

ANALISIS TERHADAP FAKTOR-FAKTOR YANG MEMPENGARUHI MAHASISWA GAGAL STUDI MENGGUNAKAN TEKNIK KLASIFIKASI

Yudi Kristyawan¹, Lambang Probo Sumirat²

E-mail : ¹yudi.kristyawan@unitomo.ac.id , ²lambang@unitomo.ac.id

^{1,2}Jurusan Informatika, Fakultas Teknik, Universitas Dr. Soetomo

Abstrak

Ketidakseimbangan jumlah mahasiswa yang masuk dan jumlah kelulusan yang disebabkan mahasiswa gagal studi pada Fakultas Teknik Universitas Dr Soetomo sudah menjadi permasalahan yang serius. Belum adanya penelitian yang dilakukan untuk mencari faktor-faktor penyebab mahasiswa gagal studi, menyebabkan penanganan masalah ini masih sulit untuk dilakukan. Tujuan dari penelitian ini adalah untuk mencari faktor-faktor yang menyebabkan mahasiswa mengalami gagal studi. Metode yang digunakan pada penelitian ini adalah dengan cara mengumpulkan data histori sistem informasi pada Universitas Dr Soetomo selama 7 tahun sebagai suatu rangkaian proses yang dibagi menjadi beberapa tahapan yaitu tahap cleaning data untuk membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, selanjutnya tahap integrasi data untuk menggabungkan data dari berbagai basis data ke dalam satu basis data baru, dilanjutkan dengan transformasi data yaitu proses penggabungan data ke dalam format yang sesuai untuk diproses dalam data mining, selanjutnya tahap data mining yang merupakan proses utama menggunakan teknik klasifikasi decision tree untuk menemukan pengetahuan berharga dan tersembunyi dari data dan tahap terakhir interpretasi data dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Hasil analisis dengan menggunakan teknik klasifikasi yang diperoleh menunjukkan bahwa faktor-faktor yang berpengaruh pada kegagalan mahasiswa adalah faktor cuti mahasiswa pada saat program skripsi, dan faktor lain terbesar yang menyebabkan mahasiswa gagal studi sebelum mengambil skripsi.

Kata kunci: mahasiswa gagal studi, klasifikasi.

1. PENDAHULUAN

Teknologi informasi dan komunikasi (TIK) saat ini sudah merupakan bagian yang tidak dapat dipisahkan dari kegiatan manusia modern dan memegang peranan penting dalam berbagai aspek kehidupan. TIK dalam bidang pendidikan memberikan dampak yang sangat positif dalam berbagai hal, di antaranya sebagai pendukung pengambilan keputusan siswa berprestasi [1], untuk memprediksi masa studi mahasiswa [2]. Peranan TIK akan semakin meningkat seiring perkembangan jaman dan hingga akhirnya menjadi elemen dasar dalam kehidupan manusia modern.

Dalam menjalankan misi pendidikan, setiap perguruan tinggi di Indonesia baik negeri maupun swasta memiliki permasalahannya sendiri-sendiri. Salah satu permasalahan yang selalu muncul di setiap perguruan tinggi adalah masalah mahasiswa gagal studi atau drop out. Hal ini disebabkan karena tolak ukur kesuksesan misi pendidikan suatu perguruan tinggi diukur oleh jumlah mahasiswanya yang gagal studi. Permasalahan mahasiswa yang gagal studi ini menarik perhatian peneliti untuk memecahkan persoalan dari berbagai sudut pandang. Sebagian peneliti ada yang berusaha mengidentifikasi faktor mahasiswa putus kuliah [3], mengklasifikasikan mahasiswa yang berpotensi drop out [4], memprediksi masa studi mahasiswa [2], memprediksi mahasiswa drop out [5], mencari faktor-faktor yang mempengaruhi terlambatnya penyelesaian studi [6], dan bahkan menganalisa pola kelulusan mahasiswa [7].

