

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan masih menjadi tantangan serius di Indonesia karena memiliki karakteristik yang kompleks dan multidimensional. Berdasarkan data dari Badan Pusat Statistik (BPS), tingkat kemiskinan nasional cenderung stagnan pada kisaran 9–10 %, meskipun menunjukkan tren penurunan dari tahun ke tahun. Namun, terdapat ketimpangan yang tajam antara kabupaten dan kota dalam hal jumlah penduduk miskin [1]. Masalah kemiskinan ini tidak semata-mata disebabkan oleh ketidakmampuan ekonomi, melainkan juga dipengaruhi oleh ketidakadilan dalam akses pendidikan dan layanan kesehatan. Salah satu studi menjelaskan bahwa biaya transportasi serta berbagai biaya tidak langsung lainnya menjadi hambatan bagi masyarakat miskin untuk mengakses layanan kesehatan, sehingga memperburuk ketertinggalan mereka dalam aspek tersebut [2]. Selain itu, terdapat temuan lain yang menunjukkan bahwa tingkat pendidikan yang rendah memiliki hubungan yang erat dengan tingginya risiko kemiskinan di Indonesia selama periode tahun 2015 hingga 2023 [3].

Di sejumlah daerah miskin, masih banyak dijumpai kekurangan infrastruktur dasar seperti jalan yang rusak, pasokan listrik yang tidak stabil, serta sanitasi yang buruk. Kondisi ini secara signifikan menghambat mobilitas ekonomi, akses pendidikan, dan layanan kesehatan masyarakat. Studi terkait ketimpangan antarwilayah juga menunjukkan bahwa disparitas ini semakin diperburuk oleh ketidakseimbangan pembangunan infrastruktur dan pendidikan. Oleh karena itu, penyusunan kebijakan pengentasan kemiskinan tidak seharusnya hanya mengacu pada indikator pendapatan semata. Diperlukan pendekatan yang lebih komprehensif dengan mempertimbangkan pemetaan wilayah berdasarkan ketersediaan infrastruktur, akses terhadap layanan dasar, dan tingkat partisipasi ekonomi masyarakat [4]. Untuk mendukung hal tersebut, sangat dibutuhkan metode analisis yang mampu mengklasifikasikan wilayah secara sistematis berdasarkan tingkat kemiskinan dan dimensi-dimensinya, sehingga program intervensi dapat dirancang secara lebih tepat sasaran dan efektif.

Seiring perkembangan teknologi informasi, teknik data mining menjadi semakin penting dalam mendukung pengambilan keputusan berbasis data, terutama di sektor publik dan pemberdayaan sosial. Salah satu metode klasifikasi yang sederhana namun terbukti efektif adalah Naive Bayes Classifier. Metode ini menggunakan pendekatan probabilistik dalam

menghitung peluang bahwa suatu data termasuk dalam sebuah kelas berdasarkan fitur-fitur yang dimilikinya. Kelebihan utama metode ini adalah efisiensi tinggi dalam memproses dataset besar serta hasil klasifikasi yang cukup akurat, meskipun asumsi independensi antar fitur tidak selalu terpenuhi [5].

Selain keunggulannya dalam efisiensi dan kecepatan komputasi, metode Naive Bayes juga sangat sesuai diterapkan dalam konteks klasifikasi sosial ekonomi karena kemampuannya dalam menangani data kategori dan hasil prediksi yang interpretable. Dalam penelitian klasifikasi wilayah berdasarkan tingkat kemiskinan, metode ini dapat memanfaatkan berbagai atribut sosial ekonomi seperti pengeluaran per kapita, indeks pembangunan manusia (IPM), partisipasi angkatan kerja, serta variabel pendidikan dan kesehatan untuk menentukan kelas kemiskinan suatu daerah [6]. Hasil klasifikasi yang dihasilkan dari metode ini dapat menjadi dasar yang kuat dalam menyusun strategi kebijakan publik yang lebih inklusif dan adaptif terhadap kondisi riil masyarakat. Dengan integrasi data mining dan pendekatan probabilistik, proses analisis dapat dilakukan secara lebih objektif dan terukur, mendukung tata kelola pembangunan yang responsif terhadap persoalan ketimpangan wilayah.

