

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kinerja akademik mahasiswa merupakan salah satu indikator utama keberhasilan proses pendidikan di perguruan tinggi [1]. Identifikasi dini terhadap mahasiswa yang berpotensi mengalami penurunan prestasi menjadi instrumen krusial agar institusi dapat memberikan intervensi preventif secara tepat waktu, seperti bimbingan akademik, pendampingan, atau dukungan finansial [1], [2]. Kebutuhan ini mendorong berkembangnya pemanfaatan data pendidikan melalui pendekatan *learning analytics* dan *educational data mining (EDM)* di lingkungan pendidikan tinggi [2]. Meskipun demikian, sebagian besar implementasi analitika pembelajaran saat ini masih bersifat deskriptif dan belum sepenuhnya bergeser ke arah prediktif yang mampu mengantisipasi capaian mahasiswa di masa depan [2].

Selain itu, model prediksi tradisional kerap kali hanya bertumpu pada data akademik internal kampus, seperti aktivitas pada *learning management system (LMS)* [3]. Padahal, banyak faktor krusial yang memengaruhi keberhasilan belajar justru berada di luar area kampus [3]. Bukti empiris menunjukkan bahwa variabel non-kognitif dan faktor eksternal seperti kondisi sosial-ekonomi keluarga, dukungan orang tua, lingkungan belajar, fasilitas, gaya hidup, dan manajemen waktu memiliki hubungan timbal balik yang sangat signifikan dengan keberhasilan akademik mahasiswa [3], [1], [4]. Oleh karena itu, memisahkan dan memfokuskan pemodelan prediktif murni pada faktor eksternal merupakan langkah strategis untuk memahami capaian akademik dari perspektif yang lebih utuh [1], [3]. Hubungan antara faktor eksternal dan kinerja akademik ini bersifat kompleks dan nonlinear, sehingga pendekatan pemodelan algoritmik seperti *ensemble learning* (misalnya **Random Forest**, **CatBoost**,

XGBoost) dan *deep learning* (seperti Feedforward Neural Network / FNN dan Long Short-Term Memory/LSTM) terbukti lebih unggul dalam hal akurasi dibandingkan metode statistik parametrik tradisional [1], [5], [6].

Namun, dalam konteks komputasi pendidikan, tingginya akurasi model pada data latih (*training data*) yang dihasilkan oleh arsitektur *ensemble* maupun *deep learning* belumlah cukup [7]. Algoritma prediktif dalam dunia pendidikan menghasilkan keputusan-keputusan yang berdampak langsung pada manajemen intervensi mahasiswa [7]. Sayangnya, banyak model canggih saat ini terjebak dalam "**krisis generalisasi**" (*generalizability crisis*), di mana model tersebut tampak sangat akurat saat diuji pada lingkungan internal laboratorium (*site-specific*), tetapi performanya menurun tajam ketika diterapkan pada skenario nyata lintas lokasi, wilayah geografis, atau kelompok data yang berbeda [7], [8], [9]. Model-model prediktif ini juga rentan mengalami pergeseran pola data seiring berjalannya waktu (*temporal drift* atau *concept drift*) akibat perubahan karakteristik mahasiswa dari tahun ke tahun [10], [11], [12].

Kesenjangan metodologis tersebut menunjukkan bahwa sebuah model prediksi kinerja akademik tidak boleh dibekukan secara statis (*frozen model*) tanpa adanya pengujian ketahanan secara berkala [12]. Probabilitas prediksi yang dihasilkan harus diuji keandalannya secara ketat melalui metode *temporal validation* (**validasi berbasis waktu**) dan **validasi lintas cohort (lintas angkatan)** [11], [12]. Melalui validasi lintas cohort dan berkala (*model refitting*), institusi dapat menilai sejauh mana model yang dikembangkan mampu mempertahankan performa prediksi dan kalibrasi yang konsisten ketika dihadapkan pada angkatan mahasiswa baru atau periode akademik yang berbeda [11], [12]. Selain itu, evaluasi komparatif performa tidak boleh terjebak pada metrik akurasi standar yang cenderung memberikan hasil terlalu optimis (*optimistic bias*) pada dataset yang tidak seimbang (*imbalanced dataset*), melainkan harus melibatkan metrik yang

lebih objektif seperti *Balanced Accuracy* dan *Matthews Correlation Coefficient (MCC)* untuk menjamin perbandingan yang adil (fair comparison) di seluruh kuadran matriks konfusi [13], [14], [15]. Melalui analisis stabilitas dan degradasi inilah, tingkat ketidakpastian serta risiko prediksi dapat dipetakan secara transparan, sehingga model tersebut dapat digunakan sebagai sistem pendukung keputusan yang bertanggung jawab di perguruan tinggi [7], [11].

