

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pendidikan merupakan komponen terpenting dalam peningkatan kualitas sumber daya manusia. Kualitas pendidikan yang terjamin menjadi faktor utama penentu kemajuan suatu bangsa terkhususnya di Indonesia, kualitas pendidikan di Indonesia masih mengalami berbagai macam kendala dan permasalahan terutama pada sistem pendidikannya [1]. Sistem pendidikan Indonesia yang sampai saat ini masih perlu mengalami perbaikan serta pemerataan pendidikan di Indonesia juga harus menjadi fokus utama dalam peningkatan kualitas sumber daya manusia Indonesia.

Ketidakmerataan pendidikan di Indonesia juga menjadi masalah yang krusial karena kesenjangan sosial-ekonomi yang menyebabkan banyaknya masyarakat yang tidak mengenyam pendidikan karena faktor ketidakmampuan dan kesanggupan dalam mencukupi biaya pendidikan yang mahal [1].

Salah satu upaya yang dilakukan pemerintah Indonesia dalam menangani ketidakmerataan akses pendidikan yaitu melalui program bantuan, salah satunya adalah Program Indonesia Pintar (PIP) [2]. Adapun tujuan dicetuskannya program ini ialah sebagai upaya untuk meningkatkan mutu pendidikan anak usia sekolah agar menekan terjadinya kemungkinan putus sekolah serta sebagai langkah untuk memenuhi kebutuhan setiap siswa dalam pembelajaran [2][3].

Namun, dalam pelaksanaannya program bantuan PIP seringkali menghadapi kendala serta tantangan, khususnya dalam penentuan kelayakan penerima bantuan. Proses seleksi program penerima bantuan yang masih dilakukan secara administratif menyebabkan program Indonesia pintar (PIP) seringkali diberikan kepada pihak-pihak yang notabennya mampu dalam segi ekonomi atau dengan kata lain program ini sering tidak tepat sasaran [4].

Berdasarkan data peserta didik jenjang SMA di Kota Jambi tahun ajaran 2025 yang digunakan, ada beberapa kekeliruan dari data yang mana siswa harusnya layak mendapatkan bantuan program justru pada hasil penilaiannya tidak layak sebagai penerima. Sedangkan, siswa yang berada pada kataegori ekonomi mampu justru pada penilaiannya layak untuk mendapatkan bantuan program ini. Ketidakakuratan serta jumlah data siswa yang besar menjadi dasar utama kebijakan yang diambil seringkali subjektif dan tidak transparan, proses pengolahan data yang masih dilaksanakan berbasis administratif dan manual sehingga menyulitkan petugas verifikator melakukan proses penilaian dan berisiko menghasilkan keputusan yang tidak akurat[5].

Pesatnya perkembangan teknologi khususnya pada cabang ilmu komputer yang dapat menjadi solusi berbagai permasalahan, salah satunya adalah penerapan *Data Mining* dalam penentuan kelayakan siswa penerima program PIP [6]. Dengan perkembangan teknologi memungkinkan penerapan metode *Data Mining* untuk menyelesaikan pengolahan data. Karena metode ini dapat melihat serta menggali informasi serta pola dengan cepat dalam suatu kumpulan data dan penilaian dapat dilakukan secara objektif dan transparan [7].

Data Mining merupakan cara dalam menemukan informasi berharga yang tersembunyi dari sekumpulan data dan menemukan pola menarik yang tidak diketahui sebelumnya [8]. Tujuan dari proses ini adalah untuk mendapatkan pemahaman baru yang nantinya dapat digunakan untuk membantu proses pengambilan keputusan. *Data Mining* juga disebut dengan istilah penambangan data atau eksplorasi data, karena memiliki fokus pada penemuan pola yang penting dari suatu data yang kompleks. Menurut Han dan Kamber *Data Mining* adalah proses memilih atau “menggali” informasi yang merujuk pada cara menganalisis data dari sudut pandang yang berbeda dan menyusunnya menjadi informasi yang bermanfaat [9].

