

Klasifikasi Waktu Kelulusan Mahasiswa Menggunakan Metode *Decision Tree*

(Studi Kasus : Fakultas Ilmu Hukum Universitas Ichsan Gorontalo)

Geby Anggriani Mundok¹, Amiruddin¹, Zulfrianto Y. Lamasigi²

¹Fakultas Ilmu komputer, Program Studi Teknik Informatika, Ichsan Gorontalo, Kota Gorontalo, Indonesia
Email: 1gebyanggriani@email.com, zulfriyanto.dsn.unisan@gmail.com

Abstrak - kelulusan mahasiswa merupakan salah satu indikator keberhasilan perguruan tinggi. Waktu kelulusan mahasiswa merupakan periode waktu yang dibutuhkan mahasiswa untuk menyelesaikan studinya. Idealnya, mahasiswa dapat lulus tepat waktu sesuai dengan masa studi yang ditetapkan. Namun, pada kenyataannya, tidak semua mahasiswa dapat lulus tepat waktu. Peneliti ini bertujuan untuk mengklasifikasi waktu kelulusan mahasiswa apakah lulus tepat waktu atau tidak dengan menggunakan metode decision tree. Data yang digunakan dalam penelitian ini adalah data lulusan Fakultas Ilmu Hukum Universitas Ichsan Gorontalo dari tahun 2015 hingga 2017 ada. Adapun atribut yang digunakan dalam klasifikasi ini terdiri dari: IPS1, PIS2, IPS3, IPS4, jenis kelamin, keterangan lulus. Pada penelitian ini dilakukan juga optimalisasi model dengan melakukan seleksi atribut, pemangkasan pohon serta pengukuran kedalaman pohon. Hasil penelitian menunjukkan bahwa metode decision tree dapat memprediksi waktu kelulusan mahasiswa dengan akurasi 92% dengan menghasilkan sebanyak 9 rule keputusan. Hal ini menunjukkan bahwa metode decision tree dapat menjadikan salah satu solusi untuk memprediksi waktu kelulusan mahasiswa, sehingga metode ini dapat menjadi salah satu solusi untuk membantu program studi dalam meningkatkan keberhasilan mahasiswa dalam menyelesaikan studinya

Kata Kunci: klasifikasi, waktu kelulusan, mahasiswa, data mining, decision tree, lulus tepat

Abstract A student's graduation time is one indicator of the university's success. Student graduation time is the period required for students to complete their studies. Ideally, students can graduate on time under a specified study period. However, not all students can graduate on time. This study is aimed at classifying whether students graduate on time using the decision tree method. The data employed in this study are the Faculty of Law's graduates of 2015 - 2017, Universitas Ichsan Gorontalo. The attributes used in this classification consist of IPS1, PIS2, IPS3, IPS4, gender, and graduation information. In this study, model optimization performed is by selecting attributes, pruning trees, and measuring inside the tree. The results of this study show that the decision tree method can predict student graduation times with 92% accuracy by producing nine (9) decision rules. It indicates that the decision tree method can be a solution for predicting student graduation times, so it can be a solution to help study programs increase student success in completing their studies.

Keywords: classification, graduation time, students, data mining, decision tree, accurate graduation

1. PENDAHULUAN

Perkembangan teknologi di zaman pada saat ini semakin berkembang terutama pada Pendidikan. Seiring meningkatnya mutu Pendidikan di tingkat sekolah tinggi atau universitas maka di butuhkan laporan yang dapat mendukung peningkatan prestasi akademik mahasiswa. Laporan ini dapat di peroleh dari datanilai mata kuliah, dan indek prestasi mahasiswa dan masih banyak lagi data-data yang dapat di dimanfaatkan untuk mendapatkan informasi yang dapat mendukung peningkatan prestasi akademik mahasiwa.

