

Klasifikasi Mahasiswa Berpotensi *Drop Out* Menggunakan Algoritma C4.5 di Politeknik Negeri Subang

Amelia Ramadhani¹, Reza Fazarany Noor², Dwi Vernanda³, Tri Herdiawan⁴

¹Jurusan Teknologi Informasi dan Komputer, Program Studi Sistem Informasi, Politeknik Negeri Subang, Subang, Indonesia

Email: ^{1*}amelia.10108008@student.polsub.ac.id, ²reza.10108046@student.polsub.ac.id, ³nanda@polsub.ac.id, ⁴tri@polsub.ac.id

Email Utama: reza.10108046@student.polsub.ac.id

Abstrak– Penelitian ini dimulai dengan memperkenalkan konteks masalah, yaitu tingkat keberhasilan dan kegagalan mahasiswa di perguruan tinggi, khususnya di Politeknik Negeri Subang. Kegagalan mahasiswa, yang dapat menyebabkan drop out, menjadi fokus utama karena dapat mempengaruhi mutu pendidikan dan akreditasi perguruan tinggi. Politeknik Negeri Subang, didirikan pada tahun 2014, memiliki jurusan Teknologi Informasi dan Komputer dengan program studi Sistem Informasi (SI). Data mahasiswa SI untuk tahun akademik 2022/2023 menjadi dasar penelitian ini. Tingkat keberhasilan mahasiswa menjadi tantangan di lingkungan perguruan tinggi yang sangat kompetitif. Identifikasi faktor penyebab kegagalan mahasiswa di Politeknik Negeri Subang menjadi kunci untuk mengatasi tantangan ini dan terus meningkatkan kualitas pendidikan. Buku panduan akademik Politeknik Negeri Subang membagi alasan drop out menjadi tiga poin utama, termasuk nilai indeks prestasi semester, jumlah SKS, dan nilai akademik. Penelitian ini bertujuan untuk melakukan klasifikasi pada kasus drop out menggunakan algoritma C4.5 guna mengidentifikasi faktor yang paling berpengaruh. Dalam konteks ini, data mining, khususnya metode klasifikasi dengan algoritma C4.5, dianggap sebagai pendekatan yang efektif untuk menganalisis data mahasiswa yang berpotensi drop out. Penelitian ini juga merujuk pada penelitian terdahulu yang menggunakan algoritma C4.5 untuk klasifikasi mahasiswa berpotensi drop out. Selanjutnya, penelitian ini menetapkan lima atribut sebagai variabel independen, yaitu IPK, SKS, jenis tempat tinggal, jenis transportasi, dan status mahasiswa. Metode penelitian mencakup beberapa tahap, termasuk pengumpulan data dari Sistem Informasi Akademik (SIKAD) Politeknik Negeri Subang, pra-proses data, pembagian data menjadi data training dan data testing, serta klasifikasi data menggunakan algoritma Decision Tree C4.5. Penelitian ini mengambil data mahasiswa tahun akademik 2022/2023 dengan total 976 data. Proses pra-proses data melibatkan data selection, data cleaning, dan data transformation. Setelah pembagian data, dilakukan klasifikasi menggunakan algoritma Decision Tree C4.5. Penelitian ini mengevaluasi performa model dengan menggunakan data testing dan menghasilkan confusion matrix. Hasil evaluasi menunjukkan akurasi sebesar 98.50%, dengan precision untuk Drop Out dan Tidak Drop Out masing-masing 93.18% dan 100.00%. Recall dari data Drop Out dan Tidak Drop Out adalah 64.44% dan 100.00%. Pohon keputusan merupakan output dari penelitian ini, yang dapat memberikan pemahaman lebih lanjut mengenai faktor-faktor yang mempengaruhi mahasiswa drop out. Penelitian ini diharapkan dapat memberikan kontribusi positif dalam upaya mengurangi angka drop out di Politeknik Negeri Subang dan perguruan tinggi pada umumnya.

Kata Kunci: *Drop Out*, Data Mining, Algoritma C4.5, Decision Tree, Kualitas Pendidikan

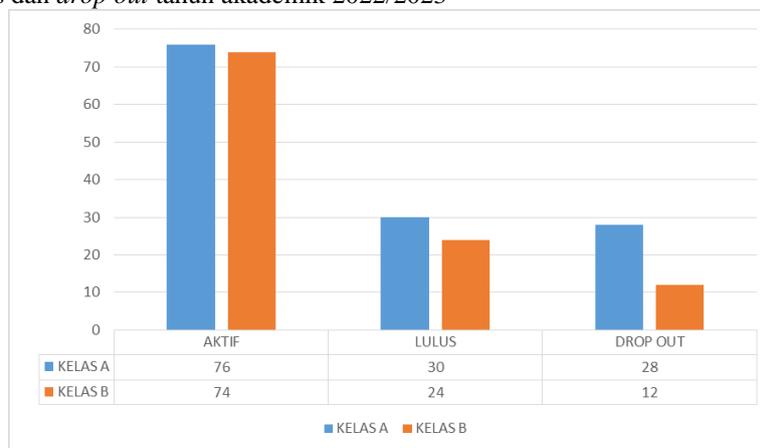
Abstract– This research begins by introducing the context of the problem, namely the success and failure rates of students in higher education, especially at the Subang State Polytechnic. Student failure, which can lead to dropping out, is the main focus because it can affect the quality of education and higher education accreditation. Subang State Polytechnic, founded in 2014, has a Department of Information and Computer Technology with an Information Systems (SI) study program. SI student data for the 2022/2023 academic year is the basis for this research. The level of student success is a challenge in a very competitive higher education environment. Identifying the factors that cause student failure at Subang State Polytechnic is the key to overcoming this challenge and continuing to improve the quality of education. The Subang State Polytechnic academic guidebook divides the reasons for dropping out into three main points, including semester achievement index, number of credits, and academic grades. This research aims to classify drop out cases using the C4.5 algorithm to identify the most influential factors. In this context, data mining, especially the classification method with the C4.5 algorithm, is considered an effective approach for analyzing data on students who have the potential to drop out. This research also refers to previous research which used the C4.5 algorithm to classify students who have the potential to drop out. Furthermore, this research determines five attributes as independent variables, namely GPA, SKS, type of residence, type of transportation, and student status. The research method includes several stages, including data collection from the Subang State Polytechnic Academic Information System (SIKAD), data preprocessing, dividing data into training data and testing data, as well as data classification using the C4.5 Decision Tree algorithm. This research took student data for the 2022/2023 academic year with a total of 976 data. The data preprocessing process involves data selection, data cleaning, and data transformation. After dividing the data, classification is carried out using the Decision Tree C4.5 algorithm. This research evaluates model performance using testing data and produces a confusion matrix. The evaluation results show an accuracy of 98.50%, with precision for Drop Out and No Drop Out of 93.18% and 100.00% respectively. Recall of Drop Out and Non-Drop Out data is 64.44% and 100.00%. The decision tree is the output of this research, which can provide further understanding regarding the factors that influence student dropout. It is hoped that this research can make a positive contribution in efforts to reduce the drop out rate at Subang State Polytechnic and universities in general.

