

BAB I PENDAHULUAN

1.1 Latar Belakang

Banyak negara berkembang yang berjuang melawan kemiskinan, yang merupakan masalah rumit yang menghalangi mereka mencapai kesejahteraan [1]. Kemiskinan masih menjadi masalah serius di Indonesia. Dalam upaya memerangi kemiskinan, pemerintah telah lama menawarkan bantuan sosial (bansos)[2], dan sejak tahun 2007, Program Keluarga Harapan (PKH) telah diterapkan untuk mempercepat inisiatif ini dan mengurangi kesenjangan sosial[3]. Pengumpulan data yang lebih tepat diperlukan untuk mengidentifikasi rumah tangga mana yang berhak menerima bantuan sembako, seperti penelitian Damuri dkk. menjelaskan bahwa bantuan sosial seringkali tidak tepat sasaran dalam praktiknya [4].

Menurut hasil dari wawancara penulis dengan Bapak Doedy Budhi Purwanto, S.IP, Kepala Bidang Perlindungan dan Jaminan Sosial Dinas Sosial Kabupaten Pekalongan, pencalonan dan pengelompokan calon peserta PKH di tingkat desa masih dilakukan secara manual. Pendekatan ini kurang efektif dan efisien karena pemerintah desa harus memasukkan data, menetapkan prioritas, dan memverifikasi kriteria penerima satu per satu. Proses ini sering diulangi karena kurangnya sumber daya, yang dapat mengakibatkan kesalahan identifikasi. Prosedur manual ini kurang efektif, memakan waktu, dan rawan kesalahan.

Untuk mengatasi problem tersebut dapat dilakukan dengan metode data mining[5]. Data mining adalah teknik komputasi yang memanfaatkan gabungan statistika, matematika, dan pembelajaran mesin (*machine learning*) untuk membantu pengambil keputusan dalam menyelesaikan masalah secara efektif[6].

Penelitian ini mengusulkan penggunaan algoritma Naive Bayes dan Random Forest karena keunggulan masing-masing. Pemilihan Naive Bayes dan Random Forest dalam penelitian ini didasarkan pada karakteristik data yang dianalisis. Naive Bayes efektif untuk data dengan fitur banyak dan berdistribusi kategori, serta memiliki kecepatan komputasi tinggi[7]. Random Forest unggul dalam akurasi

dengan menggabungkan banyak decision tree, tahan terhadap overfitting, dan mengidentifikasi fitur berpengaruh[8].

Pada Penelitian E. Fitriani [9] algoritma C4.5 memiliki akurasi yang lebih tinggi (91,25%) dibandingkan dengan Naive Bayes (87,11%) dalam klasifikasi penerima bantuan sosial. Lalu penelitian oleh A. Nur dkk [10] hasil dari penelitian ini menunjukkan bahwa algoritma Naive bayes dapat melakukan klasifikasi secara sempurna dilihat dari nilai AUC sempurna 1,000 (excellent classification) akurasi 99.87% dibanding dengan C.45 dalam penentuan kelayakan PKH. Algoritma Random Forest digunakan dalam penelitian D. Haidar dkk[11] hasil evaluasi melalui metrik ROC AUC yang menunjukkan akurasi hingga 99%, menandakan kinerja model yang sangat baik dalam penentuan penerimaan bantuan PKH.

Penelitian-penelitian sebelumnya tentang data mining dalam konteks bantuan sosial [9], [10], [11] umumnya belum mempertimbangkan secara mendalam dampak ketidakseimbangan dataset terhadap performa model klasifikasi. Penelitian ini bertujuan untuk mengisi celah tersebut dengan secara khusus menangani isu imbalanced dataset yang teridentifikasi dalam data penelitian ini, di mana proporsi kategori "layak" mencapai 77,7% dan kategori "tidak layak" 22,3%. Dengan menerapkan teknik SMOTE [12], penelitian ini berupaya meminimalkan bias akibat ketidakseimbangan data dan meningkatkan akurasi prediksi kelayakan penerimaan bantuan, terutama pada kategori "tidak layak." Selain itu, penelitian ini juga akan melakukan klasifikasi kelayakan masyarakat dalam mendapatkan bantuan Program Keluarga Harapan (PKH) di Desa Kaligawe dengan menerapkan dua metode, yaitu Naive Bayes dan Random Forest, serta melakukan analisis perbandingan antara kedua model untuk memastikan tingkat akurasi dari masing-masing metode.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan rumusan masalah dalam penelitian ini adalah :

