

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Fenomena ketimpangan kinerja akademik mahasiswa masih menjadi perhatian utama dalam sistem pendidikan tinggi di Indonesia. Kinerja akademik yang diukur melalui Indeks Prestasi Kumulatif (IPK), ketepatan waktu kelulusan, dan capaian akademik lainnya sering kali menunjukkan perbedaan signifikan antar mahasiswa meskipun mereka berada dalam lingkungan pendidikan dan kurikulum yang sama. Berdasarkan Data Kemendikbudristek menunjukkan bahwa rata-rata IPK lulusan sarjana Indonesia pada tahun 2022 berada pada angka 3,00 [1], sedangkan tingkat putus kuliah (drop-out) mencapai 4,02% [2]. Temuan ini mengindikasikan bahwa masih terdapat perbedaan dalam capaian akademik dan keberhasilan studi mahasiswa di berbagai perguruan tinggi di Indonesia. Oleh karena itu, diperlukan pendekatan yang lebih sistematis dan berbasis data untuk memprediksi dan memahami pola kelulusan mahasiswa sejak dini.

Prediksi kelulusan mahasiswa berdasarkan faktor internal seperti IPK per semester, jumlah mata kuliah yang diambil, dan performa akademik dari waktu ke waktu menjadi salah satu topik penting dalam pengembangan sistem analitik pendidikan. Melalui penerapan kecerdasan buatan, khususnya machine learning, berbagai model prediksi telah digunakan untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan atau dropout. Model seperti Random Forest, XGBoost, LightGBM, CatBoost, dan Multilayer Perceptron (MLP) terbukti mampu mempelajari hubungan non-linear dan kompleks antar variabel akademik internal mahasiswa [3]. Meskipun demikian, sebagian penelitian sebelumnya masih menggunakan pendekatan model individual sehingga hasil prediksi sering kali belum optimal karena memerlukan data yang cukup besar [4]. Hal ini disebabkan karena setiap algoritma memiliki karakteristik dan keunggulan yang berbeda dalam menangani data akademik, baik dari sisi distribusi, kompleksitas, maupun jenis fitur.

Untuk menjawab tantangan tersebut, penelitian ini mengusulkan pendekatan hybrid ensemble yang menggabungkan beberapa model pembelajaran mesin, yaitu XGBoost, LightGBM, CatBoost, Random Forest, dan MLPClassifier. Pendekatan ini memungkinkan setiap model berkontribusi dalam proses prediksi berdasarkan kekuatan masing-masing, kemudian probabilitas yang dihasilkan digabungkan untuk membentuk keputusan akhir yang lebih stabil dan konsisten dibandingkan penggunaan model tunggal [5]. XGBoost dikenal unggul dalam menangani data tabular dengan kemampuan optimasi gradien yang efisien, LightGBM memiliki kecepatan pelatihan tinggi dengan akurasi kompetitif, CatBoost efektif dalam mengelola data kategorikal tanpa memerlukan encoding yang kompleks, Random Forest memberikan kemampuan generalisasi melalui proses agregasi banyak pohon keputusan, sedangkan MLPClassifier mampu menangkap pola non-linear yang tersembunyi dalam data akademik mahasiswa [6]. Dengan mengombinasikan kelima algoritma tersebut, pendekatan Hybrid Ensemble diharapkan mampu menghasilkan prediksi kelulusan yang lebih akurat, *generalizable*, serta lebih tahan terhadap variasi distribusi data.

Namun, penggunaan model machine learning yang kompleks menimbulkan tantangan baru dalam hal interpretabilitas. Model ensemble cenderung berperan sebagai “black box” karena proses pengambilan keputusannya sulit dilacak secara langsung. Dalam pendidikan, transparansi hasil prediksi sangat penting agar universitas dapat memahami faktor-faktor internal yang berkontribusi terhadap kelulusan mahasiswa dan merumuskan strategi akademik yang sesuai. Untuk mengatasi permasalahan interpretabilitas tersebut, penelitian ini mengintegrasikan pendekatan Explainable Artificial Intelligence (XAI) menggunakan metode SHapley Additive exPlanations (SHAP).

