

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perguruan tinggi memiliki peran strategis dalam memastikan mahasiswa dapat menyelesaikan studi sesuai dengan durasi yang telah ditetapkan. Tingkat kelulusan tepat waktu sering digunakan sebagai salah satu indikator kinerja institusi karena mencerminkan efektivitas proses pembelajaran serta dukungan akademik yang tersedia [1]. Namun, perguruan tinggi dihadapkan pada kenyataan bahwa kinerja akademik mahasiswa bersifat beragam dan dinamis, sehingga tidak semua mahasiswa mampu mempertahankan performa yang stabil hingga akhir masa studi. Variasi ini menjadi sumber utama munculnya permasalahan keterlambatan kelulusan dan putus studi [2].

Permasalahan tersebut semakin terlihat pada periode pembelajaran yang terdampak pandemi Covid-19. Perubahan sistem pembelajaran yang terjadi secara masif pada tahun 2020, ketika sekitar 94% perguruan tinggi di Indonesia beralih ke pembelajaran daring, menuntut mahasiswa untuk beradaptasi dengan lingkungan akademik yang berbeda dari sebelumnya [3]. Dampak kondisi ini tercermin pada data PDDIKTI tahun 2022 yang menunjukkan peningkatan proporsi mahasiswa berstatus nonaktif dari 6,8% pada 2020 menjadi 10,3% pada 2022. Peningkatan ini mengindikasikan adanya gangguan pada stabilitas capaian akademik mahasiswa yang tidak selalu dapat diidentifikasi secara dini melalui mekanisme pemantauan konvensional [4].

Seiring dengan meningkatnya kompleksitas permasalahan akademik, pendekatan pembinaan yang diterapkan oleh sebagian besar perguruan tinggi masih bersifat reaktif. Pemantauan kinerja mahasiswa umumnya mengandalkan proses administratif rutin, seperti evaluasi nilai dan Indeks Prestasi Semester di akhir periode tertentu. Intervensi akademik baru dilakukan ketika penurunan kinerja sudah terjadi secara signifikan. Kondisi ini selaras dengan temuan bahwa lebih dari 70% mahasiswa yang mengalami penurunan IPS selama dua semester berturut-turut

tidak memperoleh pembinaan sejak tahap awal, sehingga masalah akademik cenderung terakumulasi sebelum ditangani [5].

Keterbatasan pendekatan reaktif tersebut menunjukkan adanya kesenjangan antara kebutuhan pemantauan akademik yang bersifat berkelanjutan dan praktik yang berjalan di lapangan [6]. Tanpa dukungan sistem yang mampu mendeteksi potensi risiko sejak dini, dosen pembimbing akademik sering kali berada pada posisi untuk menangani masalah yang sudah berkembang. Situasi ini menegaskan pentingnya pendekatan berbasis data yang mampu membantu institusi pendidikan dalam memahami pola kinerja mahasiswa sebelum permasalahan akademik mencapai tahap kritis [7].

Permasalahan umum tersebut juga tercermin di Universitas Multimedia Nusantara, khususnya pada Program Studi Sistem Informasi. Program studi ini memiliki kurikulum yang menuntut konsistensi kinerja akademik mahasiswa sejak semester awal hingga akhir masa studi. Ketepatan dalam memantau perkembangan akademik menjadi faktor penting untuk memastikan mahasiswa dapat lulus sesuai target. Namun, proses pemantauan dan bimbingan yang masih mengandalkan identifikasi manual menyebabkan adanya jeda waktu antara munculnya indikasi penurunan performa dengan pelaksanaan intervensi akademik [8].

Kondisi tersebut menjadi semakin relevan ketika dikaitkan dengan karakteristik kohort mahasiswa angkatan 2020 hingga 2024. Angkatan ini menjalani proses pembelajaran pada masa transisi akibat pandemi Covid-19, mulai dari pembelajaran jarak jauh penuh hingga penerapan skema *asinkron* [8]. Perubahan sistem pembelajaran yang terjadi dalam rentang waktu tersebut menghasilkan variasi kondisi akademik yang berbeda dibandingkan angkatan sebelumnya. Variasi ini meningkatkan risiko keterlambatan kelulusan dan putus studi apabila tidak diantisipasi melalui mekanisme pemantauan yang lebih cermat dan bersifat prediktif.