Permasalahan yang di hadapi oleh fakultas hukum adalah jumlah mahasiwayang masuk dan keluar setiap tahun tidak sebanding dan akan mempengaruhi akreditasi prodi tersebut dengan berdasarkan adanya system yang dapat digunakan untuk mengetahui waktu kelulusan mahsiswa, apabila waktu kelulusandapat di ketahui sejak dini maka pihak akademik dapat menerapkan suatu kebijakan untuk meminimalisir jumlah mahasiwa yang tidak lulusan tepat waktusuai dengan masa studinya. Dalam perguruan tinggi mahasiwa merupakan assetyang paling penting

Adapun penelitian yang terkait dengan klasifikasi waktu kelulusan mahasiwa yaitu penelitian dari Reza Maulana, Devi Kumalasari dengan judul “ Analisis Komparasi Agoritma Klasifikasi *Data Mining* Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika”. Penelitian tersebut dilakukan dengan

menggunakan metode K-NN, *Decision Tree*, *Naïve Bayes*, *Random Forest* dan *Random Tree* untuk melakukan klasifikasi data mining, untuk memprediksi status kelulusan mahasiswa

Menggunakan algoritma C4.5 dengan pendekatan *data mining* adalah cara terbaik untuk menangani data-data ini. Nama, mahasiswa, nim, jenis kelamin, jumlah SKS, IPK, dan alamat adalah atribut yang digunakan. Berdasarkan karakteristik tersebut, prosedur pengolahan data dapat membantu institusi pendidikan dalam upaya mendorong mahasiswa untuk meningkatkan nilai IPK dan dapat mendukung akreditasi Fakultas Hukum Universitas Ichsan Gorontalo

Ada beberapa algoritma C4.5 yang digunakan dalam pendekatan klasifikasi data mining yang menggambarkan kategorisasi data sebagai pohon keputusan. Manfaat utama dari metode C4.5 adalah bahwa metode ini dapat menghasilkan pohon keputusan yang cukup akurat dan efisien dalam menangani jenis karakteristik diskrit dan numerik

2. TINJAUAN PUSTAKA

2.1 Decision tree

Struktur data dengan simpul dan sisi disebut pohon. Simpul akar, simpul cabang/internal, dan simpul daun adalah tiga jenis simpul yang membentuk sebuah pohon [13].

Dalam sebuah pohon keputusan, simpul internal dan simpul akar diberi label dengan nama atribut, rusuk diberi label dengan nilai atribut potensial, dan simpul daun diberi label dengan berbagai kelas. Pohon keputusan adalah representasi langsung dari teknik klasifikasi untuk jumlah kelas yang terbatas [13].

2.2. Klasifikasi

Menurut etimologinya, istilah "klasifikasi" berasal dari kata kerja "to classify", yang berarti mengkategorikan atau mengatur segala sesuatu pada posisi yang tepat. Frasa ini menggambarkan teknik untuk mengatur data secara metodis [11].

Menurut *Harrolds Librarians Glossary*, Towa P. Hamakonda, dan J.N.B Tairas (2022) [12], menjelaskan dalam *Harrolds Librarians Glossary* bahwa kategorisasi adalah pengelompokan secara sistematis dan logis terhadap benda-benda, konsep-konsep, buku-buku, atau hal-hal lain ke dalam kelas-kelas atau kelompok-kelompok tertentu berdasarkan kualitas-kualitas yang sama.

Metode C4.5, yang dapat digunakan untuk menghasilkan pohon keputusan, merupakan algoritme yang paling populer karena mengungguli algoritme lainnya dalam banyak hal. Algoritma ini juga akan menghasilkan rule yang nantinya akan mempermudah dalam proses menggali sebuah data dan informasi [14].

1. Validasi (*Validation*)

Validasi adalah sebuah proses pembagian data awal menjadi dua bagian data, *data training* dan *data testing*. *Data training* adalah data yang akan diolah menggunakan algoritma klasifikasi dan *data testing* adalah data yang akan digunakan dalam proses pengujian menggunakan program komputer. Proses validasi dalam klasifikasi terbagi menjadi dua proses yaitu *cross validation* dan *split validation*. *Cross validation* melakukan pembagian data secara acak sedangkan *split validation* melakukan pembagian data menjadi dua bagian berdasarkan jumlah data.

2. Implementasi Algoritma Klasifikasi

Proses klasifikasi melakukan pengolahan data dengan mengimplementasikan algoritma klasifikasi seperti *support vector machine*, *decision tree*, *k-means* dan *naive bayes*.

