

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Perkembangan teknologi informasi dan komunikasi yang pesat telah membawa perubahan signifikan dalam berbagai bidang, termasuk pendidikan. Salah satu inovasi teknologi yang telah memberikan kontribusi besar adalah *machine learning*. *Machine learning* adalah sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia [1]. Oleh karena itu, hal ini telah membuka peluang baru dalam menganalisis dan mengklasifikasi berbagai fenomena, termasuk performa pelajar.

Pendidikan merupakan aspek penting dalam pembangunan sumber daya manusia. Dalam konteks ini, memahami faktor-faktor yang mempengaruhi performa pelajar dan mengklasifikasi hasil belajar mereka menjadi sangat penting. Performa pelajar merupakan tingkat kemampuan seorang pelajar dalam aktivitas akademik dan non-akademik. Performa pelajar dapat mencakup berbagai aspek yang menggambarkan pencapaian akademik, keterampilan, dan perilaku belajar mereka. Adapun hal yang berkaitan dengan performa pelajar, seperti nilai, partisipasi dalam kelas, keterampilan belajar, serta kontribusi dalam kegiatan ekstrakurikuler.

Penelitian ini menggunakan Algoritma *C4.5* dan *Naïve Bayes* dalam mengklasifikasi performa pelajar berdasarkan data yang ada, dengan menggunakan data historis tentang pelajar, seperti nilai IPK/GPA, jumlah kehadiran, jumlah waktu

belajar, ekstrakurikuler, dan faktor-faktor lainnya. Algoritma *C4.5* dan *Naïve Bayes* merupakan dua metode yang sering digunakan dalam proses klasifikasi karena keunggulannya masing-masing. Algoritma *C4.5* dikenal karena kemampuannya membangun pohon keputusan yang mudah dipahami dan memiliki performa yang baik dalam menangani data dengan atribut numerik maupun kategorikal. Sementara itu, *Naïve Bayes* merupakan algoritma berbasis probabilistik yang terkenal karena kesederhanaannya dan efisiensi dalam menangani *dataset* berukuran besar. Selain Algoritma *C4.5* dan *Naïve Bayes*, ada beberapa metode klasifikasi yang dapat digunakan seperti *Support Vector Machine* (SVM) dan *Random Forest*. Dengan demikian, penelitian ini membandingkan Algoritma *C4.5* dan *Naïve Bayes* agar dapat memberikan rekomendasi metode yang paling sesuai untuk mengklasifikasi performa pelajar.

Ada beberapa penelitian yang relevan dengan penulis, seperti pada penelitian yang dilakukan oleh Rinna Rachmatika dan Achmad Bisri [2], peneliti melakukan Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa dengan metode *Decision Tree* menghasilkan akurasi sebesar 77.87% dan *Naïve Bayes* menghasilkan akurasi sebesar 72.79%. Selanjutnya pada penelitian yang dilakukan oleh Abdul Rahman [3], peneliti melakukan Klasifikasi Performa Akademik Siswa menggunakan metode *Decision Tree* menghasilkan akurasi sebesar 83.89% dan *Naïve Bayes* menghasilkan akurasi sebesar 85.97%.

Beberapa studi telah menunjukkan kontribusi yang signifikan pada *machine learning* dalam evaluasi performa pelajar. Sebagai contoh, *machine learning* dapat digunakan untuk :

- a. Personalisasi Pembelajaran: Dengan menganalisis data performa pelajar, algoritma *machine learning* dapat memberikan rekomendasi materi atau strategi pembelajaran yang sesuai dengan kebutuhan individu.
- b. Identifikasi Risiko Akademik: *Machine learning* dapat mendeteksi pelajar yang berisiko mengalami kegagalan akademik lebih awal dengan menganalisis pola aktivitas belajar dan nilai mereka.
- c. Evaluasi Otomatis: Algoritma *machine learning*, seperti *Natural Language Processing* (NLP), dapat digunakan untuk menilai jawaban uraian atau esai secara otomatis, mengurangi waktu penilaian manual.
- d. Peningkatan Efektivitas Pengajaran: Dengan memberikan wawasan tentang tren dan pola pembelajaran siswa, *machine learning* membantu pendidik merancang strategi pengajaran yang lebih efektif.