Berdasarkan kondisi tersebut, Program Studi Sistem Informasi UMN memerlukan sebuah pendekatan yang mampu memberikan gambaran awal mengenai potensi status kelulusan mahasiswa. Pendekatan ini diharapkan tidak

hanya berfungsi untuk mengidentifikasi mahasiswa yang telah mengalami penurunan kinerja, tetapi juga mampu memprediksi kemungkinan hasil studi di masa mendatang. Dengan demikian, intervensi akademik dapat dilakukan lebih awal dan lebih terarah sesuai dengan karakteristik masing-masing mahasiswa.

Dalam beberapa tahun terakhir, pendekatan *machine learning* telah digunakan untuk memprediksi kinerja akademik mahasiswa dengan berbagai tingkat keberhasilan. Sejumlah penelitian menunjukkan bahwa model klasifikasi mampu mencapai tingkat akurasi yang tinggi, tetapi masih menghadapi keterbatasan, seperti ketergantungan pada fitur yang bersifat agregat, keterbatasan interpretasi model, serta tantangan dalam menangani *dataset* berukuran kecil dan tidak seimbang [9], [10]. Keterbatasan tersebut menunjukkan bahwa pemilihan algoritma dan strategi pemodelan perlu disesuaikan dengan karakteristik data pendidikan tinggi.

Oleh karena itu, penelitian ini difokuskan pada pengembangan model prediksi status kelulusan mahasiswa menggunakan tiga algoritma, yaitu CatBoost, *Logistic Regression* dengan regularisasi L1 (Lasso), dan *K-Nearest Neighbor* (KNN). Pemilihan ketiga algoritma tersebut didasarkan pada pertimbangan metodologis yang berkaitan dengan karakteristik data akademik dan tujuan prediksi yang bersifat multikelas. CatBoost dipilih karena mampu menangani hubungan non-linear antar fitur serta relatif stabil pada I berukuran terbatas dan tidak seimbang, yang umum ditemukan pada data akademik program studi. *Logistic Regression* dengan regularisasi L1 digunakan sebagai representasi model linear yang memungkinkan evaluasi kontribusi fitur secara lebih sederhana melalui mekanisme seleksi fitur implisit, sehingga relevan untuk membandingkan sejauh mana pola linear dapat menjelaskan status kelulusan mahasiswa. Sementara itu, *K-Nearest Neighbor* digunakan sebagai pendekatan non-parametrik berbasis kedekatan data untuk menilai kemampuan pemodelan tanpa asumsi bentuk fungsi prediksi tertentu. Dengan menggunakan ketiga model tersebut, penelitian ini memungkinkan perbandingan yang terstruktur antara pendekatan ensemble non-linear, model linear terregularisasi, dan metode berbasis *instance* pada konteks data akademik multikelas yang memiliki distribusi kelas tidak seimbang. Status

kelulusan mahasiswa diklasifikasikan ke dalam empat kategori, yaitu Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, Lulus Lebih Awal, dan Drop Out. Untuk mengurangi pengaruh ketidakseimbangan kelas pada proses pelatihan model, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan pada data pelatihan sebelum tahap pemodelan dilakukan.

Tujuan utama dari penelitian ini adalah membangun model prediksi yang mampu mengklasifikasikan status kelulusan mahasiswa secara akurat dan konsisten, sekaligus relevan untuk diterapkan pada lingkungan Program Studi Sistem Informasi Universitas Multimedia Nusantara. Secara praktis, hasil penelitian ini diarahkan untuk mendukung pengembangan purwarupa sistem peringatan dini yang dapat membantu dosen pembimbing akademik dalam mengidentifikasi mahasiswa berisiko sejak tahap awal. Dengan dukungan prediksi berbasis data, proses pembinaan akademik diharapkan dapat berjalan lebih terarah dan preventif, khususnya dalam menghadapi dinamika akademik pasca pandemi.

