

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Pertumbuhan data besar dan informasi telah menghasilkan perubahan penting dalam sektor pendidikan, terutama dalam penggunaan data akademis untuk membantu pengambilan keputusan yang berlandaskan bukti. Penerapan machine learning dalam bidang pendidikan memungkinkan analisis data siswa secara lebih sistematis dan akurat, terutama dalam memprediksi hasil belajar [1], [2]. Dalam konteks Educational Data Mining (EDM), prediksi performa akademik menjadi isu strategis karena berkaitan langsung dengan upaya peningkatan kualitas pembelajaran serta deteksi dini terhadap siswa yang berisiko mengalami kegagalan akademik [1], [3], [12], [13].

Permasalahan yang masih sering terjadi di lingkungan pendidikan adalah proses evaluasi akademik yang bersifat reaktif. Identifikasi siswa dengan performa rendah umumnya dilakukan setelah hasil evaluasi akhir diperoleh, sehingga intervensi pembelajaran sering kali terlambat diberikan. Kondisi ini menunjukkan adanya kebutuhan akan sistem prediksi yang mampu mendeteksi potensi penurunan prestasi sejak dini. Tanpa adanya model prediktif yang akurat dan andal, institusi pendidikan akan kesulitan merancang strategi pembelajaran yang adaptif dan preventif.

Sejumlah studi telah menunjukkan bahwa algoritma pembelajaran mesin dapat meningkatkan ketepatan prediksi jika dibandingkan dengan metode statistik tradisional, khususnya pada data yang memiliki karakteristik multidimensi dan rumit [12], [14]. Dataset Kinerja Siswa yang memuat atribut akademik dan non-akademik seperti nilai, latar belakang demografis, serta faktor pendukung pembelajaran, menyediakan peluang untuk mengembangkan model prediksi yang lebih komprehensif [4], [5], [16], [17]. Namun demikian, pemilihan algoritma yang tepat masih menjadi tantangan utama, karena setiap algoritma memiliki karakteristik, keunggulan, dan keterbatasan yang berbeda [8], [19].

Support Vector Machine dan K-Nearest Neighbor adalah metode pengelompokan yang diterapkan di studi ramalan hasil akademis [6], [7], [18]. SVM dikenal memiliki kemampuan membentuk batas keputusan optimal dan efektif dalam menangani data berdimensi tinggi, sedangkan KNN merupakan metode berbasis kedekatan jarak yang sederhana namun sensitif terhadap skala data dan pemilihan parameter [15], [21]. Meskipun beberapa penelitian komparatif telah dilakukan, sebagian besar belum secara spesifik menguji kedua algoritma

tersebut pada dataset yang sama dengan pendekatan evaluasi yang komprehensif, termasuk analisis akurasi, presisi, recall, F1-score dan juga kurva ROC serta nilai AUC.

Kesenjangan inilah yang menjadi dasar pengangkatan topik penelitian ini. Diperlukan analisis komparatif yang terukur dan sistematis untuk mengetahui algoritma mana yang lebih efektif dalam memprediksi hasil belajar siswa pada Dataset Kinerja Siswa. Selain akurasi, studi ini juga mempertimbangkan keseimbangan antara kinerja model dan kemampuannya untuk melakukan generalisasi pada data baru.

Dengan cara itu, studi ini memiliki kepentingan dan sumbangan nyata dalam menawarkan rekomendasi algoritma yang paling tepat untuk membantu pengembangan sistem peringatan dini akademik. Selain itu, penelitian ini sejalan dengan tuntutan pendidikan di era digital dan Society 5.0, di mana pemanfaatan data dan teknologi menjadi landasan utama dalam peningkatan mutu pembelajaran dan pengambilan kebijakan pendidikan yang lebih adaptif dan berbasis data [11], [19], [22].