3. Apply Model

Apply model digunakan untuk penerapan model klasifikasi yang telah dilatih sebelumnya menggunakan *data training* yang juga akan digunakan dalam pengujian *data testing*.

4. Performance

Performance adalah hasil dari pengujian yang dilakukan dalam proses klasifikasi. Hasil yang diukur pada tahap *performance* seperti akurasi, *precision*, dan *recall*.

2.3 Algoritma C4.5

Metode C4.5, yang dapat digunakan untuk menghasilkan pohon keputusan, merupakan algoritme yang paling populer karena mengungguli algoritme lainnya dalam banyak hal. Algoritma ini juga akan menghasilkan rule yang nantinya akan mempermudah dalam proses menggali sebuah data dan informasi [14]

Metode pohon keputusan C4.5 melalui beberapa langkah untuk membuat pohon keputusan, termasuk :

mengumpulkan data pelatihan. Data awal yang telah diklasifikasikan ke dalam kelompok tertentu dapat digunakan sebagai data pelatihans. Data Training adalah data yang nantinya akan diolah dengan menggunakan algoritma C4.5 dalam pemodelan pohon keputusan dan rule [15].

2.4 Confusion Marix

Confusion matrix adalah sebuah *tool* untuk mengevaluasi model dari klasifikasi untuk memprediksi objek yang bernilai benar atau salah. *Matrix* dari hasil prediksi yang nantinya akan dibandingkan dengan kelas asli inputan atau informasi nilai aktual dan prediksi pada klasifikasi [16]

3. METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

Penelitian ini termasuk dalam kategori penelitian terapan. Menurut jenis informasi yang dianalisis, penelitian ini bersifat kuantitatif. Dan penelitian ini bersifat konfirmatori bila dievaluasi dari perspektif perlakuan data.

Fakultas Hukum Universitas Ichsan Gorontalo adalah tempat studi kasus untuk penelitian ini. Dengan menggunakan metode pohon keputusan, topik penelitian ini adalah prediksi kelulusan mahasiswa Fakultas Hukum Universitas Ichsan Gorontalo. Universitas Ichsan Gorontalo akan menjadi lokasi penelitian ini.

Data primer yang di gunakan pada penelitian ini adalah data mahasiswa fakultas hukum yang di kumpulkan menggunakan Teknik observasi dan wawancara sedangkan data sekunder di kumpulkan menggunakan referensi jurnal dari penelittitan sebelumnya dan perpustakaan.

3.2 Pengumpulan Data

Penelitian ini menggunakan pengumpulan data dengan jenis data primer yang dikumpulkan dengan teknik dokumentasi, wawancara, dan observasi yang didapatkan langsung di lapangan.

3.4 Pra Pengolahan Data

Pemilihan dan transformasi data dilakukan pada data yang akan diolah nantinya. Hal ini dilakukan untuk menetapkan standar atribut yang selanjutnya akan dijadikan sebagai tolok ukur untuk mengukur kontribusi setiap atribut terhadap klasifikasi data.

3.5 Validasi

Tujuan validasi adalah untuk memisahkan data awal menjadi data training dan data testing. Data testing adalah data yang akan digunakan dalam proses pengujian dengan menggunakan program komputer, sedangkan data training adalah data yang akan diproses dengan menggunakan metode klasifikasi. Tergantung pada jumlah data, pendekatan validasi yang digunakan pada tahap ini adalah split validation, yang membagi data menjadi dua bagian.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Dataset atau data penelitian yang digunakan pada penelitian ini merupakan data private yang di ambil dari BAAK Universitas Ichsan Gorontalo terkait aktifitas kuliah mahasiswa dari tahun akademik 2017-2017 adapun dataset yang di kumpulkan dapat dilihat pada tabel 4.1.

4.2 Hasil Pemodelan

1. Seleksi Data

Berdasarkan dataset pada tabel 4.1 di atas dapat dilihat bahwa terdapat perbedaan data yang cukup signifikan, sehingga dalam penerapan metode *Decision Tree* data yang dibagi lima yaitu dataset untuk jenis kelamin, IPS1, IPS2, IPS3, dan IPS4 serta tambahan satu atribut yaitu keterangan lulus sebagai keterangan lulus sebagai keterangan keluaran.