1.2 Rumusan Masalah

Berikut adalah rumusan masalah yang akan dibahas dalam penelitian ini:

1. Bagaimana kinerja model *machine learning* dalam memprediksi status kelulusan mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara ke dalam empat kelas utama?
2. Bagaimana perbandingan performa tiga model prediksi, yaitu CatBoost, *Logistic Regression* dengan regularisasi L1 (Lasso), dan *K-Nearest Neighbor* (KNN), berdasarkan hasil evaluasi model pada data akademik mahasiswa?
3. Bagaimana dan sejauh mana model prediktif terbaik tersebut mampu memberikan hasil interpretasi melalui pendekatan *Explainable AI* menggunakan metode SHAP dalam menjelaskan faktor-faktor yang memengaruhi prediksi status kelulusan mahasiswa?

1.3 Tujuan Penelitian

Berikut adalah tujuan penelitian yang akan dibahas dalam penelitian ini:

- 1 Mengevaluasi kinerja model *machine learning* dalam memprediksi status kelulusan mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara ke dalam empat kelas, yaitu Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, Lulus Lebih Awal, dan Drop Out?
- 2 Mengevaluasi perbandingan performa tiga model prediksi, yaitu CatBoost, Logistic Regression dengan regularisasi L1 (Lasso), dan K-Nearest Neighbor (KNN), berdasarkan hasil evaluasi model pada data akademik mahasiswa?
- 3 Menganalisis sejauh mana model-model prediktif tersebut mampu memberikan hasil interpretasi melalui pendekatan Explainable AI menggunakan metode SHAP dalam menjelaskan faktor-faktor yang memengaruhi prediksi status kelulusan mahasiswa?

1.4 Urgensi Penelitian

Berikut adalah urgensi penelitian yang terbagi menjadi 2 yaitu sisi akademik dan teknologi, diantaranya:

1. Urgensi Akademik

Urgensi akademik penelitian ini berkaitan langsung dengan kebutuhan Universitas Multimedia Nusantara khususnya Program Studi Sistem Informasi dalam meningkatkan efektivitas pemantauan dan pembinaan akademik mahasiswa. UMN memiliki target kelulusan yang terstruktur dan kurikulum yang menuntut konsistensi capaian akademik sejak semester awal hingga akhir masa studi. Namun, dinamika kinerja mahasiswa yang beragam menyebabkan tidak seluruh mahasiswa dapat mempertahankan performa akademik yang stabil selama masa studi. Kondisi ini menuntut adanya pendekatan pemantauan yang mampu mengidentifikasi potensi permasalahan akademik secara lebih dini.

Kondisi akademik tersebut menjadi semakin relevan pada kohort mahasiswa UMN angkatan 2020 hingga 2024 yang menjalani proses pembelajaran pada masa pandemi Covid-19. Perubahan sistem pembelajaran dari tatap muka ke pembelajaran daring, kemudian beralih ke skema *asinkron*, menghasilkan variasi pola capaian akademik yang berbeda

dibandingkan angkatan sebelumnya. Data nasional yang menunjukkan peningkatan mahasiswa berstatus nonaktif dari 6,8% pada 2020 menjadi 10,3% pada 2022 memberikan pemahaman bahwa risiko akademik pada periode tersebut bersifat nyata dan juga berpotensi terjadi di lingkungan UMN. Oleh karena itu, diperlukan kajian yang mampu memberikan gambaran prediktif terhadap status kelulusan mahasiswa dengan mempertimbangkan dinamika tersebut.

Urgensi akademik penelitian ini juga terletak pada upayanya untuk mendukung proses pembinaan akademik di UMN agar tidak hanya bergantung pada evaluasi capaian akhir, seperti Indeks Prestasi Semester. Dengan memprediksi status kelulusan mahasiswa ke dalam empat kelas, yaitu Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, Lulus Lebih Awal, dan Drop Out, penelitian ini menyediakan kerangka analisis yang lebih komprehensif. Kerangka ini relevan untuk membantu dosen pembimbing akademik UMN dalam memahami kondisi mahasiswa secara lebih menyeluruh dan berkelanjutan.

