

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tingginya risiko kegagalan akademik dan tingkat putus kuliah di jenjang pendidikan tinggi menjadi perhatian utama bagi institusi pendidikan. Banyak mahasiswa mengalami kesulitan dalam menyelesaikan studi mereka karena berbagai faktor, yang pada akhirnya berdampak pada rendahnya tingkat kelulusan dan meningkatnya angka putus kuliah. Kondisi ini menuntut adanya upaya yang lebih efektif dalam mendeteksi dan mengatasi potensi kegagalan akademik sejak dini. Teknik *machine learning* dapat digunakan untuk menganalisis berbagai faktor yang mempengaruhi keberhasilan akademik mahasiswa. Faktor-faktor tersebut meliputi riwayat akademik, data demografi, serta kondisi sosial-ekonomi yang dapat memberikan gambaran mengenai kemungkinan seorang mahasiswa mengalami kesulitan dalam menyelesaikan studinya. Dengan menerapkan metode ini, institusi pendidikan dapat memperoleh wawasan yang lebih mendalam mengenai pola dan kecenderungan mahasiswa dalam menjalani perkuliahan. Model prediktif yang dikembangkan melalui *machine learning* dapat membantu dalam mengidentifikasi mahasiswa yang berisiko mengalami kegagalan akademik sejak awal perkuliahan. Dengan mendeteksi mahasiswa yang memiliki kemungkinan tinggi untuk tidak menyelesaikan studi mereka, langkah-langkah preventif dapat segera dilakukan guna meminimalkan angka putus kuliah. Hal ini memungkinkan institusi pendidikan untuk mengambil kebijakan yang lebih tepat sasaran dalam mendukung keberhasilan akademik mahasiswa[1].

Data yang digunakan dalam penelitian ini mencakup berbagai aspek akademik yang relevan, seperti tingkat kehadiran, nilai tugas, ujian tengah semester (UTS), dan ujian akhir semester (UAS). Berdasarkan data ini, mahasiswa dapat diklasifikasikan ke dalam tiga kategori utama, yaitu lulus, masih terdaftar, atau putus kuliah. Klasifikasi ini memberikan gambaran yang lebih jelas mengenai progres akademik mahasiswa dan memungkinkan penyusunan strategi intervensi yang lebih terarah.

Penelitian terdahulu menitikberatkan pada prediksi kelulusan mahasiswa secara tepat waktu sebagai faktor menguntungkan bagi mahasiswa dan perguruan tinggi. Salah satu elemen penting yang memengaruhi kelulusan tepat waktu adalah prestasi mahasiswa dalam setiap mata kuliah, yang secara signifikan memengaruhi rata-rata IPK mereka. Dalam rangka meningkatkan efisiensi penentuan predikat kelulusan, penelitian ini menerapkan metode Klasterisasi menggunakan algoritma *K-Means Clustering*. Dengan tiga kategori kelulusan (nilai bagus,

cukup, dan tidak lulus), hasil perhitungan dataset menggunakan *K-Means* menunjukkan tingkat akurasi sebesar 81%. Meskipun sistem mampu mengklasifikasikan data secara benar, perlu diakui bahwa pengelompokannya belum optimal, dengan beberapa data yang terklasifikasi dengan tidak tepat. Studi ini menggunakan input data, termasuk kehadiran mahasiswa, nilai tugas, UTS, dan UAS, untuk menghasilkan output berupa pengelompokan data berdasarkan algoritma *K-Means*[2]. Pada penelitian lainnya telah mencoba untuk mengklasifikasikan kelulusan mahasiswa pada suatu mata kuliah di Universitas XYZ dengan membandingkan kinerja algoritma SVM dan KNN melalui analisis kuantitatif. Meskipun KNN menunjukkan keunggulan dalam beberapa parameter, SVM terbukti lebih unggul dalam menetapkan nilai ambang kelulusan. Evaluasi menggunakan empat poin, yaitu nilai AUC, akurasi, *error*, dan nilai ambang kelulusan, menunjukkan bahwa KNN lebih efektif dalam mengatasi noise data, sementara SVM memberikan hasil yang lebih konsisten dengan ekspektasi penelitian. Namun, perlu dicatat bahwa penelitian ini memiliki kelemahan terkait dengan penggunaan data training yang mengandung anomali atau *noise*, serta asumsi bahwa ketiga penilaian (tugas, ujian tengah semester, dan ujian akhir semester) sudah terwakili dengan baik, yang dapat memengaruhi hasil perbandingan antara kedua algoritma tersebut[3].

