

# EVALUASI MODEL *DECISION TREE* C4.5 GUNA PREDIKSI POSIBILITAS RESIKO OBESITAS

Mochammad Yusa<sup>1</sup>, Wahyu Sindu<sup>2</sup>

<sup>1</sup> Magister Teknik Informatika, <sup>2</sup> Teknik Informatika, STMIK Pontianak  
<sup>1</sup> STMIK AMIKOM Yogyakarta, Jl. Ring Road Utara Condong Catur, Sleman  
<sup>2</sup> STMIK Pontianak Jl. Merdeka Barat No. 372, Pontianak  
<sup>1</sup> mohammad.yusa@gmail.com, <sup>2</sup> wahyusinduprasetya@gmail.com

## Abstrak

Obesitas merupakan suatu masalah kesehatan yang efek negatif dengan konsekuensi panjang bagi pengidapnya, keluarga, maupun orang lain. Data Riskesdas menunjukkan bahwa sebanyak 21,7% penduduk dewasa Indonesia yang berusia lebih dari 18 tahun mengidap obesitas. Obesitas juga merupakan akar dari penyakit-penyakit berbahaya lainnya seperti penyakit jantung koroner (PJK) dan cenderung menjadi diabetogenik yang akan mengkontaminasi penyakit-penyakit berbahaya seperti hiperlipidemia, penyakit hati dan kantong empedu, osteoartritis, kanker, dan penyakit saluran pernapasan. Selain menyebabkan masalah fisiologis obesitas juga menyebabkan masalah emosional dan psikologis seperti berkurangnya kepercayaan diri karena penampilan fisik 'kurang menarik'. Data Mining Data mining merupakan suatu metode atau teknik yang digunakan untuk memecahkan suatu permasalahan data menggunakan implementasi-implementasi untuk menemukan solusi data. Model *Decision Tree* Algoritma C4.5 merupakan salah satu algoritma yang digunakan proses penggalian data (*data mining*) yang sangat kuat dan terkenal. Pada penelitian ini, kami mengusulkan sebuah evaluasi model *Decision Tree* Algoritma C4.5 untuk proses klasifikasi sebagai solusi prediktif kemungkinan terjadinya obesitas. Kemudian Model ini dievaluasi dan divalidasi menggunakan WEKA 3.7.4 guna menghitung nilai pernanci model yang digunakan.

**Kata kunci** : Obesitas, Data Mining, *Decision Tree*, Algoritma C4.5, WEKA, Informatika Medis

## 1. Pendahuluan

Obesitas merupakan suatu masalah kesehatan yang efek negative dengan konsekuensi panjang bagi pengidapnya, keluarga, maupun orang lain. Hasil Riset Kesehatan Dasar (Riskesdas) Indonesia tahun 2010 menunjukkan angka kelebihan berat badan dan obesitas pada penduduk dewasa di atas usia 18 tahun besarnya 21,7%, dimana 11,7% (27,7 juta jiwa) adalah obesitas. Jadi, terdapat peningkatan cukup tinggi angka penduduk yang mengalami obesitas di Indonesia[16].

Kenaikan berat badan akibat konsumsi lemak berlebihan akan berdampak buruk bagi tekanan darah dan akan mengarahkan ke penyakit hipertensi. Selanjutnya hipertensi dan kegemukan ini dua-duanya menjadi penyumbang faktor resiko munculnya penyakit jantung koroner (PJK) dan cenderung menjadi diabetogenik (menyebabkan dia-betes), terutama bila sudah berlangsung lama. Obesitas meningkatkan resiko menderita hiperlipidemia, penyakit hati dan kantong empedu, osteoartritis, kanker, dan penyakit saluran pernapasan. Penderita obesitas juga beresiko lebih tinggi menderita encok dan tidur mendengkur dibandingkan dengan orang yang berat tubuhnya normal. Selain itu obesitas dapat memperpendek

harapan hidup penderitanya. Selain menyebabkan masalah fisiologis obesitas juga menyebabkan masalah emosional dan psikologis seperti berkurangnya kepercayaan diri karena penampilan fisik 'kurang menarik'[14].

