

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kinerja akademik mahasiswa merupakan salah satu indikator utama keberhasilan proses pendidikan dan institusi di perguruan tinggi [14]. Identifikasi dini terhadap mahasiswa yang berpotensi mengalami penurunan kinerja akademik menjadi sangat penting agar institusi dapat memberikan intervensi preventif secara tepat waktu, seperti bimbingan akademik, konseling, maupun dukungan finansial. Kebutuhan krusial ini mendorong pesatnya perkembangan pemanfaatan data pendidikan melalui pendekatan *educational data mining* (EDM) dan *learning analytics* di lingkungan pendidikan tinggi [22].

Salah satu representasi metrik kinerja akademik yang paling umum digunakan adalah Indeks Prestasi Kumulatif (IPK). Selama beberapa dekade terakhir, evaluasi dan prediksi IPK umumnya didominasi oleh penilaian faktor-faktor prasyarat akademis internal, seperti rekam jejak nilai sekolah menengah, hasil ujian masuk, atau data log aktivitas mahasiswa pada *Learning Management System* (LMS) [16]. Akan tetapi, pemodelan analitika yang hanya bertumpu pada data akademik internal cenderung gagal menangkap gambaran utuh dinamika kehidupan belajar mahasiswa. Realitas empiris di lapangan menunjukkan bahwa fluktuasi IPK mahasiswa tidak terjadi di ruang hampa kognitif; sebagian besar faktor yang sangat mendikte kelulusan dan keberhasilan mahasiswa justru berada di luar kampus. Faktor eksternal non akademis seperti kondisi sosial-ekonomi keluarga, dukungan moral, beban pekerjaan sampingan (part-time), kesehatan psikososial, hingga ketersediaan fasilitas logistik belajar kerap menjadi variabel tak kasat mata yang menjadi akar penyebab utama (*root cause*) penurunan performa mahasiswa [10], [13].

Hubungan antara spektrum faktor eksternal ini dengan kinerja akademik bersifat sangat kompleks, asimetris, dan tidak selalu linear. Oleh karena itu, pendekatan pemodelan statistik parametrik tradisional seperti regresi linier seringkali gagal menangkap pola interdependensi dari kumpulan survei keseharian mahasiswa [17]. Seiring dengan lompatan kemajuan Kecerdasan Buatan (AI),

arsitektur Deep Learning bermunculan sebagai solusi algoritmik yang tangguh. Algoritma Feedforward Neural Network (FNN) umum digunakan untuk menambang representasi pola tersembunyi dari data tabular multidimensi secara spasial tanpa memerlukan rekayasa fitur manual yang rumit [1], [2]. Di sisi lain, arsitektur Long Short-Term Memory (LSTM), yang sejatinya dikhususkan untuk pemrosesan data deret waktu (*time-series*), menawarkan potensi heuristik relasional yang besar apabila sekumpulan faktor kehidupan mahasiswa diterjemahkan secara artifisial menjadi suatu sekuens atau hierarki tematik kronologis [18].

Tantangan mendasar yang sering dijumpai dalam pengumpulan data performa perguruan tinggi masa kini adalah distribusi nilai yang sangat condong ke atas (*top heavy distribution*) atau dikenal dengan istilah inflasi nilai. Ketika sebuah institusi memiliki populasi mahasiswa di mana mayoritas memperoleh IPK di atas rata-rata (misalnya terpusat di angka  $\geq 3,50$ ), dataset akan menderita ketidakseimbangan kelas (*class imbalance*) yang sangat ekstrem [11]. Jika dipaksakan menggunakan algoritma machine learning konvensional, model hanya akan menghasilkan tebakan bias terhadap kelas mayoritas. Oleh karena itu, terdapat justifikasi teoretis yang mendesak untuk membingkai masalah ini ke dalam bentuk klasifikasi biner menggunakan teknik pruning (pemangkasan zona ambigu pada batas IPK tertentu) untuk menghasilkan ketajaman margin pemisah antar kelas [3].

Selain masalah imbalance, terdapat hipotesis akademis mengenai keberadaan *Signal Ceiling* sebuah batasan intrinsik komputasional di mana sekumpulan variabel survei faktor eksternal murni (*noisy data*) tidak memiliki daya prediktif independen yang cukup kuat untuk menebak probabilitas kelulusan melebihi angka persentase tertentu, apabila hanya diolah secara spasial datar menggunakan model seperti FNN [1]. Oleh karena itu, penelitian ini dirancang secara sistematis untuk memecahkan kebuntuan *Signal Ceiling* tersebut melalui pengaplikasian representasi "Sekuens Tematik" pada jaringan memori LSTM, yang dikawal oleh lapisan Governance Cross Validation Engine guna memastikan bahwa kinerja prediksi model benar-benar andal, terkalibrasi, dan bebas dari bias yang menyesatkan [3].

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian MBKM ini adalah sebagai berikut:

1. Bagaimana merancang, memodifikasi, dan mengoptimasi arsitektur jaringan saraf Feedforward Neural Network (FNN) dan Long Short-Term Memory (LSTM) untuk memprediksi kategori capaian IPK mahasiswa berdasarkan dataset faktor eksternal?
2. Sejauh mana implementasi teknik Pruned Binary Classification dan class weighting efektif dalam menangani isu ketidakseimbangan kelas (class imbalance) ekstrem pada dataset akademik?
3. Apakah terdapat ambang batas prediktif maksimal (Signal Ceiling) pada penggunaan faktor eksternal murni dalam pemodelan FNN, dan sejauh mana representasi Sekuens Tematik pada LSTM mampu mematahkan batasan akurasi tersebut?