Penggunaan metode LSTM (*Long Short-Term Memory*) dapat membantu mengatasi masalah data training yang memiliki anomali atau *noise*[4][5]. LSTM memiliki kemampuan deteksi anomali dalam data deret waktu dan memodelkan hubungan temporal pada penilaian seperti tugas, ujian tengah semester, dan ujian akhir semester. Hal ini membantu mengidentifikasi pola kompleks dan mengurangi dampak dari data tidak biasa. Meskipun LSTM dapat efektif, pemilihan metode harus disesuaikan dengan karakteristik dataset dan tujuan penelitian. Evaluasi perbandingan dengan SVM dan KNN diperlukan untuk memastikan keefektifan dan keakuratan penerapan metode LSTM dalam konteks penelitian tertentu. Seperti pada penelitian yang dilakukan untuk mengeksplorasi penggunaan *Long Short-Term Memory* (LSTM) dan *Educational Data Mining* (EDM) untuk memprediksi kinerja akademik mahasiswa pada mata kuliah pemrograman di dua universitas di Peru. Dengan mengevaluasi model pada data yang tidak seimbang, hasil

menunjukkan bahwa LSTM dengan GAN memiliki akurasi, *recall*, presisi, dan *F1-Score* tertinggi. Penelitian ini juga membandingkan teknik penyeimbangan data dan membahas potensi penelitian masa depan serta memberikan wawasan tentang penggunaan *machine learning* dalam EDM. Hasil eksperimen menunjukkan bahwa LSTM, khususnya dengan GAN, unggul dengan akurasi tertinggi, mencapai 98,3% pada minggu ke-8 [5].

Dalam upaya mengoptimalkan performa model LSTM, penelitian ini mengimplementasikan teknik *GridSearch* untuk melakukan penyetelan *hyperparameter*. *GridSearch* memberikan peluang eksplorasi menyeluruh terhadap berbagai kombinasi parameter, seperti jumlah unit LSTM, tingkat *dropout*, *learning rate*, dan ukuran *batch*. Melalui pendekatan ini, diperoleh konfigurasi parameter terbaik yang mampu menghasilkan akurasi model yang maksimal. Meskipun metode ini memerlukan waktu komputasi yang lebih besar dibandingkan pendekatan lain seperti *Random Search*, *GridSearch* dianggap lebih sistematis dan dapat diandalkan dalam menentukan pengaturan optimal model [19]. Keefektifan teknik ini juga didukung oleh studi lain, misalnya dalam penerapannya pada model *Random Forest* untuk deteksi malware [20].

Penelitian ini bertujuan untuk menerapkan model LSTM dalam memprediksi kelulusan mahasiswa, dengan harapan dapat memperoleh hasil prediksi yang lebih akurat dibandingkan pendekatan sebelumnya. Selain itu, penelitian ini juga menelaah dampak dari proses penyetelan parameter terhadap kinerja model, guna mengidentifikasi konfigurasi optimal yang mendukung performa terbaik. Dengan demikian, hasil penelitian ini diharapkan dapat berkontribusi dalam peningkatan efisiensi pengelolaan akademik serta memberikan sistem peringatan dini bagi mahasiswa yang berisiko tidak lulus, sehingga program studi dapat merancang strategi intervensi yang lebih tepat sasaran..

1.2 Rumusan Masalah

1.2.1 Permasalahan

Pemanfaatan data akademik secara optimal masih menjadi tantangan karena belum adanya proses analitik yang efektif untuk mendukung intervensi dini terhadap mahasiswa yang berisiko gagal atau putus kuliah. Meskipun beberapa model prediksi telah dikembangkan dalam penelitian sebelumnya,

model-model tersebut masih memiliki keterbatasan dalam hal akurasi dan ketahanan terhadap *noise* pada data, sehingga belum sepenuhnya dapat diandalkan dalam memberikan prediksi yang akurat. Oleh karena itu, diperlukan eksplorasi lebih lanjut terhadap metode *machine learning* yang lebih optimal guna meningkatkan akurasi serta keandalan model prediksi, sehingga institusi pendidikan dapat mengambil langkah preventif yang lebih tepat dalam mendukung keberhasilan akademik mahasiswa.

1.2.2 Metode Usulan

Penelitian ini menggunakan model **Long Short-Term Memory (LSTM)** untuk menganalisis pola data akademik mahasiswa dalam bentuk deret waktu, sehingga dapat menangkap hubungan temporal yang mempengaruhi keberhasilan studi. Untuk mengoptimalkan performa model, diterapkan teknik **GridSearch** guna melakukan *tuning* parameter, termasuk jumlah unit LSTM, *dropout rate*, *learning rate*, dan ukuran *batch*. Selain itu, penelitian ini juga mengevaluasi dan membandingkan model prediksi berdasarkan akurasi serta ketahanannya terhadap *noise*, guna memastikan keandalan model dalam berbagai kondisi data. Data akademik historis digunakan sebagai dasar pembelajaran model, dengan harapan dapat meningkatkan akurasi prediksi terhadap mahasiswa yang berisiko gagal atau putus kuliah, sehingga memungkinkan institusi pendidikan mengambil langkah intervensi yang lebih efektif.