Penerapan algoritma Naive Bayes dalam analisis kemiskinan telah banyak dilakukan di Indonesia. Salah satu penelitian menunjukkan bahwa metode ini berhasil mengklasifikasikan tingkat kemiskinan di berbagai kabupaten dan kota dengan tingkat akurasi mencapai $92,4\% \pm 2,99\%$ [7]. Penelitian lain kemudian mengembangkan pendekatan ini lebih lanjut dengan menggabungkan Naive Bayes dan Particle Swarm Optimization, yang terbukti mampu meningkatkan akurasi klasifikasi dari $59,75\%$ menjadi $82,93\%$ untuk wilayah-wilayah miskin di Pulau Jawa dan Sumatera [8]. Temuan-temuan tersebut menunjukkan bahwa Naive Bayes memiliki potensi kuat sebagai dasar pemetaan wilayah berdasarkan tingkat kemiskinan. Dengan dukungan hasil klasifikasi yang objektif dan akurat, para pemangku kebijakan dapat merancang program intervensi sosial secara lebih tepat sasaran dan efisien.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi kemiskinan di Indonesia dengan memanfaatkan data indikator sosial-ekonomi dari kabupaten/kota. Indikator yang digunakan meliputi persentase penduduk miskin, indeks pembangunan manusia, rata-rata lama sekolah, tingkat pengangguran terbuka, pengeluaran per kapita, akses terhadap sanitasi dan air minum layak, serta PDRB per kapita. Data ini dipilih karena merepresentasikan kondisi sosial dan ekonomi yang beragam di setiap wilayah. Untuk mengelompokkan data, penelitian ini

menerapkan metode diskretisasi menggunakan pendekatan kuartil yang membagi setiap indikator ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi, sehingga pola distribusi data menjadi lebih terstruktur dan mudah dianalisis.

Penelitian ini secara khusus mengusulkan penerapan metode Naïve Bayes untuk melakukan klasifikasi wilayah ke dalam dua kategori utama, yaitu “miskin” dan “tidak miskin”, berdasarkan sejumlah indikator sosial ekonomi yang tersedia. Pemilihan metode Naïve Bayes didasarkan pada beberapa pertimbangan teknis, antara lain kesederhanaan algoritma, efisiensi komputasi, dan kemampuannya dalam menangani dataset dengan dimensi tinggi [9]. Naïve Bayes bekerja berdasarkan prinsip probabilistik, di mana setiap fitur pada data diasumsikan saling independen terhadap satu sama lain dalam menentukan kelas. Meskipun asumsi independensi ini sering kali tidak sepenuhnya terpenuhi dalam data sosial ekonomi, berbagai studi telah membuktikan bahwa Naïve Bayes tetap mampu menghasilkan performa klasifikasi yang baik. Dengan menggunakan pendekatan ini, penelitian ini bertujuan membangun sebuah sistem yang mampu secara otomatis memprediksi status kemiskinan suatu wilayah dengan hanya mengandalkan data-data makro yang tersedia, seperti pengeluaran per kapita, IPM, angka pengangguran, dan indikator lainnya. Keunggulan metode ini terletak pada kemampuan untuk menghitung dan memetakan probabilitas kemiskinan secara menyeluruh, sehingga dapat memberikan gambaran klasifikasi yang tidak hanya binari, tetapi juga bersifat probabilistik dan interpretatif.

Sebagai bagian dari upaya untuk memastikan bahwa model yang dikembangkan benar-benar dapat diandalkan, penelitian ini juga menekankan pada evaluasi performa model secara komprehensif. Evaluasi dilakukan menggunakan beberapa metrik penting seperti akurasi, presisi, recall, dan F1-score. Akurasi mengukur seberapa tepat model dalam mengklasifikasikan seluruh data, sementara presisi dan recall memberikan gambaran tentang ketepatan dan sensitivitas model terhadap kelas tertentu. F1-score digunakan sebagai indikator keseimbangan antara presisi dan recall, terutama penting ketika distribusi kelas tidak seimbang. Evaluasi ini dilakukan tidak hanya sekali, tetapi juga melalui beberapa skenario pembagian data (misalnya 60%, 70%, dan 80% data latih) untuk melihat kestabilan model. Diharapkan, model yang dibangun tidak hanya unggul dari sisi teknis, tetapi juga relevan secara sosial, terutama dalam mendukung kebijakan publik berbasis data. Hasil akhir dari penelitian ini diharapkan dapat dimanfaatkan oleh pemerintah daerah, lembaga perencanaan

pembangunan, maupun organisasi non-pemerintah sebagai dasar untuk mengidentifikasi wilayah prioritas yang memerlukan intervensi kebijakan yang lebih tajam dan terarah, sehingga program pengentasan kemiskinan dapat dilaksanakan secara lebih efektif, efisien, dan terukur.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Permasalahan ini muncul karena masih tingginya disparitas kemiskinan antar wilayah serta perlunya pendekatan objektif berbasis data dalam mengidentifikasi wilayah miskin.

