

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan permasalahan ekonomi dan sosial yang kompleks di Indonesia yang memerlukan perhatian khusus untuk mengatasinya. Masalah utamanya adalah kurangnya pendekatan berbasis data yang efektif untuk mengklasifikasikan tingkat kemiskinan di Indonesia[1].Aspek penting dalam mendukung strategi penanggulangan kemiskinan adalah tersedianya data kemiskinan. Pemerintah Indonesia telah melakukan berbagai upaya untuk menanggulangi kemiskinan, namun penurunan tingkat kemiskinan secara signifikan masih menjadi tantangan. Hal ini disebabkan oleh banyaknya faktor yang mempengaruhi kemiskinan, termasuk ketimpangan distribusi pendapatan, akses terbatas ke pendidikan dan layanan kesehatan, serta kondisi infrastruktur yang belum merata di berbagai wilayah. Pada dasarnya, terdapat berbagai aspek dalam mengartikannya, umumnya dalam standar kemiskinan yang sering digunakan adalah aspek ekonomi. Hal ini menunjukkan perlunya pendekatan yang lebih sistematis dan berbasis data untuk memahami dan menanggulangi masalah kemiskinan[2].

Di era yang modern sekarang, pastinya teknologi informasi dan ketersediaan data sudah berkembang sangat pesat, metode berbasis data seperti pembelajaran mesin (*Machine Learning*) mulai banyak digunakan di semua kalangan untuk analisis dan pengambilan keputusan. Berbagai metode dalam pembelajaran mesin (*Machine Learning*) seperti *Decision Tree*, serta *Support Vector Machine* (SVM). Sebuah penelitian dilakukan untuk menerapkan metode *Support Vector Machine* dalam Implementasi *Support Vector Machine* pada klasifikasi status kemiskinan rumah tangga dengan metode *Support Vector Machine* (SVM) dan *classification and regression trees* (cart) menggunakan GUI R, dengan proporsi data *training*:data *testing* 80:20 menunjukkan tingkat akurasi untuk *Support Vector Machine* (SVM) sebesar 89,94%, serta hasil menggunakan metode CART sebesar 89,31%[3].

Berdasarkan hasil dari perbandingan penggunaan metode selain metode yang digunakan di dalam penelitian ini (Naïve Bayes) menghasilkan akurasi mendekati 90% menggunakan metode *Support Vector Machine* dan CART, lalu untuk penggunaan metode *decision tree* sebesar 93% yang sama-sama membahas tentang lingkup kemiskinan. Algoritma *Naïve Bayes* adalah salah satu metode dalam *Machine Learning* yang di mana sangat efektif untuk mengidentifikasi dan mengklasifikasikan data, sehingga dapat digunakan pada penelitian terkait, yaitu faktor-faktor yang mempengaruhi tingkat kemiskinan di suatu wilayah atau daerah di Indonesia. *Naïve Bayes* adalah salah satu metode klasifikasi yang berbasis pada teori probabilitas. Metode ini menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain. Penelitian yang menggunakan metode *Naïve Bayes* sudah banyak dilakukan di berbagai bidang, penelitian di bidang kesehatan yaitu prediksi kemungkinan penyakit liver menggunakan algoritma klasifikasi *Naïve Bayes*, hasil akurasi algoritma klasifikasi *Naïve Bayes* menghasilkan akurasi sebesar 91% penelitian terkait di bidang kesehatan dengan prediksi kemungkinan penyakit liver menggunakan algoritma klasifikasi *Naïve Bayes* dapat disimpulkan bahwa akurasi yang dihasilkan dalam menggunakan metode *Naïve Bayes* cukup baik[4]. Penelitian di bidang Pendidikan tentang Klasifikasi performa akademik siswa menggunakan metode *Decision Tree* dan *Naïve Bayes* menghasilkan akurasi untuk *Naïve Bayes* sebesar 89.57%, dan hasil akurasi menggunakan *Decision Tree* sebesar 83.89%[5]. Dalam kasus penelitian ini yaitu klasifikasi tingkat kemiskinan di Indonesia, metode *Naïve Bayes* dapat digunakan untuk mengklasifikasi berbagai faktor yang mempengaruhi status kemiskinan, seperti persentase penduduk miskin (P0) menurut kabupaten dan kota (persen), rata-rata lama sekolah penduduk 15+ (tahun), pengeluaran per kapita disesuaikan (ribu rupiah/orang/tahun), indeks pembangunan manusia, umur harapan hidup (tahun), persentase rumah tangga yang memiliki akses terhadap sanitasi layak, persentase rumah tangga yang memiliki akses terhadap air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, PDRB (Produk Domestik Regional Bruto) atas dasar harga konstan menurut