2. Urgensi Teknologi

Dari perspektif teknologi, urgensi penelitian ini berkaitan dengan kebutuhan UMN akan pemanfaatan *machine learning* yang sesuai dengan karakteristik data akademik internal. Data akademik di lingkungan UMN, khususnya pada Program Studi Sistem Informasi, memiliki jumlah sampel yang terbatas dan distribusi kelas yang tidak seimbang. Kondisi ini menuntut pemilihan algoritma yang tepat agar model prediksi dapat bekerja secara stabil dan konsisten. Penggunaan CatBoost, *Logistic Regression* dengan regularisasi L1 (Lasso), dan *K-Nearest Neighbor* dalam penelitian ini memberikan gambaran komparatif mengenai performa berbagai pendekatan pemodelan pada kondisi data yang realistis di UMN.

Selain performa, penerapan teknologi di lingkungan akademik UMN juga menuntut transparansi hasil analisis. Dosen pembimbing akademik memerlukan informasi yang tidak hanya menunjukkan hasil prediksi, tetapi juga menjelaskan faktor-faktor yang memengaruhi prediksi tersebut. Oleh karena itu, penerapan pendekatan *Explainable AI*

menggunakan metode SHAP menjadi aspek penting dalam penelitian ini untuk mendukung interpretasi model secara jelas dan terstruktur. Interpretasi ini memungkinkan hasil prediksi untuk lebih mudah dipahami dan dimanfaatkan dalam proses pembinaan akademik di UMN.

Urgensi teknologi penelitian ini juga terkait dengan pengembangan sistem pendukung keputusan akademik di UMN. Dengan mengombinasikan evaluasi performa model dan interpretasi hasil prediksi, penelitian ini menyediakan landasan teknis bagi pengembangan purwarupa sistem peringatan dini yang relevan dengan kebutuhan Program Studi Sistem Informasi UMN. Pendekatan ini mendukung pemanfaatan teknologi machine learning sebagai alat bantu pengambilan keputusan akademik yang berbasis data dan dapat diintegrasikan ke dalam proses pengelolaan akademik di lingkungan UMN.

1.5 Luaran Penelitian

Luaran utama dari penelitian ini adalah artikel ilmiah yang akan dipublikasikan pada jurnal nasional terakreditasi Sinta 2 atau Sinta 3, atau alternatifnya pada prosiding konferensi nasional bereputasi di bidang sistem informasi, *machine learning*, maupun *educational data mining*. Luaran penelitian diharapkan dapat memberikan kontribusi di antaranya:

1. Menyebarluaskan hasil penelitian kepada komunitas akademik dan praktisi, khususnya di bidang kecerdasan buatan dan analisis pendidikan, sebagai kontribusi ilmiah yang dapat menjadi acuan bagi penelitian lanjutan maupun pengembangan sistem pendukung keputusan akademik.
2. Memberikan kontribusi ilmiah terkait penerapan *explainable AI* dalam membandingkan tiga model prediksi (CatBoost, Lasso, dan KNN) untuk memperjelas pengaruh faktor eksternal terhadap kinerja akademik mahasiswa.
3. Menjadi bukti nyata ketercapaian penelitian tingkat sarjana dalam menghasilkan luaran yang terpublikasi pada jurnal atau konferensi bereputasi nasional.

1.6 Manfaat Penelitian

Berikut adalah manfaat penelitian yang didapatkan dari penelitian ini, diantaranya:

1. Manfaat Akademik

Secara akademik, penelitian ini memberikan kontribusi terhadap pengembangan kajian prediksi kinerja dan status kelulusan mahasiswa di lingkungan pendidikan tinggi. Penelitian ini memperluas pendekatan yang umum digunakan dengan menerapkan klasifikasi multikelas untuk memprediksi status kelulusan mahasiswa ke dalam empat kategori, yaitu Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, Lulus Lebih Awal, dan Drop Out. Pendekatan ini memberikan sudut pandang yang lebih komprehensif dibandingkan penelitian yang hanya berfokus pada satu indikator akademik atau klasifikasi biner.

