

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pendidikan merupakan elemen fundamental dalam pembangunan suatu bangsa. Melalui pendidikan, individu dibekali dengan pengetahuan, keterampilan, serta nilai-nilai moral yang menjadi dasar bagi kehidupan yang produktif dan bertanggung jawab. Khususnya pada jenjang pendidikan dasar, proses pembentukan karakter dan pondasi intelektual dimulai, sehingga akses terhadap pendidikan yang merata dan berkualitas menjadi sangat krusial [1]. Oleh karena itu, kualitas pendidikan dasar harus terus ditingkatkan agar mampu menghasilkan sumber daya manusia yang unggul dan berdaya saing tinggi. Hal ini sejalan dengan upaya pemerintah dalam menciptakan sistem pendidikan yang merata dan berkualitas di seluruh wilayah Indonesia

Salah satu upaya pemerintah dalam meningkatkan akses dan kualitas pendidikan dasar adalah melalui program beasiswa. Beasiswa merupakan bantuan biaya pendidikan yang diberikan kepada peserta didik oleh lembaga pemerintah, swasta, maupun organisasi non-profit, dengan tujuan mendukung keberlanjutan proses pembelajaran. Beasiswa biasanya diberikan berdasarkan kriteria tertentu, seperti prestasi akademik, kondisi ekonomi, dan latar belakang sosial [2].

Salah satu program beasiswa yang cukup besar cakupannya di Indonesia adalah Program Indonesia Pintar (PIP). PIP memberikan bantuan tunai pendidikan kepada siswa jenjang SD hingga SMA/SMK yang memenuhi syarat ekonomi dan

sosial tertentu. PIP merupakan bentuk komitmen pemerintah dalam menekan angka putus sekolah dan mendukung pemerataan pendidikan dasar [2]. Melalui program beasiswa tersebut, siswa dapat memenuhi kebutuhan pendidikan mereka, seperti buku, alat tulis, tas, sepatu, seragam sekolah, dan perlengkapan lainnya. Namun, dalam implementasinya, penentuan penerima PIP masih sering menemui permasalahan.

Namun dalam implementasinya, PIP masih menghadapi berbagai permasalahan di lapangan. Penelitian di empat daerah kunjungan kerja Presiden pada tahun 2017 mencatat adanya inclusion error dan exclusion error, di mana siswa yang tidak memenuhi kriteria tercatat sebagai penerima, sedangkan siswa yang layak justru tidak terakomodasi. Hal ini terkait dengan lemahnya sistem verifikasi data serta belum optimalnya integrasi data antar instansi [3]

Permasalahan ketidaktepatan sasaran dalam implementasi PIP masih menjadi sorotan di berbagai daerah, salah satunya sebagaimana diberitakan terjadi Kabupaten Merangin, Provinsi Jambi, di mana diduga terdapat siswa dari keluarga mampu yang menerima bantuan, sementara siswa yang secara ekonomi kurang mampu justru tidak terdaftar sebagai penerima. Kondisi ini menunjukkan adanya potensi ketidakakuratan dalam proses pendataan dan penentuan kelayakan penerima beasiswa.

Masalah lain yang muncul adalah proses seleksi penerima beasiswa di tingkat sekolah dasar masih banyak dilakukan secara manual, mengandalkan rekomendasi guru atau pengisian formulir tanpa analisis data secara mendalam. Hal ini berpotensi menimbulkan ketidaktepatan sasaran, yaitu siswa yang seharusnya

layak tidak mendapat bantuan, sedangkan siswa yang kurang layak justru menerima beasiswa. Seleksi manual juga rawan bias, kurang efisien, dan sulit diterapkan pada jumlah siswa yang besar [4]. Hal ini disebabkan oleh kurangnya sistem pendukung keputusan yang objektif dan sistematis

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis teknologi yang mampu meningkatkan akurasi dan efisiensi dalam proses seleksi penerima beasiswa. Salah satu solusi yang dapat digunakan adalah pendekatan data mining. Data mining merupakan proses mengekstraksi informasi yang berguna dari sejumlah besar data dengan memanfaatkan metode statistik, matematika, serta kecerdasan buatan, dalam penelitian ini dilakukan kombinasi metode *K-Means Clustering* dan *Naïve Bayes Classification*.