Metode SHAP dipilih karena memiliki dasar teori yang kuat dalam teori permainan (game theory), di mana setiap fitur diperlakukan sebagai kontributor yang memberikan pengaruh proporsional terhadap hasil prediksi. Dibandingkan dengan metode interpretasi lain seperti Local Interpretable Model-Agnostic

Explanations (LIME) dan Permutation Feature Importance, SHAP menawarkan tingkat konsistensi dan stabilitas yang lebih tinggi dalam menjelaskan kontribusi fitur pada model yang kompleks [7]. Selain itu, SHAP mampu memberikan interpretasi pada berbagai jenis algoritma, baik linear, tree-based, maupun neural network, sehingga sangat sesuai digunakan dalam pendekatan hybrid ensemble yang terdiri atas kombinasi model heterogen. Keunggulan lainnya adalah kemampuannya menjelaskan prediksi secara global maupun lokal, sehingga tidak hanya memperlihatkan fitur paling berpengaruh terhadap keseluruhan model, tetapi juga terhadap prediksi masing-masing mahasiswa [8].

Dengan demikian, penerapan SHAP dalam penelitian ini tidak hanya berfungsi untuk meningkatkan transparansi hasil prediksi, tetapi juga memberikan pemahaman yang lebih mendalam mengenai kontribusi setiap variabel akademik internal terhadap keputusan model. Pendekatan Hybrid Ensemble yang dikombinasikan dengan SHAP diharapkan mampu menghasilkan model prediksi kelulusan yang akurat, mudah dijelaskan, serta relevan dalam konteks deteksi dini mahasiswa berisiko. Meskipun berbagai penelitian sebelumnya telah menggunakan machine learning untuk memprediksi kelulusan mahasiswa, sebagian besar masih terbatas pada model tunggal, cakupan dua kelas prediksi, dan hanya dilakukan pada satu titik waktu. Berdasarkan tinjauan terhadap penelitian sebelumnya, hanya sebagian kecil studi yang mempertimbangkan prediksi multi-kelas, dan belum ditemukan penelitian yang menerapkan pendekatan multi-semester dengan interpretabilitas berbasis SHAP dalam satu kerangka prediksi. Oleh karena itu, terdapat celah penelitian yang jelas terkait kebutuhan model prediksi kelulusan yang bersifat longitudinal, multi-kelas, serta dapat dijelaskan secara transparan, sehingga penelitian ini dilakukan untuk mengisi gap tersebut dan memberikan kontribusi empiris maupun praktis bagi institusi pendidikan dalam pengambilan keputusan berbasis data.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, penelitian ini difokuskan untuk menjawab beberapa pertanyaan utama terkait pengembangan model prediksi kelulusan mahasiswa dengan pendekatan Hybrid Ensemble yang menggabungkan XGBoost, LightGBM, CatBoost, Random Forest dan MLPClassifier yang diintegrasikan dengan metode interpretabilitas SHAP (SHapley Additive Explanations), yaitu:

1. Bagaimana performa model individual (XGBoost, LightGBM, CatBoost, Random Forest dan MLPClassifier) dalam memprediksi status kelulusan mahasiswa berdasarkan data akademik dan biodata internal?
2. Apakah penerapan metode Hybrid Ensemble berbasis *voting ensemble* dapat meningkatkan akurasi dan F1-score dibandingkan model individual?
3. Faktor-faktor (fitur) apa yang paling berpengaruh terhadap prediksi status kelulusan mahasiswa berdasarkan hasil interpretasi model menggunakan SHAP (SHapley Additive Explanations)?

## **1.3 Batasan Masalah**

Agar penelitian ini terfokus dan sesuai dengan tujuan yang ingin dicapai, maka ruang lingkup kajian dibatasi pada hal-hal berikut:

1. Objek penelitian terbatas pada mahasiswa Program Studi Sistem Informasi Universitas Multimedia Nusantara (UMN) angkatan 2020 hingga 2024, dengan data akademik yang diperoleh secara internal dari institusi terkait dan data faktor eksternal yang diperoleh dari penyebaran survei.
2. Variabel yang digunakan dalam penelitian ini dibatasi pada faktor internal mahasiswa, yang meliputi data akademik seperti Indeks Prestasi Kumulatif (IPK) per semester, jumlah mata kuliah yang diambil, serta performa akademik dari waktu ke waktu. Faktor eksternal seperti kondisi sosial ekonomi, dukungan keluarga, atau lingkungan belajar tidak dimasukkan sebagai fitur dalam pemodelan prediktif, namun tetap digunakan pada tahap

analisis deskriptif dan interpretatif (SHAP) untuk memberikan konteks tambahan terhadap kondisi akademik mahasiswa serta mendukung pembahasan dan perumusan rekomendasi kebijakan, tanpa memengaruhi proses pelatihan dan evaluasi model.