Selain itu, penelitian ini menyajikan perbandingan performa tiga algoritma machine learning, yaitu CatBoost, *Logistic Regression* dengan regularisasi L1 (Lasso), dan *K-Nearest Neighbor*, pada dataset akademik yang berukuran kecil dan tidak seimbang. Hasil penelitian ini diharapkan dapat menjadi referensi ilmiah terkait pemilihan metode prediksi yang sesuai untuk data akademik di pendidikan tinggi, khususnya pada konteks program studi dengan karakteristik serupa dengan Program Studi Sistem Informasi UMN.

Penelitian ini juga memberikan kontribusi dalam aspek interpretabilitas model melalui penerapan *Explainable AI* menggunakan metode SHAP. Pendekatan ini mendukung pemahaman yang lebih mendalam mengenai kontribusi fitur akademik terhadap hasil prediksi, sehingga hasil penelitian tidak hanya berorientasi pada performa model, tetapi juga pada pemahaman proses pengambilan keputusan model secara ilmiah.

2. Manfaat Praksis

Dari sisi praktis, penelitian ini memberikan manfaat langsung bagi Universitas Multimedia Nusantara, khususnya Program Studi Sistem Informasi, dalam mendukung proses pembinaan akademik mahasiswa.

Model prediksi yang dihasilkan dapat digunakan sebagai dasar untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan kelulusan atau berisiko *drop out* sejak tahap awal masa studi. Informasi ini dapat membantu dosen pembimbing akademik dalam merencanakan intervensi yang lebih terarah dan sesuai dengan kondisi masing-masing mahasiswa.

Selain itu, hasil interpretasi model melalui SHAP memberikan penjelasan mengenai faktor-faktor akademik yang berkontribusi terhadap risiko kelulusan mahasiswa. Informasi ini dapat dimanfaatkan oleh pihak program studi untuk mengevaluasi kebijakan akademik, pola pengambilan mata kuliah, serta mekanisme pendampingan yang telah berjalan. Dengan demikian, penelitian ini mendukung pengambilan keputusan akademik yang lebih berbasis data dan relevan dengan kondisi nyata mahasiswa UMN.

Penelitian ini juga dapat menjadi landasan awal bagi pengembangan sistem peringatan dini di lingkungan UMN. Dengan adanya gambaran prediktif mengenai status kelulusan mahasiswa, program studi memiliki referensi tambahan dalam memantau perkembangan akademik mahasiswa secara berkelanjutan, tanpa hanya bergantung pada evaluasi capaian akhir semester.

3. Manfaat Pengembangan Teknologi

Dalam pengembangan teknologi, penelitian ini memberikan manfaat berupa penerapan dan evaluasi *machine learning* yang sesuai dengan karakteristik data akademik perguruan tinggi. Penggunaan CatBoost, *Logistic Regression* L1, dan *K-Nearest Neighbor* memberikan gambaran mengenai kelebihan dan keterbatasan masing-masing algoritma dalam menangani data berukuran kecil dan tidak seimbang, seperti yang ditemukan pada data akademik Program Studi Sistem Informasi UMN.

Selain itu, integrasi pendekatan *Explainable AI* menggunakan SHAP menunjukkan bagaimana teknologi interpretasi model dapat diterapkan secara praktis dalam sistem pendukung keputusan akademik.

Pendekatan ini mendukung pengembangan solusi teknologi yang tidak hanya berfokus pada akurasi, tetapi juga pada transparansi dan kemudahan pemahaman hasil analisis oleh pengguna non-teknis, seperti dosen pembimbing akademik.

Manfaat pengembangan teknologi dari penelitian ini juga terletak pada penyediaan kerangka teknis yang dapat dikembangkan lebih lanjut menjadi purwarupa sistem peringatan dini berbasis *machine learning*. Kerangka ini dapat menjadi dasar untuk integrasi dengan sistem akademik yang telah ada di UMN, serta membuka peluang pengembangan lanjutan pada skala program studi atau fakultas dengan karakteristik data yang serupa.

