clustering*, dan *association* [8].

Pada dasarnya *data mining* merupakan disiplin ilmu yang memiliki tujuan utama untuk menggali, menemukan, dan menambang data berdasarkan informasi yang dimiliki. *Data Mining* memiliki istilah lain yang disebut *Knowledge Discovery in Database* (KDD). KDD adalah sebuah proses yang terdiri dari serangkaian proses integrasi yang berurutan, dan *data mining* adalah bagian dari salah satu proses yang ada dalam KKD. Adapun langkah-langkah proses dalam KKD dapat diilustrasikan seperti berikut [9]:

- Data Clearance*
- Data Integration*
- Data Selection*.
- Data Transformation*
- Data Mining*
- Pattern evaluation*
- Knowledge Presentation*

3.1.2 Decision Tree

Decision Tree atau pohon keputusan adalah sebuah struktur pohon seperti diagram alir, di mana setiap node/simpul internal mewakili atribut pengujian, setiap cabang mewakili hasil pengujian, label kelas diwakili oleh simpul daun atau node terminal. Diberikan *tuple/sample* X, nilai atribut dari *tuple* tersebut diuji terhadap pohon keputusan [10]. Sebuah jalur akan ditelusuri dari akar sampai ke node terminal yang menyimpan prediksi kelas untuk *tuple*. Merubah pohon keputusan menjadi suatu aturan klasifikasi sangat mudah. Pembelajaran pada pohon keputusan menggunakan pohon keputusan itu sendiri sebagai model prediktif untuk memetakan pengamatan tentang suatu item menuju ke suatu kesimpulan mengenai nilai target item tersebut. Hal ini adalah salah satu pendekatan dengan pemodelan prediktif yang digunakan dalam statistik, *data mining*, dan *machine learning*. Model pohon di mana variabel target dapat mengambil sekumpulan nilai yang terbatas disebut dengan pohon klasifikasi. Dalam struktur pemodelan pohon ini, daun mewakili label kelas dan cabang mewakili sambungan fitur yang mengarah ke label kelas tersebut [11].

3.1.3 Algoritma C4.5

Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma C4.5 memiliki input berupa training samples dan samples. Training samples adalah contoh data yang akan dimanfaatkan dalam membangun sebuah tree yang telah diuji kebenarannya. Samples adalah field data yang akan digunakan sebagai parameter dalam proses klasifikasi data. Algoritma ini telah di aplikasikan secara baik kedalam program yang melibatkan ribuan kasus yang dideskripsikan dalam ratusan fitur [12].

3.1.4 Algoritma PART

Algoritma PART (*Projective Adaptive Resonance Theory*) merupakan algoritma yang dikembangkan oleh Cao dan Wu pada tahun 2014[13]. Algoritma PART adalah algoritma yang telah terbukti efektif dalam pengelompokan data set yang memiliki ruang dimensi yang tinggi. Algoritma PART berdasarkan pada asumsi jika model persamaan PART (yang memiliki skala besar dan sangat terganggu dengan sistem persamaan diferensial yang digabungkan dengan mekanisme reset) memiliki kinerja komputasi yang cukup teratur.

4. HASIL DAN PEMBAHASAN

Penelitian dilakukan di Universitas Darwan Ali dengan memanfaatkan data nilai mahasiswa yang ada pada Sistem Informasi Manajemen milik Universitas. Data ditarik dari Sistem Informasi Manajemen untuk kemudian dilakukan proses untuk mendapatkan pola berupa rule yang akan menjadi landasan dalam penelitian untuk prediksi hasil belajar mahasiswa. Adapun data yang diperoleh dapat dilihat pada tabel 1.