Penelitian ini dilaksanakan pada konteks mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara dengan memanfaatkan faktor eksternal yang diekstraksi dari biodata, transkrip akademik, dan kuesioner. Berdasarkan uraian di atas, penelitian ini secara khusus melakukan analisis generalisasi lintas batas, validasi temporal lintas cohort, serta evaluasi stabilitas komparatif terhadap performa model ensemble learning dan deep learning [1], [5], [7]. Melalui evaluasi multi-metrik yang ketat (termasuk Hybrid Score berbasis MCC dan F1-macro), studi komparatif ini diharapkan mampu memberikan kontribusi ilmiah baru mengenai batas-batas ketahanan, kelayakan, serta visualisasi tata kelola risiko model prediksi akademik berbasis faktor eksternal di lingkungan pendidikan tinggi [13], [14], [15].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini difokuskan pada pengembangan, pengujian stabilitas, dan validasi internal keandalan model komparatif, yang dirumuskan sebagai berikut:

1. Bagaimana perbandingan performa prediktif antara pendekatan *ensemble learning* (Optimized Random Forest) dan *deep learning* (FNN, BiLSTM berbasis Sekuens Tematik) dalam memprediksi kinerja akademik mahasiswa berbasis faktor eksternal secara adil

- (*fair comparison*) menggunakan metrik *Balanced Accuracy* dan *Matthews Correlation Coefficient* (MCC) [13], [14], [15]?
2. Sejauh mana model prediksi kinerja akademik mahasiswa berbasis faktor eksternal yang dikembangkan mampu mempertahankan konsistensi performanya ketika divalidasi secara internal melalui skema temporal lintas *cohort* (lintas angkatan mahasiswa) guna mengantisipasi risiko kegagalan generalisasi temporal [7], [11], [12]?
 3. Bagaimana karakteristik fluktuasi stabilitas, tingkat reliabilitas probabilitas, dan pola degradasi performa model (*performance drift*) yang terjadi pada arsitektur *ensemble learning* dan *deep learning* seiring dengan adanya pergeseran pola data eksternal mahasiswa antar-generasi [8], [10], [12]?

1.3 Batasan Masalah

Agar proses evaluasi dan pengujian generalisasi model tetap terarah serta sesuai dengan ruang lingkup yang telah ditetapkan, batasan masalah dalam penelitian ini diatur sebagai berikut:

1. Objek penelitian dibatasi pada mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara angkatan 2020 hingga 2024 untuk memfasilitasi analisis data lintas generasi akademik [1].
2. Variabel prediktor yang dievaluasi murni hanya menggunakan faktor eksternal mahasiswa yang bersifat non-kognitif (mencakup dimensi kondisi sosial-ekonomi keluarga, dukungan orang tua, lingkungan belajar, fasilitas fisik, gaya hidup, dan manajemen waktu) [1], [3], [4]. Nilai akademik sekuensial internal kampus (seperti aktivitas harian pada *learning management system*/LMS) tidak dilibatkan sebagai prediktor langsung untuk memastikan

evaluasi independen terhadap pengaruh lingkungan luar kampus [3], [13].

