

## I. PENDAHULUAN

### 1.1 Latar Belakang

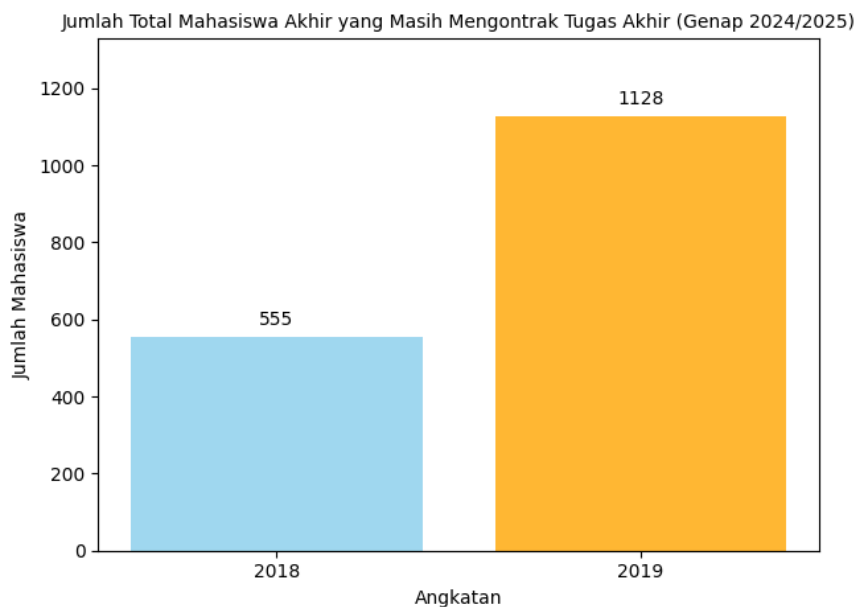
Dalam dunia pendidikan Indonesia, termasuk Perguruan Tinggi saling berkompetisi untuk mencetak lulusan yang berkualitas (Kurniasih & Isyara, 2023). Salah satu indikator kualitas perguruan tinggi dapat dilihat dari grade akreditasi yang diberikan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) maupun Lembaga Akreditasi Mandiri Pendidikan Tinggi (LAM-PT). Akreditasi yang baik diperoleh perguruan tinggi yang memenuhi kriteria-kriteria yang telah ditentukan, salah satunya adalah tingkat kelulusan tepat waktu (BANPT, 2019).

Namun, dalam pelaksanaannya, tidak semua mahasiswa mampu menyelesaikan studi tepat waktu. Banyak mahasiswa yang lulus terlambat, bahkan ada yang mengalami kegagalan akademik. Salah satu bentuk kegagalan akademik yang sering terjadi adalah *drop out* (DO), yaitu penghentian atau pemutusan hubungan studi mahasiswa dengan perguruan tinggi. *Drop out* biasanya terjadi karena berbagai faktor yang telah ditentukan oleh pihak universitas (Bahri & Midyanti, 2023). Beberapa alasan utama mahasiswa mengalami *drop out* adalah indeks prestasi kumulatif (IPK) yang rendah, seperti IPK kurang dari 2,00 pada 2 – 3 semester pertama, IPK kurang dari 2,00 pada semester 4 dengan minimal 40 SKS dan IPK kurang dari 2,00 pada semester 8 dengan minimal 60 SKS. Mahasiswa juga dapat mengalami *drop out* jika tidak mampu menyelesaikan masa studinya dalam waktu maksimal 14 semester (Peraturan Rektor No 31 Tahun 2024 Tentang Peraturan Akademik, 2024).

Masa studi mahasiswa telah diatur dalam Peraturan Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Tinggi tentang Sistem Pendidikan Tinggi yang menyatakan bahwa untuk memenuhi standar kompetensi lulusan bagi mahasiswa program sarjana (S1) beban wajib yang harus ditempuh adalah paling sedikit 144 SKS (satuan kredit semester) dengan lama studi selama 4-5 tahun (8-10 semester) (Permendikbud No. 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi (SNPT), 2014). Oleh karena itu masa studi mahasiswa menjadi salah satu hal yang penting yang perlu diperhatikan oleh Perguruan Tinggi. Meskipun demikian, masih terdapat mahasiswa yang belum mampu menyelesaikan studinya sesuai dengan ketentuan waktu yang telah ditetapkan.

Berdasarkan data statistik Universitas Jambi (<https://dashboard.unja.ac.id>) yang diakses pada tanggal 19 Februari 2025, pada semester genap tahun akademik 2024/2025, terdapat mahasiswa tingkat akhir dari angkatan 2018 dan

2019 yang masih mengontrak tugas akhir. Mahasiswa pada angkatan 2018 yang masih mengontrak tugas akhir tercatat sebanyak 555 orang, yang terdiri dari 287 laki-laki dan 268 perempuan. Sementara itu, pada angkatan 2019, terdapat 1.128 mahasiswa dengan rincian 581 laki-laki dan 547 perempuan. Jumlah mahasiswa angkatan 2018 dan 2019 yang masih mengontrak tugas akhir menunjukkan bahwa sebagian besar mahasiswa akhir belum dapat menyelesaikan studinya, fenomena ini berpotensi meningkatkan terjadinya drop out.



