

BAB I

PENDAHULUAN

I.1. Latar Belakang Masalah

Berdasarkan data dari Badan Pusat Statistik, jumlah sekolah di Indonesia terus mengalami peningkatan setiap tahun. Ini termasuk peningkatan signifikan pada jumlah Sekolah Menengah Kejuruan (SMK). Sumatera utara memiliki 270 SMK Negeri dan 703 SMK Swasta dengan jumlah siswa 137 972 untuk SMK Negeri dan 170 921 untuk SMK Swasta (Statistik, 2023). Dari data BPS diatas maka dapat disimpulkan bahwa Peningkatan ini mencerminkan upaya pemerintah dan masyarakat dalam menyediakan akses pendidikan yang lebih luas dan beragam bagi siswa di seluruh negeri. Data tersebut juga mengindikasikan persaingan antar SMK (Sekolah Menengah Kejuruan) juga semakin tinggi dan ketat yang menuntut masing-masing SMK (Sekolah Menengah Kejuruan) harus mempertahankan kelangsungan hidup dengan bersaing mendapatkan jumlah mahasiswa yang banyak. Salah indikator dari perkembangan sebuah institusi pendidikan swasta adalah dengan banyaknya jumlah siswa pada sekolah tersebut. Tentu saja dalam peningkatan jumlah siswa tidak terlepas dari tingkat kepuasan siswa pada sekolah tersebut. Selain fasilitas dan guru maka faktor yang paling menunjang berkualitasnya sekolah adalah pelayanan administrasi.

Di era informasi saat ini, pelayanan dihadapkan pada tuntutan kepentingan berbagai pihak, seperti: pimpinan, pemegang saham (yayasan), guru dalam dan siswa di SMK (Sekolah Menengah Kejuruan) tersebut. Untuk kepentingan ini

diperlukan adanya manajemen yang baik dan efektif sehingga “sasaran” penerima layanan akan tingkat kepuasan siswa merupakan suatu keberhasilan bagi sekolah. Oleh karena itu dalam memprediksi tingkat kepuasan siswa diperlukan hal yang lebih spesifik untuk memusatkan perhatian pada tingkat kepuasan siswa terhadap pelayanan administrasi.

Data Mining merupakan salah satu disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar. Proses dalam melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data disebut dengan data mining (Sinaga et al., 2022). *Huge of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data*. Peranan utama data mining yaitu: Estimasi, Prediksi, Klasifikasi, Klustering, dan Asosiasi. Salah satu penerapan data mining adalah prediksi. Prediksi adalah suatu proses di mana algoritma mencoba untuk memahami pola atau tren yang muncul secara berulang dalam data. Tujuannya adalah untuk mengklasifikasikan atau membagi data menjadi kategori atau kelompok berdasarkan perilaku atau nilai yang diharapkan di masa depan. Dengan kata lain, algoritma akan memahami pola-pola yang telah terjadi dalam data dan menggunakannya untuk meramalkan apa yang mungkin terjadi selanjutnya (Alita et al., 2021).

Naive Bayes adalah algoritma klasifikasi yang menggunakan probabilitas untuk memprediksi kelas dari data. Ia menghitung probabilitas kelas berdasarkan atribut-atributnya, dan memilih kelas dengan probabilitas tertinggi sebagai prediksi.

Algoritma ini bekerja dengan asumsi bahwa semua atribut independen, meskipun ini sering tidak akurat dalam dunia nyata. Meskipun demikian, *Naive Bayes* sering memberikan hasil yang baik dan digunakan dalam banyak bidang seperti analisis teks, klasifikasi email, dan lainnya (Narayan, 2020). Penelitian dengan judul Deteksi Sarung Samarinda Menggunakan Metode *Naive Bayes* Berbasis Pengolahan Citra (Septiarini et al., 2021) diperoleh tingkat akurasi sebesar 98.4%. Pada penelitian lain dengan memanfaatkan algoritma *Naive Bayes* yaitu Perbandingan Optimasi *Feature Selection* pada *Naive Bayes* untuk Klasifikasi Kepuasan Airline Passenger (Yoga Religia & Amali, 2021) diperoleh tingkat akurasi sebesar 86.13%.

Algoritma C4.5 adalah cara membuat pohon keputusan untuk mengklasifikasikan data. Algoritma tersebut memilih atribut yang paling berguna untuk memisahkan data menjadi kelompok berdasarkan gain informasi. Proses ini berulang hingga terbentuk pohon lengkap atau aturan-aturan keputusan. Algoritma C4.5 mampu menangani data yang beragam dan memberikan wawasan dari data tersebut (Fersellia et al., 2023). Pada penelitian (Hairani et al., 2020) menjelaskan bahwa algoritma C4.5 mampu memprediksi lebih baik dari data test dibandingkan dengan data sampling. Beberapa penelitian tentang algoritma C4.5 antara lain analisis metode C4.5 untuk mengetahui faktor kepuasan mahasiswa terhadap pembelajaran daring didapatkan hasil bahwa kesimpulan algoritma C4.5 telah berhasil diterapkan untuk menganalisa faktor kepuasan mahasiswa terhadap pembelajaran daring, adapun hasil yang diperoleh dari data kuesioner dan penerapan algoritma C4.5 dapat diperoleh informasi bahwa aspek terpenting untuk

memperoleh kepuasan pembelajaran daring adalah fasilitas interaksi mahasiswa dosen dan mahasiswa dengan mahasiswa, aspek terpenting selanjutnya adalah pembelajaran daring yang dapat memfasilitas kelengkapan objek pembelajaran, adapun akurasi prediksi kepuasan pembelajaran daring yang didapatkan dari data uji diperoleh tingkat akurasi sebesar 75%.

Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun, adalah salah satu lembaga pendidikan yang saat ini sedang berkembang yang ada di Kabupaten Labuhanbatu. SMK Swasta Siti Banun memiliki kualifikasi tersendiri dalam memilih pengajar yang akan menjadi guru di sekolah. Untuk mendukung SMK (Sekolah Menengah Kejuruan) unggul yang diminati oleh masyarakat, SMK Swasta Siti Banun juga tetap berbenah di sisi administrasi. Kerap sekali pelayanan administrasi terabaikan karena tidak dianggap bagian dari kualitas, sementara layanan administrasi merupakan ujung tombak dalam mengontrol dan menjalankan sistem. Manajemen yang baik memiliki administrasi yang tertata rapi, tertib, dan dilakukan secara konsisten. Dengan berkualitasnya layanan administrasi secara tidak langsung menjadi lokomotif perubahan di SMK Swasta Siti Banun dalam mencapai misi nya.

Dari hasil penelitian sebelumnya, terdapat beberapa kesimpulan yang menyoroti kelebihan dari penerapan algoritma *Naïve Bayes* dalam konteks analisis data. Metode ini telah ditemukan memiliki sejumlah keunggulan yang berkontribusi pada efektivitasnya dalam berbagai situasi yaitu *Naïve Bayes* terbukti efektif dalam menghadapi data yang berskala besar. Dalam penanganan dataset yang luas, algoritma ini tetap mampu menghasilkan hasil yang akurat karena penghitungan

probabilitasnya yang efisien. Hal ini menjadikannya pilihan yang cocok untuk situasi di mana volume data menjadi tantangan. Selanjutnya, *Naïve Bayes* memiliki kemampuan unik untuk mengenali pola dalam data. Khususnya dalam analisis teks, seperti dalam analisis sentimen di media sosial, *Naïve Bayes* dapat secara efektif mengidentifikasi pola yang mendalam, membantu dalam pemahaman konteks yang lebih luas (Muafa, 2022). Sedangkan dari penelitian mengenai algoritma C4.5 dapat disimpulkan bahwa algoritma C4.5 baik digunakan dalam proses menggali data (*data mining proses*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (Sari Oktapia Ningse et al., 2022).

Adapun yang menjadi hasil atau kontribusi dari penelitian ini adalah dapat membantu pihak sekolah atau manajemen untuk mengevaluasi terhadap pelayanan administrasi, sehingga dapat menjadi pertimbangan bagi pihak sekolah atau manajemen untuk meningkatkan pelayanan kepada para siswa serta diharapkan dapat memberikan dampak positif bagi layanan administrasi dan memaksimalkan informasi dan meminimalisir penilaian secara *subjektif* serta mengetahui perbandingan algoritma mana antara *Naive Bayes* dan C4.5 yang memiliki kinerja dan nilai akurasi lebih baik. Berdasarkan permasalahan tersebut maka penulis mengusulkan sebuah penulisan penelitian yang berjudul **“Analisis Perbandingan Algoritma Klasifikasi Naïve Bayes Dan C4.5 Untuk Kepuasam Siswa Terhadap Pelayanan Administrasi (Studi Kasus SMK Swasta Siti Banun Sigambal)”**.

I.2. Rumusan Masalah

Perumusan masalah yang dihasilkan adalah :

1. Menentukan klasifikasi prediksi kepuasan siswa SMK Swasta Siti Banun terhadap pelayanan administrasi.
2. Perbandingan antara dua metode, yaitu metode algoritma klasifikasi *Naïve Bayes* dengan C4.5, sehingga menghasilkan prediksi kepuasan siswa terhadap pelayanan administrasi melalui kuesioner yang diberikan kepada siswa kelas XII (Duabelas) di empat jurusan yaitu Rekayasa Perangkat Lunak, Teknik Komputer dan Jaringan, Otomatisasi Tata Kelola Perkantoran dan Akuntansi, yang keseluruhan nya berjumlah 126 siswa.
3. Bagaimana mengukur kinerja dan akurasi yang dihasilkan dari metode algoritma klasifikasi *Naïve Bayes* dan Algoritma C4.5.