3. Kategori target atau kelas prediksi dibatasi pada empat label utama, yaitu Dropout, Lulus Tepat Waktu, Tidak Lulus Tepat Waktu, dan Lulus Lebih Awal, yang merepresentasikan status akhir kelulusan mahasiswa berdasarkan durasi studi.
4. Algoritma yang digunakan dalam penelitian ini dibatasi pada model-model individual XGBoost, LightGBM, CatBoost, Random Forest, dan MLPClassifier, yang kemudian digabungkan menggunakan pendekatan Hybrid Ensemble berbasis Voting Ensemble (hard voting dan soft voting).
5. Evaluasi model dilakukan menggunakan metrik klasifikasi, yaitu akurasi dan F1-score, untuk menilai tingkat performa dan keseimbangan model dalam memprediksi setiap kelas.
6. Ruang lingkup penelitian dibatasi pada skala akademik Universitas Multimedia Nusantara, sehingga hasil penelitian ini belum tentu dapat digeneralisasikan secara penuh ke institusi pendidikan lain yang memiliki karakteristik data dan sistem akademik yang berbeda.
7. Analisis interpretabilitas model prediksi dibatasi pada penerapan metode Explainable Artificial Intelligence (XAI) menggunakan SHapley Additive exPlanations (SHAP), sehingga teknik interpretasi lain di luar SHAP tidak dibahas dalam penelitian ini.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Penelitian ini bertujuan untuk menjawab permasalahan yang telah dirumuskan, dengan fokus utama pada pengembangan model prediksi kelulusan mahasiswa berbasis pendekatan Hybrid Ensemble berbasis voting XGBoost, LightGBM, CatBoost, dan MLPClassifier yang diintegrasikan dengan metode Explainable Artificial Intelligence (XAI) menggunakan

SHapley Additive exPlanations (SHAP). Tujuan khusus dari penelitian ini adalah:

1. Menganalisis performa masing-masing model individual, yaitu XGBoost, LightGBM, CatBoost, Random Forest dan MLPClassifier, dalam memprediksi status kelulusan mahasiswa berdasarkan data akademik dan biodata internal.
2. Mengembangkan dan mengevaluasi model hybrid ensemble berbasis voting yang menggabungkan XGBoost, LightGBM, CatBoost, dan MLPClassifier dengan Logistic Regression sebagai meta-learner, untuk mengetahui sejauh mana pendekatan ini dapat meningkatkan akurasi dan F1-score dibandingkan model individual.
3. Mengidentifikasi dan menjelaskan faktor-faktor internal yang paling berpengaruh terhadap hasil prediksi status kelulusan mahasiswa melalui penerapan metode SHAP (SHapley Additive Explanations) sebagai pendekatan interpretabilitas model.

### **1.4.2 Manfaat Penelitian**

Penelitian ini diharapkan memberikan manfaat yang signifikan baik secara praktis maupun teoritis. Secara praktis, penelitian ini dapat membantu institusi pendidikan, dosen, dan mahasiswa dalam meningkatkan efektivitas proses akademik melalui sistem prediksi kelulusan berbasis pendekatan hybrid ensemble yang akurat dan mudah diinterpretasikan. Sementara itu, secara teoritis, penelitian ini diharapkan dapat memperkaya literatur di bidang *educational data mining* dan *learning analytics*, khususnya terkait penerapan model ensemble dan neural network dalam memprediksi kelulusan mahasiswa serta integrasinya dengan metode Explainable Artificial Intelligence (XAI).

#### **1.4.2.1 Manfaat Praktis**

Manfaat praktis dalam penelitian ini antara lain:

- a) Memberikan alat bantu prediktif bagi institusi pendidikan, khususnya Universitas Multimedia Nusantara, untuk mendeteksi mahasiswa berisiko tidak lulus tepat waktu atau dropout, sehingga

tindakan pembinaan akademik dapat dilakukan secara lebih dini dan tepat sasaran.

- b) Membantu dosen dan pembimbing akademik dalam memahami pola akademik mahasiswa berdasarkan data internal seperti IPK, jumlah mata kuliah, serta performa per semester, dengan dukungan interpretasi model melalui metode SHAP (SHapley Additive Explanations) yang transparan.
- c) Memberikan wawasan kepada mahasiswa mengenai faktor-faktor akademik yang paling berpengaruh terhadap kelulusan, sehingga mereka dapat melakukan evaluasi diri dan menyusun strategi belajar yang lebih efektif.
- d) Menyediakan dasar pertimbangan bagi pengambil kebijakan di perguruan tinggi dalam merancang sistem monitoring akademik dan program intervensi berbasis data yang terukur.