Keywords: *Drop Out*, C4.5 Algorithm, Decision Tree, Data Mining, Quality of Education

1. PENDAHULUAN

Tingkat keberhasilan mahasiswa dan tingkat kegagalan mahasiswa di perguruan tinggi sangat erat kaitannya dengan mutu pendidikan yang diberikan. Salah satu tanda kegagalan mahasiswa adalah ketidak mampuan mereka untuk menyelesaikan studi sesuai dengan waktu yang ditentukan, sehingga dapat mengakibatkan *drop out*. *Drop out* (DO) atau pemberhentian status mahasiswa adalah kondisi di mana perguruan tinggi mencabut status kemahasiswaan seseorang karena beberapa alasan yang telah ditetapkan. Untuk mengurangi risiko *drop out*, perguruan tinggi bisa mengimplementasikan kebijakan atau peraturan terkait standar maksimum dalam pembelajaran.

Politeknik Negeri Subang yang didirikan pada tahun 2014 adalah perguruan tinggi negeri pertama di Kabupaten Subang dan mempunyai 4 jurusan. Politeknik Negeri Subang merupakan Pendidikan Tinggi Vokasi Program Diploma III yang bertujuan untuk menghasilkan tenaga ahli, berpengalaman, mempunyai lapangan kerja dan mampu bersaing dalam bidang studinya dalam kurun waktu 3 tahun atau 6 semester. Jurusan Teknologi Informasi dan Komputer adalah jurusan yang melakukan pemanfaatan dan pemeliharaan atas sumber daya secara efektif untuk bidang informasi demi mewujudkan tujuan yang telah ditentukan agar dapat memperoleh hasil yang diharapkan secara maksimal. Sistem Informasi (SI) merupakan salah satu program studi yang ada di Jurusan Teknologi Informasi dan Komputer. Berikut jumlah mahasiswa di program studi Sistem Informasi (SI) yang berstatus aktif, lulus dan *drop out* tahun akademik 2022/2023



Gambar 1. Data Mahasiswa Sistem Informasi Status Aktif, Lulus, Drop Out

Pentingnya tingkat keberhasilan mahasiswa dan identifikasi faktor penyebab kegagalan mahasiswa di Politeknik Negeri Subang merupakan sebuah tantangan kompetitif dalam meningkatkan kualitas pendidikan. Saat ini, Politeknik Negeri Subang beroperasi dalam lingkungan yang sangat kompetitif, di mana setiap perguruan tinggi berusaha keras untuk terus memperbaiki manajemen mereka guna meningkatkan mutu pendidikan dan meraih akreditasi yang lebih baik. Oleh karena itu, sangat penting untuk mengidentifikasi faktor-faktor yang memengaruhi kegagalan mahasiswa di Politeknik Negeri Subang agar dapat mengatasi tantangan ini dan terus berinovasi dalam meningkatkan kualitas pendidikan[1]. Berdasarkan buku panduan akademik Politeknik Negeri Subang alasan mahasiswa *Drop Out* terbagi tiga poin yaitu;

- 1) *Drop Out* karena IPS < 1,75 untuk semester 1 dan 2
- 2) *Drop Out* karena IPS < 2 untuk semester 3 ke atas
- 3) *Drop Out* karena jumlah SKS Nilai D > 8 atau memiliki nilai E

Kasus *drop out* juga dapat terjadi karena berbagai faktor seperti rendahnya kemampuan akademik, faktor biaya, domisili saat menempuh studi dan lain sebagainya. Untuk itu dalam penelitian ini akan dilakukan klasifikasi pada kasus *drop out* agar dapat diketahui faktor apa yang paling berpengaruh terhadap kasus tersebut sehingga dapat diambil kebijakan untuk mencegah ataupun mengurangi potensi mahasiswa yang berpotensi *drop out*.

Berdasarkan uraian di atas, peluang mahasiswa *drop out* dapat menimbulkan permasalahan yang dapat menurunkan mutu pendidikan dan akreditasi perguruan tinggi. Penting untuk melakukan analisis terhadap data mahasiswa yang telah *drop out* guna memahami karakteristik mereka secara dini. Jurusan Teknologi Informasi dan Komputer memiliki jumlah data mahasiswa yang cukup besar, sehingga dapat dimanfaatkan dalam penerapan data *mining*. Data *mining* merupakan metode untuk menggali informasi dari data yang berjumlah besar. Salah satu metode yang digunakan dalam data *mining* adalah klasifikasi, yang bertujuan untuk mengekstrak dan memprediksi kategori label untuk setiap data[2]. Dalam hal ini, algoritma klasifikasi C4.5 dapat digunakan untuk mencari model yang dapat membedakan kelas data, sehingga model tersebut dapat digunakan untuk memprediksi kelas dari data yang belum diketahui. Dengan memanfaatkan C4.5, analisis data dapat menjadi lebih efektif dan membantu dalam mengidentifikasi mahasiswa yang berpotensi *drop out*[3].