Obesitas ditimbulkan oleh beberapa faktor-faktor penyebab. Menurut Albu-Shamah dan Zhan, faktor-faktor tersebut adalah faktor genetic, faktor fisik, faktor mental, faktor emosional, mental, dan psikologis, faktor sosial, faktor asupan makanan, dan faktor lainnya[1]. Namun beberapa penelitian tidak menjadikan semua kriteria faktor-faktor penyebab obesitas. Dalam penelitian yang dilakukan oleh S. Mana dan J. Bewkes hanya mengambil prevelensi terhadap 2 faktor penyebab saja yaitu aktivitas fisik anak, dan lingkungan rumah dan metode yang digunakan adalah metode *fuzzy signature*. Dalam penelitiannya faktor tersebut menjadi parameter dalam komputasi[13].

Data mining merupakan suatu metode atau teknik yang digunakan untuk memecahkan suatu permasalahan data menggunakan implementasi-implementasi untuk menemukan solusi data[11]. Dalam kasus obesitas, kami merasa ini sangat penting untuk dilakukan perhitungan klasifikasi data untuk memudahkan proses penemuan solusi

prediktif terhadap data mengingat jumlah kematian yang terus naik.

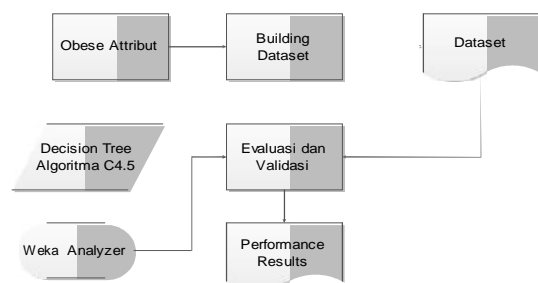
Dalam paper ini, kamu mengusulkan klasifikasi data terhadap dataset menggunakan metode *decision tree* algoritma C4.5. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Proses pada *decision tree* adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*[3]. Decision Tree merupakan teknik data mining yang didasarkan pada *divide-and-conquer* untuk klasifikasi suatu masalah[19]. Metode ini juga merupakan klasifikasi dari proses representasi struktur *tree* dimana setiap *node* mempresentasikan atribut, dan cabangnya mempresentasikan nilai dari atribut, serta daunnya (*leaves*) mempresentasikan kelas[6]. Algoritma C4.5 telah banyak digunakan sebagai implementasi pencarian pola klasifikasi prediktif di bidang-bidang tertentu misalnya untuk kasus bisnis[8][10][17] dan akademis[2][5] serta medis[9].

*Preview includes metode penelitian, data collection, building data training, analisis using Weka, Results.*

## 2. Metode dan Materi

### 2.1 Kerangka Penelitian

Kami menawarkan sebuah model (Gambar 1.) pemecahan masalah untuk mengidentifikasi atau klasifikasi obesitas menggunakan atribut-atribut berdasarkan faktor-faktor penyebab obesitas menurut penelitian dari Albu-Shamah dan Zhan[1]. Kemudian atribut tersebut digunakan untuk membangun *dataset* yang kemudian record-record yang dihasilkan dianalisis menggunakan model Decision Tree Algoritma C4.5. Model Tersebut kemudian dievaluasi menggunakan software analisis WEKA versi 3.7.4.



Gambar 1. Model Penelitian

### 2.2 Obesitas

Obesitas atau kegemukan mempunyai pengertian yang berbeda-beda bagi setiap orang.

Terkadang kita sering dibuat bingung dengan pengertian obesitas dan overweight, padahal kedua istilah tersebut mempunyai pengertian yang berbeda. Obesitas adalah suatu kondisi kelebihan berat tubuh akibat tertimbunnya lemak, untuk pria dan wanita masing-masing melebihi 20% dan 25% dari berat tubuh dan dapat membahayakan kesehatan. Sementara overweight (kelebihan berat badan, kegemukan) adalah keadaan dimana BB seseorang melebihi BB normal[15]. Perhitungan IMT dapat dilakukan menggunakan rumus sebagai berikut:

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

Keterangan :