Tabel 1. Data Nilai

No	NPM	Nama	Prodi	IPS1	IPS2	IPS3	IPS4	Jenis Kelamin	Status Keluar
1	13542xxx	Mhs 1	S1 Budidaya Perairan	3,45	3,38	3,88	3,78	Laki-laki	Tepat
2	13542xxx	Mhs 2	S1 Budidaya Perairan	3,14	3,44	3,38	3,52	Perempuan	Tepat
3	13542xxx	Mhs 3	S1 Budidaya Perairan	3,32	3,25	3,63	3,65	Laki-laki	Tepat
4	13542xxx	Mhs 4	S1 Budidaya Perairan	3,14	3,13	2,88	2,52	Perempuan	Tepat
5	13542xxx	Mhs 5	S1 Agribisnis	3	2,57	3,14	3,29	Laki-laki	Tepat
6	13542xxx	Mhs 6	S1 Agribisnis	3,67	3	3,14	3,76	Perempuan	Tepat
7	13542xxx	Mhs 7	S1 Agribisnis	3,67	1,71	3,43	3,86	Perempuan	Tepat
8	13542xxx	Mhs 8	S1 Agribisnis	3,67	3,29	3,71	3,9	Perempuan	Tepat
9	13542xxx	Mhs 9	S1 Agribisnis	3,67	3,29	3,14	3,81	Laki-laki	Tepat
10	13542xxx	Mhs 10	S1 Teknik Sipil	3,5	3,04	3,41	3	Laki-laki	Tepat
11	13542xxx	Mhs 11	S1 Teknik Sipil	2,94	2,96	3,29	3,23	Perempuan	Tepat
12	13542xxx	Mhs 12	S1 Teknik Sipil	3,17	3,17	3,41	3,32	Laki-laki	Tepat
13	13542xxx	Mhs 13	S1 Sistem Informasi	3,54	3,04	3,04	3,29	Laki-laki	Tepat
14	13542xxx	Mhs 14	S1 Sistem Informasi	3,25	3,17	3,04	3,04	Laki-laki	Tepat
15	13542xxx	Mhs 15	S1 Sistem Informasi	3,25	3,08	3,08	2,88	Perempuan	Tepat
16	13542xxx	Mhs 16	S1 Sistem Informasi	3,46	3,42	3,46	2,21	Laki-laki	Tepat
....
011	7612xxxx	Mhs1011	S1 Manajemen	2,74	2	2,67	2,36	Laki-laki	Telat
012	6572xxxx	Mhs1012	S1 Sistem Informasi	3,75	4	3,54	3,58	Perempuan	Tepat

Selanjutnya, sebelum proses klasifikasi dilakukan dengan data yang telah melalui *Pre-processing*. Data yang telah melewati tahap *Pre-processing* data selanjutnya akan ke proses klasifikasi. Jumlah data yang ditarik dari Sistem Informasi Manajemen adalah sejumlah 1102 data. Adapun tahapan proses *Pre-Processing* meliputi:

a. Data Clearance

Pada tahapan ini dilakukan proses pembersihan data untuk menghilangkan *Noise*, duplikasi data dan data yang inkonsisten. Pada tahapan ini juga dilakukan proses perbaikan data dan proses enrichment dengan menambahkan data dari sumber lain.

b. Data Integration

Pada tahap ini dilakukan proses penggabungan data atau integrasi data. Hal ini dilakukan karena penarikan data dilakukan dengan semester yang berbeda untuk setiap attribut, sehingga perlu dilakukan proses integrasi data

c. Data Selection

Pada tahapan ini dilakukan proses seleksi data yang bertujuan untuk mengambil data yang relevan yang akan dimanfaatkan untuk proses analisis pada *data mining*. Dalam tahap ini diciptakan himpunan data target dan pemilihan himpunan data dimana penemuan data akan dilakukan. Sistem Informasi Manajemen sendiri memiliki data yang cukup besar sehingga perlu dilakukan seleksi fitur yang relevan untuk penelitian. Dalam penelitian ini, attribut dataset yang digunakan adalah IPS semester 1, IPS semester 2, IPS Semester 3, IPS semester 4 dan jenis kelamin mahasiswa. Data NPM serta Nama mahasiswa dihilangkan karena bukan merupakan data yang diperhitungkan dan hanya bersifat sebagai identitas. Data yang telah di lakukan seleksi dapat dilihat pada tabel 2.

