

Penerapan Algoritma C4.5 Untuk Klasifikasi Tumbuh Kembang Balita

Muhamad Rizky Abadi¹, Sofi Defiyanti², Nina Sulistiowati³

^{1,2,3}Universitas Singaperbangsa Karawang,
Jl. H.S Ronggowaluyo, Telukjambe, Karawang
Email: 1441177004045@student.unsika.ac.id¹, sofi.defiyanti@staff.unsika.ac.id²,
nina.sulistio@staff.unsika.ac.id³

Abstrak. Sejak lahir sampai dengan usia lima tahun, anak seharusnya diukur tinggi dan berat badannya secara teratur untuk mengetahui pertumbuhannya. Cara ini dapat membantu untuk mengetahui lebih awal tentang gangguan pertumbuhan, sehingga segera dapat diambil tindakan tepat secepat mungkin. Hasil pengukuran dapat mengetahui apakah seorang anak terlalu cepat bertambah pertumbuhannya dibandingkan usianya atau tidak bertambah pertumbuhannya. Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan *training data* yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3. Penelitian ini menggunakan algoritma C4.5 untuk klasifikasi tumbuh kembang balita yang akan menghasilkan Indeks Massa Tubuh menurut Umur (IMT/U) berupa kategori status gizi sangat kurus, kurus, normal, gemuk. Metode yang digunakan KDD (*Knowledge Discovery in Database*) menggunakan *tools* WEKA 3.8.2. Hasil pengolahan data akan diuji dengan *test options* menggunakan *10-fold cross validation* yang akan menghasilkan tingkat akurasi, *precision*, *recall*, *f-measure*. Hasil penelitian ini menunjukkan bahwa algoritma C4.5 dengan *test options* menggunakan *10-fold cross validation* memiliki hasil akurasi tertinggi yang didapat sebesar 90,4901%, *precision* 0,897, *recall* 0,905 dan *f-measure* sebesar 0,900. Hasil evaluasi menghasilkan model pohon keputusan tumbuh kembang balita

Kata kunci: *algoritma c4.5, data mining, klasifikasi, pohon keputusan, tumbuh kembang balita*

1. Pendahuluan

Tumbuh kembang balita dapat dilihat dari tinggi dan berat badannya, tumbuh kembang tersebut dapat menjadi tolak ukur pertumbuhan fisik balita yang ideal. Tumbuh kembang terjadi begitu cepat saat masa balita, maka dari itu memantau tumbuh kembang balita setiap bulan di Posyandu atau dokter sangat penting dilakukan. Mengenali akan tanda balita yang ideal bagi orang tua merupakan hal yang harus diperhatikan dan tidak boleh diremehkan, karena dengan mengetahui akan ciri-ciri balita ideal akan bisa membantu mencegah tumbuh kembang balita yang buruk[1].

Data Mining didefinisikan sebagai proses menemukan dan membaca pola dalam data sehingga dapat digunakan sebagai alat untuk membantu menemukan informasi yang sangat penting dari data tersebut[2]. Algoritma C4.5 merupakan teknik yang digunakan untuk melakukan proses klasifikasi, clustering, serta proses prediksi. Algoritma C4.5 menggunakan teknik untuk membagi ruang pencarian masalah menjadi himpunan masalah. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan[3].

Berdasarkan permasalahan yang dibahas, bagaimana menerapkan algoritma C4.5 dalam mengklasifikasikan tumbuh kembang balita dan mengevaluasi model yang dihasilkan oleh algoritma C4.5 berdasarkan akurasi, *precision*, *recall* dan *f-measure*. Penelitian ini menggunakan algoritma C4.5 dalam mengklasifikasikan tumbuh kembang balita untuk pencegahan tumbuh kembang balita yang buruk.

2. Landasan Teori

2.1 Data Mining

Arti dari data mining adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa memperbaiki pengambilan keputusan di masa depan[4]. Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap[5]. Tahap-tahap data mining yaitu pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), transformasi data (*data transformation*), *data mining*, *evaluation*, *knowledge*.