**Gambar 1.** Jumlah Mahasiswa Akhir yang Masih Mengontrak Tugas Akhir

Sehubungan dengan hal tersebut, diperlukan suatu metode yang mampu mengklasifikasikan status risiko *drop out* mahasiswa dengan memanfaatkan klasifikasi dalam *data mining*. Klasifikasi adalah tipe analisis data yang dapat membantu orang menentukan kelas label dari sampel yang ingin di klasifikasi (Hendrian, 2018). Klasifikasi merupakan metode *supervised learning*, yaitu metode yang mencoba menemukan hubungan antara atribut masukan dan atribut target. Tujuan klasifikasi adalah untuk meningkatkan keandalan hasil yang diperoleh dari data. Salah satu metode yang sering digunakan adalah algoritma *decision tree*. Metode ini melibatkan konstruksi pohon keputusan yang terdiri dari kumpulan node keputusan. Setiap cabang pohon mengarah ke node lain, baik berupa node keputusan lanjutan maupun node daun yang menandai akhir proses keputusan (Ginantra et al., 2021).

*Decision tree* memiliki kelebihan dalam mempermudah pemahaman dan interpretasi hasil prediksi melalui struktur pohon keputusan (Rahayu et al., 2024), tetapi *decision tree* memiliki kelemahan yaitu rentan terhadap *overlapping*,

terutama ketika fitur yang digunakan jumlahnya sangat banyak sehingga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan (Ginantra et al., 2021). Penerapan teknik seleksi fitur dapat mengatasi kelemahan ini dengan mengurangi jumlah fitur yang kurang relevan, sehingga meningkatkan kinerja model (Rahayu et al., 2024)

Penelitian sebelumnya yang menggabungkan metode *decision tree* dan seleksi fitur dilakukan oleh Hariyanti et al. (2024) yang membandingkan seleksi fitur ANOVA, *Univariate* dan *chi-square* dimana hasil dari penelitian ini didapat bahwa seleksi fitur terbaik dihasilkan oleh metode ANOVA yang mampu meningkatkan performa model klasifikasi, namun dari sisi waktu komputasi, penggunaan seleksi fitur ANOVA membutuhkan waktu yang lebih banyak dibandingkan metode lainnya saat proses pengujian.

Pada penelitian yang dilakukan oleh (Adnyana et al., 2021) menggunakan algoritma *Random Forest* dan algoritma ID3 dengan penerapan seleksi fitur *Principal Component Analysis* (PCA) mendapatkan hasil bahwa penggunaan PCA justru mengurangi akurasi dari prediksi dibandingkan hasil prediksi yang dilakukan tanpa menggunakan PCA. Hal ini menunjukkan bahwa dalam kasus ini, PCA tidak mampu menjadi teknik seleksi fitur tunggal dikarenakan PCA memiliki beberapa kelemahan yang tidak bisa ditangani.

Penelitian lainnya juga dilakukan oleh Putri et al. (2024) dalam klasifikasi status pemulangan pasien cedera kepala yang menggunakan algoritma C4.5 dengan seleksi fitur *chi-square* dan *mutual information*. Pada penelitian ini terbukti bahwa fitur mampu dalam menyeleksi fitur yang kurang relevan, sehingga hanya fitur yang memiliki pengaruh signifikan yang digunakan dalam penerapan model klasifikasi. Seleksi fitur menggunakan *mutual information* juga terbukti mampu meningkatkan akurasi model dari 88.57% menjadi 91.42%.

Dalam membangun pohon keputusan, beberapa algoritma yang digunakan adalah ID3, CART, C4.5, dan C5.0 (Prasetyo, 2014). ID3 menggunakan konsep *information gain* untuk memilih atribut terbaik, sementara C4.5 yang merupakan pengembangan ID3, memiliki kemampuan dalam menangani *missing value* dan atribut kontinu, membuatnya lebih fleksibel (Ginantra et al., 2021). Algoritma C5.0, yang dikembangkan dari ID3 dan C4.5, memiliki keunggulan fase *boosting* di proses terakhir, yang menjadikan C5.0 lebih unggul dibandingkan algoritma sebelumnya (Karlitasari et al., 2023). Dalam beberapa pengujian, C5.0 juga menunjukkan akurasi yang lebih baik dibanding pendahulunya (Aesyi et al., 2021).