3. Variabel target penelitian merupakan klasifikasi biner dari kategori indeks prestasi kumulatif (*Clean-Margin GPA Category*) guna memisahkan kelompok mahasiswa secara polaritatif dan tegas, bukan merupakan nilai indeks prestasi dalam skala kontinu [5], [15].
4. Terdapat perbedaan batasan ukuran sampel operasional pasca-praproses antara kedua jalur pemodelan, yaitu sebanyak 131 sampel rekam jejak mahasiswa pada jalur *Ensemble Learning* dan 136 sampel mahasiswa pada jalur *Deep Learning* [1], [4]. Perbedaan ini dipicu oleh batasan pemotongan spesifik tiap arsitektur (*Pruned Binary Classification*), di mana jalur *Ensemble* mengeliminasi sampel pada rentang IPK 3,50 hingga $<3,70$ [7], sedangkan jalur *Deep Learning* memangkas data ekstrem bawah ($IPK \leq 1,00$) untuk menjaga konvergensi gradien serta mereduksi zona ambiguitas (*grey zone*) pada rentang IPK 3,50–3,67 [4], [16]. Alasan utama (*rationale*) eliminasi rentang IPK 3,50 hingga $<3,67$ berakar dari aturan kampus di mana $IPK \geq 3,67$ merupakan ambang kelulusan *With Honors/Cum Laude* (Pujian), sedangkan nilai di bawah 3,50 mewakili kategori umum/risiko intervensi. Membiarkan sampel pada rentang 3,50–3,67 ikut dilatih sebagai target biner akan menciptakan *overlap* fitur sosiokultural yang sangat tinggi karena karakteristik mahasiswanya hampir identik secara statistik [7], [15].
5. Berdasarkan dampak pemangkasan (*results obtained*) dari penerapan *Pruned Binary Classification* (131 vs 136 sampel), eliminasi zona ambigu ini berhasil mempertegas batas keputusan (*decision boundary*), mereduksi *false certainty bias*, serta meningkatkan skor diskriminasi berimbang (*Matthews Correlation Coefficient/MCC*) secara signifikan dibandingkan jika rentang

tersebut dipaksa dimasukkan secara biner kaku ($IPK \geq 3,50$ vs $<3,50$) [13], [15]. Selain itu, insight utama yang diperoleh menunjukkan bahwa sensitivitas algoritma terhadap manipulasi data praproses bersifat *architecture-specific*, di mana jaringan saraf tiruan sangat rentan terhadap pencilan (*outliers*) yang merusak stabilitas pembaharuan bobot (*weight updates*), sementara model *Ensemble* berbasis pohon keputusan (*decision tree*) lebih tahan (*robust*) terhadap nilai ekstrem namun membutuhkan garis pemisah margin (*clean-margin*) yang lebih lebar [7], [15].

6. Ruang lingkup komparasi algoritma dibatasi pada perbandingan performa antara pendekatan ensemble learning (*Optimized Random Forest* berbasis *Nested Cross-Validation*) [3], [4] dan pendekatan *deep learning* (*Feedforward Neural Network/FNN* serta *Bidirectional LSTM* berbasis Sekuens Tematik lintas 8 domain sosiokultural) [13], [14] yang telah dikembangkan secara spesifik oleh Mahasiswa 1 dan Mahasiswa 2.
7. Metode validasi performa model difokuskan pada pengujian eksternal melalui validasi sekuensial berbasis waktu (*temporal validation*) dan validasi lintas *cohort* (antar-angkatan) untuk mensimulasikan kinerja model saat menghadapi data angkatan mahasiswa yang lebih baru, bukan sekadar menggunakan validasi silang acak standar (*random cross-validation*) yang rentan memicu bias optimis [7], [11], [12].
8. Metrik evaluasi utama yang digunakan untuk menilai keadilan perbandingan (*fair comparison*) model di bawah kondisi ketidakseimbangan data (*mild class imbalance*) alami dibatasi pada *Matthews Correlation Coefficient* (MCC), *Balanced Accuracy*, *F1-score macro*, *precision*, dan *recall* [13], [15].
9. Analisis stabilitas dibatasi pada pemetaan perilaku penurunan akurasi model (*performance drift* atau *concept drift*) seiring waktu [10], [12], kuantifikasi galat probabilistik (*Brier Score* dan

Reliability Error), serta visualisasi peta risiko prediktif (*Risk Governance Funnel*) [7], di mana sifat hubungan pemodelan adalah prediktif-asosiatif tanpa ditujukan untuk menetapkan hubungan kausalitas struktural murni antara faktor eksternal dan capaian prestasi akademik mahasiswa [1], [6].