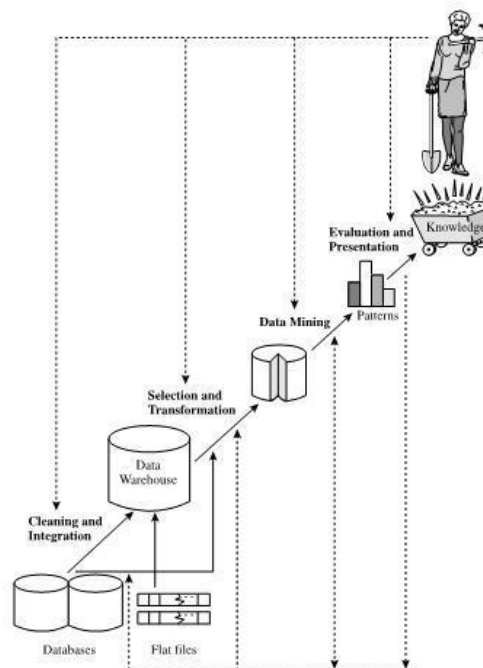
Universitas Dr Soetomo Surabaya sudah lebih dari 10 tahun memanfaatkan TIK dalam pengelolaan sistem manajemen perguruan tinggi dan sistem akademiknya. Sistem akademik berisi data mahasiswa, kurikulum, dosen, kartu rencana studi, jadwal kuliah, nilai, kartu hasil studi dan alumni. Karena sudah diterapkan sangat lama, maka data yang tersimpan sudah sangat banyak dan bersifat historis. Namun, keberadaan data yang berlimpah ini belum dimanfaatkan untuk menggali pengetahuan yang tersembunyi di dalamnya.

Pemanfaatan data histori yang berlimpah pada basis data sistem informasi perguruan tinggi Universitas Dr. Soetomo dapat dijadikan sebagai sumber pengetahuan baru dengan menggunakan aplikasi data mining untuk melakukan penggalian informasi yang tersembunyi yang bertujuan mencari faktor-faktor yang berpengaruh pada mahasiswa gagal studi. Faktor-faktor yang dimaksud adalah data pribadi mahasiswa yang berupa: indeks prestasi mahasiswa, program studi, shift kuliah, jurusan SLTA, skripsi, cuti dan lain sebagainya.

Seperti telah diuraikan diatas, bahwa penelitian ini mengeksplorasi data histori basis data sistem informasi perguruan tinggi yaitu berupa informasi atau pengetahuan yang digunakan sebagai pendukung pembuatan keputusan strategis dalam merespon permasalahan mahasiswa gagal studi di Universitas Dr Soetomo khususnya di fakultas teknik. Sehingga penelitian ini menjawab pertanyaan “Faktor-faktor apa yang paling dominan mempengaruhi mahasiswa gagal studi khususnya di Fakultas Teknik Universitas Dr. Soetomo?”

2. METODOLOGI

Tahapan yang digunakan dalam penelitian ini sama dengan tahapan proses penemuan pengetahuan dalam database yang terdiri dari cleaning and integration, selection and transformation, data mining, dan evaluation and presentation yang disadur dari [8] seperti ditunjukkan pada gambar 1 dibawah ini.



Gambar 1. Tahapan proses penemuan pengetahuan dalam database

2.1 Cleaning and Integration

Pada tahap ini, data yang sudah didapatkan dilakukan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh,

memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining akan dibuang. Pembersihan data ini sudah barang tentu akan mempengaruhi performansi teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

Sedangkan *data integration* atau integrasi data merupakan proses menggabungkan atau menyatukan beberapa data dari berbagai sumber *database* yang berbeda baik dalam satu server atau beberapa server ke dalam sebuah penyimpanan tunggal seperti gudang data (*data warehouse*).

2.2 Selection and Transformation

Seleksi data merupakan proses meminimalkan jumlah data yang akan digunakan untuk proses penemuan pengetahuan dengan tetap merepresentasikan data aslinya. Pengumpulan data-data operasional perlu dilakukan sebelum langkah penggalian informasi dimulai. Pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada variabel atau sampel data, dimana penemuan pengetahuan akan dilakukan. Hasil seleksi data yang akan digunakan untuk proses penemuan pengetahuan, disimpan dalam file terpisah dari database operasional sehingga tidak membahayakan data operasional. Beberapa metode seleksi yang digunakan dalam proses seleksi diantaranya adalah metode *sampling*, yaitu menyeleksi subset representatif dari populasi data yang besar, *denoising*, yaitu suatu proses untuk menghilangkan noise dari data yang akan ditransformasikan dan *feature extraction*, yaitu suatu proses membuka spesifikasi data yang signifikan dalam bentuk konteks tertentu.