IMT : Indeks Massa Tubuh (kg/m)

Batas ambang batas IMT (Table 1.)ditenentukan dengan merujuk ketentuanb FAO/WHO, yang membedakan batas ambang untuk laki-laki dan perempuan.Batas ambang normal laki-laki 20,1-25,0 dan untuk perempuan 18,7-23,8. Menurut WHO WPR / IASO / IOTF untuk wilayah Asia Pasifik, klasifikasi berat badan lebih dan obesitas berdasarkan IMT terbagi menjadi berat badan kurang, kisaran normal, dan berat badan lebih dengan derajat beresiko, obes I, dan obes II [18]

Tabel 1. Kategori ambang batas IMT

Kategori	IMT(Kg/m)
Berat badan kurang	<18,5
Berat badan normal	18,5-22,9
Berat badan lebih	23,0
Beresiko	23,0 - 24,9
Obes I	25,0 - 29,9
Obes II	30,0

### 2.3 Algoritma C4.5

Menurut Kusriani dan Luthfi [11], Algoritma C4.5 adalah algoritma yang digunakan untuk membentuk suatu pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode ini mengubah fakta ata data yang besar menjadi suatu pola yang berbentuk *Decision tree* yang

Algoritma C4.5 juga merupakan pengembangan dari algoritma ID3 dimana telah terjadi perubahan atau perbaikan di beberapa bagian. Algoritma ini dapat menghandel beberapa masalah yang tidak dapat ditangani oleh algoritma ID3 seperti *numeric attributes*, *missing values*, *noisy data*, dan aturan yang menghasilkan aturan dari *tree*[19].

Secara umum dalam membangun klasifikasi dengan *Decision Tree* yang menggunakan

Algoritma C4.5, melalui beberapa tahapan [12] sebagai berikut:

- a. Mempersiapkan data training.
- b. Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus :

$$Entropy(S) = \sum_{j=1}^n - P_j \log_2 P_j$$

(1)

dimana  $S$  adalah himpunan (*dataset*) kasus,  $n$  adalah banyaknya partisi  $S$ , dan  $P_j$  adalah probabilitas yang di dapat dari kelas dibagi total kasus.

- c. Kemudian hitung nilai *gain* menggunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

(2)

Dimana nilai ( $A$ ) adalah semua nilai yang mungkin dari atribut  $A$ , dan  $S_i$  adalah *subset*

dari  $y$  dimana  $A$  mempunyai nilai  $i$ .

- d. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
- e. Proses partisi pohon keputusan akan berhenti saat :
  - 1) Semua *Record* dalam simpul  $n$  mendapat kelas yang sama,
  - 2) Tidak ada atribut di dalam *record* yang dipartisi lagi, dan
  - 3) Tidak ada *record* di dalam cabang yang kosong

## 2.4 Confusion Matrix

Dalam proses evaluasi dan validasi, kami menggunakan perhitungan berdasarkan pengujian *confusion Matrix*. *Confusion matrix* memberikan keputusan yang diperoleh dalam training dan testing, *confusion matrix* memberikan penilaian performance klasifikasi berdasarkan objek dengan benar atau salah [6]. *Confusion matrix* berisi informasi aktual (*actual*) dan informasi terprediksi (*predicted*) pada sistem klasifikasi. *Confusion Matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel matrix seperti pada tabel 2. Pada Tabel tersebut dapat kita lihat bahwa jika dataset terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [4]. Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*.

**Tabel 2. Confusion Matrix**

Correct Classification	Classified as	
	+	-
+	True_positive	False_negative
-	False_positive	True_negative

## 3. Analisis dan Pembahasan

### 3.1 Atribut Penyebab Obesitas

Meskipun terdapat pengaruh genetic dan hormonal pada penambahan berat badan, kemungkinan penambahan berat badan juga muncul ketika kita mengkonsumsi banyak kalori dari makanan daripada kalori yang dibakar melalui aktivitas fisik dan olahraga[1].

