

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining adalah salah satu teknik dan proses untuk menemukan suatu pola dan pengetahuan dari data yang berjumlah besar (Han dkk., 2011). Proses yang dilakukan untuk menemukan pengetahuan tersebut terdiri dari beberapa tahap yaitu : (1) *data cleaning*, yaitu untuk menghapus data yang tidak dipakai dan data yang tidak konsisten; (2) *data integration*, tahap untuk menggabungkan berbagai sumber data; (3) *data selection*, mengambil data yang relevan untuk dianalisis dari basisdata; (4) *data transformation*, yaitu proses perubahan data menjadi bentuk yang seharusnya untuk diolah dengan menganalisis ringkasan atau jumlah total agregasi; (5) *data mining*, yaitu sebuah proses penting di mana metode kecerdasan diterapkan dengan tujuan untuk mengolah pola dari data; (6) *pattern evaluation*, tahap ini untuk mengidentifikasi pola-pola menarik yang menjelaskan mengenai ukuran dasar pengetahuan yang ada; (7) *knowledge presentation*, yaitu visualisasi dan teknik representasi *knowledge* digunakan untuk menyajikan *knowledge* yang telah diolah untuk pengguna.

Klasifikasi adalah salah satu tugas dari data mining yang bertujuan untuk memprediksi label kategori benda yang tidak diketahui sebelumnya, dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau fitur (Parvin dkk., 2008). Menurut Han dkk (2011) klasifikasi adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting. Klasifikasi memprediksi kategori (*discrete, unordered*) ke dalam label *class*. Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau *class* data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa *if-then-rules, decision tree*, formula matematis atau *neural network*.

Salah satu teknik klasifikasi yang paling dasar dan sederhana adalah K-Nearest Neighbor (KNN) (Parvin dkk, 2008). KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Menurut Bhatia dan Vandana (2010) KNN memiliki beberapa keunggulan dan kelemahan, keunggulannya yaitu :1) pelatihan sangat cepat; 2) sederhana dan mudah dipelajari; 3) tahan terhadap data pelatihan yang memiliki derau; dan 4) efektif jika data pelatihan besar. Sedangkan kekurangan dari KNN adalah: 1) nilai k menjadi bias; 2) komputasi yang kompleks; 3) keterbatasan memori; dan 4) mudah tertipu dengan atribut yang tidak relevan. Dengan segala kelebihan dan kekurangan yang dimiliki oleh algoritma KNN, telah banyak permasalahan yang menggunakan algoritma ini sebagai algoritma untuk klasifikasi. Salah satu masalah yang dapat diselesaikan dengan algoritma ini adalah klasifikasi status gizi.

Masalah kekurangan dan kelebihan gizi pada orang dewasa (usia 18 tahun keatas) merupakan masa penting, karena selain mempunyai resiko penyakit-penyakit tertentu, juga dapat mempengaruhi produktifitas kerja. Oleh karena itu pemantauan keadaan tersebut perlu dilakukan oleh setiap orang secara berkesinambungan. Malnutrisi merupakan suatu keadaan patologis akibat kekurangan atau kelebihan zat gizi, baik secara relatif maupun absolut. Malnutrisi adalah istilah umum untuk suatu kondisi medis yang disebabkan oleh pemberian atau cara makan yang tidak tepat atau tidak mencukupi (Indra dan Wulandari, 2013). Pada dasarnya, konsumsi makanan bertujuan untuk mencapai status gizi optimal. Upaya penyediaan pangan agar tercapai status gizi optimal dapat dilakukan dengan mengonsumsi karbohidrat, lemak, protein, vitamin, dan mineral sesuai dengan angka kecukupan gizi dalam rangka proses metabolisme, transformasi, dan interaksinya dengan zat lain demi tercapainya keseimbangan energi tubuh (Indra dan Wulandari, 2013).

Status gizi dapat diukur secara antropometri. Penilaian antropometri secara umum digunakan untuk melihat ketidakseimbangan asupan energi dan protein.

Indeks antropometri yang sering digunakan, yaitu : berat badan terhadap umur (BB/U), tinggi badan terhadap umur (TB/U), dan berat badan terhadap tinggi badan (BB/TB) (Depkes RI, 2004). Standar rujukan yang digunakan untuk penentuan klasifikasi status gizi dengan antropometri menggunakan SK Menkes nomor 1995/Menkes/SK/XII/2010.