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, tujuan utama dari penelitian ini adalah untuk mengembangkan dan memvalidasi secara internal model prediksi indeks prestasi mahasiswa berbasis faktor eksternal dengan pendekatan *ensemble* dan *deep learning* [7]. Secara spesifik, penelitian ini ditujukan untuk:

1. Mengevaluasi dan membandingkan efektivitas performa prediktif secara adil (*fair comparison*) antara arsitektur *ensemble learning* (Optimized Random Forest) dan *deep learning* (*Feedforward Neural Network*/FNN serta *Bidirectional LSTM* berbasis Sekuens Tematik lintas 8 domain sosiokultural) dalam memprediksi kategori biner indeks prestasi mahasiswa menggunakan metrik berimbang seperti *Balanced Accuracy* dan *Matthews Correlation Coefficient* (MCC) [13], [14], [15].
2. Menganalisis ketahanan, keadilan perbandingan, dan kemampuan generalisasi temporal dari model prediktif yang dikembangkan ketika dihadapkan pada skenario dunia nyata melalui skema validasi temporal lintas *cohort* (antar-angkatan mahasiswa) [7], [11], [12].
3. Memetakan fluktuasi dispersi stabilitas, kualitas kalibrasi probabilistik (*Brier Score* dan *Reliability Error*), serta pola degradasi performa model (*performance drift* atau *concept drift*) seiring berjalannya waktu akibat adanya pergeseran pola data eksternal mahasiswa antar-generasi guna mengidentifikasi batas kelayakan operasional sistem prediksi [8], [10], [12].

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi signifikan yang terbagi ke dalam tiga aspek utama:

1. **Manfaat Teoretis:** Memberikan kontribusi ilmiah pada kajian *educational data mining* (EDM) dan *learning analytics*, khususnya mengenai pemetaan batas-batas generalisasi model prediktif yang murni memanfaatkan faktor eksternal sosiokultural non-kognitif [2], [3], [6]. Penelitian ini memperluas literatur mengenai cara mengatasi krisis generalisasi (*generalizability crisis*) dan bias optimis (*optimistic bias*) pada pemodelan data tabular berskala kecil di lingkungan pendidikan tinggi [7], [9], [15].
2. **Manfaat Praktis:** Menyediakan dasar rekomendasi dan alat audit yang objektif bagi institusi, khususnya dosen wali dan pengelola program studi, untuk memahami masa pakai model (*model lifespan*) seiring pergeseran angkatan [12]. Melalui deteksi dini degradasi performa dan pemantauan sebaran ketidakpastian (*Risk Governance Funnel*), institusi dapat mengetahui kapan sebuah model prediktif harus dikalibrasi atau dilatih ulang (*model refitting*) sebelum luaran probabilitasnya digunakan sebagai basis intervensi preventif bagi mahasiswa berisiko [1], [7], [12].
3. **Manfaat Metodologis:** Menyajikan kerangka kerja evaluasi komparatif (*dual-modeling pipelines*) yang ketat, transparan, dan dapat direproduksi (*reproducible*), mulai dari standarisasi pengujian menggunakan metrik yang tahan terhadap ketidakseimbangan kelas (*class imbalance*), penerapan alur validasi temporal sekuensial yang bebas dari kebocoran data (*data leakage*), hingga pemanfaatan analisis sensitivitas (*Permutation Feature Importance*) untuk mengevaluasi konsensus fitur antar-arsitektur model [6], [13], [17].

1.5 Sistematika Penulisan

Laporan penelitian ini disusun ke dalam lima bab utama dengan sistematika penulisan yang terstruktur sebagai berikut:

- **BAB I PENDAHULUAN**

Memuat latar belakang penelitian yang menyoroti krisis generalisasi (*generalizability crisis*) dan bias optimis (*optimistic bias*) pada model prediksi [7], [9], rumusan masalah, batasan masalah, tujuan dan manfaat penelitian yang berorientasi pada ketahanan model lintas batas [7], serta sistematika penulisan laporan penelitian.