Salah satu penelitian yang berkaitan yaitu penelitian menurut Nasrullah tahun 2018 dalam klasifikasi mahasiswa berpotensi *drop out* menggunakan algoritma C4.5 yang terdiri dari 9 atribut penelitian. Dalam penelitian tersebut atribut paling berpengaruh terhadap potensi *drop out* ialah nilai IPS semester IV dengan menghasilkan 17 *rule* mahasiswa berpotensi *drop out*.

Penelitian lain yang serupa juga dilakukan oleh Putra tahun 2017 mengenai prediksi mahasiswa *drop out* menggunakan algoritma C4.5 dengan 4 atribut penelitian. Hasil penelitian menunjukkan bahwa mahasiswa yang memiliki jumlah SKS dan jumlah semester yang tidak baik berpotensi mengalami *drop out*. Kemudian pada penelitian Wahyuni, Saragih dan PeranginAngin tahun 2018 dalam menganalisa mahasiswa *drop out* menggunakan algoritma decision tree C4.5 dengan 6 atribut penelitian. Atribut paling berpengaruh terhadap mahasiswa yang mengalami *drop out* ialah asal sekolah[4].

Dari penelitian tersebut dapat dilihat bahwa algoritma C4.5 memiliki akurasi yang baik dalam melakukan klasifikasi. Oleh karena itu, penulis akan melakukan “Klasifikasi Mahasiswa Berpotensi *Drop out* Menggunakan Algoritma C4.5 di Politeknik Negeri Subang” bertujuan untuk mengevaluasi hasil dari penerapan algoritma C4.5 dalam melakukan klasifikasi terhadap mahasiswa yang memiliki potensi untuk mengalami *drop out*. Dengan demikian, penelitian ini diharapkan dapat menghasilkan sistem atau aplikasi yang dapat digunakan oleh Jurusan Teknologi Informasi dan Komputer serta seluruh Jurusan di Politeknik Negeri Subang dalam mengelompokkan mahasiswa yang berpotensi *drop out*. Selain itu, penelitian ini juga diharapkan dapat memberikan panduan bagi pengambil kebijakan dalam merumuskan strategi terkait permasalahan *drop out*. Dengan menggunakan algoritma C4.5, diharapkan dapat ditemukan pola dan karakteristik yang dapat memprediksi kemungkinan seorang mahasiswa mengalami *drop out*. Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi positif dalam upaya mengurangi angka *drop out* di lingkungan akademik[5]. Penelitian ini memiliki 5 atribut yaitu, IPK, SKS, Jenis Tempat Tinggal, Jenis Transportasi, dan Status Mahasiswa. Perbedaan dengan penelitian sebelumnya pada penelitian ini atribut yang paling berpengaruh adalah IPK yang paling berpengaruh di Politeknik Negeri Subang.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Adapun tahapan penelitian yang dilakukan oleh peneliti disajikan pada Gambar 2 di bawah ini.

2.2 Pengumpulan Data

Data dalam penelitian ini merupakan data sekunder yang diperoleh dari Sistem Informasi Akademik (SIKAD) Politeknik Negeri Subang. Data yang digunakan didasarkan pada Politeknik Negeri Subang yaitu:

- 1) *Drop Out* karena $IPS < 1,75$ untuk semester 1 dan 2,
- 2) *Drop Out* karena $IPS < 2$ untuk semester 3 ke atas,
- 3) *Drop Out* karena jumlah SKS Nilai $D > 8$ atau memiliki nilai E.

Data yang diambil adalah data yang didapatkan dari rekap data mahasiswa tahun akademik 2022/2023. Sasaran yang utama yaitu nilai semester I – semester VI dan IPK. Data rekap mahasiswa dikumpulkan dengan sejumlah 976 data mahasiswa yang terkumpul.

2.3 Praproses Data

Praproses data merupakan proses untuk persiapan data sebelum dilakukannya proses klasifikasi. Pada penelitian ini dilakukan praproses data yaitu data *selection*, data *cleaning*, dan data *transformation*.

1) Data *Selection*

Data *Selection* adalah proses seleksi data yang sesuai untuk dianalisis yang diambil dari sumber data atau database, seleksi data dilakukan karena tidak semua data yang ada pada database digunakan[6].

2) Data *Cleaning*

Data *cleaning* merupakan Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan pada data, seperti kesalahan cetak. juga dilakukan proses memperkaya data yang sudah ada dengan data atau informasi yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal[7].

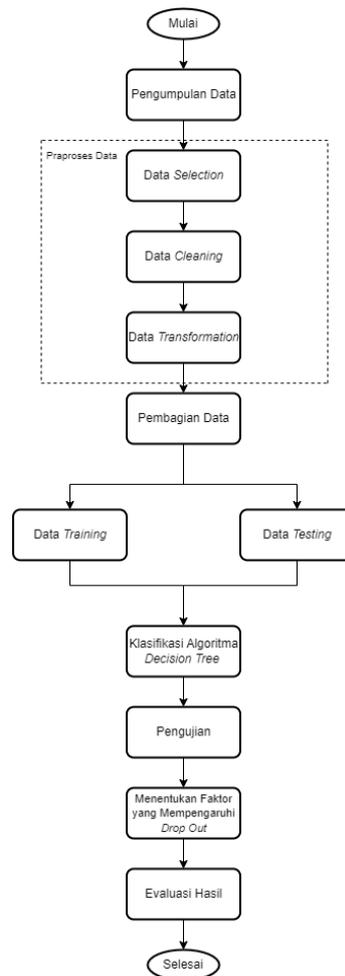
3) Data *Transformation*

Data *transformation* dilakukan untuk mengubah bentuk dan format data. Hal ini tentunya sangat membantu memudahkan pengguna dalam proses mining ataupun memahami hasil yang didapat. Dalam proses data *transformation* bisa dilakukan dengan *centering*, *normalization*, dan *scaling*[8].