#### **1.4.2.2 Manfaat Teoritis**

Manfaat teoritis dalam penelitian ini antara lain:

- a) Memperkaya kajian akademik mengenai pemanfaatan pendekatan *Hybrid Ensemble* berbasis voting yang menggabungkan berbagai algoritma pembelajaran mesin seperti XGBoost, LightGBM, CatBoost, Random Forest, dan MLPClassifier, serta terintegrasi dengan metode XAI SHAP dalam bidang *educational data mining* dan *learning analytics*.
- b) Memberikan kontribusi pada pengembangan metodologi analisis data pendidikan yang mampu mengintegrasikan data tabular (statis) dengan data sekuensial (dinamis) sekaligus menyajikan interpretasi yang transparan.
- c) Menjadi rujukan bagi penelitian selanjutnya dalam mengeksplorasi pendekatan *Hybrid* untuk memprediksi kinerja akademik maupun bidang lain yang memiliki karakteristik data serupa.

- d) Mendorong perkembangan literatur mengenai integrasi *machine learning* dengan *explainable AI* pada bidang pendidikan, sehingga hasil prediksi dapat lebih dipercaya dan aplikatif.

## 1.5 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran menyeluruh mengenai isi dan alur pembahasan dari keseluruhan penelitian. Tujuan penyusunan sistematika ini adalah untuk mempermudah pembaca dalam memahami tahapan penelitian yang dilakukan secara sistematis, mulai dari perumusan masalah hingga kesimpulan akhir. Skripsi ini terdiri atas lima bab utama dengan uraian sebagai berikut:

### BAB I : PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Seluruh komponen dalam bab ini disusun untuk memberikan gambaran umum penelitian dan menjadi dasar pengembangan model prediksi kelulusan mahasiswa berbasis pendekatan Hybrid Ensemble (Voting Ensemble) XGBoost, LightGBM, CatBoost, Random Forest dan MLPClassifier yang diintegrasikan dengan metode Explainable Artificial Intelligence (XAI) menggunakan SHAP.

### BAB II : LANDASAN TEORI

Bab ini memuat teori-teori, konsep-konsep, dan hasil penelitian terdahulu yang relevan dengan topik penelitian. Pembahasan dalam bab ini mencakup teori mengenai kelulusan mahasiswa dan faktor-faktor internal yang memengaruhinya, konsep dasar machine learning dan algoritma yang digunakan (XGBoost, LightGBM, CatBoost, Random Forest dan MLPClassifier), teori tentang pendekatan ensemble learning, serta konsep Explainable Artificial Intelligence (XAI) dengan fokus pada SHapley Additive Explanations (SHAP). Bab ini berfungsi sebagai dasar teoretis dalam merancang model prediksi yang akurat sekaligus dapat diinterpretasikan.

### **BAB III : METODOLOGI PENELITIAN**

Bab ini menjelaskan metode penelitian yang digunakan untuk mencapai tujuan penelitian. Pembahasannya mencakup jenis dan pendekatan penelitian, sumber data dan teknik pengumpulan data, karakteristik data akademik mahasiswa Sistem Informasi Universitas Multimedia Nusantara angkatan 2020–2024, serta tahapan pengolahan dan pembersihan data. Selanjutnya dijelaskan proses feature engineering, penerapan model individual (XGBoost, LightGBM, CatBoost, dan MLPClassifier), perancangan model hybrid ensemble berbasis voting, integrasi metode SHAP untuk interpretasi model, serta metode evaluasi model.

### **BAB IV : ANALISIS DAN HASIL PENELITIAN**

Bab ini menyajikan hasil analisis data dan kinerja model berdasarkan metodologi yang telah ditetapkan. Pembahasan meliputi deskripsi data akademik mahasiswa, hasil pengolahan dan eksplorasi data, pengujian performa masing-masing model individual, serta pengembangan dan evaluasi model hybrid ensemble berbasis voting. Bab ini juga menampilkan hasil perbandingan performa model individual dan hybrid, interpretasi hasil menggunakan SHAP untuk mengidentifikasi faktor-faktor internal yang paling berpengaruh terhadap kelulusan, serta analisis terhadap kelebihan dan keterbatasan model yang dihasilkan.

### **BAB V : SIMPULAN DAN SARAN**

Bab ini berisi simpulan hasil penelitian yang diperoleh berdasarkan analisis dan pembahasan pada bab sebelumnya. Simpulan disusun untuk menjawab rumusan masalah serta menunjukkan pencapaian tujuan penelitian. Selain itu, bab ini juga memuat saran-saran yang ditujukan untuk penelitian selanjutnya, seperti pengembangan model hybrid ensemble dengan algoritma lain, eksplorasi metode interpretabilitas tambahan di luar SHAP, dan pemanfaatan variabel akademik baru guna meningkatkan akurasi dan generalisasi model prediksi kelulusan mahasiswa.