Tabel 2. Data setelah proses seleksi

Prodi	IPS1	IPS2	IPS3	IPS4	Jenis Kelamin	Status Keluar
S1 Budidaya Perairan	3,45	3,38	3,88	3,78	Laki-laki	Tepat
S1 Budidaya Perairan	3,32	3,25	3,63	3,65	Laki-laki	Tepat
S1 Agribisnis	3	2,57	3,14	3,29	Laki-laki	Tepat
S1 Agribisnis	3,67	3,29	3,14	3,81	Laki-laki	Tepat
S1 Teknik Sipil	3,5	3,04	3,41	3	Laki-laki	Tepat
S1 Teknik Sipil	3,17	3,17	3,41	3,32	Laki-laki	Tepat
S1 Sistem Informasi	3,54	3,04	3,04	3,29	Laki-laki	Tepat
S1 Sistem Informasi	3,25	3,17	3,04	3,04	Laki-laki	Tepat
S1 Sistem Informasi	3,46	3,42	3,46	2,21	Laki-laki	Tepat
....
S1 Manajemen	2,74	2	2,67	2,36	Laki-laki	Telat
S1 Sistem Informasi	3,75	4	3,54	3,58	Perempuan	Tepat

d. Data Transformation

Pada tahap ini dilakukan proses transformasi data kedalam bentuk yang ideal untuk dilakukan penambangan. Transformasi dilakukan untuk attribut yang memiliki nilai banyak tak beraturan. Pada penelitian ini attribut yang dianggap memiliki nilai banyak tak beraturan yaitu pada attribut indek prestasi semester (IPS). Nilai index prestasi semester terdapat pada 4 kolom yaitu IPS1, IPS2, IPS3, dan IPS4. Pada keempat kolom ini dilakukan proses dikritisasi data yang juga diproses dengan menggunakan aplikasi Weka. Proses Dikritisasi data adalah teknik yang digunakan untuk membagi rentang attribut sehingga akan mengurangi jumlah nilai attribut. Rentang nilai pada proses diskritisasi dilakukan dengan pengaturan default pada aplikasi weka. Hasil dikritisasi data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Dikritisasi Data

Prodi	IPS1	IPS2	IPS3	IPS4	Jenis Kelamin	Status Keluar
S1 Budidaya Perairan	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Laki-laki	Tepat
S1 Budidaya Perairan	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Perempuan	Tepat
S1 Budidaya Perairan	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Laki-laki	Tepat
S1 Budidaya Perairan	≥ 2.955	≥ 2.69	< 2.945	$2.1 - 2.894$	Perempuan	Tepat
S1 Agribisnis	≥ 2.955	< 2.69	≥ 2945	≥ 2895	Laki-laki	Tepat
S1 Agribisnis	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Perempuan	Tepat
S1 Agribisnis	≥ 2.955	< 2.69	≥ 2945	≥ 2895	Perempuan	Tepat
S1 Agribisnis	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Perempuan	Tepat
S1 Agribisnis	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Laki-laki	Tepat
S1 Teknik Sipil	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Laki-laki	Tepat
....
....
S1 Manajemen	< 2.955	< 2.69	< 2.945	$2.1 - 2.894$	Laki-laki	Telat
S1 Sistem Informasi	≥ 2.955	≥ 2.69	≥ 2945	≥ 2895	Perempuan	Tepat

e. Data Mining

Pada tahap ini dilakukan proses penambangan data, yaitu proses dalam penggunaan suatu metode tertentu dalam mengekstrak pola-pola yang terdapat dalam data. Proses pengolahan data dilakukan dengan menggunakan test options *cross validation* dengan *folds 10*. Pemrosesan dilakukan dua kali dengan menggunakan algoritma C.45 dan algoritma PART. Jumlah data yang dilakukan ekstraksi polanya adalah sejumlah 1012 data. Proses penambangan data dilakukan dengan menggunakan aplikasi Weka.