Berikut ini adalah penyebab-penyebab obesitas yang diklasifikasikan menurut Albu-Shamah dan Zhan[1]:

Pertama adalah *Faktor geneti*. Tidak selalu kelebihan badan dapat diidentikan dengan mengkonsumsi banyak makanan. Penelitian yang dilakukan pada tahun 1980an-1990an di US menunjukkan bahwa orang kurus makan lebih banyak daripada orang gemuk tetapi tetap kurus sementara orang gemuk makan sedikit tetapi tetap menjadi gemuk. Phenomena inilah yang dinamakan gen yang sudah mendarah daging yang dapat diidentifikasi menjadi salah satu penyebab obesitas.

Kedua adalah *Faktor fisik*. Studi di Amerika juga menunjukkan bahwa hal yang paling negative dan dapat menyebabkan obesitas adalah orang yang menonton TV lebih dari 3 jam dalam sehari.

Ketiga adalah *Faktor emosional, mental dan psikologi*. Faktor ini juga dapat mempengaruhi obesitas. Orang yang sedang menderita stress, marah atau masalah emosional lainnya cenderung akan mengkonsumsi makanan dengan porsi yang banyak.

*Faktor sosial* juga berpengaruh terhadap kecendrungan posibilitas terjadinya obesitas. Faktor sosial mempunyai pengaruh terhadap jenis makanan dan porsi makanan yang diambil oleh seseorang. Misalnya, di dalam lingkungan rumah, orang tua mereka tergolong hobi untuk menyediakan makanan dengan porsi yang besar, mau tidak mau si anak pasti akan disajikan makanan dengan porsi yang besar juga. Dengan begitu asupan makan yang dikonsumsi akan berpengaruh terhadap kalori yang dikonsumsi.

Faktor selanjutnya yang berpengaruh adalah *Faktor asupan makana*. Sering memakan makanan yang mempunyai nilai kalori besar seperti Junk food, yang mengandung gula, dan mengandung kalori banyak merupakan suatu penyebab terjadinya kegemukan dan obesitas.

Faktor terakhir adalah *Faktor-faktor lainnya* yang dapat mempengaruhi kemunculan obesitas.

Kebiasaan-kebiasan seperti makan dengan cepat, rakus, dan tergesa-gesa terhubung dengan asupan energy yang digunakan, ketahanan tubuh yang terhubung dengan system kerja tubuh. Faktor ini mempunyai pengaruh yang tidak terlalu signifikan dalam perhitungan obesitas tapi dapat juga berpengaruh terhadap kenaikan berat badan yang signifikan. Maka dari itu, bagi expert, mungkin faktor ini tidak terlalu diperhitungkan karena mempunyai porsi yang kecil dalam dampak kenaikan Indeks massa tubuh (IMT).

Faktor-faktor di atas kemudian dijadikan *instances* sebagai acuan dalam fokus penelitian ini. Tabel 3. Adalah contoh atribut data yang akan digunakan dalam penelitian ini.

**Tabel 3. Faktor, Value, dan Tipe data**

No	Faktor-Faktor	Values	Tipe Data
1	Jenis Kelamin	laki-laki (L), Perempuan (P)	Nominal
2	Aktivitas Fisik	Tinggi, Normal, Rendah	Polinomial
3	Faktor Genetik	Ya, Tidak	Nominal
4	Faktor Emosional, Mental, Psikologis	Tinggi, Normal, Rendah	Nominal
5	Faktor Sosial	Ya, Tidak	Nominal
6	Asupan Makanan	Tinggi, Normal, Kurang	Polinomial
7	Faktor Lainnya	Ya, Tidak	Nominal
8	Output	Resiko, Tidak beresiko	Nominal

### 3.2 Data training

Agar penelitian ini dapat terfokus dalam penemuan pola-pola berbentuk keputusan, kami menggunakan data simulasi yang digunakan sebagai acuan pembentukan pola pohon keputusan. Data tersebut ditunjukkan pada Tabel 4. Data tersebut sesuai dengan faktor dan *value* yang sebagai *instances* atau *attribut* yang ditunjukkan pada Tabel 3.