4) Data *Mining*

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar. Data mining adalah bagian

dari proses KDD (Knowledge Discovery in Databases) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, data mining, dan evaluasi hasil[9].



Gambar 2. Tahapan Penelitian

2.4 Pembagian Data

Pada tahap ini, data akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*.

- 1) Data *training* merupakan sekumpulan data yang memiliki atribut label atau kelas yang digunakan oleh mesin untuk mengenal karakteristik kumpulan data sehingga menghasilkan sebuah pola atau model data[10].
- 2) Data *testing* adalah sekumpulan data yang juga memiliki label atau kelas yang digunakan untuk menguji ketepatan pola atau model dalam mengklasifikasi data *testing*[11].

2.5 Klasifikasi Data

Pada tahap ini dilakukan klasifikasi data menggunakan algoritma *Decision tree* C4.5. Algoritma C4.5 merupakan salah satu solusi pemecahan kasus yang sering digunakan dalam pemecahan masalah pada teknik klasifikasi. Keluaran dari algoritma C4.5 itu berupa sebuah *decision tree* layaknya teknik klasifikasi lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. Adapun penjelasan tentang Algoritma C4.5 itu sendiri yaitu salah satu algoritma C4.5 induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). Input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Jika suatu set data mempunyai beberapa pengamatan dengan *missing value* yaitu *record* dengan beberapa nilai *variable* tidak ada, jika jumlah pengamatan terbatas maka atribut dengan *missing value* dapat diganti dengan nilai rata-rata dari *variable* yang bersangkutan. Algoritma C4.5 merupakan salah satu metode untuk membuat *decision tree* berdasarkan *training* data yang telah disediakan. Algoritma C4.5 dimunculkan oleh Ross Quinlan dan dengan adanya algoritma C4.5 dapat mengatasi *missing value*, *continue* data, dan *running*[12]. Tahapan algoritma C4.5 dalam membuat *decision tree* sebagai berikut:

- 1) Memilih attribute sebagai *root*

- 2) Membuat cabang untuk setiap nilai
- 3) Mendistribusikan kasus pada cabang
- 4) Mengulangi proses pada setiap cabang sehingga seluruh *instance* pada cabang mempunyai kelas yang sama.

Nilai gain tertinggi dari *attribute* yang ada, digunakan untuk pemilihan *root attribute* sesuai

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^N Entropy(S_i) \quad (1)$$

Dimana:

S : Kumpulan kasus

A : Attribute

N : Jumlah partisi attribute A

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

Sr : Jumlah kasus dalam S

2.6 Pengujian

Pada tahap ini dilakukan pengujian menggunakan aplikasi *RapidMiner* untuk menguji performa model klasifikasi *Decision Tree* yang telah dibangun, serta untuk mengidentifikasi sejauh mana model ini dapat mengelompokkan data dengan akurat dan mengungkap pola-pola yang relevan. Pengujian ini merupakan tahap penting untuk memvalidasi hasil analisis dan memastikan keakuratan interpretasi yang akan dilakukan terhadap hasil klasifikasi data.

2.7 Menentukan Faktor Yang Mempengaruhi Drop Out

Faktor yang mempengaruhi mahasiswa *drop out* akan ditentukan berdasarkan implementasi algoritma dengan hasil performa paling baik.

2.8 Evaluasi Hasil

Setelah semua tahap selesai dilakukan, maka dapat ditarik kesimpulan dari penelitian tersebut mengenai klasifikasi mahasiswa berpotensi *drop out* dari algoritma *Decision Tree* menggunakan teknik *confusion matrix* dengan menghasilkan nilai performa berupa *accuracy*, *precision*, dan *recall*. Hasil yang diperoleh harus dievaluasi untuk menentukan apakah model yang dihasilkan memiliki akurasi dan kesesuaian dengan data asli atau tidak.

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan mengenai tahapan penelitian untuk memperoleh hasil klasifikasi mahasiswa berpotensi drop out menggunakan metode algoritma C4.5. Adapun tahapan tersebut adalah tahap persiapan dataset, penentuan variabel, penentuan *rule* dan klasifikasi. Dataset yang digunakan adalah data mahasiswa yang diperoleh dari rekap data mahasiswa Jurusan Teknologi Informasi dan Komputer tahun akademik 2022/2023.

3.1 Pengumpulan Data

Data yang dikumpulkan dalam penelitian ini ialah rekap data mahasiswa tahun akademik 2022/2023 yang diperoleh dari Ketua Program Studi Sistem Informasi Politeknik Negeri Subang. Variabel atau atribut penelitian yang digunakan yaitu data akademik mahasiswa Jurusan Teknologi Informasi dan Komputer Program Studi Sistem Informasi tahun akademik 2022/2023.

Variabel penelitian tersebut dapat dilihat pada tabel berikut:

Tabel 1. Variabel Penelitian

No.	Variabel	Keterangan
1.	Indeks Prestasi Kumulatif (IPK)	Nilai prestasi belajar secara kumulatif yang memiliki rentang nilai 0-4
2.	Jumlah Satuan Kredit Semester (SKS)	Jumlah mata kuliah yang telah di kontrak.
3.	Jenis Tempat Tinggal	Menjelaskan tempat tinggal mahasiswa saat ini yaitu

		bersama orang tua, wali, kos atau sebagainya.
4.	Jenis Transportasi	Menjelaskan jenis transportasi mahasiswa yang digunakan, selama masa studi yaitu jalan kaki, angkutan umum, atau sepeda motor.
5.	Status Mahasiswa	Sebagai class target yang menjelaskan bahwa mahasiswa berpotensi dropout atau tidak dropout.

a. Indeks Prestasi Kumulatif (IPK)

