

# BAB I

## PENDAHULUAN

### 1.1 LATAR BELAKANG MASALAH

Ketersediaan data yang berlimpah yang dihasilkan dari penggunaan teknologi informasi di hampir semua bidang kehidupan menimbulkan kebutuhan untuk dapat memanfaatkan informasi dan pengetahuan yang terkandung di dalam limbah data tersebut, yang kemudian melahirkan *data mining*. Klasifikasi *data mining* merupakan salah satu cara penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi biasanya digunakan untuk memprediksi kelas pada suatu label tertentu, yaitu dengan mengklasifikasi data (membangun model) berdasarkan training set dan nilai-nilai (label kelas) dalam mengklasifikasikan atribut tertentu dan menggunakannya dalam mengklasifikasi data yang baru [1].

Untuk dapat membandingkan serta memetakan mutu dari setiap satuan pendidikan, perlu dilakukan akreditasi bagi setiap lembaga dan program pendidikan. Proses ini dilakukan secara berkala dan terbuka dengan tujuan membantu dan memberdayakan satuan pendidikan agar mampu mengembangkan sumber daya dalam mencapai tujuan pendidikan nasional [2].

Ketepatan dalam pengklasifikasian objek sangat penting, metode klasifikasi yang baik adalah metode yang menghasilkan kesalahan kecil. Ada beberapa metode pengklasifikasian salah satunya adalah *naive bayes*. Keuntungan penggunaan *naive bayes* adalah bahwa metode ini hanya membutuhkan jumlah data yang kecil untuk

menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasinya.

Setelah melakukan penelitian tersebut maka data akan dimasukkan ke file yang berbentuk kertas yang akan diinput oleh petugasnya. Data dengan skala besar tersebut butuh waktu lama dan panjang dalam melakukan pengelompokan data. Data yang sudah ada akan dikelompokkan kembali dan juga akan dihitung total pengelompokannya. Maka dari itu penerapan *data mining* ini dapat membantu instansi dalam melakukan pengelompokan dengan rumusan dan juga penerapannya.

*Data mining* adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau mencari dari data yang terdapat pada basis data [3].

Setiap algoritma memiliki bisa menghasilkan klasifikasi yang berbeda. Algoritma terbaik dapat ditentukan dari data yang akan diklasifikasi. Terdapat beberapa metode dan teknik yang digunakan dalam *data mining* diantaranya C4.5, Naïve Bayes, SVM, dan KNN [4].

Algoritma C4.5 termasuk ke dalam pohon keputusan. Struktur sebuah pohon keputusan seperti pada *flowchart*, dimana setiap simpul internal (simpul bukan daun) melakukan pengujian pada atribut, masing-masing cabang merupakan sekumpulan hasil, dan masing-masing simpul daun (atau simpul terminal) menjadi label kelas [4].

Pada SVM dapat mengklasifikasikan data linier dan non linier. Data input merupakan nilai variable variabel prediktor sedangkan *output* merupakan variabel target dimana saling bergantung. Metrik untuk konsep fungsi klasifikasi "terbaik" dapat diwujudkan secara geometris [4].

Klasifikasi Bayes merupakan klasifikasi secara statistik, klasifikasi ini dapat memprediksi peluang keanggotaan kelas seperti probabilitas suatu tupel merupakan milik kelas tertentu. Pada Naïve Bayes berasumsi bahwa efek dari nilai atribut pada kelas tertentu independen dari nilai-nilai dari atribut lainnya [4].

Algoritma k Nearest Neighbor ini merupakan pendekatan mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama. Disebut juga dengan pembelajar yang malas (*lazy learners*) karena hanya melihat kedekatan dengan tetangga (*neihbor*) [4].