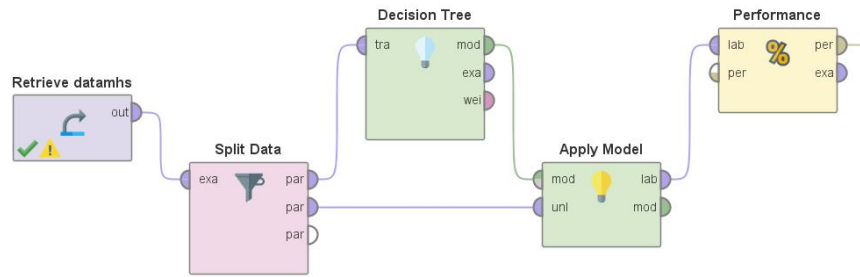
Transformasi data adalah upaya-upaya yang dilakukan dengan tujuan utama merubah skala pengukuran data asli menjadi data bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari tujuan sehingga data siap untuk dianalisis. Beberapa contoh misalnya adalah merubah data numerik menjadi data kategorik atau merubah dari beberapa variabel yang sudah ada dibuat dalam satu variabel komposit yang baru.

2.3 Data Mining

Data mining didefinisikan sebagai penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar[9]. Data mining, sering juga disebut sebagai bagian dari *knowledge discovery in database* (KDD). KDD merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar[10]. Pada tahapan data mining ini dilakukan proses ekstraksi data yang besar untuk menemukan pengetahuan yang belum tersampaikan sebelumnya, dengan menggunakan teknik klasifikasi. Teknik klasifikasi yang digunakan adalah metode *decision tree* untuk mendapatkan pola yang menarik dari data set yang besar. Alasan utama penggunaan metode *decision tree* karena sudah terbukti handal dalam *data mining* adalah karena adanya sejumlah data yang besar agar menghasilkan informasi dan pengetahuan yang berguna. Tabel data hasil transformasi diproses dengan menggunakan teknik klasifikasi menggunakan model *decision tree*.

Sebelum memulai proses mining, dilakukan pemanggilan data yang akan diolah. Untuk melakukan proses data mining ini digunakan aplikasi *open source* RapidMiner versi 9.3 yang mempunyai banyak sekali koleksi-koleksi algoritma *machine learning* yang dapat digunakan untuk menggeneralisasi dari data percobaan. Tahapan pengolahan data ini diawali dengan cara mengambil file excel ke dalam lingkungan RapidMiner. Selanjutnya dilakukan proses membagi data menjadi 2 bagian, dimana bagian pertama digunakan sebagai data pelatihan dan data yang kedua digunakan sebagai data uji. Berikutnya data pelatihan dimasukkan kedalam metode *decision tree* untuk diolah. Data pelatihan diumpungkan ke metode *decision tree* untuk diproses selanjutnya. Kemudian keluaran dari *decision tree* diumpungkan ke dalam *apply model* bersamaan dengan data kedua hasil split data. Lalu keluaran dari *apply model* diumpungkan ke dalam modul

performance untuk dievaluasi. Sebagai penggambaran, untuk lebih jelasnya dapat dilihat di gambar 2.



Gambar 2. Tahapan-tahapan pengolahan data mining

2.4 Evaluation and Presentation

Tahap ini merupakan hasil dari teknik yang diterapkan pada data mining berupa pola-pola yang khas maupun model prediksi evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Apabila ternyata hasil yang diperoleh belum sesuai hipotesa, maka diambil beberapa alternatif lain seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*, kemudian mencoba metode *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

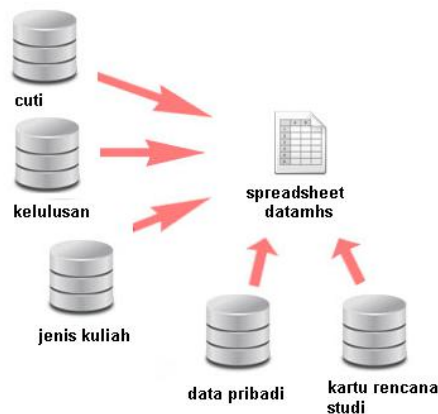
Tahap terakhir dari proses *data mining* berupa penyajian pengetahuan dari suatu metode yang digunakan untuk memperoleh pengetahuan dan bagaimana memformulasikan keputusan dari hasil analisis yang didapat agar dapat dimengerti dan dipahami oleh pengguna. Dalam tahap presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining*[11].

3. HASIL DAN PEMBAHASAN

Hasil penelitian analisis terhadap faktor-faktor yang mempengaruhi mahasiswa gagal studi akan dijelaskan per sub bab sesuai dengan tahapan penelitian yang dianut.