2. Preprocessing

Pada tahapan ini dilakukan pengecekan data yang kosong atau data yang tidak konsisten terhadap hasil akhir atau kelas output. Berikut contoh sampel data yang tidak konsisten

3. Transformasi

Pada tahapan ini dilakukan transformasi yaitu data jenis kelamin diubah dalam bentuk nilai 1 dan 2 agar mempermudah proses pemodelan dengan metode *decision tree* sampel hasil transformasi

4. Proses Data Mining

Setelah dilakukan tahapan diatas, maka selanjutnya dilakukan proses data mining terdiri dari penentuan data atribut NO, Jenis Kelamin, IPS1, IPS2, IPS3, IPS4, dan Keterangan, pembagian data training dan data testing dengan perbandingan 70:30 sehingga didapat Data Training sebanyak (375) dan Data Testing sebanyak (161) data. Setelah itu dilakukan proses pemodelan dengan metode *decision tree*

5. Evaluasi

Setelah dilakukan proses data mining perlu dilakukan evaluasi terhadap model yang sudah dibuat dengan melihat hasil akurasi dan visualisasi hasil pemodelan a. evaluasi kinerja model evaluasi kinerja model *decision tree* dengan hasil perhitungan

6. Perbaikan Permodelan

Berdasarkan hasil evaluasi kinerja model di atas, belum didapatkan hasil yang maksimal sehingga perlu dilakukan perbaikan atau optimalisasi model

Ada beberapa algoritma C4.5 yang digunakan dalam pendekatan klasifikasi data mining yang menggambarkan kategorisasi data sebagai pohon keputusan. Manfaat utama dari metode C4.5 adalah bahwa metode ini dapat menghasilkan pohon keputusan yang cukup akurat dan efisien dalam menangani jenis karakteristik diskrit dan numerik

4.3 Pembahasan Kinerja Model

Pemodelan *Decision tree* dalam klasifikasi waktu kelulusan mahasiswa dilakukan uji coba pemodelan dilakukan sebanyak dua kali pengujian, setelah dilakukan optimalisasi.

1. kinerja model sebelum optimalisasi

Setelah dilakukan pemodelan *decision tree* selanjutnya dilakukan pengujian model dengan menggunakan data testing sebanyak 161 adapun hasil pengujian akurasi model menggunakan confusion matrix

2. kinerja model setelah optimalisasi

Setelah dilakukan pemodelan *decision tree* selanjutnya dilakukan pengujian model kedua dengan menggunakan data testing sebanyak 161 adapun hasil pengujian akurasi model menggunakan confusion matrix

4.4 Pembahasan Hasil Pemodelan

Hasil pemodelan *Decision Tree* pada Klasifikasi Waktu Kelulusan Mahasiswa dilakukan 2 kali tahapan uji coba yaitu pemodelan yang dilakukan seperti seleksi atribut, mengukur kedalaman pohon dan pemangkasan pohon. Adapun hasil masing-masing uji coba tersebut diuraikan pada pembahasan.

4.5 Hasil Pemodelan Setelah Optimalisasi

Hasil pemodelan *Decision Tree* setelah dilakukan optimalisasi model terkait seleksi atribut, pemangkasan pohon, dan mengukur kedalaman pohon.

Pada optimalisasi seleksi data atribut terpilih atribut IPS1, IPS2, IPS3, IPS4 sesuai hasil pengukuran menggunakan metode *Decision Tree*

Pada optimalisasi pemangkasan pohon dilakukan pemangkasan IPS > 3.72 dijadikan kelasnya Menjadi Tepat Waktu semua, untuk IPS < 2.96 dijadikan kelasnya menjadi Tidak Tepat Waktu.

