

BAB I PENDAHULUAN

1.1 Latar Belakang

Penduduk kurang mampu merupakan suatu masalah yang dialami beberapa Negara, termasuk Indonesia. Masalah kemiskinan adalah masalah lama yang belum terselesaikan. Meskipun sudah begitu banyak bantuan bagi penduduk kurang mampu yang di berikan oleh pemerintah, seperti halnya Jaminan Kesehatan Nasional, Kartu Indonesia Sehat (KIS), Kartu Indonesia Pintar (KIP) serta jenis bantuan lainnya, namun hal itu masih belum mampu mengatasi sepenuhnya masalah kemiskinan di Indonesia. Salah satu contohnya, pada tahun 2016 di Kabupaten Jember masih tercatat 265.000 jiwa penduduk kurang mampu dimana klasifikasi penduduk kurang mampu tersebut didasarkan pada beberapa parameter yang diantaranya terdiri dari pendapatan, jumlah makan sehari, pendidikan, luas bangunan, program yang diterima, status kepemilikan rumah, sumber penerangan dan MCK.

Dari penelitian yang dilakukan oleh Haditsah Annur (2018) menyatakan bahwa kinerja algoritma *Naïve Bayes* terhadap dataset yang telah diambil diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori bagus (*good*). Namun, penelitian tersebut masih memiliki kekurangan. Kekurangan penelitian tersebut salah satunya adalah tidak ada *pre-processing* untuk melengkapi data masukan yang kosong (*missing value*). Sedangkan di lapangan, data survei atau data di lapangan sering kali tidak lengkap atau yang sering disebut *missing value* atau data kosong yang terdapat di antara beberapa parameter. Data kosong seringkali terjadi akibat tidak adanya respon terhadap item-item yang dibutuhkan. Kurangnya data di salah satu parameter sangat berpengaruh pada parameter-parameter yang lain. Kurang lengkapnya data ini selalu menjadi permasalahan umum dalam melakukan survey dan menyebabkan data penelitian menjadi tidak lengkap.

Untuk mengklasifikasi data penduduk kurang mampu tersebut bisa digunakan beberapa metode, misalnya metode *Naïve Bayes*. *Naïve Bayes* adalah metode sederhana klasifikasi probabilistik yang menghitung serangkaian probabilitas dengan menambahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma teorema bayes digunakan dalam metode ini dan mengasumsikan semua atribut independen yang diberikan oleh nilai pada variabel kelas (Patil dan Sherekar, 2013).

Peneliti merasa perlu untuk melakukan laporan tentang definisi di atas yaitu tentang *missing value*, untuk memprediksi *missing value* pada data penduduk kurang mampu tersebut. Untuk perbaikan *missing value* dalam data penduduk tersebut, akan menggunakan metode *k-nearest neighbor*. Penanganan missing data dengan KNN (*K-Nearest Neighbor*) diawali dengan menentukan sejumlah tetangga terdekat atau observasi terdekat yang disimbolkan dengan K, kemudian menghitung jarak terkecil dari setiap observasi yang tidak mengandung missing data (Martha Shantika dkk, 2018).

Dengan memperbaiki missing value menggunakan metode k-nearest neighbor yang dikombinasikan dengan klasifikasi *naïve bayes* dalam pengambilan keputusan, diharapkan dapat membantu mengatasi permasalahan *missing value* dalam data penduduk kurang mampu di Kabupaten Jember sekaligus mengetahui pengaruh data yang kosong bila dibandingkan parameter-parameter lain terhadap pengambilan keputusan.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka permasalahan yang dapat diambil dalam tugas akhir ini adalah sebagai berikut:

1. Berapa tingkat akurasi *k-nearest neighbor* untuk mengatasi *missing value* dan tingkat akurasi *naïve bayes* untuk proses klasifikasi?
2. Apakah terdapat perbandingan selisih antara hasil *naïve bayes* tanpa *k-nearest neighbor* dengan *naïve bayes* menggunakan *k-nearest neighbor*?

1.3 Batasan Masalah

Mengacu pada masalah yang telah dirumuskan, maka batasan masalah yang akan dibahas dalam pembuatan tugas akhir ini adalah sebagai berikut:

1. Data yang digunakan adalah data penduduk kurang mampu tahun 2017 yang diperoleh dari Dinas Sosial di Kabupaten Jember.
2. Parameter yang digunakan adalah Pendidikan, Status Rumah, Program Yang Diterima, Luas Bangunan, Fasilitas MCK, Sumber Penerangan, Jumlah Makan Sehari dan Pendapatan.
3. Untuk data yang bersifat *continue* akan dilakukan pengelompokan agar dapat bersifat diskrit.

1.4 Tujuan

Adapun tujuan yang ingin dicapai dari tugas akhir yang dilakukan yaitu:

1. Untuk mengetahui akurasi *k-nearest neighbor* untuk mengatasi *missing value* dan mengetahui akurasi *naïve bayes* untuk klasifikasi.
2. Untuk mengetahui hasil selisih *naïve bayes* tanpa *k-nearest neighbor* dengan *naïve bayes* menggunakan *k-nearest neighbor*.

1.5 Manfaat

Adapun manfaat yang diharapkan dari tugas akhir ini antara lain sebagai berikut:

1. Bagi Pemerintah
Manfaat dari tugas akhir ini diharapkan agar dapat digunakan dan mempermudah pemerintah dalam pengklasifikasi status penduduk kurang mampu untuk analisa yang lebih efisien.
2. Bagi Peneliti
Untuk membuktikan bahwa *k-nearest neighbor* dapat melengkapi *missing value* dan metode *naïve bayes* untuk klasifikasi penduduk kurang mampu.
3. Bagi Peneliti Selanjutnya
Untuk menjadi dasar bagi penelitian selanjutnya dalam pengklasifikasian sebuah data menggunakan metode *naïve bayes*.