4.1 Klasifikasi dan Pengujian dengan Algoritma C4.5

Pengolahan data dilakukan dengan melakukan proses perhitungan menggunakan algoritma C4.5. Tabel akurasi untuk proses perhitungan Algoritma C.45 dapat dilihat pada tabel 4.

Tabel 4. Tabel Akurasi C4.5

TP Rate	FP Rate	Precision	Recall	F-Measures	Class
0.955	0.707	0.853	0.955	0.901	Tepat
0.293	0.045	0.602	0.293	0.394	Telat
0.830	0.582	0.806	0.830	0.806	

Berdasarkan tabel 4, diketahui jika *True Positive Rate* untuk class tepat adalah 0.955 dengan *False Positive Rate* sebesar 0.707. Sedangkan untuk class telat, *True Positive Rate* bernilai 0.293 dan *False Positive Rate* sebesar

0.045. Diketahui dari perhitungan dengan algoritma C45 menunjukkan nilai akurasi senilai 83,004%

Rule yang dihasilkan oleh pengolahan data menggunakan algoritma C4.5 adalah:

Jika IPS4 < 2.1 maka Kelulusan Telat,

Jika IPS4 = 2.1 – 2.895, maka perhatikan

Jika IPS1 < 2.955, maka:

Prodi S1 Budidaya Perairan: Tepat

Prodi S1 Agribisnis: Telat

Prodi S1 Teknik Sipil perlu memperhatikan

Jenis Kelamin Laki-laki: Tepat

Jenis Kelamin Perempuan: Telat

Prodi S1 Sistem Informasi: Telat

Prodi D3 Manajemen Informatika: Tepat

Prodi S1 Manajemen perlu memperhatikan

Jenis Kelamin Laki-laki: Telat

Jenis Kelamin Perempuan: Tepat

Jika IPS1 > 2.955, maka Tepat

Jika IPS4 > 2.895, maka Tepat

4.2 Klasifikasi dan Pengujian dengan Algoritma PART

Sebagai banding, dilakukan kembali pengolahan data dengan menggunakan algoritma yang berbeda, yaitu Algoritma PART. Tabel 5 menunjukkan hasil akurasi dengan menggunakan algoritma PART.

Tabel 5. Tabel Akurasi PART

TP Rate	FP Rate	Precision	Recall	F-Measures	Class
0.939	0.649	0.861	0.939	0.899	Tepat
0.351	0.061	0.573	0.351	0.435	Telat
0.828	0.538	0.807	0.828	0.811	

Berdasarkan tabel 5, diketahui jika *True Positive Rate* untuk class tepat adalah 0.939 dengan *False Positive Rate* sebesar 0.649. Sedangkan untuk class telat, *True Positive Rate* bernilai 0.351 dan *False Positive Rate* sebesar 0.061. Diketahui dari perhitungan dengan menggunakan Algoritma PART nilai Akurasi senilai 82.8063 %

Rule yang dihasilkan oleh pengolahan data menggunakan algoritma PART adalah:

Jika IPS4 = 2.895 AND IPS3 = 2.945, maka Tepat

Jika IPS1 = 2.955 AND IPS4 = 2.895 AND Prodi S1 Manajemen, maka Tepat

Jika IPS1 = 2.955 AND IPS4 = 2.895 AND

Prodi S1 Sistem Informasi, maka Tepat

Jika $IPS1 = 2.955$ AND $IPS4 = 2.1-2.895$ AND Prodi D3 Manajemen Informatika, maka Tepat

Jika $IPS4 < 2.1$ AND Jenis Kelamin Laki-laki AND $IPS1 < -2.955$, maka Telat

Jika $IPS2 > 2.69$ AND $IPS4 = 2.1-2.895$ AND Prodi S1 Teknik Sipil AND Jenis Kelamin Laki-laki, maka Tepat