## 1.3 Batasan Penelitian

Agar penelitian ini tetap terarah, mendalam, dan relevan dengan ruang lingkup sumber daya waktu pelaksanaan program MBKM, batasan masalah ditetapkan sebagai berikut:

1. Objek dan sumber data penelitian adalah 169 responden mahasiswa dari Program Studi Sistem Informasi Universitas Multimedia Nusantara (UMN).
2. Variabel prediktor yang digunakan murni hanya berasal dari 40 faktor eksternal (seperti logistik finansial, keluarga, pekerjaan, fasilitas, dan manajemen kegiatan luar). Variabel faktor akademik internal (seperti rasio kehadiran absensi atau nilai historis mata kuliah) sengaja dikecualikan agar model dapat dievaluasi murni berdasarkan kemampuannya menangkap dinamika keseharian di luar akademis.
3. Variabel target penelitian dibingkai sebagai klasifikasi biner dalam memprediksi kategori IPK (CumLaude vs non-CumLaude. Hal ini dilakukan dengan eksklusi dan pemangkasan (*pruning*) data IPK yang

tidak masuk jangkauan relevansi model, yaitu menghapus sampel anomali ekstrem dengan  $IPK \leq 1,0$  serta membuang 38 sampel pada zona IPK ambigu (3,50 hingga 3,67) sehingga menyisakan 131 sampel data bersih yang divalidasi.

4. Pemodelan algoritma yang menjadi subjek perbandingan komparatif dalam penelitian ini dibatasi pada variasi arsitektur *Feedforward Neural Network* (FNN) dan variasi *Long Short Term Memory* (termasuk *BiLSTM + Attention*).

#### 1.4 Tujuan Penelitian

Tujuan utama dari pelaksanaan penelitian dan penyusunan laporan akhir ini adalah:

1. Mengembangkan dan memvalidasi internal model prediksi IPK berbasis faktor eksternal dengan FNN dan LSTM. Model ini dikembangkan sebagai instrumen prediktif yang dapat diandalkan oleh pengelola program studi.
2. Menganalisis dampak perlakuan pra-pemrosesan data menggunakan metode *Pruned Binary Classification* terhadap perbaikan stabilitas metrik evaluasi model di tengah ketidakseimbangan kelas.
3. Mengidentifikasi batas fenomena *Signal Ceiling* pada arsitektur spasial FNN, serta membuktikan secara empiris keunggulan *feature engineering* berbasis *Sekuens Tematik* pada mekanisme *gate memory* LSTM.

#### 1.5 Urgensi Penelitian

Urgensi penelitian ini terletak pada kebutuhan esensial instansi perguruan tinggi terhadap Sistem Peringatan Dini (*Early Warning System*) yang presisi dan proaktif. Pendekatan institusional yang reaktif terhadap evaluasi akademik seperti memberikan sanksi administratif, pemanggilan, atau Surat Peringatan (SP) baru *setelah* mahasiswa mengalami kegagalan di akhir semester sudah tidak lagi relevan dan memakan biaya psikososial yang sangat besar. Dengan mengetahui batasan akurasi komputasional dari pengaruh masalah keseharian (finansial, beban magang, gangguan mental), institusi dapat merancang intervensi yang tepat sasaran bagi

kelompok mahasiswa yang rentan secara logistik sebelum masalah tersebut terakumulasi menjadi penurunan IPK yang fatal secara permanen.

## **1.6 Luaran Penelitian**

Hasil dari pelaksanaan MBKM Penelitian ini ditargetkan untuk menghasilkan luaran berupa:

1. Laporan MBKM Penelitian komprehensif terkait rancang bangun pipeline Deep Learning pada data mahasiswa.
2. Draf publikasi artikel ilmiah yang membahas mengenai batasan klasifikasi Deep Learning berbasis faktor non-akademik, untuk diajukan ke jurnal atau prosiding bereputasi.
3. Artefak kode komputasi (*Jupyter Notebook/Python Script*) yang terdokumentasi.

## **1.7 Manfaat Penelitian**

### **1.7.1 Manfaat Teoritis**

Memberikan kontribusi dan memperkaya literatur pada ranah *Educational Data Mining* (EDM), khususnya terkait penemuan empiris mengenai fenomena kebuntuan Signal Ceiling pada survei kuesioner mahasiswa. Studi ini juga menawarkan validasi teoretis kebaruan (*novelty*) bahwa dataset tabular survei non-temporal dapat disimulasikan dan ditata ulang menjadi hierarki matriks sekuensial (dimensi waktu semu) guna memicu aktivasi kapabilitas *attention memory* pada algoritma LSTM.

### **1.7.2 Manfaat Praktis**

Menyediakan landasan model algoritmik (purwarupa *BiLSTM + Attention*) bagi institusi, khususnya dosen wali akademik dan pengelola program studi, untuk mengembangkan mekanisme deteksi dini profil mahasiswa berisiko. Intervensi tidak lagi berpatokan murni pada nilai, tetapi mempertimbangkan konteks holistik keseharian mahasiswa.

### **1.7.3 Manfaat Metodologis**

Menyajikan alur kerangka kerja pengembangan model klasifikasi *Deep Learning* berskala industri yang terstandardisasi melalui *Governance*

*Cross-Validation Engine*, yang membedah keandalan kalibrasi ambang batas (*threshold*), stabilitas, serta risiko ketidakpastian prediksi.