Tujuan dari pembentukan data ini adalah menciptakan himpunan data target, pemilihan data, dan memfokuskan subset *variabel*, dimana penelusuran pola akan dilakukan. Terdapat 20 dataset yang kami gunakan sebagai simulasi dalam proses pembuatan pola pohon keputusan. Data ini kemudian akan dievaluasi dan divalidasi. Data ini juga merupakan dasar pemfokusan model evaluasi.

**Tabel 4. Data Training**

No	Jenis Kelamin	Aktivitas Fisik	Faktor Genetik	Faktor Emosional, Mental, Psikologis	Faktor Sosial	Asupan Makanan	Faktor Lainnya	Output
1	P	Normal	Ya	Tidak	Ya	Normal	Tidak	Tidak Beresiko
2	L	Tinggi	Ya	Ya	Tidak	Normal	Tidak	Beresiko
3	L	Normal	Tidak	Ya	Tidak	Normal	Tidak	Tidak Beresiko
4	L	Rendah	Ya	Tidak	Tidak	Normal	Ya	Beresiko
5	P	Rendah	Ya	Ya	Tidak	Normal	Ya	Tidak Beresiko
6	P	Rendah	Ya	Tidak	Tidak	Rendah	Tidak	Tidak Beresiko
7	P	Normal	Tidak	Ya	Tidak	Rendah	Tidak	Beresiko
8	P	Normal	Tidak	Ya	Ya	Tinggi	Ya	Beresiko
9	L	Normal	Tidak	Tidak	Ya	Normal	Ya	Tidak Beresiko
10	L	Normal	Ya	Tidak	Ya	Normal	Ya	Tidak Beresiko
11	P	Normal	Tidak	Tidak	Ya	Rendah	Tidak	Tidak Beresiko
12	L	Rendah	Tidak	Tidak	Ya	Normal	Ya	Tidak Beresiko
13	L	Rendah	Ya	Tidak	Ya	Rendah	Tidak	Beresiko
14	P	Tinggi	Tidak	Ya	Ya	Rendah	Tidak	Tidak Beresiko
15	P	Rendah	Ya	Tidak	Tidak	Rendah	Ya	Tidak Beresiko
16	L	Normal	Ya	Ya	Tidak	Normal	Ya	Beresiko
17	L	Tinggi	Tidak	Ya	Ya	Tinggi	Tidak	Beresiko
18	P	Rendah	Tidak	Ya	Ya	Rendah	Ya	Tidak Beresiko
19	P	Tinggi	Ya	Tidak	Tidak	Tinggi	Ya	Beresiko
20	P	Tinggi	Ya	Tidak	Tidak	Tinggi	Ya	Beresiko

### 3.3 Analisis Data

Pada tahap ini, kami melakukan evaluasi dan analisis terhadap pohon keputusan yang dibentuk. Pada tahap ini kami melakukan uji coba menggunakan evaluasi *Matrix Confusion* untuk mengetahui nilai *accuracy*, *recall*, dan *precision*.

**Tabel 5. Evaluasi pada dataset**

<i>Kappa statistic</i>	0.596
<i>Mean absolute error (MAE)</i>	0.2629
<i>Root mean squared error (RMSE)</i>	0.3625
<i>Relative absolute error (RAE)</i>	53.0537%
<i>Root relative squared error (RRSE)</i>	72.8684%
<i>Coverage of cases (0.95 level)</i>	100%
<i>Mean rel. region size (0.95 level)</i>	80%
<i>Total Number of Instances</i>	20

Berdasarkan Tabel 5. Diatas, hasil percobaan menunjukkan bahwa model yang dibangun menggunakan Algoritma C4.5 bahwa dari jumlah total *training set* diklasifikasikan 16 records atau 80% *Correctly Classified Instances* dan 4 records atau 20 % *Incorrectly Classified Instances*.

Hasil percobaan menunjukkan statistik Kappa (*k*) sebesar 0.596 yang menunjukkan bahwa korelasi antar output. Nilai *Error Absolut* rata-rata (MAE) adalah 0.2629 yang berarti nilai rata-rata kesalahan sistem dalam proses peramalan. RSME menghasilkan nilai 0.3625. sedangkan untuk nilai RAE adalah 53.0537% dan nilai RRSE adalah 72.8684%.