2.2 Klasifikasi

Proses yang terjadi dalam klasifikasi adalah proses penggolongan data ke dalam *variable* target atau *variable* tujuan dengan membangun sebuah model penyelesaian dengan memperhatikan atribut yang paling berpengaruh[6].

2.3 Algoritma C4.5

Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3[7]. Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan-aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap node

dan syarat percabangan akan membentuk suatu kondisi atau suatu if, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk[8].

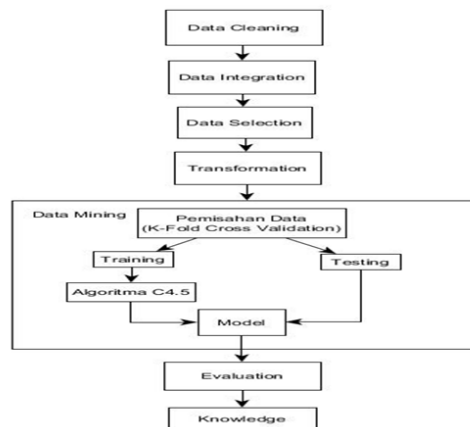
3. Metodologi Penelitian

3.1 Dataset

Data yang didapat adalah data keseluruhan hasil pemeriksaan rutin bulanan pada Posyandu Nusa Indah Palumbonsari Kabupaten Karawang, data didapat berdasarkan observasi. Data yang digunakan yaitu data pada tahun 2014 sampai dengan tahun 2016. Jumlah keseluruhan data sebanyak 1367 data dengan 5 atribut yaitu Umur, Jenis Kelamin, Berat Badan, Tinggi Badan dan Indeks Massa Tubuh (IMT).

3.2 Metode

Untuk melakukan komparasi terhadap algoritma C4.5 pada penelitian ini, digunakan tahapan-tahapan dari metode KDD (*Knowledges Discovery in Databases*). Berikut alur penelitian penerapan algoritma C4.5 untuk klasifikasi tumbuh kembang balita:



Gambar 1 Alur Penelitian

4. Hasil dan Pembahasan

4.1 *Data Cleaning*

Tahap ini dilakukannya pembersihan data yang akan diinputkan dari duplikasi data, *missing value*, memperbaiki kesalahan pada data agar data yang akan diinputkan bisa digunakan dengan baik.

4.2 *Data Integration*

Data pertumbuhan balita ini berasal dari satu sumber yaitu Posyandu Nusa Indah, data yang didapatkan dari tahun 2014 sampai 2016. Data keseluruhan sebanyak 1367 data. Data keseluruhan sudah terintegrasi sehingga tidak dilakukan proses *Data Integration*.

4.3 *Data Selection*

Dataset yang dipilih berjumlah 5 atribut yaitu Umur, Jenis Kelamin, Berat Badan, Tinggi Badan dan Indeks Massa Tubuh (IMT).

4.4 *Transformation*

Dataset Class Target yang digunakan adalah data Indeks Massa Tubuh (IMT) yang dikategorikan berdasarkan Status Gizi Sangat Kurus, Kurus, Normal dan Gemuk.

Data Indeks Massa Tubuh (IMT) didapat dari perhitungan berat badan dan tinggi badan. Indeks massa tubuh berupa numerik diperoleh dengan rumus :

$$IMT = \frac{\text{Berat Badan (Kg)}}{\text{Tinggi Badan}^2 \text{ (Meter)}} \quad (1)$$

Nilai *Z-Score* pada data pertumbuhan balita digunakan pada proses persiapan transformasi data.

$$Z - Score = \frac{\text{Nilai Individu Subjek} - \text{Nilai Median Baku Rujukan}}{\text{Nilai Simpangan Baku Rujukan}} \quad (2)$$

Nilai *Z-Score* Indeks Massa Tubuh (IMT) menjadi tipe data nominal. Transformasi ini dilakukan sesuai **Tabel 1**.