Terdapat berbagai metode dalam *Data Mining* yang dapat digunakan untuk analisis dan klasifikasi data. Dari berbagai macam metode klasifikasi dalam *Data Mining* ada beberapa metode yang sudah terbukti keakuratannya pada beberapa penelitian yang telah dilakukan sebelumnya diantaranya adalah *Naïve Bayes*, *Decision Tree* (J48), dan *Random Forest*.

Naïve Bayes adalah salah satu algoritma yang banyak digunakan pada pengolahan *Data Mining*. Karena sangat efektif dalam menganalisis data, membantu proses pengambilan keputusan. Penelitian terdahulu menunjukkan bahwa performa algoritma *Naïve Bayes* memberikan akurasi dan kinerja dan tingkat akurasi yang baik. *Naïve Bayes* dapat menjadi solusi dalam analisis data untuk menentukan apakah siswa layak mendapatkan atau menerima Program Indonesia Pintar (PIP).

Penggunaan *Naïve Bayes* juga sering di komparasi dengan Algoritma *Decision Tree* (J48). Algoritma *Decision Tree* (J48) adalah algoritma yang menampilkan hasil berupa struktur pohon seperti diagram alir, setiap cabang menunjukkan hasil dari pengujian dan daun pohon menggambarkan penyebaran kelas [10]. Algoritma *Decision Tree* (J48) juga sering digunakan dalam proses klasifikasi karena kemampuannya dalam menampilkan hasil analisis dalam bentuk pohon keputusan yang mudah dipahami. Namun, Algoritma *Decision Tree* (J48) memiliki beberapa kekurangan, metode ini cenderung mengalami *overfitting* apabila struktur pohon yang dihasilkan terlalu kompleks atau data yang digunakan tidak seimbang.

Untuk mengatasi kelemahan tersebut, diterapkan juga metode atau algoritma *Random Forest* sebagai pengembangan dari *Decision Tree*. *Random Forest* adalah metode algoritma dalam *Data Mining* yang penggunaannya untuk mengklasifikasikan suatu kumpulan data. Metode ini membangun model prediktif yang melibatkan teknik *bagging multi-model* yang bekerja secara kolaboratif [11]. Dengan kompleksitas dari algoritma *random forest* ini model yang dibangun memiliki tingkat keberhasilan yang tinggi sehingga hasil akurasi yang didapat cenderung akurat dibandingkan metode algoritma lainnya.

Pada penelitian ini, diterapkan 3 metode atau algoritma yaitu, *Naïve Bayes*, Algoritma *Decision Tree* (J48), dan *Random Forest* yang digunakan sebagai solusi dalam menganalisis data siswa untuk menentukan kelayakan penerima Program Indonesia Pintar (PIP). Penggunaan ketiga metode tersebut untuk membandingkan metode mana yang memiliki tingkat akurasi yang lebih tinggi dan diharapkan hasil

yang didapat bisa membantu dalam proses penentuan kelayakan penerima bantuan agar penilaian dilakukan secara cepat, objektif, dan tepat sasaran.

Namun, pada data yang didapat ditemukan permasalahan ketidakseimbangan kelas (*Class Imbalance*). *Class Imbalance* adalah jumlah data antara kelas “Ya” dan “Tidak” mengalami ketidakseimbangan. Maka dari itu, digunakanlah teknik SMOTE (*Synthetic Minority Oversampling Technique*). SMOTE bekerja dengan menduplikasi data secara sintesis sehingga permasalahan ketidakseimbangan data dapat teratasi [12].

Selain itu, dalam meningkatkan hasil akurasi, digunakan juga teknik seleksi fitur yaitu, *Gain Ratio Attribute Evaluation*. Seleksi Fitur digunakan untuk menghilangkan atau mengurangi fitur-fitur yang kurang atau tidak relevan pada saat proses klasifikasi. *Gain Ratio* adalah salah satu teknik seleksi fitur yang menggunakan nilai *Gain* untuk menghilangkan nilai yang bersifat bias pada data yang akan digunakan, sehingga fitur yang tersisa adalah fitur terbaik diantara fitur lainnya. Dengan penerapan teknik seleksi fitur *gain ratio* diharapkan dapat meningkatkan akurasi model yang dibangun [13][14].