pengeluaran (rupiah), dengan data yang lengkap dan relevan serta jumlah dari data yang banyak, Naïve Bayes dapat menghasilkan hasil akurasi yang akurat untuk mengklasifikasikan tingkat kemiskinan dalam kategori miskin atau tidak miskin. Keunggulan metode Naïve Bayes dalam menangani dataset besar dengan fitur/atribut yang banyak serta menggunakannya dengan sederhana dan cepat menjadikannya metode Naïve Bayes cocok untuk mengklasifikasikan tingkat kemiskinan.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang yang telah dijelaskan di atas permasalahan yang ingin diselesaikan dalam penelitian ini yaitu hasil dari mengklasifikasikan tingkat kemiskinan yang ada di Indonesia dengan metode Naïve Bayes.

1.2.2 Metode Usulan

Metode yang akan digunakan dalam penelitian ini yaitu Algoritma *Naïve Bayes*.

1.2.3 Pertanyaan Penelitian

Berdasarkan masalah di dalam penelitian dan metode yang disajikan, pertanyaan penelitian sebagai berikut :

Berapa hasil akurasi pada klasifikasi tingkat kemiskinan di Indonesia menggunakan metode *Naïve Bayes*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan dalam penelitian yaitu untuk memperoleh hasil akurasi pada klasifikasi tingkat kemiskinan di Indonesia.

1.3.2 Manfaat Penelitian

1. Mengetahui hasil kinerja dari metode *Naïve Bayes* dalam melakukan klasifikasi

2. Memberikan manfaat kepada pemerintah untuk mengetahui tingkat kemiskinan di Indonesia melalui hasil dari penelitian.

1.4 Batasan Masalah

Berdasarkan pada perumusan masalah yang sudah disebutkan di atas, maka batasan masalah penelitian ini adalah sebagai berikut:

1. Metode yang digunakan pada pengklasifikasian ini adalah Algoritma *Naïve Bayes*,
2. Dataset yang digunakan dalam penelitian ini diambil dari dataset yang ada di website Kaggle. Dataset tersebut mempunyai data sebanyak 514 data.
3. Dataset yang diklasifikasi yaitu dataset yang mencakup sebanyak 10 atribut dan 514 data.

1.5 Sistematika Penulisan

BAB I PENDAHULUAN

Pada BAB I berisi tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan

BAB II KAJIAN PUSTAKA

Pada BAB II berisi tentang teori yang dibutuhkan untuk penelitian, yaitu penerapan metode *Naïve Bayes* dalam kasus klasifikasi tingkat kemiskinan di Indonesia, serta beberapa penelitian sebelumnya

BAB III METODOLOGI PENELITIAN

Pada BAB III berisi tentang langkah-langkah atau proses dalam memecahkan masalah yang sedang diteliti menggunakan metode yang telah ditetapkan.

BAB IV HASIL DAN PEMBAHASAN

Pada BAB IV berisi tentang penjelasan hasil implementasi dan algoritma klasifikasi serta hasil evaluasi dari klasifikasi.

BAB V PENUTUP

Pada BAB V berisi tentang penjelasan mengenai kesimpulan dari hasil klasifikasi beserta saran penelitian