

BAB I

PENDAHULUAN

1.1 Latar Belakang

Puskesmas Tegaldlimo adalah salah satu unit pelaksana kesehatan yang ada di daerah kecamatan dan mempunyai fungsi salah satunya melakukan upaya pemenuhan akan gizi dan kesehatan anak usia bawah lima tahun (balita). Sebelumnya penulis sudah mendapat izin dari Dinas Kesehatan Kabupaten Banyuwangi dan diteruskan ke Puskesmas Tegaldlimo untuk mendapatkan data gizi balita. Di Puskesmas Tegaldlimo sampai 2018 tercatat ada sebanyak 1186 balita, disini ada empat status gizi yang menurut penulis sangat jauh perbedaannya yaitu “gizi baik”, “gizi kurang”, “gizi lebih”, ”gizi buruk”, dengan rincian gizi baik sebanyak 1101 balita, gizi kurang sebanyak 39 balita, gizi lebih sebanyak 39 balita dan gizi buruk sebanyak 7 balita. Dengan adanya klasifikasi terhadap gizi balita maka dapat mempermudah dan membuat waktu pengerjaan lebih efisien. sehingga penulis mendapat suatu permasalahan yaitu data “gizi baik” jauh lebih banyak daripada data “gizi kurang”, “gizi lebih”, “gizi buruk”.

Proses klasifikasi dengan berbagai algoritma *machine learning*, yaitu algoritma yang mempelajari pengelompokan data berdasarkan pengolahan data yang telah ada sebelumnya, yang mempunyai tujuan untuk mendapatkan target kelas yang akurat. Namun kenyataannya muncul suatu permasalahan dalam proses klasifikasi tersebut ketika salah satu kelasnya mempunyai jumlah data yang jauh lebih banyak pada *training dataset*-nya, yaitu kumpulan data yang dijadikan bahan pengolahan data. Permasalahan tersebut disebut juga dengan *imbalance dataset problem*. Kondisi tersebut akan berpengaruh pada proses klasifikasi data yang akan dilakukan untuk menentukan kelas suatu data. Jika kondisi data tidak seimbang (*imbalance*) maka kecenderungan kelas data tidak stabil karena data akan lebih condong ke bagian data yang memiliki komposisi data lebih besar (*Majority class*) (Riswanto, dkk, 2007).

Untuk menyelesaikan permasalahan tersebut salah satu pendekatan yang sangat populer adalah dengan metode sampling. Metode *sampling* secara umum bekerja untuk menyeimbangkan data sehingga komposisi data yang akan dihasilkan akan berbentuk sama besar. Terdapat dua metode sampling yaitu, *Random Over Sampling* dan *Random Under Sampling*. (Riswanto, dkk, 2007). Tugas Akhir ini akan menggunakan metode *random over-sampling* dalam proses *rebalance* data, yaitu proses penyeimbangan data. Dan selanjutnya untuk memperoleh hasil pengukuran nilai akurasi, analisis ini menggunakan metode *Naive Bayes (Nb)* sebagai algoritma klasifikasi.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah

1. Bagaimana tingkat akurasi algoritma *Random Over Sampling* untuk mengatasi terjadinya ketidakseimbangan kelas?
2. Bagaimana hasil analisa penggunaan algoritma *Random Over Sampling* dalam permasalahan klasifikasi balita?

1.3 Tujuan

1. Mengetahui tingkat akurasi algoritma *Random Over Sampling* untuk mengatasi terjadinya ketidakseimbangan kelas.
2. Mengetahui hasil analisa penggunaan algoritma *Random Over Sampling* dalam klasifikasi balita.

1.4 Batasan Masalah

Untuk mempermudah penelitian, maka dibutuhkan batasan masalah untuk penerapannya. Adapun beberapa batasan masalahnya adalah :

1. Atribut yang digunakan adalah usia, berat badan, jenis kelamin, dan tinggi badan balita.

2. Klasifikasi dilakukan dengan menggunakan tools bantuan Anaconda dan menggunakan bahasa pemrograman *Python*.
3. Data yang digunakan adalah data dari Puskesmas Tegaldlimo tahun 2018.

1.5 Manfaat

Membantu mengatasi masalah imbalanced data gizi balita serta membantu proses klasifikasi gizi balita pada Puskesmas Tegaldlimo.