Meskipun C5.0 memiliki keunggulan dalam proses *boosting*, diperlukan proses seleksi fitur untuk meningkatkan kinerjanya. Seleksi fitur bertujuan untuk mengurangi jumlah atribut yang tidak relevan, sehingga dapat meningkatkan efisiensi proses klasifikasi dan akurasi hasil (Colanus et al., 2022). Dalam penelitian ini, metode *mutual information* dipilih sebagai metode seleksi fitur karena memiliki keunggulan yaitu dapat mengukur semua jenis hubungan antara atribut termasuk hubungan non linear, *mutual information* juga tidak terpengaruh oleh perubahan seperti rotasi data, selama perubahan tersebut tidak mengubah urutan asli data. Keunggulan ini menjadikan *mutual informasi* efektif untuk berbagai jenis data dan transformasi (Vergara & Estévez, 2014). Dengan demikian, penelitian ini menggunakan algoritma C5.0 dengan metode seleksi fitur *mutual information* dalam proses pengklasifikasian status mahasiswa berisiko *drop out*.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini berjudul “Klasifikasi Status Mahasiswa Berisiko *Drop Out* Mahasiswa Menggunakan *Decision Tree* C5.0 dengan Seleksi Fitur di Universitas Jambi”. Penelitian ini diharapkan mampu membangun model klasifikasi yang akurat dan efisien dalam mendeteksi mahasiswa yang berisiko *drop out*.

## 1.2 Rumusan Masalah

Dari latar belakang diatas, dapat dirumuskan beberapa permasalahan sebagai berikut :

1. Bagaimana performa algoritma *decision tree* C5.0 dalam mengklasifikasikan status mahasiswa yang berisiko *drop out* di Universitas Jambi?
2. Apakah seleksi fitur *mutual information* dapat meningkatkan akurasi algoritma *decision tree* C5.0 dalam mengklasifikasikan status mahasiswa yang berisiko *drop out*?
3. Fitur-fitur apa saja yang paling berpengaruh terhadap klasifikasi status mahasiswa yang berisiko *drop out* berdasarkan hasil seleksi fitur *mutual information*?

## 1.3 Tujuan Penelitian

Pada uraian latar belakang dan perumusan masalah yang telah disajikan, tujuan dari penelitian ini antara lain adalah:

1. Menganalisis performa algoritma *decision tree* C5.0 dalam mengklasifikasikan status mahasiswa yang berisiko *drop out* di Universitas Jambi.
2. Mengukur pengaruh seleksi fitur *mutual information* terhadap peningkatan akurasi algoritma *decision tree* C5.0 dalam mengklasifikasikan status mahasiswa yang berisiko *drop out*.

3. Mengidentifikasi fitur-fitur yang paling berpengaruh terhadap klasifikasi status mahasiswa yang berisiko *drop out* berdasarkan hasil seleksi fitur *mutual information*.

#### **1.4 Manfaat Penelitian**

Adapun beberapa manfaat yang diharapkan dalam penelitian ini yaitu:

1. Memberikan pemahaman tentang performa algoritma *decision tree* C5.0 dalam mengklasifikasikan status mahasiswa yang berisiko *drop out*, sehingga algoritma ini dapat digunakan secara optimal dalam analisis data akademik.
2. Memberikan wawasan tentang seleksi fitur *mutual information* dalam meningkatkan akurasi algoritma C5.0, yang dapat menjadi pertimbangan dalam memilih metode terbaik untuk analisis data serupa.
3. Mengidentifikasi fitur-fitur yang berpengaruh terhadap prediksi mahasiswa yang berisiko *drop out*, sehingga pihak akademik dapat lebih fokus pada faktor-faktor yang signifikan dalam upaya pencegahan *drop out*.
4. Menjadi acuan bagi pihak akademik dalam menyusun strategi pencegahan *drop out* dan mengidentifikasi mahasiswa yang berisiko *drop out*.

#### **1.5 Batasan Masalah**

Batasan masalah yang terkait dengan penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data mahasiswa Program Sarjana (S1) angkatan 2018 hingga 2023 pada semester ganjil tahun akademik 2024/2025 di Universitas Jambi.
2. Data dibagi menjadi dua kelompok, yaitu data mahasiswa angkatan 2022–2023 dan data mahasiswa angkatan 2018–2021.
3. Atribut data yang digunakan diperoleh dari sistem SIAKAD dan ELISTA yang disediakan oleh UPA TIK sesuai dengan ketersediaan data.
4. Teknik Random Undersampling digunakan untuk menangani permasalahan ketidakseimbangan kelas pada data.
5. Seleksi fitur dilakukan menggunakan metode Mutual Information.
6. Model machine learning yang digunakan adalah algoritma Decision Tree C5.0.
7. Evaluasi model dilakukan menggunakan Confusion Matrix, dengan metrik pengukuran berupa nilai accuracy, precision, recall, dan F1-score.