Jika $IPS2 > 2.69$ AND $IPS4 = 2.1-2.895$ AND $IPS1 > 2.955$, maka Tepat

Jika $IPS4 < 2.1$ AND Jenis Kelamin Laki-laki AND $IPS3 < 2.945$, maka Telat

Jika $IPS4 < 2.1$ AND Jenis Kelamin Perempuan, maka Telat

Jika Jenis Kelamin Perempuan AND Prodi S1 Manajemen, maka Tepat

Jika $IPS1 > 2.955$ AND Prodi S1 Agribisnis, maka Tepat

Jika $IPS1 > 2.955$, maka Tepat

Jika Prodi S1 Teknik Sipil AND $IPS4 > 2.895$, maka Tepat

Jika Prodi S1 Sistem Informasi AND Jenis Kelamin Laki-laki, maka Telat

Jika $IPS3 > 2.945$, maka Tepat

Jika Prodi D3 Manajemen Informatika AND $IPS2 > 2.69$, maka Tepat

Jika Prodi S1 Manajemen, maka Telat

Jika Prodi S1 Agribisnis AND $IPS2 < 2.69$, maka Telat

Jika Prodi S1 Agribisnis AND Jenis Kelamin Laki-laki, maka Telat

4.3 Analisis Hasil

Setiap Algoritma pada data mining memiliki tingkat performa dalam mengolah data. Performa suatu algoritma data mining dipengaruhi beberapa hal, salah satunya adalah kondisi dari suatu data yang akan diolah. Data yang akan diolah atau proses *mining data* harus

terbebas dari data tidak lengkap, noisy dan lain sebagainya.

Tingkat performa yang dilakukan dalam penelitian ini yaitu dengan melihat nilai akurasi dari kedua Algoritma yang digunakan untuk mengolah data. Diketahui bahwa hasil pengolahan data menunjukkan performa akurasi Algoritma C4.5 lebih tinggi dibanding dengan Algoritma PART. Perbandingan performa kedua Algoritma dapat dilihat pada tabel 6 dibawah ini.

Tabel 6 perbandingan Performa Algoritma

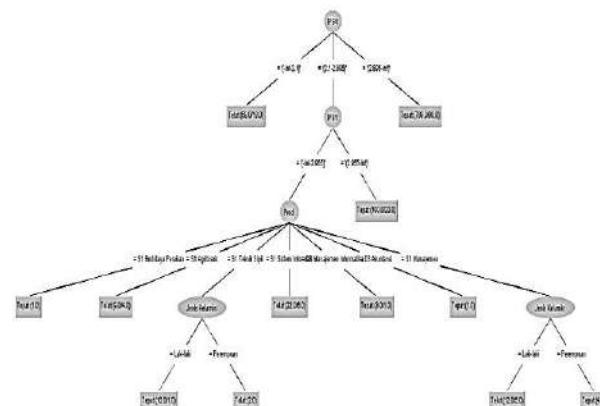
Algoritma	Akurasi	Level
C4.5	83,004%	1
PART	82.8063 %	2

Performa Akurasi [14] dapat dibagi menjadi 5 kategori yaitu sebagai berikut :

1. 90% - 100% = *excellent classification*;
2. 80% - 90% = *good classification*;
3. 70% - 80% = *fair classification*;
4. 60% - 70% = *poor classification*;
5. 50% - 60% = *failure*.

Berdasarkan Klasifikasi yang ditunjukan pada Tabel 6 menunjukkan bahwa nilai akurasi yang dihasilkan pada kedua Algoritma masuk pada kategori *good classification*. Nilai akurasi pada Algoritma C4.5 adalah 83,004%. Kemudian nilai akurasi pada Algoritma PART adalah 82.8063 %.

Pada penelitian ini Algoritma C4.5 memiliki nilai Akurasi tertinggi. Berdasarkan Rule yang dihasilkan dari klasifikasi menggunakan Algoritma C4.5 memperoleh sebuah pohon keputusan. Pohon keputusan yang dihasilkan dari klasifikasi menggunakan Algoritma C4.5 dapat dilihat pada gambar 2 dibawah ini.