Berdasarkan permasalahan diatas serta beberapa penelitian yang relevan yang dapat menguatkan penelitian ini, maka penulis tertarik untuk melakukan penelitian yang berjudul **“Perbandingan Algoritma C4.5 Dan Naive Bayes Dalam Machine Learning Untuk Klasifikasi Performa Pelajar”**. Dengan demikian, penerapan *machine learning* menggunakan Algoritma C4.5 dan Naive Bayes diharapkan tidak hanya mampu mengklasifikasi performa pelajar dengan akurasi yang baik, tetapi juga memberikan *insight* yang berharga bagi pendidik dan pengelola sekolah maupun kampus untuk meningkatkan kualitas pendidikan secara keseluruhan.

1.2 PERUMUSAN MASALAH

Dari permasalahan pada latar belakang yang telah dipaparkan penulis diatas, maka didapat rumusan masalah, yaitu bagaimana mengklasifikasi performa pelajar menggunakan metode Algoritma *C4.5* dan *Naïve Bayes*?

1.3 BATASAN MASALAH

Pembatasan masalah pada penelitian yang penulis lakukan bertujuan agar ruang lingkup penelitian menjadi lebih terarah. Adapun batasan masalah dalam penelitian ini, yaitu :

1. Data yang diperlukan adalah *dataset* yang bernama “Student Performance Data” yang berjumlah 2.392 data. (<https://www.kaggle.com/datasets/rabieelkharoua/students-performance-dataset>).
2. Variabel yang digunakan dalam penelitian ini adalah faktor-faktor yang memberikan informasi terkait upaya atau proses seorang pelajar.
3. Alat bantu penelitian yang akan digunakan yaitu *Python*.

1.4 TUJUAN PENELITIAN

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Untuk mengklasifikasi data performa pelajar.
2. Untuk mendapatkan perbandingan akurasi terbaik dari dua metode dalam mengklasifikasi performa pelajar.
3. Untuk mendapatkan atribut yang berpengaruh sebagai faktor utama dalam mengklasifikasi performa pelajar.

4. Untuk menghasilkan pohon keputusan yang mempermudah dalam mengklasifikasi performa pelajar.

1.5 MANFAAT PENELITIAN

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Memberikan kontribusi keilmuan pada penelitian bidang *machine learning* khususnya untuk klasifikasi performa pelajar.
2. Mengetahui seberapa besar tingkat perbandingan akurasi Algoritma *C4.5* dan *Naïve Bayes* dalam klasifikasi performa pelajar.
3. Mengetahui faktor utama pada performa pelajar berdasarkan pohon keputusan yang dihasilkan dari Algoritma *C4.5*.
4. Menambah wawasan bagi penulis dalam mengklasifikasi data performa pelajar.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan ini menggambarkan secara umum mengenai apa yang akan penulis bahas dalam setiap bab dari laporan ini. Adapun sistematika penulisan tesis ini terdiri dari beberapa bab, yaitu :

BAB I : PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI DAN TINJAUAN PUSTAKA

Bab ini membahas teori dan pendapat ahli yang terkait dengan masalah yang dianalisis. Teori yang digunakan terkait *machine learning*, *supervised learning*, Algoritma *C4.5*, Algoritma *Naïve Bayes*, klasifikasi, *python*, dan performa pelajar.

BAB III : METODE PENELITIAN

Bab ini menjelaskan alur penelitian, bahan penelitian, serta alat bantu yang digunakan untuk penelitian ini.

BAB IV : HASIL PENELITIAN DAN PEMBAHASAN

Bab ini menguraikan proses *machine learning* menggunakan dua metode yaitu Algoritma *C4.5* dan *Naïve Bayes* terhadap data-data performa pelajar. Lalu pada bab ini akan ditampilkan hasil dari analisis data dan bentuk visualisasi analisis dari *python*.

BAB V : PENUTUP

Bab ini berisi kesimpulan dan saran dari hasil penelitian yang telah dilakukan, tentunya sebagai wawasan terutama bagi pendidik dan pengelola sekolah maupun kampus.