Algoritma *K-Means Clustering* adalah salah satu teknik pengelompokan data yang termasuk dalam kategori *unsupervised learning*. Algoritma ini bertujuan untuk membagi sekumpulan data ke dalam sejumlah klaster (k) berdasarkan kemiripan fitur antar data. Setiap data akan dimasukkan ke dalam satu klaster yang memiliki pusat (*centroid*) terdekat. Proses ini dilakukan secara iteratif untuk meminimalkan jarak antara data dan pusat klaster, sehingga menghasilkan pengelompokan yang optimal. Kelebihan dari K-Means adalah kesederhanaan dan efisiensinya dalam menangani dataset berukuran besar[5]. Dalam konteks ini, *K-Means* digunakan untuk mengelompokkan siswa ke dalam kategori “layak” dan “tidak layak” menerima beasiswa berdasarkan atribut-atribut seperti alat transportasi, penerima KPS, pekerjaan ayah penghasilan ayah, pekerjaan ibu,

penghasilan ibu, layak PIP, kebutuhan khusus, jumlah saudara kandung, dan jarak rumah ke sekolah.

Sementara itu, Naïve Bayes Classification merupakan metode *supervised learning* yang didasarkan pada Teorema Bayes. Algoritma ini bekerja dengan menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan nilai atribut yang dimilikinya. Naïve Bayes dikenal memiliki kemampuan klasifikasi yang cepat dan akurat, terutama pada data berskala besar dan dengan fitur yang bersifat independen [6]. Naive Bayes merupakan algoritma pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma ini menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas [7]. Selain itu Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [8]. Keuntungan penggunaan Naive Bayes adalah 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 [7].

Penelitian ini menggabungkan dua algoritma, yaitu K- Means Clustering dan Naïve Bayes. Dalam hal ini, K-Means Clustering tidak berperan sebagai algoritma utama klasifikasi, melainkan digunakan pada tahap awal untuk

mengoptimasi atribut melalui proses pengelompokan data berdasarkan kemiripan karakteristik. Hasil pengelompokan tersebut kemudian dimanfaatkan oleh algoritma Naïve Bayes sebagai dasar untuk melakukan klasifikasi akhir terhadap kelayakan penerima beasiswa. Dengan demikian, penelitian ini berfokus pada penerapan kombinasi algoritma K-Means dan Naïve Bayes untuk mengklasifikasikan siswa yang layak atau tidak layak menerima beasiswa, dengan harapan dapat menghasilkan model prediksi yang lebih akurat dan efisien.

Penelitian terdahulu menunjukkan bahwa kombinasi beberapa algoritma data mining dapat meningkatkan efektivitas dalam proses analisis dan pengambilan keputusan. Usman dan Paraga [9] mengimplementasikan algoritma K-Means, Naïve Bayes, dan Decision Tree untuk memprediksi penjualan bahan bakar minyak (BBM), di mana K-Means digunakan untuk clustering data penjualan, sedangkan Naïve Bayes dan Decision Tree digunakan untuk prediksi. Hasilnya, kombinasi tersebut mampu mengoptimalkan distribusi BBM dan meningkatkan kepuasan pelanggan. Penelitian lain oleh Best, Foo, dan Tian [10] juga menerapkan pendekatan hybrid dengan menggabungkan K-Means dan Naïve Bayes untuk deteksi anomali pada sistem Internet of Things (IoT), dan berhasil meningkatkan akurasi deteksi hingga mencapai 90%– 100%. Sementara itu, Kurniawan dan Susanto [11], menggunakan kombinasi kedua algoritma tersebut dalam analisis sentimen Pemilihan Presiden (Pilpres) 2019, di mana K-Means berfungsi untuk clustering data dan Naïve Bayes digunakan untuk classification. Hasil pengujian terhadap 100 dan 150 data uji menunjukkan akurasi rata-rata sebesar 93,35% dengan error rate sebesar 6,66%.