Tabel 1 Kategori & Ambang Batas Pertumbuhan Balita

Indeks	Kategori Status Gizi	Ambang Batas (Z-Score)
Indeks Massa Tubuh Menurut	Sangat Kurus	< -3 SD

Umur (IMT/U)	Kurus	-3 SD sampai dengan < -2 SD
	Normal	-2 SD sampai dengan 2 SD
	Gemuk	> 2 SD

4.5 Data Mining

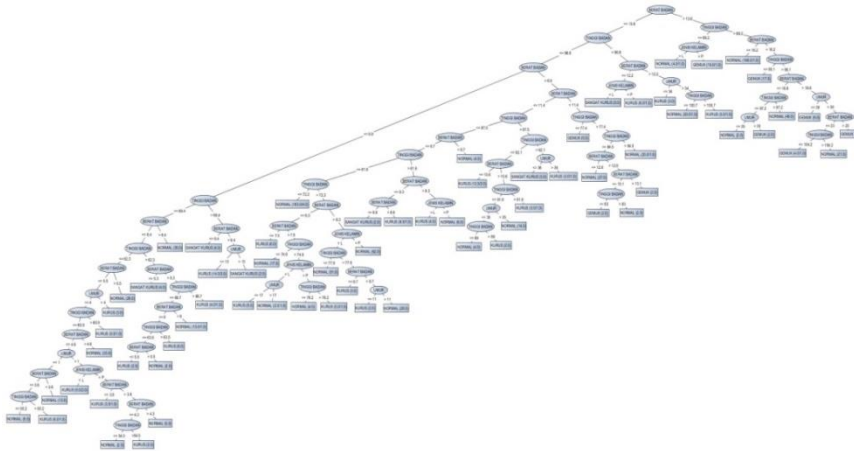
Untuk skenario yang digunakan algoritma C4.5 adalah *10-fold cross validation*. Hasil yang diharapkan adalah nilai akurasi yang baik untuk model klasifikasi tumbuh kembang balita berdasarkan indeks massa tubuh. Dalam melakukan evaluasi akurasi data digunakan *confussion matrix*.

4.6 Evaluation

Hasil dari algoritma C4.5 dengan skenario pengujian *10-fold cross validation* yang menghasilkan nilai akurasi, *precision*, *recall*, dan *f-measure* dapat dilihat pada **Tabel 2**

Algoritma C4.5	Akurasi	Precision	Recall	F-Measure
<i>10-fold cross validation</i>	90,4901%	0,897	0,905	0,900

Berdasarkan hasil evaluasi pada pemodelan, diketahui bahwa algoritma C4.5 dengan skenario pengujian *10-fold cross validation* menghasilkan model klasifikasi tumbuh kembang balita berdasarkan indeks massa tubuh. Tahap evaluasi menghasilkan nilai akurasi, *precision*, *recall*, *f-measure*. Oleh karena itu tahap evaluasi telah sesuai dengan tahap data mining dalam mengklasifikasikan tumbuh kembang balita.



Gambar 2 Pohon Keputusan Tumbuh Kembang Balita

4.7 Knowledge

Hasil dari pohon keputusan model klasifikasi tumbuh kembang balita berdasarkan indeks massa tubuh diterapkan kedalam bahasa pemrograman *visual basic*, program aplikasi dibuat menggunakan *microsoft visual basic 2010*. *Input* program aplikasi berupa umur, jenis kelamin, berat badan, tinggi badan dan *output* berupa hasil indikator pertumbuhan balita berdasarkan indeks massa tubuh yang terdiri dari sangat kurus, kurus, normal dan gemuk.



Gambar 3 Tampilan Aplikasi Tumbuh Kembang Balita

Gambar 4 Tampilan Hasil Aplikasi Tumbuh Kembang Balita

4.8 Pembahasan

Algoritma C4.5 dalam klasifikasi tumbuh kembang balita dapat diterapkan. Penerapan algoritma C4.5 dengan metode KDD (*Knowledge Discovery in Database*) yang terdiri dari beberapa tahapan yaitu *data cleaning*, *data integration*, *data selection*, *transformation*, *data mining*, *evaluation*, *knowledge*. Menggunakan *tools* WEKA versi 3.8.2 pengujian *10-fold cross validation* menghasilkan akurasi sebesar 90,4901%, *precision* sebesar 0,897, *recall* 0,905, *f-measure* sebesar 0,900 dan dari hasil evaluasi dapat diketahui bahwa algoritma C4.5 dengan pengujian *10-fold cross validation* menghasilkan model klasifikasi tumbuh kembang balita berdasarkan indeks massa tubuh.

5. Kesimpulan

Penerapan algoritma C4.5 dengan metode KDD (*Knowledge Discovery in Database*) dengan menggunakan *tools* WEKA 3.8.2 merupakan salah satu strategi dalam menyelesaikan penelitian ini. Metode KDD terdiri dari beberapa tahapan yaitu *data cleaning*, *data integration*, *data selection*, *transformation*, *data mining*, *evaluation*, *knowledge*. Pengolahan data diuji dengan *test options* menggunakan *10-fold cross validation*.

Hasil Algoritma C4.5 klasifikasi tumbuh kembang balita akurasi tertinggi yang didapat sebesar 90,4901%, *precision* 0,897, *recall* 0,905 dan *f-measure* sebesar 0,900.

Penelitian selanjutnya diharapkan dapat menggunakan pengujian *test options* yang berbeda agar dapat mengetahui apakah hasil evaluasi yang berbeda dan menemukan kesimpulan yang baru. Diharapkan menggunakan algoritma lain yang lebih baik dalam mengklasifikasikan tumbuh kembang balita.

Mengaplikasikan hasil pemodelan kedalam bahasa pemrograman berbeda yang lebih memudahkan dalam mengklasifikasikan tumbuh kembang balita.

6. Referensi

- [1] K. K. RI, *Profil Kesehatan Indonesia Tahun 2016*. 2017.
- [2] N. Purwati, C. Agustina, and G. B. S, "Komparasi Algoritma C.45 Dan Backpropagation Untuk Klasifikasi Status Gizi Balita Berdasarkan Indeks Antropometri Bb / U Dan BB / PB," *Sentra Penelit. Eng. dan Edukasi*, vol. 9, no. 3, pp. 26–33, 2017.
- [3] A. Putra, "Solusi Prediksi Mahasiswa Drop Out Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Bina Darma," *J. SIMETRIS*, vol. 8, no. 1, pp. 177–184, 2017.
- [4] R. Rahwali, S. Hansun, and Y. W. Wiratama, "Prediksi Kelayakan Masuk Jurusan IPA Siswa Sekolah Menengah Atas Menggunakan C4.5 (Studi Kasus: SMA Tarakanita Gading Serpong)," *J. Telemat.*, vol. 10, no. 2, pp. 12–26, 2017.
- [5] G. C. Sutradana, "Penerapan Data Mining Untuk Analisis Pengaruh Lama Studi Mahasiswa Teknik Informatika UIN Sunan Kalijaga Yogyakarta Menggunakan Metode Apriori," *JISKa*, vol. 1, no. 3, pp. 153–162, 2017.
- [6] R. Kumara and C. Supriyanto, "Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5," pp. 1–10, 2014.
- [7] H. Widayu, S. D. Nasution, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Media Inform. Budidarma*, vol. Vol 1, No, no. June, p. 7, 2017.
- [8] B. Hermanto, A. Sn, and F. P. Putra, "Analisis Kinerja Decision Tree C4.5 Dalam Prediksi Potensi Pelunasan Kredit Calon Debitur," *Inovtek Polbeng Seri Inform.*, vol. 2, no. 2, 2017.