1. Apakah proses klasifikasi kelayakan penerima bantuan PKH dapat diotomatisasi untuk mengatasi kendala proses manual yang memakan waktu dan rentan terhadap kesalahan?
2. Bagaimana dampak penerapan teknik SMOTE terhadap performa algoritma Naive Bayes dan Random Forest dalam menangani dataset yang tidak seimbang?
3. Algoritma mana, antara Naive Bayes dan Random Forest, yang memberikan akurasi lebih tinggi dalam proses klasifikasi kelayakan penerima bantuan PKH di Desa Kaligawe?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memastikan ruang lingkup yang terfokus, yaitu:

1. Lokasi penelitian dibatasi hanya di Desa Kaligawe Kecamatan Karangdadap Kabupaten Pekalongan.
2. Data yang digunakan adalah data penerima PKH di Desa Kaligawe tahun 2024.
3. Data yang digunakan mencakup atribut-atribut tertentu yang dianggap relevan, seperti tingkat penghasilan, jumlah tanggungan, dan kriteria lainnya, sesuai dengan standar dari Program Keluarga Harapan.
4. Penelitian hanya membandingkan algoritma Naive Bayes dan Random Forest untuk proses klasifikasi.
5. Penanganan ketidakseimbangan data dilakukan menggunakan metode Synthetic Minority Oversampling Technique (SMOTE).
6. Evaluasi performa algoritma dilakukan menggunakan metrik akurasi, *confusion matrix*, dan *ROC-AUC*.

7. Asumsi penelitian adalah bahwa data yang digunakan valid dan representative untuk menggambarkan kondisi kelayakan masyarakat dalam mendapatkan bantuan dana PKH.
8. Penelitian ini tidak mempertimbangkan faktor-faktor politik lokal yang mungkin mempengaruhi proses pengambilan Keputusan terkait pemberian bantuan dana PKH.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengembangkan model klasifikasi yang mampu mengotomatisasi proses pengelompokan kelayakan penerima bantuan PKH di Desa Kaligawe.
2. Menganalisis dampak penerapan teknik SMOTE terhadap performa algoritma Naive Bayes dan Random Forest dalam menangani dataset tidak seimbang.
3. Membandingkan kinerja algoritma *Naive Bayes* dan *Random Forest* dalam klasifikasi kelayakan penerima bantuan PKH berdasarkan metrik akurasi, *confusion matrix*, dan *ROC-AUC*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini terbagi menjadi dua, yaitu:

1. Secara Teoritis
Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan dalam bidang data mining, khususnya dalam penerapan algoritma klasifikasi untuk masalah social seperti penentuan kelayakan penerima bantuan.
2. Secara Praktis
Penelitian ini dapat memberikan manfaat langsung bagi desa dan pihak terkait dalam mengelola dan menentukan kelayakan calon penerima bantuan PKH secara lebih cepat dan akurat. Dengan adanya solusi berbasis teknologi, diharapkan distribusi bantuan sosial menjadi lebih tepat sasaran, mengurangi potensi kesalahan manusia dalam proses manual, dan menjadi

model pendukung keputusan yang dapat diterapkan pada program bantuan sosial lainnya.

1.6 Sistematika Penulisan

Berisi sistematika penulisan skripsi yang memuat uraian secara garis besar isi skripsi untuk tiap-tiap bab :

BAB I PENDAHULUAN

Bagian ini merupakan pengantar pokok dari permasalahan yang akan dipecahkan dan gambaran dari penelitian ini secara keseluruhan. Adapun yang dibahas pada bab ini adalah latar belakang, rumusan masalah, batasan masalah, tujuan masalah, manfaat penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian teori yang relevan dengan penelitian, termasuk konsep dasar data mining, algoritma Naïve Bayes dan Random Forest serta studi-studi terkait tentang klasifikasi kelayakan bantuan PKH.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan desain penelitian komparatif untuk membandingkan akurasi dan efisiensi algoritma Naive Bayes dan Random Forest dalam klasifikasi kelayakan penerima bantuan PKH di Desa Kaligawe. Penelitian ini menggunakan data sekunder dari instansi pemerintah yang mencakup variabel demografis dan ekonomi masyarakat. Tahapan penelitian meliputi pengumpulan data, preprocessing, pemodelan dengan algoritma, dan evaluasi model menggunakan metrik seperti akurasi dan precision. Pengolahan data dilakukan dengan Python dan library scikit-learn untuk menganalisis kinerja algoritma.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil penelitian berdasarkan perbandingan

kinerja algoritma Naive Bayes, dan Random Forest dalam klasifikasi kelayakan penerima bantuan PKH, serta pembahasan mengenai temuan-temuan penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan serta saran-saran untuk penelitian lebih lanjut atau untuk penerapan di lapangan.