Sedangkan dari hasil percobaan juga menghasilkan *Confusion Matrix* seperti yang ditunjukkan pada Tabel 6. *Confusion Matrix* tersebut selanjutnya akan digunakan untuk menghitung nilai *TP rate* atau nilai *record* positif yang diklasifikasikan sebagai positif, *FP rate* nilai *record* positif yang diklasifikasikan sebagai positif, *Precision*, *Recall*, dan *ROC area*.

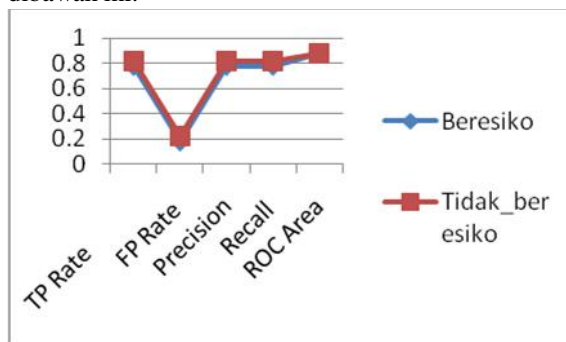
**Tabel 6. Confusion Matrix**

Output	Correctly Instances	indicator	Resiko	
			True	Tidak_beresiko
Beresiko	16	+	7	2
Tidak_beresiko	4	-	2	9

**Tabel 7. Hasil Evaluasi**

Output	TP Rate	FP Rate	Precision	Recall	ROC Area
Beresiko	0,778	0,182	0,778	0,778	0,879
Tidak_beresiko	0,818	0,222	0,818	0,818	0,879

Dari hasil pengujian dengan skema atau skenario kondisi parameter dengan proporsi 80% dan 20 untuk *correctly* dan *incorrectly classified instances* (Tabel 7.) didapatkan hasil *TP rate* untuk *output Beresiko* sebesar 0,778 dan *TP rate output Tidak\_beresiko* sebesar 0,818. Dan *FP rate beresiko* sebesar 0,182 dan tidak beresiko sebesar 0,222. Dari hasil pengujian juga didapatkan nilai yang sama antara nilai *precision* dan *Recall* untuk *output beresiko* dan tidak beresiko secara berturut-turut adalah 0,778 dan 0,818. Sedangkan Nilai Area ROC berada pada nilai 0,879 yang berarti klasifikasi tersebut berada pada kondisi *Good Classification*[6]. Grafik dari hasil pengujian dapat dilihat pada Gambar 2 dibawah ini.



Gambar 2. Grafik Evaluasi data training

Tahap selanjutnya adalah proses perhitungan akurasi dari algoritma C4.5 dalam kasus obesitas dari dataset sebanyak 20 records (*data training*). Menurut Han dan Kamber[7], cara menghitung akurasi (*AC*) dapat menggunakan rumus berikut ini:

$$Accuracy = \frac{True_{neg} + True_{pos}}{True_{neg} + False_{neg} + True_{pos} + False_{pos}} * 100 \quad (3)$$

Dari hasil perhitungan menggunakan rumus diatas maka didapatkan nilai akurasi (*AC*) model *Decision Tree* Algoritma C4.5 adalah sebesar 80%.

#### 4. Kesimpulan

Metode Klasifikasi *Decision Tree* menggunakan algoritma C4.5 guna memprediksi obesitas mempunyai hasil yang sangat baik. dari hasil pengujian yang dilakukan, Model ini mempunyai nilai akurasi yang berada pada angka 80%. Nilai *Precision* dan *Recall* yang dihasilkan dari hasil pengujian cukup tinggi yaitu 0,778 untuk hasil *output beresiko* dan 0,818 *output Tidak\_beresiko*. Area ROC yang ditunjukkan pada hasil ujicoba juga berada pada *range Good Classification* yaitu 0,878. Jadi kami menyimpulkan bahwa Metode Model Klasifikasi *Decision Tree* menggunakan Algoritma C4.5 mempunyai performa yang baik dalam memberikan solusi prediktif terhadap kemungkinan terjadinya obesitas.

#### Daftar Pustaka:

- [1] Albu-Shamah, Ahmad & Justin Zhan, 2013. *Towards Obesity Causes, Prevalence and Prevention*. IEEE Computer Society
- [2] Andriani, Anik, 2013, *Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: Amik "Bsi Yogyakarta"*, Proceeding SENTIKA 2013, 9 Maret 2013, hal. 163-168, ISSN: 2089-9815, Yogyakarta.
- [3] Basuki, A. and Syarif, I., 2003, *Decision Tree*, Surabaya: Politeknik Elektronika Negeri Surabaya - ITS.
- [4] Bramer, Max, 2007, *Principles of Data Mining*, London: Springer.
- [5] Ginting, S.L., Zarman, Wendi, dan Hamidah, Ida, 2014, *Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik*, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014, Yogyakarta, 15 November 2014, ISSN: 1979-911X, Hal. A263-A272.
- [6] Gorunescu, F., 2011, *Data Mining Concept Model Technique*, India: Springer.
- [7] Han, J. dan Kamber, M., 2001, *Data Mining, Concepts and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann.

- [8] Julianto,Windy, Yunitarini,Rika ,Dan Sophan, Mochammad K.,2014, *Algoritma C4.5 Untuk Penilaian Kinerja Karyawan*, Jurnal Scan Vol. IX Nomor 2 Juni 2014, ISSN : 1978-0087, Hal. 33-39.
- [9] Karolis,Minas A., dan Moutiris,Joseph A., 2010, *Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees*, IEEE Transactions On Information Technology In Biomedicine, Vol. 14, NO. 3, May 2010. hal. 559-566.
- [10] Kharis, Irma, Rosa delima, dan Joko Purwadi, 2013, *Generator pohon Keputusan Dengan Menerapkan Algoritma C4.5 Untuk Program Konsultasi*, Jurnal Informatika Vol. 9. 1 April 2013. Hal. 1-10.
- [11] Kusriani & luthfi, E.T., 2009, *Algoritma Data Mining*, Yogyakarta:Andi publishing.
- [12] Larose, D. T, 2005, *Discovering Knowledge in Data*, New Jersey: John Willey & Sons, Inc.
- [13] Manna, S. Dan Jewkes, A.M. 2014. *Understanding early childhood obesity risks: An empirical study using fuzzy signatures*. IEEE International Conference
- [14] Dinas Kesehatan Yogyakarta, 2015, *Obesitas, Faktor Resiko Berbagai Penyakit. Anda Obesitas?, Awas Bahaya Mengancam!*. Berita dan Artikel Dinas Kesehatan Yogyakarta, [http://dinkes.jogjaprovo.go.id/berita/detail\\_berita/549-obesitas-faktor-resiko-berbagai-penyakit-anda-awas-bahaya-mengancam](http://dinkes.jogjaprovo.go.id/berita/detail_berita/549-obesitas-faktor-resiko-berbagai-penyakit-anda-awas-bahaya-mengancam), Diakses 1 Mei 2015.
- [15] Rimbawan dan Albiner Siagian. 2004. *Indeks Glikemik Pangan*. Bogor : Penebar Swadaya
- [16] Badan Penelitian dan Pengembangan Kesehatan, 2010, *Riset Kesehatan Dasar (Riskesdas 2010)*, Kementerian Kesehatan Republik Indonesia, Jakarta.
- [17] Santosa, Budi T.,2014, *Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan*, Jurnal Ilmiah Fakultas Teknik LIMIT'S Vol. 10 No.1, Hal 1-10. ISSN: 0216 - 1184
- [18] Sugondo, S., 2006, *Obesitas*. In: *Sudoyo, AW., Setiyohadi, B., Alwi, I., Simadibrata, MK., Setiati, S.*, ed. Buku Ajar Ilmu Penyakit Dalam. Pusat Penerbit Departemen Ilmu Penyakit Dalam Fakultas Kedokteran Universitas Indonesia, Jakarta: 1919-1925
- [19] Witten I. H., Frank E. and Hall, M. A. , 2011, *DATA MINING - Practical Machine Learning Tools and Techniques* (3rd ed), Elsevier Inc.