Penggunaan nilai IPK sebagai variabel dalam penelitian ini didasarkan pada penelitian Sawitri (2019) dengan judul Analisis Penentuan Putus Studi (*Drop Out*) Mahasiswa Sekolah Pascasarjana Universitas Sumatera Utara menggunakan Algoritma Data *Mining* C4.5. Variabel yang digunakan dalam penelitian tersebut ialah Lama Studi, Beban Studi, Status, IPK dan Putus Studi. Dari penelitian tersebut didapat hasil bahwa variabel yang paling berpengaruh adalah IPK.

b. Jumlah Satuan Kredit Semester (SKS)

Penggunaan SKS Total sebagai variabel didasarkan pada penelitian Putra (2017) dengan judul Solusi Prediksi Mahasiswa *Dropout* Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma dengan variabel yang digunakan yaitu IPK, Jumlah SKS yang telah di tempuh, semester dan status perkuliahan mahasiswa. Hasil dalam penelitian tersebut menyatakan bahwa salah satu penyebab potensi *drop out* ialah mahasiswa yang memiliki jumlah SKS yang tidak baik

c. Jenis Tempat Tinggal

Variabel jenis tempat tinggal merupakan variabel yang berhubungan dengan faktor motivasi keluarga. Dalam penelitian ini jenis tempat tinggal yang dimaksud ialah bagaimana pengaruh mahasiswa yang tinggal bersama orang tua dan yang tidak tinggal bersama orang tua terhadap potensi mahasiswa *drop out*. Untuk penggunaan variabel tersebut didasarkan pada penelitian Hidayati (2010) dengan judul Faktor–Faktor yang Berpengaruh Terhadap Keberhasilan Studi Mahasiswa Prodi D.III Kebidanan FIK UMSurabaya. Hasil dalam penelitian tersebut menyatakan bahwa faktor paling berpengaruh terhadap keberhasilan studi mahasiswa ialah faktor motivasi dari orang tua.

d. Jenis Transportasi

Variabel Transportasi merupakan variabel yang berhubungan dengan faktor keterlambatan selama jam kuliah berdasarkan domisili yang ditempuh selama masa studi menggunakan jenis transportasi paling berpengaruh.

e. Status Mahasiswa

Status mahasiswa merupakan variabel yang digunakan untuk menentukan *class* target yang terdiri dari class DO (*Drop Out*) dan Tidak *Drop Out*. Penentuan class target didasarkan pada Peraturan Rektor Nomor 02 Tahun 2017 mengenai Evaluasi Hasil Belajar.

3.2 Praproses Data

Praproses data merupakan proses mempersiapkan data sebelum dilakukannya proses klasifikasi. Praproses data dalam penelitian ini terdiri dari data *selection*, data *cleaning*, dan data *transformation*.

a. Data *Selection*

Data *selection* atau seleksi data adalah proses pemilihan data yang benar- benar diperlukan dan relevan untuk proses klasifikasi. Data yang diperoleh tidak seluruhnya digunakan, hanya data yang sesuai dengan proses klasifikasi yang digunakan. Dalam penelitian ini akan dilakukan beberapa proses seleksi data diantaranya menentukan atribut target yaitu status mahasiswa yang dikelompokkan kedalam dua *class* yakni *Drop Out* dan Tidak *Drop Out*. Data yang diperlukan untuk menentukan atribut status mahasiswa ialah IPK, SKS, jenis tempat tinggal dan jenis transportasi.

b. Data *Cleaning*

Data *cleaning* merupakan tahap pembersihan data dimana dalam penelitian ini terdapat data kosong (*missing value*) perlu dihapus atau dihilangkan karena keberadaannya dalam penelitian ini dianggap tidak diperlukan. Pada tahap ini dilakukan data *preprocessing* dengan merubah status mahasiswa dikeluarkan menjadi *Drop Out* dan status mahasiswa lulus menjadi Tidak *Drop Out*. Kemudian melakukan penanganan *missing value* dengan imputasi, yaitu mengisi data jenis tempat tinggal yang kosong sebanyak 109 *record*, mengisi data jenis transportasi yang kosong sebanyak 60 *record*. Karena datanya bersifat kategorikal, maka pengisian dilakukan dengan data yang populer. Dari seluruh data yang diperoleh sehingga total data setelah dilakukan *Cleaning Data* menjadi 976 data.

c. *Data Transformation*

Data *transformation* merupakan proses mengubah data menjadi bentuk yang di perlukan. Dalam penelitian ini, peneliti mengubah data IPK dan SKS.

3.3 Pembagian Data

Setelah dilakukan praproses data, selanjutnya akan dilakukan pembagian data. Pada tahap ini, data akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Pembagian data *training* dan data *testing* ini berdasarkan atribut target yang telah memiliki *class* data. Data *training* merupakan data yang digunakan untuk melatih algoritma. Tujuannya agar algoritma dapat mempelajari pola dari data yang diberikan. Sedangkan data *testing* merupakan data yang digunakan untuk melihat performa dari algoritma yang telah dilatih. Dalam penelitian ini data akan di bagi dengan proporsi 80% data *training* dan 20% data *testing*. Berikut jumlah data setelah dilakukan pembagian.

Tabel 2. Pembagian Data

Klasifikasi	Jumlah Data	Data Training (80%)	Data Testing (20%)
Drop Out	188	150	38
Tidak Drop Out	788	630	158
Total	976	780	196

Dari tabel diatas diketahui bahwa jumlah data *training* yang dihasilkan adalah 780 data dan data *testing* berjumlah 196 data.

3.4 Klasifikasi Menggunakan Algoritma *Decision Tree*

Klasifikasi pada algoritma *Decision Tree* dilakukan dengan beberapa tahap, mulai dari menentukan variabel sampai dengan penarikan kesimpulan. Adapun tahapan-tahapan klasifikasi adalah sebagai berikut:

a. Menentukan Data *Training*

Data *Training* digunakan sebagai data acuan dalam klasifikasi ini. Data yang digunakan sebanyak 780 data mahasiswa Program Studi Sistem Informasi Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Subang.

