

# Optimasi Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Menggunakan Algoritma C4.5 Adaboost Classification

Wiwid Wahyudi

Progdi Desain Grafis STEKOM Semarang, [wi2d.wahyudi@gmail.com](mailto:wi2d.wahyudi@gmail.com)

Jl. Majapahit605, Semarang, telp/fax : 024-6717201-02

## ARTICLE INFO

Article history:

Received 30 November 2019

Received in revised form 2 Desember 2019

Accepted 10 Desember 2019

Available online 12 Desember 2019

## ABSTRACT

Infant health can be known one of them through the assessment of nutritional status. In general, Body Mass Index (BMI) has been used as a method for measuring the nutritional status of children. If there are two children who have same body weight and height, they may have different nutritional status. Whenever this occurs, the use of BMI for measuring the nutritional status shall be deemed less accurate. The anthropometry will be vital in measuring the nutritional status. The

guidelines for determining the nutritional status Anthropometry parameters are selected and recommended which includes an assessment of the age, weight, body length or height. This research aims to build a model of C4.5 adaboost so it can recognize patterns and be able to classify the nutritional status of children into five classes: normal, fat, very fat, thin and very thin. The variables used in this classification is Gender, Age (Months), Weight (kg) Height (cm). C4.5 (decision tree) Method has a good performance in dealing with the classification of nutritional status but the C4.5 has a weakness in the class imbalance. Adaboost is one of boosting methods that could reduce imbalances class by giving weight to the level of classification error which may alter the distribution of data. Experiments carried out by applying the adaboost method C4.5 to obtain optimal results and a good degree of accuracy. The experimental results obtained from C4.5 method show that accuracy is 89.53%, the error rate is 10.47%, while the results of C4.5 with adaboost show 90.23% accuracy and 9.77% error rate. It can be concluded in the classification of nutritional status of children with C4.5 and adaboost proven method to solve problems of class imbalance and improve the high accuracy and can reduce the level of classification error.

Keywords: Child Nutrition Status Classification, Anthropometrics Index, C4.5, Adaboost.

## I. PENDAHULUAN

Status gizi balita merupakan salah satu indikator yang menggambarkan tingkat kesejahteraan masyarakat. Status gizi balita dapat ditentukan melalui pemeriksaan laboratorium maupun secara antropometri. Pengukuran antropometri adalah pengukuran yang digunakan untuk menentukan keadaan gizi seseorang. Agar memperoleh hasil yang tepat, diberikan suatu patokan sebagai pedoman. Adapun pedoman antropometri bagi penentuan keadaan gizi merupakan parameter yang dipilih dan dianjurkan, yang meliputi penilaian terhadap usia dan berat badan, panjang badan, atau tinggi badan, dan lingkar lengan atas [1]. Indeks Massa Tubuh (IMT) direkomendasikan sebagai indikator yang baik untuk menentukan status gizi.

Jumlah data tentang gizi balita yang terus meningkat memerlukan beberapa metode untuk mengolah dan mengambil kesimpulan dan informasi dari data tersebut, yang diharapkan akan mampu meningkatkan kualitas data atau informasi serta efisiensi dan efektivitas pengolahan data. Sehingga pada

*Received November 23, 2019; Revised November 29, 2019; Accepted Desember 12, 2019*

akhirnya memudahkan atau membantu dalam pembuatan kebijakan terutama dalam menanggulangi masalah gizi pada balita.

Untuk mendukung hal ini dapat digunakan teknik data mining untuk menggali informasi yang berharga dari kumpulan informasi atau histori data gizi balita. Penelitian tentang gizi balita dengan menggunakan klasifikasi data mining sudah pernah dilakukan, baik komparasi beberapa klasifikasi data mining models ataupun improvement terhadap klasifikasi data mining.

Penelitian tentang status gizi balita telah banyak dilakukan dan telah dipublikasikan. Untuk melakukan penelitian ini, perlu ada kajian terhadap penelitian yang terkait sebelumnya. Agar dapat mengetahui metode apa saja yang digunakan, data seperti apa yang diproses, dan model seperti apa yang dihasilkan. Berikut beberapa ulasan tentang penelitian yang terkait. Penelitian yang dilakukan oleh [2] menggunakan metode C4.5. C4.5 merupakan algoritma klasifikasi yang sangat efektif dan efisien (kecepatan penggunaannya dan kesederhanaan untuk menangani dataset yang berisi banyak atribut dengan cara mudah dan sederhana). Datasetnya berjumlah 47 yang telah berumur 19 hingga 22 tahun. Menggunakan 5 variabel pengukuran yaitu ; tinggi badan , berat badan, jenis kelamin, lingkaran pergelangan dan lingkaran perut. Berdasarkan hasil pengujian terhadap semua data terdapat 38 yang sesuai dengan kelas yang diberikan dan 9 hasil yang tidak sesuai dengan hasil yang diberikan. Hasil dari pengujian data diperoleh nilai total kinerja sebesar 0,923 atau 93,2%

Sedangkan pada penelitian yang dilakukan oleh [3] menggunakan metode JST. JST merupakan sebuah metode yang bersifat terawasi dimana diperlukan data pelatihan dalam proses pembelajarannya. Datasetnya berjumlah 130 yang berumur dibawah lima tahun (0-59 bulan). Menggunakan 5 variabel pengukuran yaitu ; jenis kelamin , umur, berat badan, penyakit penyerta dan status ekonomi. Berdasarkan hasil pengujian terhadap semua data, diperoleh nilai optimum dengan parameter lapisan tersembunyi 8 neuron, learning rate 0,9, maksimum epoch 2000 dan momentum 0,3. Tingkat kebenaran klasifikasi sebesar 93,8 %.

Dari beberapa penelitian tentang klasifikasi status gizi balita seperti diatas. Dapat disimpulkan metode JST memiliki akurasi tertinggi dalam klasifikasi status gizi balita, kelebihan JST terdapat pada prediksi non-linier, kuat pada paralel processing dan kemampuan untuk mentoleransi kesalahan, tetapi memiliki kelemahan pada perlunya data training yang besar, over-fitting, rendahnya konvergensi, dan sifatnya yang local optimum [4]. C4.5 dapat memecahkan masalah JST yaitu menangani over-fitting, menangani atribut yang continue, memilih yang tepat untuk attribute selection, menangani training data dengan nilai atribut yang hilang, dan meningkatkan efisiensi komputasi [5], pada umumnya tingkat keberhasilan dari C4.5 difokuskan pada dataset yang relatif seimbang [6]. Maka dari itu C4.5 dengan kasus ketidakseimbangan kelas diperlukan metode yang dapat mengatasi masalah tersebut untuk meningkatkan kinerja klasifikasi C4.5 agar dapat menghasilkan kinerja yang lebih baik.

Ada tiga pendekatan untuk menangani dataset tidak seimbang, yaitu pendekatan pada level data, level algoritmik, dan menggabungkan atau memasang metode ensemble. Sedangkan pada pendekatan metode ensemble ada dua algoritma ensemble-learning yang paling populer, yaitu boosting dan bagging [7]. Pada pendekatan algoritma dan ensemble memiliki tujuan yang sama, yaitu memperbaiki algoritma pengklasifikasi tanpa mengubah data, sehingga dapat dianggap ada 2 pendekatan saja, yaitu pendekatan level data dan pendekatan level algoritma [8]. Dengan membagi menjadi 2 pendekatan dapat mempermudah fokus objek perbaikan, pendekatan level data difokuskan pada pengolahan awal data, sedangkan pendekatan level algoritma difokuskan pada perbaikan algoritma atau menggabungkan ensemble.

Secara umum algoritma boosting lebih baik dari pada bagging, tetapi tidak merata baik. Boosting telah menunjukkan dapat meningkatkan kinerja pengklasifikasi dalam banyak situasi, termasuk ketika data tidak seimbang [9]. Adaboost secara teoritis dapat secara signifikan digunakan untuk mengurangi kesalahan dari beberapa algoritma pembelajaran secara konsisten menghasilkan kinerja pengklasifikasi yang lebih baik. Algoritma boosting adalah algoritma iteratif yang memberikan bobot yang berbeda pada distribusi data pada setiap iterasi. Setiap iterasi boosting menambahkan bobot pada contoh-contoh kesalahan klasifikasi dan

menurunkan bobot pada contoh klasifikasi yang benar, sehingga secara efektif dapat merubah distribusi pada data training [10].

Pada penelitian ini berbeda dengan penelitian yang telah ada, untuk klasifikasi menggunakan C4.5, sedangkan untuk mengatasi masalah ketidakseimbangan kelas menggunakan metode AdaBost untuk menyelesaikan masalah data atau kelas yang sifatnya tidak seimbang dan nantinya hasilnya akan dibandingkan dengan metode JST.

## II. RUMUSAN MASALAH

Dari latar belakang di atas dapat diketahui rumusan masalahnya, yaitu belum diketahuinya teknik terbaik untuk penanganan imbalance data dan error klasifikasi, untuk mengoptimalkan akurasi Klasifikasi Status Gizi Balita berdasarkan Indeks Antropometri.

## III. TUJUAN PENELITIAN

Tujuan penelitian ini adalah Untuk meningkatkan akurasi pada algoritma C4,5 dalam Klasifikasi Status Gizi Balita berdasar Indeks Antropometri, dengan penelitian sebelumnya atau algoritma JST.

## III. KAJIAN PUSTAKA

### A. Data Mining

Data Mining (DM) adalah inti dari proses Knowledge Discovery in Database (KDD), melibatkan algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang tidak diketahui sebelumnya. Model ini digunakan untuk memahami fenomena dari data, analisis dan prediksi. Aksesibilitas dan banyaknya data membuat Knowledge Discovery dan Data Mining menjadi masalah yang cukup penting dan dibutuhkan.

Menurut Larose (2007) berdasarkan tugasnya, data mining dibagi menjadi 6 kelompok [7], yaitu:

1. Deskripsi  
Terkadang peneliti dan analisis secara sederhana mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data..
2. Estimasi  
Estimasi hampir sama dengan klasifikasi, tetapi variable target estimasi lebih ke arah numerik daripada kategori.
3. Prediksi  
Prediksi hampir sama dengan estimasi dan klasifikasi, tetapi dalam prediksi akan menghasilkan nilai.
4. Klasifikasi  
Dalam klasifikasi terdapat target variabel kategori.
5. Cluster  
Mengelompokkan *record*, pengamatan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Tujuan dari algoritma *cluster* adalah dengan memecahkan setiap data dalam *dataset* menjadi kelompok kelompok yang homogen. Kelompok data ini biasanya disebut sebagai *cluster*. Setiap *cluster* yang terbentuk akan terdiri dari data yang sejenis dan berbeda dengan data pada *cluster* lainnya. Pengelompokkan ini sama dengan cara kerja otak manusia, dimana ilmu pengetahuan dikelompokkan dalam setiap bidangnya. Dengan adanya pengelompokkan, data yang dapat diolah dengan lebih spesifik sesuai dengan tujuan penelitian. Pemecahan data kedalam *cluster* data juga diterapkan pada tahap pengolahan awal data dalam proses data mining, sehingga dapat diterapkan metode data mining data mining kedalam setiap *cluster* data. Proses *clustering* juga dapat mengurangi jumlah ataupun dimensi data yang diolah.

### B. Status Gizi

Status gizi adalah suatu ukuran mengenai kondisi tubuh seseorang yang dapat dilihat dari makanan yang dikonsumsi dan penggunaan zat-zat gizi di dalam tubuh. Status gizi dibagi menjadi tiga kategori, yaitu status gizi kurang, gizi normal, dan gizi lebih.

### C. Indeks Antropometri

Indeks antropometri adalah pengukuran dari beberapa parameter. Indeks antropometri bisa merupakan rasio dari satu pengukuran terhadap satu atau lebih pengukuran atau yang dihubungkan dengan umur dan tingkat gizi. Salah satu contoh dari indeks antropometri adalah Indeks Massa Tubuh (IMT) atau yang disebut dengan Body Mass Index. Antropometri gizi berhubungan dengan berbagai macam pengukuran dimensi tubuh dan komposisi tubuh dari berbagai tingkat umur dan tingkat gizi. Dalam pemakaian untuk penilaian status gizi, antropometri disajikan dalam bentuk indeks yang dikaitkan dengan variabel lain, yaitu umur, berat badan, dan tinggi badan.

### D. Algoritma C4.5

Algoritma C4.5 adalah hasil dari pengembangan algoritma ID3 (Iterative Dichotomiser) yang dikembangkan oleh [18]. Algoritma ini digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar. Sebelumnya diakhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, seorang peneliti dibidang machine learning, membuat sebuah algoritma decision tree yang dikenal dengan ID3 (Iterative Dichotomiser). Kalau ID3, pengukuran seleksi atribut ditentukan oleh Information Gain, sedangkan C4.5 pengukuran seleksi atribut ditentukan oleh GainRatio.

### E. Teknologi Biometrika

Nugroho (2008) menyatakan bahwa teknologi biometrika merupakan sebuah teknologi baru yang memiliki fungsi utama untuk mengenali manusia melalui sidik jari, mata, wajah, atau bagian tubuh yang lain. Biometrika berasal dari kata bios, yang berarti kehidupan, dan metron, yang berarti ukuran. Biometrika merupakan teknologi untuk mengenali seseorang secara unik. Dengan didukung oleh faktor biaya penggunaan teknologi yang semakin terjangkau dan fleksibilitas teknologi ini, teknologi ini dirasa akan dapat menggusur penggunaan kata sandi (password) ataupun kartu, misal kartu kredit sebagai alat otentikasi maupun identifikasi. Teknologi identifikasi ini juga disebut dengan teknologi "what you are".

### F. Algoritma Esemble Adaboost

Adaboost adalah algoritma yang ide dasarnya adalah untuk memilih dan menggabungkan sekelompok pengklasifikasi lemah untuk membentuk klasifikasi yang kuat. Parameter yang telah dipelajari tersebut akan digunakan sebagai penelitian untuk dapat meningkatkan akurasi pengklasifikasi dasar C4.5 melalui iterasi yang tepat. [14] menjelaskan teknik pembobotan pada algoritma Adaboost sebagai berikut :

Inisialisasi bobot data  $\{ W_n \}$  dengan  $W_n^{(m)}$  untuk  $n = 1, 2, \dots, N$  For  $m = 1, \dots, M$

a. Training  $Y_m(x)$  dengan meminimalkan fungsi kesalahan (*error function*) sebagai berikut :

$$J_m = \sum_{n=1}^N W_n^{(m)} I(y_m(x_n) \neq t_n)$$

b. Evaluasi Kesalahan

$$\varepsilon_m = \frac{\sum_{n=1}^N W_n^{(m)} I(y_m(x_n) \neq t_n)}{\sum_{n=1}^N W_n^{(m)}}$$

c. Dan kemudian digunakan evaluasi

$$\alpha_m = \ln \left\{ \frac{1 - \varepsilon_m}{\varepsilon_m} \right\}$$

d. Memperbaiki (Update) bobot data

$$W_n^{(m+1)} = W_n^{(m)} \exp(\alpha_m I(y_m(x_n) \neq t_n))$$

e. Membuat prediksi menggunakan model terakhir sebagai berikut:

$$Y_m(x) = \text{sign} \left( \sum_{m=1}^M \alpha_m y_m(x) \right)$$

## IV. PEMBAHASAN PENELITIAN

Dalam eksperimen ini, Awal Optimasi AdaBoost pada Algoritma C4.5 untuk Klasifikasi Status Gizi Balita berdasarkan Indeks Antropometri yaitu Penerapan Algoritma C4.5 dengan JST dan C4.5+AdaBoost pada Tool Rapid Miner kemudian didapatkan hasil yang bisa dilihat pada tabel dibawah ini :

Pengukuran C4.5 dengan Confusion Matrix akan menampilkan Class Recall dan Class Precision seperti Tabel 4.1

	True Normal	True Kurus	True Gemuk	True S. Gemuk	True S. Kurus	Class Precision
Pred. Normal	365	12	10	4	5	92,17%
Pred. Kurus	6	9	0	0	3	50,00%
Pred. Gemuk	0	0	6	2	0	75,00%
Pred. Sangat Gemuk	0	0	0	3	0	100,00%
Pred. Sangat Kurus	1	2	0	0	2	40,00%
Class Recall	98,12%	39,13%	37,50%	33,33%	20,00%	

Tabel 4.1 Hasil Akurasi C4.5 Untuk Klasifikasi Status Gizi Balita

Hasil C4.5+Adaboost dengan Confusion Matrix akan menampilkan Class Recall dan Precision seperti Tabel 4.2

	True Normal	True Kurus	True Gemuk	True S. Gemuk	True S. Kurus	Class Precision
Pred. Normal	368	10	11	4	7	92,00%
Pred. Kurus	3	12	0	0	3	66,67%
Pred. Gemuk	0	0	5	2	0	71,43%
Pred. Sangat Gemuk	0	0	0	3	0	100,00%
Pred. Sangat Kurus	1	1	0	0	0	00,00%
Class Recall	98,92%	52,17%	31,25%	33,33%	00,00%	

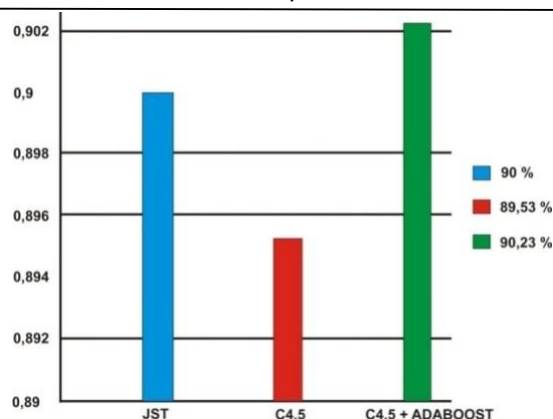
Tabel 4.2 Hasil Akurasi C4.5 +Adaboost Untuk Klasifikasi Status Gizi Balita

Pengukuran JST dengan Confusion Matrix akan menampilkan Class Recall dan Class Precision seperti Tabel 4.3

	True Normal	True Kurus	True Gemuk	True S. Gemuk	True S. Kurus	Class Precision
Pred. Normal	367	8	10	4	6	92,91%
Pred. Kurus	4	13	0	0	3	65,00%
Pred. Gemuk	0	0	4	2	0	66,67%
Pred. Sangat Gemuk	0	0	2	2	0	50,00%
Pred. Sangat Kurus	1	2	0	1	1	20,00%
Class Recall	98,66%	56,52%	25,00%	22,22%	10,00%	

Tabel 4.3 Hasil Akurasi JST Untuk Klasifikasi Status Gizi Balita

Dan pengujian akurasi yang dilakukan di dapatkan hasil seperti pada gambar 4.1 di bawah ini:



Gambar 4.1 Perbandingan C4.5 , C4.5 +AdaBoost dan JST

Pada hasil keseluruhan terlihat nilai akurasi untuk model algoritma C4.5 sebesar 89.53% dan nilai akurasi untuk model C4.5 +AdaBoost sebesar 90.23% dengan selisih akurasi 1%. Akurasi Metode C4.5+Adaboost menghasilkan tingkat akurasi yang lebih baik dibandingkan dengan menggunakan C4.5 versi standar. Hal tersebut seperti dikatakan oleh [9] bahwa adaboost dapat memberikan keuntungan, lebih efektif dan akurat dalam pengklasifikasian. Terbukti bahwa hasil pengujian algoritma C4.5 + AdaBoost memiliki nilai akurasi yang lebih baik dibandingkan dengan algoritma C4.5 tunggal dan algoritma JST.

## V. KESIMPULAN DAN SARAN

### A. Kesimpulan

Dalam penelitian ini, kami mengadopsi teknik peneliti sebelumnya untuk penambahan teknik Ensemble yaitu AdaBoost yang diterapkan pada algoritma C4.5 yang diimplementasikan dengan data status gizi balita menghasilkan akurasi yang meningkat. Sebelumnya penelitian dilakukan menggunakan algoritma C4.5 tanpa teknik adaboost, data mengalami ketidakseimbangan data pada saat data dihadapkan dengan peraturan antropometri. Dimana pada umur tertentu berat badan bayi harus sesuai umurnya, sedangkan pada saat perhitungan terkadang berat badan bayi tidak sesuai umur yang ada di real data sehingga klasifikasi hanya mencapai 89,53 %. Setelah data di beri teknik adaboost, ketidakseimbangan data bisa diatasi dimana data sesuai dengan standar peraturan antropometri dan menghasilkan akurasi klasifikasi yang meningkat sebesar 90,23%. Untuk itu pada akhir penelitian disimpulkan bahwa algoritma *C4.5 AdaBoost* ada peningkatan akurasi klasifikasi status gizi balita dibanding *C4.5* tunggal.

Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa C.45 berbasis AdaBoost mampu memberikan peningkatan akurasi klasifikasi untuk data status gizi balita dan mampu melampaui akurasi dari penelitian sebelumnya yaitu menggunakan metode JST. Sehingga Algoritma ini dapat digunakan sebagai salah satu metode untuk klasifikasi status gizi balita berdasarkan hasil pengukuran antropometri dan model sistem yang dibangun memiliki kinerja yang baik dari hasil penelitian sebelumnya.

### B. Saran

- 1) Dalam penelitian ini dilakukan menggunakan algoritma C4.5 berbasis AdaBoost untuk status gizi balita. Kedepan mungkin kita bisa mencobakan metode ini dengan dataset yang berbeda tetapi masih tentang status gizi. Misalkan dataset status gizi orang dewasa.
- 2) Melakukan uji komparasi baik pada algoritmanya maupun metode representasinya yang mungkin dapat diterapkan, sehingga dapat meningkatkan akurasi klasifikasi status gizi balita. Misalnya dengan menggunakan metode NBC Bagging atau JST AdaBoost.

**VI. DAFTAR PUSTAKA**

- [1] Kementerian Kesehatan RI Direktorat Jendral Bina Gizi dan Kesehatan Ibu dan Anak, 2011; *"Keputusan Menteri Kesehatan Republik Indonesia Nomor:1995/MENKES/SK/XII/2010 Tentang Standar Antropometri Penilaian Status Gizi Anak"*.
- [2] Kusumadewi, Sri, 2009; *"Klasifikasi Status Gizi Menggunakan Naïve Bayesian Classification"*, Yogyakarta: Jurusan Teknik Informatika, Universitas Islam Indonesia.
- [3] Reni Anggraeni, Aviarani Indrarti, 2010; *"Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri (BB/U) Menggunakan Jaringan Saraf Tiruan"*, Yogyakarta: Jurusan Teknik Informatika, Universitas Gunadarma.
- [4] Capparuccia, R., Leone, R. D., & Marchitto. E., 2007; *"Integrating Support vector machines and neural network, Neural Network, 590-597"*.
- [5] Quinlan, J.R, 1993; *"C4.5: Programs for Machine Learning"*, Morgan Kaufmann.
- [6] Yap, B. W., Rani, K. A., Rahman, H. A., Fong, S., Khairudin, Z., & Abdullah, N. N, 2014; *"An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets"*, Proceedings of the First International Conference on Advanced Data and Information Engineering.
- [7] Peng, Y., & Yao, J, 2010; *"AdaOUBoost: Adaptive Over-sampling and Under-sampling to Boost the Concept Learning in Large Scale Imbalanced Data Sets"*, Philadelphia, Pennsylvania, USA: Proceedings of the international conference on Multimedia information retrieval (pp. 111-118).
- [8] Seiffert, C., Khoshgoftaar, T. M., Hulse, J. V., & Napolitano, A, 2008; *"Resampling or Reweighting: A Comparison of Boosting Implementations"*, 20th IEEE International Conference on Tool with Artificial Intelligence, 445-451.
- [9] Kotsiantis, S. B., and Pintelas, P. E, 2009; *"Selective costing ensemble for handling imbalanced data sets"*, International Journal of Hybrid Intelligent Systems, 123-133.