## **1.2. Rumusan Masalah**

Berdasarkan konteks penelitian ini, maka rumusan masalah yang didapat adalah :

1. Bagaimana performa metode Support Vector Machine dalam memprediksi hasil belajar siswa dengan Student Performance Dataset?
2. Dengan menggunakan Dataset Kinerja Siswa, seberapa baik pendekatan K-Nearest Neighbor memprediksi hasil belajar siswa ?
3. Algoritma manakah yang lebih efektif dalam konteks klasifikasi hasil belajar siswa?

## **1.3. Batasan Masalah**

Keterbatasan penelitian ini agar tetap fokus adalah sebagai berikut:

1. Student Performance Dataset yang bersumber dari Kaggle berisi data demografis siswa serta nilai ujian (math, reading, dan writing), digunakan sebagai dataset tanpa penambahan data eksternal dan diasumsikan telah melalui pembersihan dasar.
2. Variabel target dibatasi pada hasil belajar siswa yang diklasifikasikan ke dalam kelas tertentu, seperti lulus dan tidak lulus, berdasarkan kriteria nilai ujian pada dataset.
3. Studi ini hanya membandingkan dua metode klasifikasi, Support Vector Machine dan K-Nearest Neighbor, dan tidak mencakup semua algoritma machine learning.
4. Kinerja dari model dievaluasi berdasarkan metrik klasifikasi yang umum, seperti akurasi, presisi, recall, dan F1-score, tanpa memperhitungkan indikator lain sebagai faktor utama.

## **1.4. Tujuan Penelitian**

Dari penelitian ini, terdapat tujuan sebagai berikut:

1. Mengukur performa SVM dalam memprediksi hasil belajar siswa.
2. Mengukur kinerja KNN dalam memprediksi hasil belajar siswa.
3. Menganalisis perbandingan kinerja kedua algoritma prediktif menggunakan metrik evaluasi seperti skor F1, tingkat recall, tingkat presisi, dan tingkat akurasi.

## **1.5. Manfaat Penelitian**

### **1.5.1 Bagi Peneliti**

Researchers and students interested in doing research in the areas of educational data mining and artificial intelligence can use this study as an academic resource. Furthermore, this study might provide a methodological framework for comparing classification algorithms, particularly SVM and KNN, in the context of student learning outcome prediction.

### **1.5.2 Bagi Siswa/Siswi**

Studi ini bertujuan untuk membantu siswa memperoleh manfaat tidak langsung dari penggunaan sistem prediksi hasil belajar yang lebih akurat . Karena sistem ini dapat membantu mengidentifikasi potensi akademik dan tantangan belajar siswa sejak dini , hal ini memungkinkan siswa untuk menerima dukungan dan teknik pembelajaran yang disesuaikan dengan kebutuhan individu mereka . Selain itu , diharapkan penggunaan teknologi prediktif dalam pendidikan akan meningkatkan motivasi belajar siswa dan membuat mereka lebih sadar akan kemajuan akademik mereka.

### **1.5.3 Bagi Tenaga Pendidik (Guru)**

Dengan memanfaatkan alat analisis prediktif yang disediakan oleh penelitian ini, para ahli dapat mengidentifikasi siswa yang berisiko mengalami tingkat prestasi atau motivasi akademik yang rendah. Hasil penelitian ini berpotensi meningkatkan pengetahuan para pendidik mengenai cara teknologi yang berfokus pada data dapat dimanfaatkan untuk memperkirakan dan memantau hasil pembelajaran siswa. Para pendidik dapat mengidentifikasi siswa yang berisiko mengalami penurunan prestasi akademik lebih awal dengan menggunakan model prediksi yang andal , sehingga memungkinkan mereka untuk memberikan intervensi pengajaran yang lebih tepat sasaran dan berhasil.

### **1.5.4 Bagi Instansi Sekolah**

Pengembangan sistem peringatan dini akademik di sekolah dapat didasarkan pada studi ini . Sistem ini dapat digunakan untuk mendukung sekolah dalam membuat pilihan strategis tentang pembinaan akademik siswa , meningkatkan kualitas pengajaran , dan merancang program dukungan bagi siswa yang membutuhkan perhatian ekstra.