Penelitian sebelumnya yang berjudul “Implementasi Algoritma C4.5 Dan *Naïve Bayes* Dalam Pengambilan Keputusan Untuk Program Indonesia Pintar (PIP) Di Sekolah Dasar Negeri 04 Majalangu”, yang mencapai tingkat akurasi 100% [15] serta pada penelitian “Klasifikasi Menggunakan Metode *Naive Bayes* Untuk Menentukan Calon Penerima PIP” didapatkan tingkat akurasi keseluruhan 88.89% [8].

Sementara itu, pada penelitian dengan judul “Perbandingan Algoritma Naïve Bayes, *Decision Tree*, KNN, dan *Random Forest* Untuk Memprediksi Data Penduduk Penerima BPJS Di Lampung Timur” menghasilkan tingkat akurasi berbeda pada setiap metodenya. *Naïve Bayes* menunjukkan akurasi sebesar 86.89%, KNN menunjukkan akurasi sebesar 73.78%, dan *Decision Tree* menunjukkan akurasi sebesar 68.89% serta *Random Forest* menunjukkan akurasi sebesar 71.75%. [11].

Selain ketiga penelitian tersebut terdapat pula penelitian dengan judul “Klasifikasi Kelayakan Keluarga Penerima Bantuan Langsung Tunai Menggunakan *Gain Ratio* Dan Naïve Bayes” dengan teknik seleksi fitur *gain ratio* yang mendapatkan tingkat akurasi sebesar 90,60%, presisi 96,67%, dan recall 92,06%. Pada 3 opsi pengujian Use *Training* set, 5 *Fold* Cross-Validation dan 10 *Fold* Cross-Validation [16]. Serta pada penelitian yang menerapkan Teknik SMOTE yaitu “Penyeimbangan Kelas SMOTE Dan Seleksi Fitur Ensemble Filter Pada Support Vector Machine Untuk Klasifikasi Penyakit Liver” mendapatkan akurasi sebesar 85% dan AUC sebesar 0,850 [17]. Dan pada penelitian lain yang menggunakan metode *Decision Tree* Algoritma (J48) berjudul “Perbandingan Klasifikasi Metode *Naive Bayes* dan Metode *Decision Tree* Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda” memperoleh hasil untuk metode *Naïve Bayes* 81.25% dan *Decision Tree* algoritma (J48) sebesar 87.5% [18].

Dari uraian dan pemaparan di atas, dapat disimpulkan pada penelitian ini akan menerapkan metode klasifikasi *Naïve Bayes*, *Decision Tree* (J48), dan *Random*

Forest dengan teknik SMOTE serta seleksi fitur *Gain Ratio* pada data peserta didik jenjang SMA yang didapatkan dari BPMP Provinsi Jambi. Penelitian ini berjudul **“IMPLEMENTASI METODE *DATA MINING* DENGAN TEKNIK *SMOTE* DAN SELEKSI FITUR *GAIN RATIO* UNTUK KLASIFIKASI KELAYAKAN SISWA SMA PENERIMA PIP DI KOTA JAMBI”**.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah yang telah dijabarkan sebelumnya maka dapat dirumuskan permasalahan yang akan dikaji dalam penelitian ini yaitu:

1. Bagaimana penerapan metode klasifikasi *Naïve Bayes*, algoritma *Decision Tree* (J48), dan *Random Forest* dengan teknik *SMOTE* serta seleksi fitur *Gain Ratio* dapat digunakan untuk menentukan kelayakan siswa SMA penerima PIP di Kota Jambi?
2. Bagaimana perbandingan performa ketiga metode klasifikasi dalam menentukan kelayakan siswa SMA penerima PIP di Kota Jambi?