Gambar 2. Pohon keputusan

5. KESIMPULAN

Berdasarkan 1012 jumlah data yang berisi data Prodi, Indeks Prestasi dari semester 1 hingga 4, jenis kelamin serta status kelulusan mahasiswa, diperoleh beberapa *rule* yang dapat dijadikan panduan oleh Universitas untuk mengetahui ketepatan waktu lulus mahasiswa berdasarkan Indeks Prestasi 4 semester pertama serta jenis kelamin mahasiswa. Algoritma yang direkomendasikan untuk digunakan dalam proses klasifikasi adalah algoritma C4.5 karena memiliki akurasi yang lebih baik dibandingkan algoritma PART, yaitu sebesar 83.004 %.

DAFTAR PUSTAKA

- [1] Zulhkairi, 2020. Implementasi Sistem Informasi Manajemen Akademik Sevima dalam mengoptimalkan mutu layanan akademik perguruan tinggi. *Jurnal Pendidikan dan Kependidikan* Vol. 4 No. 1. eISSN: 2656-8012
- [2] PDDIKTI, 2018. Statistik Pendidikan Tinggi. Kementerian Riset, Teknologi, dan Pendidikan Tinggi.
- [3] Jaman, J.H., 2013. Prediksi Kelulusan Mahasiswa Dengan Metode Algoritma C4.5. *Syntax* Vol. 2 No. 2 hlm. 39-43.
- [4] Mustaqbal, M.S., Firdaus, R.F., Rahmadi, H., 2015. Pengujian aplikasi menggunakan *Black Box Testing Boundary Value Analysis* (Studi Kasus: Aplikasi Prediksi Kelulusan SNMPTN). *Jurnal Ilmiah Teknologi Informasi Terapan* Vol. 1 No. 3. ISSN: 2407-3911.
- [5] Hakim, L.A.R., Rizal, A.A., Ratnasari, D., 2019. Aplikasi Prediksi Kelulusan mahasiswa Berbasis K-Nearest Neighbor (K-NN). *Jurnal Teknologi Informasi dan Multimedia* Vol. 1 No. 1, hlm. 30-36
- [6] Pambudi, R.D., Supianto, A.A., Setiawan, N.Y., 2019. Prediksi Kelulusan Mahasiswa Berdasarkan kinerja akademik menggunakan pendekatan data mining pada program studi sistem informasi Fakultas Ilmu Komputer Universitas Brawijaya. *Jurnal pengembangan Teknologi Informasi dan Ilmu Komputer* Vol. 3 No. 3. eISSN: 2548-964X
- [7] Imron, M., 2017. Penerapan Data Mining Algoritma Naïve Bayes dan Part untuk mengetahui Minat Baca Mahasiswa di perpustakaan STIMIK Amikom Purwokerto. *Jurnal Telematika* Vol. 10 No. 2. eISSN: 2442-4528
- [8] Ridwan, M., 2013. Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*. Vol. 7, No. 1.
- [9] Thomas E. 2004. *Data Mining: Definition and Decision Tree Examples*, e-book.
- [10] Sembiring, M.A., 2016. Penerapan Metode Decision Tree Algoritma C4.5 untuk memprediksi hasil belajar mahasiswa berdasarkan riwayat akademik. *Jurnal Teknologi dan Sistem Informasi* Vol. 3 No. 1. ISSN: 2407-1811
- [11] Rathee, A., Mathur, R.P., 2013. Survey on Decision Tree Classification algorithm for Evaluation of Student Performance. *International Journal of Computers & Technology*, Volume 4 No. 2. ISSN 2277-3061
- [12] Quinlan, J. 2014. C4.5: Program for Machine Learning 1st Edition. Morgan Kaufman Publishers Inc.
- [13] Cao, Y., Wu, J., 2004. Dynamics of projective Adaptive Resonance Theory Model: The Foundation of PART Algorithm. *IEEE Transactions on Neural Networks*.
- [14] [14] F. Gorunescu, Data Mining: Concepts, Models and Techniques. Berlin Heidelberg: Springer - Verlag, 2011.