3.1 Cleaning and Integration

Dari serangkaian data operasional sistem informasi akademik terdapat beberapa tabel yang perlu diintegrasikan diantaranya adalah tabel datapribadi, kartu rencana studi, jenis kuliah, cuti dan tabel kelulusan menjadi satu bentuk data *spreadsheet* seperti ditunjukkan pada gambar 3.



Gambar 3. Integrasi data

Setelah tahap integrasi data selesai dihasilkan sebanyak 2549 data 20 atribut reguler dalam bentuk file *spreadsheet* excel. Berikutnya dilakukan tahap cleaning pada file *spreadsheet* excel untuk menghilangkan data yang tidak valid, salah ketik, atau

kosong. Pada gambar 4 ditunjukkan beberapa contoh data statistik atribut yang memiliki nilai kosong (*missing*).

Name	Type	Missing	Statistics	Filter (20 / 20 attributes)
nim	Integer	0	Min: 2008410001, Max: 2018425006, Average: 2013965436.92	
namamhs	Polynomial	0	Least: sandy putra pratama (1), Most: ALIPE AL [...] JESUS (5), Values: ALIPE ALVARO	
namasmu	Polynomial	402	Least: sma (1), Most: Smk Farmasi Sekesal (2...), Values: Smk Farmasi Se	
addresssmu	Polynomial	577	Least: surabaya (1), Most: SURABAYA (89), Values: SURABAYA (89)	
citysmu	Polynomial	425	Least: ZUMALAI (1), Most: Surabaya (771), Values: Surabaya (771)	
jurusansmu	Polynomial	497	Least: Pts (Patiseri) (1), Most: IPA (1373), Values: IPA (1373), DKV	
addressmhs	Polynomial	54	Least: wolowuwu RT.01/01 (1), Most: BAIROPIT [...] -DILI (5), Values: BAIROPIT [...] E	

Gambar 4. Data statistik atribut dengan nilai kosong (*missing*).

3.2 Selection and Transformation

Setelah melalui tahap *cleaning* dan *integration* di dalam *file spreadsheet excel* sudah tidak terdapat data yang kosong. Tahap berikutnya dilakukan seleksi atau pemilihan himpunan data, pemilihan atribut reguler dan atribut target (*special*) terhadap hasil proses *cleaning* dan *integration* sebelumnya. Hasil pemilihan himpunan data terpilih sebanyak 703 data yaitu untuk mahasiswa angkatan 2008 sampai dengan 2011 yang maksimal drop out pada tahun 2018, atribut reguler sebanyak 6 atribut yaitu: mahasiswa, jurusan SLTA, gajiortu, kuliah, skripsi dan cuti. Sedangkan untuk atribut target (*special*) hanya satu atribut yaitu status. Data statistik atribut hasil proses seleksi dapat dilihat pada gambar 5.

Name	Type	Missing	Statistics	Filter (7 / 7 attributes)
Label STATUS	Polynomial	0	Least: LULUS (267), Most: TIDAK LULUS (436), Values: TIDAK LULUS (436)	
JURUSAN SLTA	Polynomial	0	Least: IPS (41), Most: IPA (662), Values: IPA (662), IPS (41)	
MAHASISWA	Polynomial	0	Least: LUAR KOTA (262), Most: DALAM KOTA (441), Values: DALAM KOTA (441)	
GAJI ORTU	Polynomial	0	Least: TINGGI (20), Most: SEDANG (378), Values: SEDANG (378), TINGGI (20)	
KULIAH	Polynomial	0	Least: PAGI (344), Most: SORE (359), Values: SORE (359), PAGI (344)	
SKRIPSI	Polynomial	0	Least: BELUM (336), Most: SUDAH (367), Values: SUDAH (367), BELUM (336)	
CUTI	Polynomial	0	Least: PERNAH (46), Most: TIDAK (657), Values: TIDAK (657), PERNAH (46)	

Gambar 5. Data statistik atribut hasil seleksi

Berikutnya dilakukan proses transformasi data dengan tujuan utama merubah skala pengukuran data asli menjadi data bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari tujuan. Hasil proses transformasi data dapat dilihat pada tabel 1 berikut.