Tabel 3. Menentukan Data Training

No.	IPK	SKS	Jenis Tempat Tinggal	Jenis Transportasi	Status Mahasiswa
1	Cukup	Memenuhi	Orang Tua	Sepeda Motor	Tidak <i>Drop Out</i>
2	Kurang	Memenuhi	Orang Tua	Sepeda Motor	<i>Drop Out</i>
3	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
4	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
5	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	<i>Drop Out</i>
6	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak <i>Drop Out</i>
7	Baik	Memenuhi	Kost	Angkutan Umum	Tidak <i>Drop Out</i>
8	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak <i>Drop Out</i>
9	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
10	Kurang	Memenuhi	Orang Tua	Jalan Kaki	<i>Drop Out</i>
11	Baik	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
12	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
13	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak <i>Drop Out</i>
14	Cukup	Memenuhi	Orang Tua	Sepeda Motor	Tidak <i>Drop Out</i>
15	Kurang	Tidak Memenuhi	Kost	Jalan Kaki	<i>Drop Out</i>

16	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
17	Kurang	Tidak Memenuhi	Kost	Sepeda Motor	Drop Out
18	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
19	Kurang	Tidak Memenuhi	Kost	Jalan Kaki	Drop Out
20	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
21	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
22	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
23	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
24	Kurang	Memenuhi	Kost	Jalan Kaki	Drop Out
25	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
26	Kurang	Memenuhi	Orang Tua	Jalan Kaki	Drop Out
27	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
28	Cukup	Memenuhi	Kost	Angkutan Umum	Drop Out
29	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
30	Baik	Memenuhi	Kost	Angkutan Umum	Tidak Drop Out
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
780	Baik	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out

b. Menentukan Data *Testing*

Setelah proses *training* dilakukan pada sebuah algoritma *machine learning*, tahap selanjutnya adalah melakukan evaluasi terhadap performa algoritma tersebut atau biasa disebut *testing*. Pada proses *testing*, performa algoritma akan diuji menggunakan data *testing*, dimana data *testing* dan data *training* merupakan data yang berbeda. Data yang digunakan dalam data *testing* adalah sebanyak 196 data mahasiswa Program Studi Sistem Informasi Jurusan Teknologi Informasi dan Komputer Politeknik Negeri Subang.

Tabel 4. Menentukan Data Testing

No.	IPK	SKS	Jenis Tempat Tinggal	Jenis Transportasi	Status Mahasiswa
1	Cukup	Memenuhi	Orang Tua	Sepeda Motor	Tidak Drop Out
2	Kurang	Memenuhi	Orang Tua	Sepeda Motor	Drop Out
3	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
4	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
5	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
6	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
7	Baik	Memenuhi	Kost	Angkutan Umum	Tidak Drop Out
8	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
9	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
10	Kurang	Memenuhi	Orang Tua	Jalan Kaki	Drop Out
11	Baik	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
12	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
13	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
14	Cukup	Memenuhi	Orang Tua	Sepeda Motor	Tidak Drop Out
15	Kurang	Tidak Memenuhi	Kost	Jalan Kaki	Drop Out
16	Kurang	Tidak Memenuhi	Orang Tua	Jalan Kaki	Drop Out
17	Kurang	Tidak Memenuhi	Kost	Sepeda Motor	Drop Out
18	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
19	Kurang	Tidak Memenuhi	Kost	Jalan Kaki	Drop Out
20	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
21	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
22	Cukup	Memenuhi	Kost	Sepeda Motor	Tidak Drop Out
23	Baik	Memenuhi	Orang Tua	Angkutan Umum	Tidak Drop Out
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
196	Baik	Tidak Memenuhi	Kost	Sepeda Motor	Drop Out

3.5 Mencari Kelas Prediksi

Setelah nilai probabilitas Tidak *Drop Out* dan *Drop Out* dihitung, tahap berikutnya adalah membandingkan kedua nilai tersebut. Jika nilai probabilitas Tidak *Drop Out* lebih besar dari probabilitas *Drop Out*, maka data akan diklasifikasikan sebagai data Tidak *Drop Out*. Sebaliknya jika probabilitas *Drop Out* lebih besar dari probabilitas Tidak *Drop Out*, maka data akan diklasifikasikan sebagai data *Drop Out*.

Label	Binominal	0	Negative Tidak Drop Out	Positive Drop Out	Value Tidak Drop Out (159), Drop Out (41)
Prediction	Binominal	0	Negative Tidak Drop Out	Positive Drop Out	Value Tidak Drop Out (156), Drop Out (44)
Confidence_Tidak Drop Out	Real	0	Min 0	Max 1	Average 0.776
Confidence_Drop Out	Real	0	Min 0	Max 1	Average 0.224
Jenis Kelamin	Polynomial	0	Least Perempuan (100)	Most Laki-Laki (100)	Value Laki-Laki (100), Perempuan (100)
IPK	Polynomial	0	Least Kurang (39)	Most Baik (98)	Value Baik (98), Cukup (63), ...[1 more]
SKS	Polynomial	0	Least Tidak Memenuhi (41)	Most Memenuhi (159)	Value Memenuhi (159), Tidak Memenuhi (41)
Jenis Tempat Tinggal	Polynomial	0	Least Orang Tua (97)	Most Kost (103)	Value Kost (103), Orang Tua (97)

Gambar 3. Kelas Prediksi

Berdasarkan gambar diatas data yang didapatkan bahwa ada 3 data yang sebenarnya Tidak *Drop Out* tapi hasil prediksinya *Drop Out*. Hal ini terjadi dikarenakan penelitian ini menggunakan parameter SKS. Mahasiswa yang sebenarnya Tidak *Drop Out* tetapi hasil prediksi *Drop Out* disebabkan karena mahasiswa tersebut sering tidak aktif mengikuti proses perkuliahan sehingga SKS tidak memenuhi walaupun IPK semester mahasiswa tersebut baik, sehingga hasilnya mahasiswa tersebut di *drop out*.

3.6 Menghitung Confusion Matriks

Hasil evaluasi klasifikasi menggunakan algoritma *decision tree* C4.5 dapat dilihat pada tabel berikut.