- **BAB II LANDASAN TEORI**

Memuat tinjauan penelitian terdahulu (*state-of-the-art*) terkait pemanfaatan prediktor sosiokultural non-kognitif [3], [1], landasan teori mengenai kinerja akademik berbasis faktor eksternal, *educational data mining* (EDM), dan *learning analytics* [2], [13]. Bab ini juga mendasari aspek teoretis dari algoritma *ensemble learning* (*Random Forest*, CatBoost, XGBoost) [1], [5], rumpun *deep learning* (*Feedforward Neural Network/FNN* serta *Bidirectional LSTM* berbasis Sekuens Tematik) [13], [14], metrik evaluasi klasifikasi yang adil menggunakan *Balanced Accuracy* dan *Matthews Correlation Coefficient* (MCC) [13], [15], serta konsep teoretis mengenai pergeseran pola data (*temporal drift* atau *concept drift*) dan degradasi performa model seiring waktu [10], [12].

- **BAB III METODOLOGI PENELITIAN**

Memuat alur pengumpulan data faktor eksternal mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara [1], tahapan dua jalur praproses spesifik arsitektur (*architecture-specific pipelines* dan rekayasa sekuens) [4], [5], serta spesifikasi desain eksperimen komparatif. Secara khusus, bab ini memaparkan metodologi pengujian generalisasi eksternal melalui skema validasi temporal lintas *cohort* (antar-angkatan) [11], [12], penentuan ambang batas keputusan (*decision threshold*), kalibrasi probabilitas menggunakan *Platt Scaling* [7], pemodelan indikator stabilitas, serta kerangka analisis kuantifikasi ketidakpastian prediksi (*Risk Governance Funnel*).

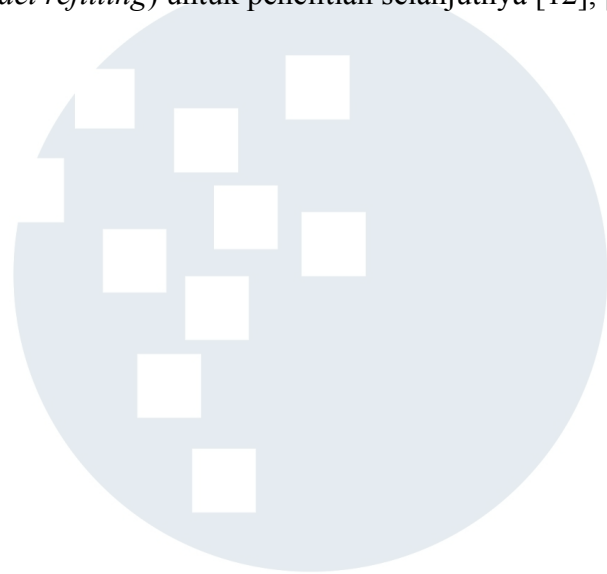
- **BAB IV HASIL DAN PEMBAHASAN**

Memuat analisis karakteristik sebaran dan integritas penyaringan partisipan (*clean-margin target reformulation*) hasil praproses [1], [5], visualisasi dan tabulasi hasil komparasi performa antara arsitektur *ensemble learning* dan *deep learning* berbasis nilai MCC dan *Balanced Accuracy* [13], [15], hasil pengujian ketahanan model melalui validasi temporal lintas *cohort* [11], [12], serta pemetaan grafik stabilitas lintas lipatan (*cross-validation stability*) [6]. Bab ini juga membedah ketimpangan sensitivitas (*recall disparity*) pada matriks konfusi, analisis sensitivitas penggerak eksternal utama (*top exogenous drivers*) [1], tingkat kesalahan kalibrasi probabilistik (*Brier Score* dan *Reliability Error*), serta ditutup dengan pembahasan mendalam mengenai masa pakai model (*model lifespan*), mitigasi risiko prediksi, dan batas

kelayakannya sebagai sistem pendukung keputusan yang bertanggung jawab [12], [7].

- **BAB V SIMPULAN DAN SARAN**

Memuat simpulan komprehensif yang menjawab seluruh rumusan masalah pengujian generalisasi dan stabilitas model komparatif, batasan eksperimen lapangan yang ditemukan, serta saran praktis bagi institusi mengenai implementasi kebijakan *Human-in-the-loop* dan jadwal berkala untuk kalibrasi atau pelatihan ulang model (*model refitting*) untuk penelitian selanjutnya [12], [7]. .



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA