

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berdasarkan hasil observasi yang dilakukan di SMA Negeri 1 Singaparna didapatkan informasi bahwa kelulusan peserta didik dipegang sepenuhnya oleh satuan pendidikan terkait. Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 4 Tahun 2018 Tentang Penilaian Hasil Belajar Oleh Satuan Pendidikan Dan Penilaian Hasil Belajar Oleh Pemerintah pada Bab VI Kelulusan Peserta Didik Dari Satuan Pendidikan Pasal 19 ayat (1) Peserta didik dinyatakan lulus dari satuan/program pendidikan setelah: a. menyelesaikan seluruh program pembelajaran; b. memperoleh nilai sikap/perilaku minimal baik; dan c. lulus ujian satuan/program pendidikan (Ujian Sekolah). (2) Kelulusan peserta didik sebagaimana dimaksud pada ayat (1) ditetapkan oleh satuan/program pendidikan yang bersangkutan. Berdasarkan peraturan tersebut, pemerintah tidak menentukan standar khusus atau standar minimal mengenai syarat kelulusan, sehingga standar khusus kelulusan siswa ditentukan oleh sekolah. Kewenangan yang diberikan kepada sekolah berpotensi menimbulkan penilaian subjektif terhadap kelulusan siswa yang berpotensi tidak kompetennya siswa yang lulus dari sekolah. Dengan demikian diperlukan perhitungan yang sistematis secara objektif tanpa peranan manusia pada proses klasifikasi kelulusan siswa.

Masalah tersebut dapat diatasi dengan perhitungan yang sistematis untuk klasifikasi kelulusan siswa. Dalam proses ini, ada banyak algoritma yang dapat

diterapkan. Klasifikasi kelulusan pada penelitian ini menggunakan algoritma *C4.5* dan *Naïve Bayes*. Algoritma *C4.5* adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa pola aturan-aturan dapat digunakan untuk memprediksi nilai yang baru. Algoritma *C4.5* sendiri merupakan pengembangan dari algoritma ID3, dimana pengembangan dilakukan dalam beberapa hal diantaranya dapat mengatasi *missing data*, data kontinu dan *pruning*. Algoritma *Naïve Bayes* dipilih karena metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari & Parveen, 2012)

Model *Machine Learning* yang telah dibuat akan dievaluasi menggunakan perhitungan *Confussion Matrix* dan nilai *Area Under The Curve* (AUC) yang dihasilkan dari kurva *Receiver Operating Characteristic* (ROC). Hal ini digunakan untuk melihat kinerja hasil dari masing-masing algoritma dan sejauh mana perbedaan akurasi yang dihasilkan algoritma *C4.5* dan *Naïve Bayes* dalam memprediksi kelulusan siswa.

Berdasarkan latar belakang yang diuraikan, penelitian ini akan membandingkan kinerja dari dua algoritma *machine learning* yaitu algoritma *C4.5* dan *Naïve Bayes*. Pada penelitian sebelumnya perbandingan algoritma antara *C4.5* dan *Naïve Bayes* pemodelan dilakukan dengan lebih sedikit data dan sedikit atribut yang digunakan dimana hal tersebut dapat berpengaruh pada akurasi yang

dihasilkan setiap algoritmanya. Oleh karena itu, pada penelitian ini akan melakukan perbandingan dari kedua algoritma tersebut dengan lebih banyak data dan atribut yang digunakan untuk memperoleh hasil algoritma yang paling maksimal dalam menentukan kelulusan siswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan penelitian ini adalah:

1. Bagaimana mengimplementasikan algoritma *C4.5* dan *Naïve Bayes* pada kelulusan siswa Sekolah Menengah Atas?
2. Bagaimana perbandingan performa yang dihasilkan dari algoritma *C4.5* dan *Naïve Bayes* untuk memprediksi kelulusan siswa Sekolah Menengah Atas menggunakan 45 atribut?

1.3 Tujuan Penelitian

Penelitian ini mempunyai tujuan sebagai berikut :

1. Mengimplementasikan algoritma *C4.5* dan *Naïve Bayes* pada kelulusan siswa Sekolah Menengah Atas.
2. Membandingkan performa *accuracy*, *precision* dan *recall* dari algoritma *C4.5* dan *Naïve Bayes* untuk memprediksi kelulusan siswa Sekolah Menengah Atas menggunakan 45 atribut.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat bagi seluruh pihak yang terkait, diantaranya :

1. Model *data mining* pada kelulusan siswa dapat membantu pengambilan keputusan pihak sekolah dalam kelulusan siswa.
2. Membuat model *Data Mining* dengan algoritma *C4.5* dan *Naïve Bayes* untuk memprediksi kelulusan siswa.
3. Dapat dijadikan sebagai referensi penelitian selanjutnya mengenai algoritma *C4.5* dan *Naïve Bayes* dalam bidang pendidikan.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut:

1. Data yang digunakan dalam penelitian ini menggunakan data nilai Rapor semester 1 sampai 6 periode tahun ajaran 2019/2020 sampai 2021/2022 dari SMA Negeri 1 Singaparna.
2. Pembuatan model *data mining* menggunakan aplikasi RapidMiner.
3. Menggunakan 45 atribut untuk membuat model dalam memprediksi kelulusan siswa.
4. Menggunakan *ROC Curve* dan *Confussion Matrix* dalam evaluasi model prediksi kelulusan siswa.

1.6 Metodologi Penelitian

Metodologi penelitian yang digunakan untuk menyelesaikan penelitian ini diantaranya:

1. *Business Understanding*, Pemahaman bisnis merupakan awal dari proses penelitian ini yang membutuhkan pengetahuan untuk memperoleh data yang mendukung proses klasifikasi kelulusan siswa. Pemahaman bisnis berkaitan dengan pengumpulan informasi serta pengetahuan mengenai studi kasus serta sumber yang berhubungan dengan penelitian ini yang didapatkan dari jurnal dan *e-proceeding* terkait metode dan algoritma yang digunakan.
2. *Data Understanding*, Pemahaman data yang menjadi objek penelitian adalah laporan nilai (Rapot) siswa. Sumber data yang digunakan dalam penelitian ini adalah data Rapor siswa kelas 12 semester 1 sampai semester 6 mulai tahun ajaran 2019/2020 sampai 2021/2022.
3. *Data Preparation*, Persiapan data dilakukan dalam penyusunan dataset yang akan digunakan sebagai input (*Data Training* dan *Data Testing*). Tahapan dalam persiapan data diantaranya *Data Cleaning*, *Integration Data*, *Data Selection*, *Data Transformation*.
4. *Modeling*, Pemodelan yang dipilih adalah dalam penelitian ini menggunakan teknik klasifikasi, menggunakan algoritma *C4.5* dan *Naïve Bayes*. *Tools* yang digunakan dalam pemodelan ini adalah *RapidMiner*.
5. *Evaluation*, Menguji model yang telah dibuat dan mengevaluasi akurasi model dan skenario terbaik berdasarkan hasil perhitungan *Confusion Matrix* dan nilai *AUC* pada *ROC Curve*.

6. *Deployment*, Hasil penelitian ini adalah analisa perbandingan dua algoritma klasifikasi dalam prediksi kelulusan siswa dengan menerapkan model *C4.5* dan *Naïve Bayes*. Mengetahui informasi tingkat kelulusan dapat memudahkan staf dalam mengelola siswa serta memberikan informasi kepada siswa supaya dapat lulus.

1.7 Sistematika Penulisan

Sistematika dalam penulisan ini disusun untuk mempermudah pembacaan dan gambaran umum tentang penelitian yang akan dilakukan dengan sistematika penulisan penelitian ini sebagai berikut :

BAB I PENDAHULUAN

Bab pendahuluan membahas tentang latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian yang ingin dicapai, manfaat penelitian, metodologi penelitian yang merupakan langkah untuk menyelesaikan masalah, dan sistematika penulisan laporan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini membahas berbagai konsep dasar teori-teori yang berkaitan dengan penelitian seperti teori mengenai metode dan algoritma yang terkait dengan penelitian ini. Bab ini juga berisi penjelasan dari penelitian sebelumnya yang relevan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan metode serta algoritma yang akan digunakan dalam proses data mining.

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan hasil dari proses yang telah dilakukan dari penelitian ini. Lebih tepatnya adalah membahas penerapan dari algoritma *C4.5* dan *Naïve Bayes* yang diterapkan pada *tools* Rapidminer.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir yang membahas tentang penarikan kesimpulan dari hasil penelitian yang telah dilakukan, keterkaitan masalah dan tujuan penelitian yang dilakukan. Kesimpulan adalah hasil akhir dari penelitian yang dilakukan, sedangkan Saran berisi tentang rekomendasi sesuai dengan keterbatasan yang ada pada sistem. Saran yang bersifat membangun untuk penelitian selanjutnya tentang penelitian ini supaya dapat menutupi kekurangan-kekurangan yang ada.