I.3. Tujuan Penelitian

Penelitian ini bertujuan untuk memecahkan permasalahan yang telah dirumuskan di atas yaitu :

1. Melakukan analisis model tingkat kepuasan siswa terhadap pelayanan administrasi yang dilihat dari hasil kuesioner 126 siswa kelas XII (Duabelas) di empat jurusan yaitu Rekayasa Perangkat Lunak, Teknik Komputer dan Jaringan, Otomatisasi Tata Kelola Perkantoran dan Akuntansi dan data dari kuesioner di uji dengan menggunakan metode algoritma *Naïve Bayes* dan algoritma C4.5.
2. Penelitian ini akan membandingkan kinerja *Naïve Bayes* dengan algoritma C4.5

3. Untuk mengetahui ketepatan tingkat akurasi yang dihasilkan dari metode algoritma klasifikasi *Naïve Bayes* dan algoritma C4.5.

I.4. Manfaat Penelitian

Beberapa manfaat yang diharapkan pada pembuatan tesis ini adalah:

1. Membantu mendapatkan informasi tentang tingkat kepuasan siswa yang nantinya dapat mempermudah proses evaluasi pelayanan administrasi di SMK Swasta Siti Banun Sigambal.
2. Dapat memberikan informasi dan masukan kepada Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal sebagai dasar untuk meningkatkan kinerja pelayanan administrasi.
3. Memperkaya literatur dan referensi tentang algoritma klasifikasi *Naïve Bayes* dan algoritma C4.5.
4. Memberikan kontribusi bagi ilmu pengetahuan khususnya pada bidang ilmu komputer mengenai algoritma *Naïve Bayes* dan algoritma C4.5, serta memberikan kontribusi bagi Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun Sigambal dalam meningkatkan kinerja pelayanan administrasi, sehingga nantinya berdampak positif terhadap layanan yang diberikan kepada siswa.

I.5. Batasan Masalah

Agar tidak menyimpang dari pokok pembahasan, penulis membuat batasan masalah yang akan dibahas antara lain :

1. Dalam penelitian ini membahas perbandingan metode algoritma klasifikasi *Naïve Bayes* dan algoritma C4.5.

2. Data yang akan dilakukan penelitian data administrator atau pelayanan administrasi di Sekolah Menengah Kejuruan (SMK) Swasta Siti Banun.
3. Hanya diperuntukan untuk siswa kelas XII (Duabelas) SMK Swasta Siti Banun di empat jurusan yaitu Rekayasa Perangkat Lunak, Teknik Komputer dan Jaringan, Otomatisasi Tata Kelola Perkantoran dan Akuntansi, yang berjumlah 126 siswa.
4. Untuk mendukung analisis data dalam penelitian ini akan menggunakan bantuan perangkat lunak *data mining* yang berbasis *open source* seperti bahasa pemrograman *Python* dengan *Library Sklearn*.

I.6. Sistematika Penulisan

Tesis ini mencoba untuk membahas pokok permasalahan secara cermat dan sistematis. Untuk itu pembahasan dibuat dan disusun dengan sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang permasalahan, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bagian tinjauan pustaka berisi kajian pustaka yang meninjau atau mengkaji sumber acuan primer yang relevan dengan mengutamakan hasil penelitian pada jurnal-jurnal ilmiah yang telah dipublikasi terdahulu. Tinjauan pustaka berfungsi untuk memperdalam permasalahan pada jurnal yang diteliti.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang metode penelitian yang dilakukan untuk penulisan tesis ini.

BAB IV HASIL

Hasil penelitian yaitu berupa pemaparan hasil pelaksanaan metode/teknik penelitian dan menyajikan data yang mendukung hasil tersebut. Bab ini membahas tentang hasil perbandingan algoritma C45 dan Naïve Bayes pada penilaian kepuasan siswa terhadap pelayanan administrasi. Penyajian data dan penjelasannya dilakukan secara berurutan dan lengkap menggunakan teks dan ilustrasi lainnya (misalnya, table dan gambar) sebagai pendukung hasil.

BAB V PEMBAHASAN

Pembahasan penelitian yaitu menerjemahkan makna dari hasil yang diperoleh untuk menjawab pertanyaan atau masalah penelitian. Fungsi lainnya adalah untuk menjelaskan pemahaman baru yang didapatkan dari hasil penelitian, yang diharapkan berguna dalam pengembangan keilmuan. Dalam penelitian tingkat lanjut, hasil pembahasan diharapkan menunjukkan kontribusi penulis terhadap pengembangan keilmuan. Dalam penelitian ini, hasil pembahasan bukan pengembangan keilmuan secara substansial, tetapi cukup terhadap pemahaman personal dalam implementasi konsep atau teori.

BAB VI KESIMPULAN DAN SARAN

Kesimpulan dan saran berisi hasil akurasi menggunakan metode algoritma klasifikasi Naïve Bayes dan C4.5 pengolahan data pengujian terhadap model dengan menggunakan bahasa pemrograman *Python* dengan *Library Sklearn*. Bagian kesimpulan berisi jawaban atas rumusan masalah pada penelitian dan pada bagian saran berupa pengembangan penelitian dan kendala yang dihadapi dalam menyusun penelitian sehingga dapat memberikan masukan bagi penelitian selanjutnya