Metode *naïve bayes* berpotensi baik dalam mengklasifikasi dokumen data yang peneliti ambil dibandingkan dengan metode yang lain dalam hal akurasi dan efisiensi komputasi. Hal inilah yang melatar belakangi penulis untuk melakukan penelitian guna memberi solusi terhadap masalah yang terjadi dengan mengangkat judul **“KLASIFIKASI AKREDITASI SEKOLAH DI PROVINSI JAMBI MENGGUNAKAN METODE NAÏVE BAYES DI BPMP PROVINSI JAMBI”**.

## **1.2 RUMUSAN MASALAH**

Berdasarkan latar belakang yang sudah dijelaskan, maka dapat diambil suatu rumusan masalah yaitu sebagai berikut “

1. Bagaimana penerapan Naïve Bayes dalam klasifikasi jumlah akreditasi sekolah di Kota Jambi?”
2. Bagaimana mengevaluasi hasil dari hasil klasifikasi akreditasi sekolah yang ada di Kota Jambi?

### **1.3 BATASAN MASALAH**

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi :

1. Data akreditasi BPMP yang menjadi objek utama dari penelitian ini pada periode Januari – September 2022.
2. Metode yang digunakan dalam penelitian ini adalah *Algoritma Naïve Bayes* dengan klasifikasi.
3. Atribut yang terdapat pada data tersebut adalah Standar Pengelolaan, Standar Isi, Standar Penilaian, Standar Pembiayaan, Standar Proses, Standar Sarpras, Standar PTK, Standar Kompetensi, Lulusan, Status, dan Bentuk Pendidikan.
4. Kategori kelas akreditasi adalah A, B, dan C.
5. Alat bantu Analisa menggunakan *WEKA*.

## **1.4 TUJUAN DAN MANFAAT PENELITIAN**

### **1.4.1 Tujuan Penelitian**

Berdasarkan permasalahan yang tertera diatas, maka dapat disimpulkan bahwa tujuan dari penelitian ini adalah :

1. Untuk menerapkan metode Naïve Bayes dalam klasifikasi akreditasi sekolah di Kota Jambi.
2. Untuk mengevaluasi keakurasian klasifikasi menggunakan metode *Naïve Bayes*.
3. Untuk memberikan informasi secara lebih rinci dan lebih terukur tentang jumlah data akreditasi sekolah di Kota Jambi

### **1.4.2 Manfaat Penelitian**

Serta manfaat yang akan didapat dari penelitian, yaitu:

1. Diharapkan penelitian ini dapat memberikan solusi dalam klasifikasi akreditasi sekolah dalam upaya meningkatkan mutu Pendidikan di Provinsi Jambi,
2. Agar dapat membantu mempermudah pihak BPMP dalam proses seleksi klasifikasi akreditasi sekolah dengan skala besar.
3. Dapat digunakan sebagai salah satu acuan untuk penelitian selanjutnya.

## **1.5 SISTEMATIKA PENULISAN**

Laporan penelitian ini dibuat dalam sistematika yang sesuai dengan kaidah penulisan ilmiah yang benar dan dibagi dalam bab-bab sebagai berikut:

## **BAB I : PENDAHULUAN**

Pada bab ini merupakan bab pendahuluan yang mengemukakan latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

## **BAB II : LANDASAN TEORI**

Pada bab ini merupakan landasan teori yang mendasari pembahasan laporan secara khusus berisi definisi-definisi yang melandasi penelitian yang didapat dengan melakukan studi pustaka sebagai dasar dalam melakukan analisis dan perancangan.

## **BAB III : METODOLOGI PENELITIAN**

Pada bab ini berisikan kerangka kerja penelitian, metode penelitian, metode pengembangan sistem serta alat bantu penelitian yang di gunakan untuk mengembangkan perangkat lunak.

## **BAB IV : ANALISIS DAN HASIL**

Di dalam bab ini adlaah melakukan tahapan-tahapan seperti menganalisis menggunakan *data mining* dengan metode *Naïve Bayes* dngan memanfaatkan data yang telah ada.

## **BAB V : KESIMPULAN**

Dalam bab ini membahas tentang kesimpulan dan saran selama melaksanakan proses penelitian