Indeks Massa Tubuh (IMT) atau *Body Mass Index* (BMI) merupakan alat atau cara yang sederhana untuk memantau status gizi orang dewasa, khususnya yang berkaitan dengan kekurangan dan kelebihan berat badan. Berat badan kurang dapat meningkatkan resiko terhadap penyakit infeksi, sedangkan berat badan lebih akan meningkatkan resiko terhadap penyakit degeneratif. Oleh karena itu, mempertahankan berat badan normal memungkinkan seseorang dapat mencapai usia harapan hidup yang lebih panjang. Pengukuran dengan IMT ini akan memberikan hasil apakah berat badan seseorang dinyatakan normal, kurus atau gemuk. IMT hanya dapat digunakan untuk orang dewasa berumur > 18 tahun dan tidak dapat diterapkan pada bayi, anak, remaja, ibu hamil, dan olahragawan. Untuk mengetahui nilai IMT, dapat dihitung dengan rumus berikut :

$$\text{IMT} = \frac{\text{Berat Badan (Kg)}}{\text{Tinggi Badan (m)}^2}$$

Nilai standar IMT untuk orang Asia adalah sebagai berikut.

Kurang dari 18,5 : Kurus

18,5 - 22,9 : Normal

Lebih dari 22,9 : digolongkan Obesitas.

Salah satu pada data mining yang cukup dikenal dengan masalah klasifikasi adalah KNN. Seperti telah disebutkan, salah satu kelemahan yang ada pada KNN adalah penentuan parameter k yang bias. Beberapa penelitian telah dilakukan untuk melakukan perbaikan pada KNN, baik dalam memperbaiki nilai akurasi KNN maupun optimasi nilai k pada KNN. Penelitian terkait peningkatan nilai akurasi KNN adalah Modified KNN (MKNN), di mana dalam MKNN ditambahkan perhitungan nilai validitas yang berguna untuk mengatasi permasalahan outlier dalam perhitungan nilai bobot pada KNN tradisional.

Namun MKNN juga memiliki kelemahan, yaitu nilai k yang bias dan komputasi yang kompleks (Parvin dkk, 2008). Sedangkan, penelitian terkait optimasi nilai k pada KNN adalah Genetic KNN (GKNN), selain dapat menentukan nilai k secara otomatis, GKNN dapat meningkatkan nilai akurasi dan dapat mengurangi kompleksitas (Suguna dan Thanushkodi, 2010). Berdasarkan latar belakang tersebut, maka diusulkan suatu penelitian untuk mengimplementasikan algoritma KNN dengan optimasi parameter k dengan algoritma genetika pada klasifikasi status gizi menggunakan pengukuran IMT.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, diperoleh rumusan masalah untuk melakukan penelitian mengenai penerapan algoritma K-Nearest Neighbor dan algoritma genetika untuk klasifikasi status gizi yaitu sebagai berikut :

1. Bagaimana merancang arsitektur algoritma genetika untuk optimasi dalam penentuan nilai k pada algoritma KNN.
2. Bagaimana melakukan implementasi untuk klasifikasi status gizi dengan melakukan optimasi menggunakan algoritma genetika dalam penentuan nilai k pada algoritma KNN.

1.3 Batasan Masalah

Agar penelitian menjadi fokus maka beberapa batasan yang akan diberikan adalah sebagai berikut :

1. Metode klasifikasi yang digunakan adalah algoritma K-Nearest Neighbor (KNN).
2. Algoritma genetika digunakan untuk mengoptimasi nilai k yang digunakan pada KNN.
3. Penentuan status gizi yang dilakukan adalah status gizi orang dewasa dengan metode pengukuran dengan Indeks Massa Tubuh (IMT).
4. Parameter yang digunakan dalam pengukuran dengan metode IMT ini adalah : berat badan (kg) dan tinggi badan (cm).

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma genetika dalam optimasi penentuan nilai k pada algoritma KNN.

1.5 Manfaat Penelitian

Penelitian mengenai penerapan algoritma K-Nearest Neighbor dan algoritma genetika untuk klasifikasi status gizi diharapkan dapat memberikan manfaat untuk mengatasi kelemahan dalam algoritma KNN dalam penentuan nilai k dan membantu untuk penentuan klasifikasi status gizi.

1.6 Target Luaran

Luaran penelitian yang diharapkan selain publikasi ilmiah pada jurnal lokal informatika maupun jurnal nasional, selain itu dapat di deseminasikan pada seminar ilmiah baik regional maupun nasional.