Penelitian lain oleh Budiman dkk. [12] mengimplementasikan algoritma Naïve Bayes untuk menentukan calon penerima beasiswa di SMK YPM 14 Sumobito Jombang, dengan hasil akurasi 90,48%, precision 96,88%, dan recall 83,33%. Selain itu, Rahayu dkk. [13] menerapkan algoritma K-Means untuk mengelompokkan calon penerima beasiswa Bidikmisi ke dalam empat klaster, yaitu sangat layak, kurang layak, dipertimbangkan, dan tidak layak menerima beasiswa.

Berdasarkan uraian diatas maka akan dilakukan penelitian tentang klasifikasi kelayakan siswa penerima beasiswa menggunakan algoritma K-Means dan Naïve Bayes, yang akan dilakukan di Sekolah Dasar Sekecamatan Tabir, adapun tahapan dilakukan dengan pengelompokan menggunakan algoritma K-Means Setelah dikelompokkan, hasilnya digunakan sebagai fitur tambahan dalam model *Naïve Bayes* untuk menentukan siapa yang akan menerima beasiswa lalu diuji dengan metode *confusion matrix*.

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana kinerja algoritma K-Means dalam mengelompokkan kelayakan siswa penerima beasiswa di tingkat sekolah dasar pada Kecamatan Tabir?

2. Bagaimana kinerja algoritma Naïve Bayes dalam mengklasifikasikan kelayakan siswa penerima beasiswa berdasarkan hasil pengelompokan K-Means?
3. Seberapa akurat kombinasi algoritma K-Means dan Naïve Bayes dalam menentukan siswa yang layak menerima beasiswa jika diuji menggunakan confusion matrix?

1.3 BATASAN MASALAH

Agar penelitian lebih terfokus dan terarah, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini dilakukan pada sekolah dasar se Kecamatan Tabir dengan jumlah siswa 4155 orang.
2. Data yang digunakan dalam penelitian data siswa yang mencakup alat transportasi, penerima KPS, pekerjaan ayah, penghasilan ayah, pekerjaan ibu, penghasilan ibu, penerima KIP, kebutuhan khusus, jumlah saudara kandung, dan jarak rumah ke sekolah
3. Metode yang digunakan adalah algoritma K-Means untuk proses clustering dan algoritma Naïve Bayes untuk klasifikasi.
4. Evaluasi hasil prediksi dilakukan menggunakan Confusion Matrix, dengan parameter akurasi, presisi, recall, dan F1-score.

1.4 TUJUAN PENELITIAN

Penelitian ini bertujuan untuk:

1. Menerapkan algoritma K-Means untuk melakukan pengelompokan kelayakan data calon penerima beasiswa
2. Menggunakan hasil pengelompokan K-Means sebagai fitur tambahan dalam model algoritma Bayes (NB) untuk meningkatkan akurasi prediksi penerima beasiswa.
3. Mengevaluasi tingkat akurasi, presisi, recall, dan F1-score dengan metode Confusion Matrix.

1.5 MANFAAT PENELITIAN

Manfaat penelitian ini diharapkan dapat memberikan kontribusi sebagai berikut:

1. Bagi Sekolah : Memberikan solusi yang lebih efisien dan akurat dalam memilih siswa yang layak menerima beasiswa, sehingga proses seleksi menjadi lebih objektif dan berbasis data.
2. Bagi Pengembangan Ilmu Pengetahuan: Menambah wawasan tentang kombinasi metode K-Means dan Naïve Bayes dalam seleksi beasiswa.
3. Bagi Penerima Beasiswa: Meningkatkan transparansi dalam proses seleksi penerima beasiswa, sehingga memberikan kesempatan yang lebih adil bagi siswa yang memenuhi kriteria.

1.6 SISTEMATIKA PENULISAN

Penulisan laporan penelitian ini disusun dengan sistematika sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang penelitian, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas teori-teori yang relevan dengan penelitian, seperti algoritma K-Means, Naïve Bayes, serta penelitian terdahulu yang berkaitan dengan topik.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, termasuk pengumpulan data, pengolahan data, implementasi algoritma, serta metode evaluasi kinerja algoritma.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi algoritma, analisis kinerja algoritma, dan pembahasan hasil perbandingan kinerja antara K-Means dan Naïve Bayes.

BAB V KESIMPULAN DAN SARAN

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan penelitian di masa depan.