1.3 BATASAN MASALAH

Untuk menghindari pembahasan tentang hal-hal di luar topik penelitian ini, maka dari itu peneliti menetapkan batasan masalah yang akan dibahas pada penelitian ini sebagai berikut:

1. Data Peserta didik jenjang SMA di Kota Jambi menjadi objek utama penelitian ini.
2. Metode Klasifikasi yang digunakan adalah *Naïve Bayes*, algoritma *Decision Tree* (J48) dan *Random Forest*.

3. Teknik Seleksi Fitur yang digunakan adalah *Gain Ratio*.
4. Teknik *Balancing* Data yang digunakan adalah SMOTE.
5. Metode validasi yang akan digunakan *K-Fold Cross Validation*.
6. WEKA akan digunakan sebagai alat bantu dalam penelitian ini.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 TUJUAN PENELITIAN

Tujuan Penelitian ini untuk memberikan penjelasan yang tepat mengenai fokus utama yang ingin dicapai dalam penelitian ini. Adapun tujuan dari penelitian ini ialah:

1. Menerapkan metode klasifikasi *Naïve Bayes*, algoritma *Decision Tree* (J48), dan *Random Forest* dengan teknik *SMOTE* serta seleksi fitur *Gain Ratio* untuk menentukan kelayakan siswa SMA penerima PIP di Kota Jambi.
2. Membandingkan tingkat akurasi masing-masing metode klasifikasi dalam menentukan kelayakan siswa penerima PIP.

1.4.2 MANFAAT PENELITIAN

Penelitian ini diharapkan dapat memberikan manfaat serta kontribusi secara teoritis maupun praktis. Adapun manfaat dari penelitian ini adalah:

1. Dapat menerapkan konsep *Data Mining* menggunakan metode Klasifikasi *Data Mining*.
2. Memberikan informasi yang dapat digunakan oleh instansi terkait untuk menentukan dan memverifikasi kelayakan siswa SMA penerima Program Indonesia Pintar (PIP) di Kota Jambi.

1.5 SISTEMATIKA PENULISAN

Laporan ini disusun secara sistematis sesuai dengan kaidah penulisan ilmiah yang benar, dan dibagi menjadi bab-bab berikut:

BAB I : PENDAHULUAN

Bab ini berisi pendahuluan yang menjelaskan latar belakang permasalahan, rumusan dan batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini berisi landasan teori yang mendasari pembahasan laporan secara khusus berisikan pendapat ahli dan teori-teori yang berhubungan dengan permasalahan yang dianalisis. Teori-teori yang digunakan antara lain *Data Mining*, metode klasifikasi, *Naïve Bayes*, algoritma *Decision Tree* (J48), *Random Forest*. Serta

Teknik SMOTE, Seleksi Fitur *Gain Ratio*, Program Indonesia Pintar (PIP), Siswa SMA dan WEKA.

BAB III : METODOLOGI PENELITIAN

Bab ini berisi kerangka kerja penelitian, data yang digunakan, alur dari penelitian, teknik pengumpulan data, metode klasifikasi, dan alat bantu penelitian yang digunakan untuk mengembangkan proses perhitungan yang dilakukan.

BAB IV : ANALISIS & INTERPRETASI

Bab ini dilakukan analisis dan interpretasi dengan menggunakan metode *Naïve Bayes*, algoritma *Decision Tree* (J48), *Random Forest* terhadap kelayakan siswa SMA penerima PIP dengan penerapan teknik SMOTE dan seleksi fitur *Gain Ratio*

BAB V : PENUTUP

Bab ini berisi penutup dari hasil keseluruhan proses penelitian. Mencakup kesimpulan yang dibuat berdasarkan hasil yang didapat sebelumnya, serta saran untuk pengembangan.

Dengan sistematika tersebut, diharapkan pembahasan dalam penelitian ini dapat tersusun secara runtut, jelas, dan mudah dipahami.