Pada optimalisasi pengukuran kedalaman didapatkan hasil kedalaman pohon 5 yang memiliki tingkat akurasi tertinggi, sehingga nilai kedalaman ini yang digunakan dalam pemodelan

5. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan terkait klasifikasi waktu kelulusan mahasiswa dapat disimpulkan bahwa :

1. Penerapan metode *Decision tree* didapatkan rule keputusan sebanyak 5 rule, dengan atribut yang terpilih yaitu: IPS1, IPS2, IPS3, IPS4 hasil didapatkan berdasarkan hasil optimalisasi model yaitu seleksi fitur, pemangkasan pohon dan pengukuran kedalaman pohon.
2. Penerapan metode *Decision Tree* pada Klasifikasi Waktu kelulusan Mahasiswa di dapatkan hasil akurasi sebesar 94% dengan demikian metode klasifikasi *decision tree* dapat digunakan untuk mengklasifikasi waktu kelulusan mahasiswa, dengan menggunakan metode ini , perguruan tinggi dapat mengidentifikasi mahasiswa yang beresiko tidak lulus tepat waktu dan memberikan intervensi yang di perlukan untuk membantu mereka menyelesaikan studinya

Berdasarkan kesimpulan laporan tersebut diatas, peneliti dapat memberikan saran untuk peneliti selanjutnya, yaitu:

1. Penelitian selanjutnya dapat dilakukan dengan menggunakan data yang lebih besar dan lebih beragam. Hal ini akan dapat meningkatkan akurasi dari metode *decision tree*.
2. Penelitian dapat juga di lakukan untuk menguji efektivitas metode *decision tree* dalam pemprediksi waktu kelulusan mahasiswa di berbagai perguruan tinggi. Hal ini akan dapat memberikan gambaran yang lebih luas tentang efektivitas metode ini.

DAFTAR PUSTAKA

- [1] R. Maulana and D. Kumalasari, "Analisis Komparasi Agoritma Klasifikasi Data Mining Untuk Prediksi Status Kelulusan Mahasiswa Akademik Bina Sarana Informatika," 2019.
- [2] A. Kadir Jaelani, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Akademi Farmasi Jember."
- [3] S. P. Y. Pramana, "Proyek Machine Learning ," 2021.
- [4] D. Dalbergio, M. N. Hayati, and Y. N. Nasution, "Klasifikasi Lama Studi Mahasiswa Menggunakan Metode Algoritma C5.0 Pada Stui Kasus Data Kelulusan Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Tahun 2017," 2019.
- [5] F. Rezkika *et al.*, "Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Menggunakan Algoritma C4.5 (Studi Kasus : Fasilkom Unsika)," 2021.
- [6] M. I. A. Akaseh, "Aplikasi Data Mining Untuk Pemilihan Jurusan IPA dan IPS untuk Siswa Kelas X Pada Sekolah Menengah Umum Negeri 1 Telaga Kabupaten Gorontalo," 2020.
- [7] "Pengelompokkan Kasus Covid-19 Di Indonesia Menggunakan Metode Density-Based Spatial Clustering Of Applicatios With Noise (DbSCAN)." [Online]. Available: <http://repository.unimus.ac.id>
- [8] H. Handayani, "Pengelompokan Indutri Mikro dan Kecil Di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM)," 2017.
- [9] E. Triyanto, H. Sismoro, and A. D. Laksito, "Implementasi Algoritma Regresi Linear Berganda Untuk Meprediksi Produksi Padi di Kabupaten Bantul," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 4, no. 2, pp. 66–75, Jul. 2019, doi: 10.36341/rabit.v4i2.666.
- [10] S. Septiawanti, "Penerapan Metode Case-Based Reasoning Untuk Prediksi Data Nasabah yang Berpotensi Membuka Simpanan Deposito," 2018.
- [11] R. B. Mahaputra, "Rancang Bangun Aplikasi Ujian Akhir Semester Online Untuk Mengukur Pencapaian Kompetensi Siswa".
- [12] D. A. Novilla, R. Goejantoro, F. Deny, and T. Amijaya, "Classification of Insurance Data Customers Using Naive Bayes Method (Case Study : PT. Prudential Life MT. Haryono Street Samarinda)," *Jurnal EKSPONENSIAL*, vol. 10, no. 2, 2019.
- [13] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [14] G. Taufik and D. Jatmika, "Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang," vol. 6, no. 1, 2021.
- [15] Z. Gustiana, "Penerapan Algoritma C 4.5 Dalam Sistem Pendukung Keputusan Evaluasi Kinerja Fasilitator Pamsimas (Studi Kasus di Kabupaten Kampar)," 2020.

- [16] K. Suhada, A. Elanda, and A. Aziz, "Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang)," 2021.