Tabel 1. Hasil proses transformasi

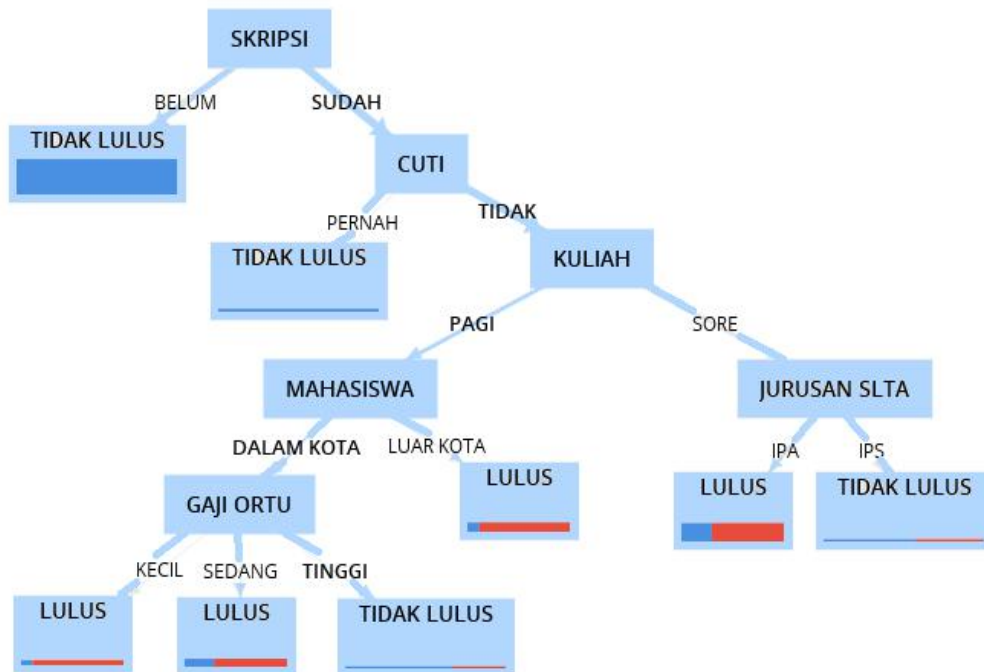
Atribut	Keterangan	Skala
Status	Status kelulusan	TIDAK LULUS/LULUS
Jurusan SLTA	Jurusan saat SLTA	IPA/IPS
Mahasiswa	Asal mahasiswa	DALAM KOTA/LUAR KOTA
Gajiortu	Gaji orang tua	KECIL/SEDANG/TINGGI
Kuliah	Masuk kuliah	PAGI/SORE
Skripsi	Mengambil skripsi	BELUM/SUDAH
Cuti	Mengambil cuti	TIDAK/PERNAH

3.3 Data Mining

Pada tahapan data mining ini dilakukan proses ekstraksi data yang besar untuk menemukan pengetahuan yang belum tersampaikan sebelumnya. Analisis terhadap faktor-faktor mahasiswa gagal studi dilakukan menggunakan teknik klasifikasi dengan metode *decision tree*. Pada proses pembuatan analisis dan pengujian model digunakan *software* RapidMiner 9.3. Data training dibagi menjadi 2 bagian, sebanyak 527 data digunakan untuk membangun sebagai model decision tree dan 175 data sebagai uji coba.

3.4 Evaluation and Presentation

Penyajian berupa visualisasi pengetahuan dari suatu teknik klasifikasi yang digunakan untuk memperoleh pengetahuan dan bagaimana memformulasikan keputusan dari hasil analisis merupakan hal yang sangat penting. Tujuan utama visualisasi adalah agar pengguna paham dan mengerti hasil analisis yang disajikan. Hasil visualisasi decision tree pada penelitian ini dapat dilihat pada gambar 6 di bawah ini.



Gambar 6. Decision tree faktor-faktor yang mempengaruhi mahasiswa gagal studi

Deskripsi dari teknik klasifikasi menggunakan metode decision tree faktor-faktor yang mempengaruhi mahasiswa gagal studi dijelaskan secara berjenjang sebagai berikut:

- SKRIPSI = BELUM: TIDAK LULUS {TIDAK LULUS=336, LULUS=0}
- SKRIPSI = SUDAH
 - | CUTI = PERNAH: TIDAK LULUS {TIDAK LULUS=13, LULUS=0}
 - | CUTI = TIDAK
 - | | KULIAH = PAGI