1.2.2 Metode Usulan

Metode Naive Bayes Classifier untuk melakukan klasifikasi wilayah berdasarkan tingkat kemiskinan. Data sosial ekonomi yang digunakan diperoleh dari Kaggle, mencakup 514 data kabupaten/kota dengan 10 indikator utama, seperti pengeluaran per kapita, rata-rata lama sekolah, indeks pembangunan manusia, dan lainnya. Data numerik dikonversi menjadi kategori (rendah, sedang, tinggi) menggunakan metode diskretisasi kuartil. Selanjutnya, klasifikasi dilakukan dengan bantuan perangkat lunak Orange Data Mining, dan hasil model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score.

1.2.3 Pertanyaan Penelitian

Berapa nilai akurasi, presisi, recall, dan F1-score dari metode Naive Bayes untuk mengklasifikasikan tingkat kemiskinan di Indonesia?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan Penelitian ini adalah:

Bertujuan untuk mengetahui nilai akurasi, presisi, recall, dan F1-score untuk memastikan keandalan hasil klasifikasi.

1.3.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini ialah untuk mengetahui nilai dari akurasi, presisi, recall, dan F1-score untuk mempermudah dalam mengklasifikasi Tingkat kemiskinan di Indonesia.

1.4 Batasan Masalah

Terdapat beberapa batasan masalah pada penelitian ini:

1. Penelitian ini hanya menggunakan data indikator sosial-ekonomi dari kabupaten/kota di Indonesia, yaitu persentase penduduk miskin, indeks pembangunan manusia, rata-rata lama sekolah, tingkat pengangguran terbuka, pengeluaran per kapita, akses terhadap sanitasi, akses air minum layak, dan PDRB per kapita.
2. Data yang digunakan merupakan data dari sumber terbuka salah satu platform yang di manfaatkan adalah kaggle.
3. Klasifikasi indikator sosial-ekonomi dilakukan menggunakan metode diskretisasi kuartil yang membagi data menjadi tiga kategori, yaitu rendah, sedang, dan tinggi.
4. Metode klasifikasi yang digunakan dalam penelitian ini terbatas pada metode Naive Bayes dan tidak membahas atau membandingkan dengan metode klasifikasi lain.
5. Evaluasi performa model klasifikasi hanya difokuskan pada pengukuran akurasi, presisi, recall, dan F1-score.

Penelitian ini tidak menganalisis secara mendalam faktor penyebab kemiskinan, melainkan hanya memfokuskan pada proses klasifikasi berdasarkan data yang tersedia.

1.5 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini disusun untuk memberikan gambaran yang sistematis dan runtut mengenai seluruh tahapan penelitian yang dilakukan. Adapun struktur penulisan skripsi ini terdiri dari lima bab utama, yaitu:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang permasalahan yang menjadi dasar penelitian, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, metodologi penelitian yang digunakan, serta sistematika penulisan secara keseluruhan.

BAB II KAJIAN PUSTAKA

Bab ini berisi uraian teori-teori dasar yang menjadi landasan penelitian, termasuk pengertian klasifikasi, metode Naïve Bayes, data mining, evaluasi model dengan confusion matrix, serta ulasan penelitian terdahulu yang relevan sebagai acuan ilmiah.

BAB III: PERANCANGAN SISTEM

Bab ini menguraikan perancangan perangkat lunak klasifikasi kemiskinan yang mencakup desain arsitektur sistem, pencarian dan pengolahan dataset, seleksi fitur dan penentuan label, implementasi metode Naïve Bayes, serta evaluasi hasil klasifikasi.

BAB IV: HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen, analisis data, evaluasi performa model berdasarkan metrik akurasi, presisi, recall, dan F1-score, serta interpretasi terhadap hasil klasifikasi yang diperoleh dengan metode Naïve Bayes.

BAB V: PENUTUP

Bab terakhir memuat kesimpulan dari hasil penelitian serta saran-saran untuk pengembangan penelitian lebih lanjut atau penerapan model dalam skenario nyata.

Setiap bab dalam skripsi ini disusun secara logis dan saling berkaitan untuk membangun narasi ilmiah yang utuh, sehingga dapat memberikan kontribusi akademik dalam penerapan metode klasifikasi pada permasalahan sosial, khususnya dalam pemetaan tingkat kemiskinan di Indonesia.