

BAB II

LANDASAN TEORI

2.1 Gizi

Gizi mempunyai peranan besar dalam daur kehidupan. Semua orang sepanjang kehidupan membutuhkan nutrien yang sama, namun dalam jumlah yang berbeda. Nutrien tertentu yang didapat dari makanan, melalui peranan fisiologis yang spesifik dan tidak tergantung pada nutrien yang lain, sangat dibutuhkan untuk hidup dan sehat (Kusharisupeni, 2007).

WHO (World Health Organization) mengartikan ilmu gizi sebagai ilmu yang mempelajari proses yang terjadi pada organisme hidup. Proses tersebut mencakup pengambilan dan pengolahan zat padat dan cair dari makanan yang diperlukan untuk memelihara kehidupan, pertumbuhan, berfungsinya organ tubuh dan menghasilkan energi.

Status gizi adalah ukuran keberhasilan dalam pemenuhan nutrisi untuk anak yang diindikasikan oleh berat badan dan tinggi badan anak. Status gizi juga didefinisikan sebagai status kesehatan yang dihasilkan oleh keseimbangan antara kebutuhan dan masukan nutrien. Penelitian status gizi merupakan pengukuran yang didasarkan pada data antropometri serta biokimia dan riwayat (Supriasa, 2002).

2.1.1 Antropometri Gizi

Antropometri adalah ilmu yang mempelajari berbagai ukuran tubuh manusia. Dalam bidang ilmu gizi digunakan untuk menilai status gizi. Ukuran yang sering digunakan adalah berat badan dan tinggi badan. Selain itu juga ukuran tubuh lainnya seperti lingkar kepala, lingkar lengan atas, lingkar perut, lingkar pinggul (Sadjaja, 2009).

Adapun beberapa syarat yang mendasari penggunaan antropometri ini adalah (Supriasa, 2002) :

1. Alatnya mudah didapat dan digunakan seperti dacin, pita, mikrotoa, dan alat pengukuran panjang bayi yang dapat dibuat sendiri.

2. Pengukuran dapat dilakukan berulang-ulang dengan mudah dan objektif.
3. Pengukuran bukan hanya dilakukan dengan tenaga khusus atau profesional, juga oleh tenaga lain yang setelah dilatih untuk itu.
4. Biaya relatif murah, karena alat mudah didapat dan tidak memerlukan bahan-bahan lainnya.
5. Hasilnya mudah disimpulkan karena mempunyai ambang batas dan baku rujukan yang sudah pasti.
6. Secara ilmiah diakui kebenarannya. Hampir semua negara menggunakan antropometri sebagai metode untuk mengukur status gizi, khususnya untuk penampisan (*Screening*) status gizi.

2.1.2 Parameter Dalam Antropometri

1. Umur

Faktor umur sangat penting dalam penentuan status gizi. Kesalahan penentuan umur dapat mengakibatkan interpretasi status gizi salah. Batasan umur yang digunakan :

- 1) Tahun umur penuh (*Completed Year*)

Contoh : 6 tahun 2 bulan, maka dihitung 6 tahun

5 tahun 11 bulan, dihitung 5 tahun

- 2) Bulan usia penuh (*Completed Month*)

Contoh : 3 bulan 7 hari, dihitung 3 bulan

2 bulan 26 hari, dihitung 2 bulan

2. Berat Badan (BB)

Merupakan ukuran antropometri terpenting dan paling sering digunakan pada bayi baru lahir. Selain itu pada masa bayi-balita berat badan dapat digunakan untuk melihat laju pertumbuhan fisik maupun status gizi.

3. Tinggi Badan (TB)

Tinggi badan merupakan antropometri yang menggambarkan keadaan pertumbuhan skeletal. Untuk bayi, pengukuran pertumbuhan linear adalah panjang badan, untuk anak yang lebih tua pengukurannya berdasarkan tinggi badan. Tinggi badan merupakan parameter paling penting bagi keadaan yang telah lalu dan keadaan sekarang. Jika umur tidak diketahui dengan tepat, serta dapat digunakan sebagai ukuran kedua yang penting, karena dengan menghubungkan BB terhadap TB faktor umur dapat dikesampingkan.

4. Lingkar Kepala

Lingkar kepala dihubungkan dengan ukuran otak dan tulang tengkorak. Ukuran otak meningkat secara cepat selama tahun pertama, kepala juga digunakan sebagai informasi tambahan dalam pengukuran umur.

2.1.3 Macam - Macam Status Gizi

Status gizi terbagi menjadi dua macam, yaitu status gizi normal dan malnutrisi (Suprisa, dkk. 2002) :

1. Status Gizi Normal

Keadaan tubuh yang mencerminkan keseimbangan antara konsumsi dan penggunaan gizi oleh tubuh (*adequate*).

2. Malnutrisi

Keadaan patologis akibat kekurangan atau kelebihan secara relatif maupun absolut satu atau lebih zat gizi. Ada empat bentuk :

- a) *Under nutrition*: kekurangan konsumsi pangan secara relatif atau absolut untuk periode tertentu.
- b) *Specific deficiency*: kekurangan zat gizi tertentu, misalnya kekurangan *iodium dan Fe* (zat besi).
- c) *Over nutrition*: kelebihan konsumsi pangan untuk periode tertentu.

- d) *Imbalance*: keadaan disproporsi zat gizi, misalnya tinggi kolesterol karena tidak imbangnya kadar LDL, HDL dan VDL.

2.2 Data Mining

Istilah lain yang memiliki makna yang sama dengan data mining adalah *knowledge discovery in database* (KDD). Definisi data mining menurut (Tan, 2006) adalah suatu proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai proses ekstraksi informasi baru yang diambil dari sekumpulan besar data yang dapat digunakan untuk membantu dalam pengambilan suatu keputusan. Sementara (Eko Prasetyo, 2014) menjelaskan pengertian data mining secara narataif yang mempunyai beberapa maksud sebagai berikut :

- a. Pencarian otomatis pola dalam basis data besar, menggunakan teknik komputasional campuran dari statistik, pembelajaran mesin, dan pengenalan pola.
- b. Pengekstrakan implisit non-trivial, yang sebelumnya belum diketahui secara potensial adalah informasi berguna dari data.
- c. Ilmu pengekstakan informasi yang berguna dari set data atau basis data besar.
- d. Eksplorasi otomatis atau semi otomatis dan analisi data dalam jumlah besar dengan tujuan untuk menemukan pola yang bermakna.
- e. Proses penentuan informasi otomatis dengan mengidentifikasi pola dan hubungan ‘tersembunyi’ dalam data.

Data mining dari pemasaran diatas dapat digambarkan sebagai proses ekstraksi sejumlah besar data yang tersimpan dalam basis data menjadi suatu informasi baru yang belum diketahui sebelumnya. Data mining sebagai suatu rangkaian proses dapat dibagi menjadi beberapa tahap sebagai berikut :

a. *Data Cleansing*

Data Cleansing merupakan tahap pembersihan data yang berguna untuk membuang data yang tidak konsisten dan *noise*. Ada kemungkinan terdapat data-data yang memiliki nilai kosong, salah pengetikan atau tidak valid saat

proses pengambilan data dari perusahaan maupun dari hasil penelitian. Hal ini akan mengurangi nilai mutu dan akurasi dari hasil data mining, sehingga perlu adanya proses pembersihan data untuk membuang data-data tersebut.

b. *Data Integration*

Data yang dibutuhkan untuk melakukan penelitian tidak selalu berasal dari suatu basis data, namun terkadang terdiri dari beberapa basis data yang terpisah atau berasal dari berkas teks, sehingga perlu dilakukan penggabungan data ke dalam satu basis data yang baru. Integrasi data harus dilakukan dengan cermat karena kesalahan dalam proses ini akan mengakibatkan hasil yang menyimpang dan dapat menyesatkan dalam pengambilan tindakan nantinya.

c. *Data Selection*

Seleksi data berguna untuk menyeleksi atau mengambil data-data yang sesuai dengan analisis.

d. *Data Transformation*

Transformasi dan pemiihan data ini bertujuan untuk menentukan kualitas dari hasil data mining, sehingga data diubah dalam bentuk yang tepat atau sesuai untuk diproses dalam data mining dengan cara meringkas atau melakukan operasi.

e. *Data Mining*

Tahap ini adalah proses pokok dimana metode kecerdasan diterapkan dalam ekstrasi informasi.

f. *Pattern Evoluation*

Evaluasi pola dilakukan untuk mengidentifikasi pola-pola menarik untuk merepresentasikan ke dalam *knowledge based*.

g. *Knowledge Presentation*

Representasi pengetahuan merupakan tahap untuk menunjukkan hasil yang diperoleh dengan cara visualisasi.

2.3 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu.

Tahapan dari klasifikasi dalam data mining menurut (Han dan Kamber, 2006) terdiri dari :

1. Pembangunan Model

Pada tahapan ini dibuat sebuah model untuk menyelesaikan masalah klasifikasi *class* atau *attribut* dalam data. Tahap ini merupakan fase pelatihan, dimana data latih dianalisis menggunakan algoritma klasifikasi, sehingga model pembelajaran direpresentasikan dalam bentuk aturan klasifikasi.

2. Penerapan Model

Pada tahapan ini model yang sudah dibangun sebelumnya digunakan untuk menentukan *class* atau *attribut* dari sebuah data baru yang *class* atau *attribut* belum diketahui sebelumnya. Tahap ini digunakan untuk memperkirakan keakuratan aturan klasifikasi terhadap data uji. Jika model dapat diterima, maka aturan data diterapkan klasifikasi data baru.

2.4 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. KNN termasuk dalam golongan *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. Nantinya kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat dengan vektornya.

Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan *attribut* dan *training sample*. *Classifier* tidak menggunakan model

apapun untuk disosokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah k obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari k obyek. Algoritma *K-Nearest Neighbor* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Metode *K-Nearest Neighbor* sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai jika kelas c merupakan klasifikasi yang paling banyak ditemukan jika k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *Euclidean Distance*.

Jarak *Euclidean* paling sering digunakan dalam menghitung jarak karena sangat cocok untuk menggunakan jarak terdekat (lurus) antara dua data. Jarak *euclidean* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek yang direpresentasikan sebagai berikut

$$D(a, b) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots\dots\dots(2.2)$$

Keterangan :

D (a,b) = Jarak *Euclidean* Data a dan Data b

X = Koordinat titik X

Y = Koordinat titik Y

Menurut Eko Prasetyo (Eko Prasetyo, 2012) dalam buku Konsep dan Aplikasi *Data Mining*, salah satu masalah yang dihadapi KNN adalah pemilihan nilai K yang tepat. Misalnya, diambil K bernilai 13, kelas 0 dimiliki oleh 7 tetangga yang jauh, sedangkan kelas 1 dimiliki oleh 6 tetangga yang lebih dekat. Hal ini mengakibatkan data uji tersebut akan terdistorsi sehingga

ikut tergabung dengan kelas 0. Hal ini karena setiap tetangga tersebut mempunyai bobot yang sama terhadap data uji, sedangkan K yang terlalu kecil bisa menyebabkan algoritma terlalu sensitif terhadap *noise*. Nilai k yang bagus dapat dipilih berdasarkan optimasi parameter, misalkan dengan *cross validation*. Kasus dimana klasifikasi diprediksi berdasarkan *training* data yang paling dekat (dengan kata lain, $k = 1$) disebut algoritma *nearest neighbors*.

Ketepatan algoritma KNN sangat dipengaruhi oleh ada atau tidaknya fitur – fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi lebih baik. Berikut ini adalah urutan proses pada algoritma *K-Nearest Neighbors* :

1. Menentukan parameter K (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak *euclidean* masing-masing obyek terhadap data sampel yang diberikan.
3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *euclidean* terkecil.
4. Mengumpulkan kategori Y (klasifikasi *nearest neighbors*).
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai *query instance* yang telah dihitung.

KNN merupakan teknik klasifikasi sederhana, tetapi mempunyai hasil kerja yang cukup bagus. Meskipun begitu, KNN juga mempunyai kelebihan dan kekurangan. Beberapa karakteristik KNN adalah sebagai berikut (Eko Prasetyo, 2012) :

1. KNN merupakan algoritma yang menggunakan seluruh data latih untuk melakukan proses klasifikasi (*complete storage*). Hal ini mengakibatkan proses prediksi yang sangat lama untuk data dalam jumlah yang sangat besar. Pendekatan lain adalah dengan menggunakan *mean* data dari setiap kelas tersebut. Hal ini memberi keuntungan kerja yang lebih cepat, tetapi hasilnya

kurang memuaskan karena model hanya membentuk *hyperplane* yang dibuat. Ada relasi pertukaran (*trade-off relation*) antara jumlah data latih pada biaya komputasi dengan kualitas batas keputusan (*desicion boundary*) yang dihasilkan.

2. Algoritma KNN tidak membedakan setiap fitur dengan suatu bobot seperti pada *Artificial Neural Network (ANN)* yang berusaha menekan fitur yang tidak mempunyai kontribusi terhadap klasifikasi menjadi 0 pada bagian bobot. KNN tidak memiliki bobot untuk masing- masing fitur.

3. Karena KNN masuk kategori *lazy learning* yang menyimpan sebagian atau semua data dan hampir tidak ada proses pelatihan, KNN sangat cepat dalam proses pelatihan (karena memang tidak ada), tetapi sangat lambat dalam proses prediksi.

4. Hal yang rumit adalah menentukan nilai K yang paling sesuai.

5. Karena KNN pada prinsipnya memilih tetangga terdekat, parameter jarak juga penting untuk dipertimbangkan sesuai dengan kasus datanya. *Euclidean* sangat cocok untuk menggunakan jarak terdekat

2.5 Menghitung Akurasi

Menurut (Olson, David, Yong Shi. 2008) tujuan evaluasi percobaan pada klasifikasi yaitu untuk mengukur keefektifan apakah sistem mengklasifikasi secara benar. Pada menghitung akurasi biasanya membutuhkan sebuah matriks yang disebut berupa *confusion matrix*. Berikut merupakan tabel *confusion matrix*.

Tabel 2.1 *Confusion Matrix*

Kelas Asli \ Kelas Hasil Prediksi	Positif	Negatif
	Positif	True Positive (TP)
Negatif	False Positive (FP)	True Negative (TN)

Keterangan **Tabel 2.1** :

1. TP adalah jumlah dari kelas positif (kelas yang mempunyai jumlah data lebih sedikit) yang benar diklasifikasikan.
2. FN adalah jumlah kelas positif yang salah diklasifikasikan kedalam kelas negatif.
3. FP adalah jumlah kelas negatif yang salah diklasifikasikan kedalam kelas positif.
4. TN adalah jumlah kelas negatif yang benar diklasifikasikan.

Berdasarkan data yang didapatkan dari *confusion matrix*, maka akan dihitung :

1. Akurasi

Akurasi digunakan untuk mengukur prosentase pengenalan secara keseluruhan dan dihitung sebagai jumlah data uji yang dikenali dengan benar, dibagi dengan jumlah seluruh data uji. Berikut rumus akur pada persamaan 2.3.

$$\begin{aligned} \text{Akurasi} &= \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \\ &= \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad \dots \dots \dots (2.3) \end{aligned}$$

2.6 Penelitian Sebelumnya

Penelitian sebelumnya oleh mahasiswa Universitas Muhammadiyah Gresik bernama Faridatul Choiriyah dengan judul **“Implementasi Metode Naïve Bayes Sebagai Penentu Status Gizi Balita (Study Kasus puskesmas Dukuh Kupang Surabaya)”**. Algoritma yang digunakan adalah naïve bayes. Data yang dijadikan inputan dalam sistem klasifikasi status gizi balita diperoleh dari data balita yang terdapat dipegawai bidan KIA puskesmas dukuh kupang surabaya, bulan september tahun 2014 sebanyak 131 balita. Atribut yang terdapat pada tabel mewakili fitur data yang digunakan meliputi jenis kelamin, usia, berat badan, tinggi badan, dan lingkar kepala. Jumlah data yang

digunakan sebanyak 131 *record* dengan kelas Baik dan Kurang. Penelitian ini diuji sebanyak 2 kali pengujian dengan rata-rata akurasi sebesar 92%.

Penelitian sebelumnya yang kedua dilakukan oleh seorang mahasiswa Universitas Muhammadiyah Gresik yang bernama Rizki Saputra dalam penelitian skripsinya yang berjudul **“Klasifikasi Penyakit Gagal Ginjal Kronis Menggunakan *K-Nearest Neighbor* (Studi Kasus : Rumah Sakit Muhammadiyah Gresik)”**. Pada penelitian ini peneliti menggunakan data sebanyak 140 data dan dilakukan sebanyak 4 kali pengujian, dimana 2 kali pengujian dengan melibatkan proses normalisasi, dan 2 kali pengujian tidak melibatkan proses normalisasi. Berdasarkan hasil pengujian sistem dengan menggunakan metode *K-Nearest Neighbor* dan metode akurasi *confusion matrix* Hasil akurasi dengan melibatkan proses normalisasi data pada skenario pertama diperoleh 93.33% dan skenario kedua diperoleh 87.50% dengan rata-rata akurasi yang didapat 90.42%. Sedangkan untuk hasil akurasi dengan tidak melibatkan proses normalisasi data pada skenario pertama diperoleh 92.50% dan skenario kedua diperoleh 91.67% dengan rata-rata akurasi yang didapat 92.09%.

Penelitian selanjutnya dilakukan oleh Nani Purwati dalam Indonesian Journal On Networking And Security Volume 5 No 14 Tahun 2016 yang berjudul **“Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Bb/U dan Bb/Tb Menggunakan Algoritma Backpropagation”**. Pada penelitian ini peneliti menggunakan dataset sebanyak 261 data balita yang diperoleh dari puskesmas meranti, variabel yang digunakan pada uji coba data tersebut adalah jenis kelamin, umur, jenis kelamin, berat badan, tinggi badan, status ekonomi. Dari hasil pengujian dataset menggunakan algoritma backpropagation didapatkan akurasi sebesar 89,50% untuk indeks Bb/U. Untuk kategori indeks Bb/Tb didapatkan akurasi sebesar 82,85%.