- | | | MAHASISWA = DALAM KOTA
- | | | | GAJI ORTU = KECIL: LULUS {TIDAK LULUS=4, LULUS=32}
- | | | | GAJI ORTU = SEDANG: LULUS {TIDAK LULUS=18, LULUS=45}
- | | | | GAJI ORTU = TINGGI: TIDAK LULUS {TIDAK LULUS=2, LULUS=1}
- | | | MAHASISWA = LUAR KOTA: LULUS {TIDAK LULUS=9, LULUS=68}
- | | KULIAH = SORE
- | | | JURUSAN SLTA = IPA: LULUS {TIDAK LULUS=50, LULUS=118}
- | | | JURUSAN SLTA = IPS: TIDAK LULUS {TIDAK LULUS=4, LULUS=3}

Tahap evaluasi dilakukan dengan cara membandingkan nilai akurasi, nilai *precision*, dan nilai *recall* sesuai dengan tabel *confusion matrix* pada gambar 7 berikut:

accuracy: 90.00%

	true TIDAK LULUS	true LULUS	class precision
pred. TIDAK LULUS	73	0	100.00%
pred. LULUS	14	53	79.10%
class recall	83.91%	100.00%	

Gambar 7. Tabel confusion matrix

4. KESIMPULAN DAN SARAN

Berdasarkan training dan pengujian data mining dengan teknik klasifikasi pada data mahasiswa teknik Universitas Dr. Soetomo menggunakan metode decision tree menghasilkan informasi bahwa faktor dominan yang paling banyak gagal studi adalah sebelum menempuh skripsi dan faktor cuti mahasiswa disaat sedang memprogram skripsi merupakan faktor dominan yang mengakibatkan gagal studi mahasiswa. Sedangkan faktor asal mahasiswa, gaji orang tua, dan jurusan saat SLTA tidak berpengaruh secara signifikan terhadap kegagalan studi mahasiswa.

Saran bagi pihak universitas agar pada data penerimaan mahasiswa baru perlu ditambahkan data-data yang berhubungan dengan kemampuan mahasiswa seperti data tes potensi akademik atau tes-tes lainnya yang berhubungan dengan kemampuan mahasiswa. Karena dalam hasil analisis tidak diketahui faktor-faktor yang mengakibatkan banyak mahasiswa gagal studi sebelum menempuh skripsi.

5. DAFTAR RUJUKAN

- [1] Haryo Kusumo, "Sistem Pendukung Keputusan Pemilihan Siswa Berprestasi Sebagai Peserta Olimpiade Sains Menggunakan Metode Simple Additive Weighting (SAW)," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 37-45, September 2018.
- [2] Siska Haryati, Aji Sudarsono, and Eko Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *Jurnal Media Infotama*, vol. 11, no. 2, pp. 130-138, September 2015.
- [3] Fadjrian Imran, Budi Susetyo, and Aji Hamim Wigena, "Identifikasi Faktor-Faktor Yang Berhubungan Dengan Mahasiswa Putus Kuliah di IPB Angkatan 2008 Menggunakan Analisis Survival," *Xplore*, vol. 2, no. 1, pp. 1-6, 2013.
- [4] Asmaul Husnah Nasrullah, "Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out," *ILKOM*, vol. 10, no. 2, pp. 244-250, Agustus 2018.
- [5] Ade Putra, "Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma," *Jurnal SIMETRIS*, vol. 8, no. 1, pp. 177-184, April 2017.

- [6] Nor Amira and Swistantoro , "Faktor-Faktor Yang Mempengaruhi Terlambatnya Penyelesaian Studi Pada Mahasiswa Fakultas Ilmu Sosial Dan Ilmu Politik Universitas Riau," *JOM FISIP*, vol. 3, no. 2, pp. 1-14, 2016.
- [7] Dwi Anggih Yosepta and Tria Aprilianto, "Analisa Pola Kelulusan Mahasiswa Pada Sekolah Tinggi Manajemen Informatika & Komputer Asia Malang Dengan Menggunakan Algoritma Iterative Dichotomiser 3 (Id3)," *Jurnal POSITIF*, vol. 3, no. 1, pp. 47 - 55, 2017.
- [8] Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining: Concepts and Techniques (Third Edition)*. Waltham: Morgan Kaufmann, 2012.
- [9] Paul Beynon-Davies, *Database Systems (3th Edition)*. New York: Palgrave Macmillan, 2004.
- [10] Budi Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [11] Jamie MacLennan, *Data Mining with Microsoft SQL Server 2008 1st Edition.*: Wiley, 2008.
- [12] Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook (Second Edition)*. London: Springer, 2010.