1.2.3 Pertanyaan Penelitian

LSTM memiliki hyperparameter dimana nilai pada setiap parameter berdampak pada hasil akurasi pada prediksi kelulusan dini mahasiswa sehingga penentuan parameter yang tepat akan berpengaruh pada tingkat keberhasilan dalam memprediksi secara dini kelulusan mahasiswa. Sehingga pertanyaan penelitian yang akan diselesaikan, adalah Bagaimana hasil dari metode **Long Short-Term Memory (LSTM)** yang dioptimalkan dengan **GridSearch** dapat meningkatkan akurasi prediksi kelulusan mahasiswa dibandingkan metode sebelumnya, serta bagaimana pengaruh tuning *hyperparameter* terhadap performa model dalam mengidentifikasi faktor risiko kegagalan akademik dan

mendukung strategi intervensi yang lebih efektif bagi institusi pendidikan?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berikut adalah tujuan dalam penelitian ini, yaitu :

1. Mengembangkan model prediksi berbasis **Long Short-Term Memory (LSTM)** yang mampu mengidentifikasi mahasiswa yang berisiko mengalami kegagalan akademik atau putus kuliah. Model ini akan dilatih menggunakan data akademik historis dan dievaluasi berdasarkan akurasi dalam memprediksi kelulusan mahasiswa. Selain itu, penelitian ini juga akan mengeksplorasi efektivitas metode **GridSearch** dalam melakukan *tuning hyperparameter*, seperti jumlah unit LSTM, *dropout rate*, *learning rate*, dan ukuran *batch*, guna mengoptimalkan performa model dalam menangani data dengan berbagai pola dan *noise*.
2. Menganalisis faktor-faktor utama yang paling berpengaruh terhadap kelulusan mahasiswa, baik dari segi akademik, demografi, maupun sosial-ekonomi. Dengan memahami faktor risiko yang signifikan, hasil penelitian ini diharapkan dapat memberikan wawasan bagi institusi pendidikan dalam merancang strategi intervensi akademik yang lebih efektif. Implementasi model prediksi ini dapat membantu institusi dalam mendeteksi mahasiswa yang memerlukan dukungan tambahan, sehingga dapat mengurangi tingkat kegagalan dan putus kuliah melalui program bimbingan, pendampingan, atau kebijakan akademik yang lebih tepat sasaran.

1.3.2 Manfaat Penelitian

Adapun manfaat dari dilakukannya penelitian ini, yaitu :

1. Penelitian ini membantu universitas dalam mengidentifikasi mahasiswa yang berisiko gagal atau putus kuliah lebih awal. Dengan model prediksi berbasis **LSTM**, institusi dapat menyusun strategi intervensi yang lebih efektif, seperti bimbingan akademik, mentoring, atau kebijakan dukungan belajar, guna meningkatkan tingkat kelulusan mahasiswa.
2. Model prediksi ini memberikan peringatan dini kepada mahasiswa

mengenai risiko akademik yang mereka hadapi. Dengan informasi ini, mereka dapat mengambil langkah preventif, seperti mengikuti kelas tambahan atau mendapatkan bimbingan akademik, sehingga meningkatkan peluang untuk menyelesaikan studi dengan sukses.

1.4 Batasan Masalah

Berikut batasan-batasan yang akan diberikan agar permasalahan tidak terlalu luas dan tidak menyimpang sehingga sesuai dengan tujuan penulis berikut ini :

1. Data yang digunakan terbatas pada informasi akademik dan demografi mahasiswa tanpa mempertimbangkan faktor psikologis atau motivasi pribadi.
2. Penelitian ini memusatkan analisis pada metode Long Short-Term Memory (LSTM).
3. Proses tuning parameter hanya dilakukan menggunakan **GridSearch**, tidak mencakup metode lain seperti *Random Search* atau *Bayesian Optimization*.

1.5 Sistematika Penulisan

Adapun sistematika penulisan yang digunakan terdiri dari beberapa bagian, yaitu :

BAB I PENDAHULUAN

Pada bab ini membahas mengenai latar belakang, permasalahan, usulan metode, pertanyaan penelitian, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini memberikan tinjauan literatur mengenai konsep dan landasan teoritis di balik pemecahan masalah pada penelitian ini.

BAB III METODE PENELITIAN

Secara umum bab ini menjelaskan tentang perancangan alur *sentiment analysis*, tahapan penelitian, arsitektur sistem, data penelitian, *pre-processing data*, penjelasan TF-IDF, tahapan klasifikasi data menggunakan metode *Random Forest Classifier*, tahapan seleksi fitur dengan *Information Gain* dan evaluasi model, dan uji coba skenario.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi hasil pengolahan data, pengujian dari hasil proses klasifikasi menggunakan metode *Random Forest* dengan penerapan seleksi fitur *Information Gain* serta penerapan pengaturan *Hyperparameter Tuning* pada *dataset* ulasan kenaikan harga BBM pada platform Twitter.

BAB V PENUTUP

Pada bab ini berisi rangkuman kesimpulan dari hasil penelitian yang telah dilakukan. Selain itu, bab ini juga mencakup saran yang disampaikan untuk keperluan pengembangan penelitian di masa mendatang.

DAFTAR PUSTAKA

Daftar pustaka berisi daftar penelitian terdahulu yang dijadikan rujukan dalam penelitian ini.