Tabel 5. Confusion Matrix

Klasifikasi	Data Sebenarnya		Class Precision
	Drop Out	Tidak Drop Out	
Drop Out	41	3	93.18%
Tidak Drop Out	0	156	100.00%
Class Recall	100.00%	98.11%	

Dari hasil klasifikasi pada tabel *confussion matrix* di atas dapat dilihat bahwa:

- True Positive* (TP) menjelaskan dimana data terklasifikasi *Drop Out*, memang benar *Drop Out*. Dalam hal ini jumlah data yang didapat sebanyak 41 data.
- False Positive* (FP) menjelaskan bahwa data yang terklasifikasi *Drop Out* ternyata Tidak *Drop Out*. Jumlah data yang didapat sebesar 0 data.
- False Negative* (FN) menjelaskan bahwa data yang terklasifikasi Tidak *Drop Out* sebenarnya adalah *Drop Out*. Data yang dihasilkan berjumlah 3 data.
- True Negative* (TN) menjelaskan dimana data yang terklasifikasi Tidak *Drop Out*, memang benar Tidak *Drop Out*. Jumlah data yang didapat sebanyak 156 data

Dari hasil *confussion matrix* yang telah diperoleh, dihasilkan nilai *accuracy*, *precision*, dan *recall*. Hasil *accuracy* yang didapat 98.50% dengan nilai *precision* untuk data *Drop Out* dan Tidak *Drop Out* masing-masing sebesar 93.18% dan 100.00%. Nilai *recall* dari data *Drop Out* dan Tidak *Drop Out* yaitu 64,44% dan 100.00%.

3.7 Pohon Keputusan

Metode pohon keputusan adalah metode yang dapat mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturannya mudah dipahami dalam bahasa alami. Pohon keputusan adalah salah satu teknik klasifikasi sebagai bagian dari ilmu data mining, yaitu ilmu yang melakukan penggalian pengetahuan (knowledge) terhadap data. Pohon keputusan menggunakan parameter yang digunakan untuk mengukur jumlah keberagaman atau keberbedaan dalam sebuah himpunan data. Kelebihan dari

penggunaan model pohon keputusan ini selain mudah dipahami juga dapat digunakan untuk menemukan aturan atau syarat-syarat yang dapat dijadikan sebagai kriteria yang berguna untuk keperluan analisa dalam suatu proses pengambilan keputusan[13]. Pohon keputusan adalah struktur yang dapat digunakan untuk membagi kumpulan data besar menjadi kumpulan catatan yang lebih kecil menggunakan beberapa aturan keputusan. Tahapan ini dilakukan uji pohon keputusan dengan menggunakan Aplikasi *RapidMiner*. Hasil uji pohon keputusan yang dilakukan oleh algoritma *Decision Tree C4.5* di aplikasi *Rapid Miner Studio* menunjukkan bahwa variabel *IPK* menjadi *root*.



Gambar 4. Pohon Keputusan

Berdasarkan hasil tersebut, *IPK* menjadi *root*. Adapun rules atau aturan yang dihasilkan dari uji pohon keputusan dapat dilihat pada Tabel 3.2 dimana diperoleh 7 rules atau aturan untuk klasifikasi mahasiswa berpotensi *Drop out* yaitu:

Tabel 6. Rules Prediksi

Rules	Deskripsi
Rules 1	IF <i>IPK</i> “Baik” Then Mahasiswa Tidak <i>Drop Out</i>
Rules 2	IF <i>IPK</i> “Kurang” Then Mahasiswa <i>Drop Out</i>
Rules 3	IF <i>IPK</i> “Cukup” dan <i>SKS</i> “Memenuhi” Then Mahasiswa Tidak <i>Drop Out</i>
Rules 4	IF <i>SKS</i> “Tidak Memenuhi” dan Jenis Transportasi “Angkutan Umum” Then Mahasiswa <i>Drop Out</i>
Rules 5	IF <i>SKS</i> “Tidak Memenuhi” dan Jenis Transportasi “Sepeda Motor” Then Mahasiswa <i>Drop Out</i>
Rules 6	IF <i>SKS</i> “Tidak Memenuhi”, Jenis Transportasi “Jalan Kaki”, dan “Jenis Tempat Tinggal “Kost” Then Mahasiswa <i>Drop Out</i>

Rules 7 IF SKS “Tidak Memenuhi”, Jenis Transportasi “Jalan Kaki”, dan Jenis Tempat Tinggal “Orang Tua” Then Mahasiswa Tidak *Drop Out*

Aturan keputusan diatas merupakan 7 aturan (*rules*) yang terbentuk dari pohon keputusan. Pada aturan pertama dimana jika IPK “Baik” maka status mahasiswa “Tidak *Drop Out*”, jika IPK “Kurang” maka status mahasiswa “*Drop Out*”, jika IPK “Cukup” dan SKS “Memenuhi” maka status mahasiswa “Tidak *Drop Out*”, jika SKS “Tidak Memenuhi” dan Jenis Transportasi “Angkutan Umum” maka status mahasiswa “*Drop Out*”, jika SKS “Tidak Memenuhi” dan Jenis Transportasi “Sepeda Motor” maka status mahasiswa “*Drop Out*”, jika SKS “Tidak Memenuhi” Jenis Transportasi “Jalan Kaki” dan “Jenis Tempat Tinggal “Kost” maka status mahasiswa “*Drop Out*”, jika SKS “Tidak Memenuhi” Jenis Transportasi “Jalan Kaki” dan Jenis Tempat Tinggal “Orang Tua” maka status mahasiswa “Tidak *Drop Out*”. Aturan-aturan ini membantu mengklasifikasi mahasiswa berdasarkan faktor-faktor seperti IPK, jumlah SKS, jenis tempat tinggal, dan jenis transportasi dapat mempengaruhi potensi mereka untuk *drop out*. Dengan memahami aturan-aturan ini, perguruan tinggi dapat mengidentifikasi dan memberikan dukungan yang tepat kepada mahasiswa yang berpotensi *drop out*, serta mengambil langkah-langkah pencegahan yang sesuai untuk meningkatkan tingkat kelulusan.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan potensi drop out mahasiswa dengan menggunakan metode algoritma C4.5. Data mahasiswa tahun akademik 2022/2023 dari Jurusan Teknologi Informasi dan Komputer Program Studi Sistem Informasi digunakan, dengan variabel penelitian meliputi Indeks Prestasi Kumulatif (IPK), Jumlah Satuan Kredit Semester (SKS), Jenis Tempat Tinggal, Jenis Transportasi, dan Status Mahasiswa. Berdasarkan penelitian sebelumnya, IPK dipilih sebagai variabel utama karena memiliki pengaruh signifikan terhadap potensi drop out. Proses pra-proses data melibatkan data selection, data cleaning, dan data transformation, dengan pembagian data menjadi training (80%) dan testing (20%). Algoritma Decision Tree C4.5 digunakan untuk klasifikasi, dengan evaluasi menggunakan confusion matrix yang menghasilkan akurasi 98.50%, precision untuk Drop Out sebesar 93.18%, dan recall 64.44%. Pohon keputusan menunjukkan bahwa IPK menjadi root, dengan 7 aturan yang membantu klasifikasi mahasiswa. Hasil ini memberikan gambaran faktor-faktor yang mempengaruhi potensi drop out, dan aturan-aturan tersebut dapat digunakan sebagai dasar untuk implementasi strategi intervensi yang lebih efektif di lingkungan pendidikan. Kesimpulannya, penelitian ini dapat berkontribusi dalam pemahaman dan prediksi potensi drop out mahasiswa serta mendukung langkah-langkah pencegahan yang tepat untuk meningkatkan tingkat kelulusan di Politeknik Negeri Subang.

UCAPAN TERIMAKASIH

Terima kasih kepada semua pihak yang telah mendukung dalam pelaksanaan penelitian ini. Terima Kasih kepada Allah SWT atas kelancaran, rahmat, dan petunjuk-Nya. Terima Kasih juga kepada kedua orang tua yang senantiasa memberikan do'a dan dukungan kepada kami. Kami juga berterima kasih kepada dosen pembimbing yang telah membimbing kami dalam menyelesaikan penelitian ini. Tak lupa juga, kami ucapkan terima kasih kepada teman-teman 3B yang telah membantu kami dalam menyelesaikan penelitian ini. Semoga hasil penelitian ini dapat memberikan wawasan baru kepada penulis dan pembaca serta bermanfaat bagi banyak orang.

REFERENCES

- [1] A. H. Nasrullah, “ILKOM Jurnal Ilmiah Volume 10 Nomor 2 Agustus 2018,” vol. 10, 2018.
- [2] S. Faisal and U. Buana Perjuangan Karawang Karawang, “TechnoXplore Jurnal Ilmu Komputer & Teknologi Informasi KLASIFIKASI DATA MINNING MENGGUNAKAN ALGORITMA C4.5 TERHADAP KEPUASAN PELANGGAN SEWA KAMERA CIKARANG,” 2019.

- [3] P. P. Haryoto, H. Okprana, and I. S. Saragih, "Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," vol. 2, no. 5, pp. 358–364, 2021, [Online]. Available: <https://ejurnal.seminar-id.com/index.php/tin>
- [4] K. Faozi, "Optimasi Algoritma C4.5 dengan Fuzzy Inference System Mamdani dalam Memprediksi Mahasiswa Berpotensi Dropout," 2022. [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>
- [5] Y. Dwi Atma and A. Setyanto, "PERBANDINGAN ALGORITMA C4.5 DAN K-NN DALAM IDENTIFIKASI MAHASISWA BERPOTENSI DROP OUT," vol. 2, no. 2, 2018.
- [6] T. Vaksinasi Pada Kecamatan Tambun Selatan Anissa Aidha Zulyani, A. Susilo Yuda Irawan, and A. Jamaludin, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan," *Journal Of Social Science Research*, vol. 3, pp. 7037–7050, 2023.
- [7] A. Suryanto, I. Alfarobi, T. A. Tutupoly, and R. Fauziahti, "OPTIMASI NAIVE BAYES MENGGUNAKAN OPTIMIZE WEIGHTS DAN STRATIFIED PADA DATA KREDIT KOPERASI," *Jurnal Mantik Penusa*, vol. 3, no. 1, pp. 211–219, 2019.
- [8] A. Junia, "SAINTEK (Jurnal Sains dan Teknologi) Implementasi Algoritma C4.5 Untuk Mengukur Tingkat Kepuasan BPJS Ketenaga Kerjaan," 2021.
- [9] C. Zai and T. Komputer, "IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA," 2022.
- [10] F. Faisal Nugraha, I. Sunandar, and C. Juliane, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [11] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 6, pp. 61–74, 2023.
- [12] D. Ramdan, F. Valentino, N. Ikhsan, and A. Sani, "JBPI-Jurnal Bidang Penelitian Informatika Penerapan Data Mining Terhadap Prediksi Mahasiswa Drop Out Pada Kampus STMIK Widuri Jakarta Dengan Metode Decision Tree C4.5," 2023. [Online]. Available: <https://ejournal.kreatifcemerlang.id/index.php/jbpi>
- [13] P. Seminar Nasional Pendidikan Matematika, U. Mulawarman ; Rizki, N. A. Mumtaza, M. F. Dewi, C. F. Syahlafandi, and D. Fendiyanto, "Implementasi Pohon Keputusan Untuk Klasifikasi Cara belajar Mahasiswa Mandiri dan Berkelompok Berdasarkan Sumber Belajarnya. Prosiding Seminar Nasional Pendidikan Matematika," *Universitas Mulawarman*, vol. 3, pp. 101–106, 2023, [Online]. Available: <https://jurnal.fkip.unmul.ac.id/index.php/psnpm>