PENENTUAN GIZI ANAK MENGGUNAKAN KOMPARASI METODE C4.5 DAN K-NEAREST NEIGHBOR (KNN)

M. Noer Fadli Hidayat
Teknik Informatika Universitas Nurul Jadid
Karanganyar Paiton Probolinggo
e-mail: fadli@unuja.ac.id

ABSTRAK

Kesehatan anak usia di bawah 2 tahun harus sempurna untuk membentuk generasi emas di masanya. Upaya yang dapat dilakukan salah satunya adalah pemenuhan asupan dan kecukupan gizi selama 1000 hari pertama kehidupan (1000 HPK). Dalam 1000 HPK anak terjadi periode emas pertumbuhan yang harus terpenuhi gizinya agar menjamin kesehatan yang baik di masa depan. Pada 1000 HPK dijelaskan bahwa gizi berkaitan dengan kebiasaan pola makan dan lingkungan selama kehamilan hingga usia 2 tahun seorang anak mempengaruhi kondisi fisiknya kelak di masa yang akan datang. Pertumbuhan anak yang normal dapat dilihat dari pengukuran fisik secara langsung atau secara antropometri. Pengukuran fisik secara langsung terdapat 3 faktor utama yaitu :1) Pengukuran Berat Badan Sesuai Umur (BB/U), 2) Pengukuran Tinggi Badan Sesuai Umur (TB/U), 3) Pengukuran Berat Badan Sesuai Tinggi Badan (BB/TB). Dari ketiga faktor tersebut harus dicocokan dengan menggunakan tabel pertumbuhan yang berlaku untuk suatu masyarakat, sehingga dapat mengevaluasi dan melihat perkembangan fisik anak. Penelitian ini akan menerapakan dua metode yang akan dicari model terbaik sebagai panutan untuk penentuan gizi anak. Metode yang akan digunakan adalah dua metode data mining yang klasifikasi yaitu C.45 dan K-Nearest Neighbor (K-NN), yang akan dicari model terbaik dari kedua metode tersebut dalam mencari tingkat keakurasian yang paling tinggi. Hasil penelitian menunjukan kinerja metode KNN dengan akurasi mencapai 93,30% lebih baik dari hasil kinerja metode C.45 dengan akurasi mencapai 78,03%.

Kata Kunci: C4.5, Data Mining, Gizi Anak, K-Nearest Neighbor (K-NN).

ABSTRACT

The health of children under 2 years of age has contributing factors in formingtheir golden age. To achive this goal, the children needs the fulfillment of nutritional intake and adequacy for the first 1000 days of life (1000 HPK). In 1000 HPK children, there is a golden period of growth where nutrition should be primarly fulfilled in order to ensure their good health. The 1000 HPK explains that nutrition is related to dietary habits and the environment during pregnancy until the age of 2 years of a child which affects his physical condition in the future. Normal child growth can be seen from direct physical measurements or anthropometry. There are 3 main factors for direct physical measurement, namely: 1) Measurement of Body Weight for Age (BW / U), 2) Measurement of Height according to Age (TB / U), 3) Measurement of body weight according to height and weight (BW / TB). These three factors must be matched using a growth table applicable to a society, so that it is possible to evaluate the physical development of the child. This research is to use two methods in which the best model will be sought as a role model for determining child nutrition. By using two classification data mining methods, namely C.45 and K-Nearest Neighbor (K-NN), this study will figure out the best model of the two methods in finding the higest level of accuracy. The study resulted that the performance of the KNN method with an accuracy of 93.30% was better than that of the C.45 method with an accuracy of 78.03%.

Keywords*: C4.5 method, Data mining, K-Nearest Neighbor, Nutrition intake.*

I. PENDAHULUAN

Kesehatan anak usia di bawah 2 tahun harus sempurna untuk membentuk generasi emas di masanya. Upaya yang dapat dilakukan salah satunya adalah pemenuhan asupan dan kecukupan gizi selama 1000 hari pertama kehidupan (1000 HPK). Dalam 1000 HPK anak terjadi periode emas pertumbuhan yang harus terpenuhi gizinya agar menjamin kesehatan yang baik di masa depan[1]. Perhitungan 1000 HPK ini dimulai sejak janin berada didalam rahim ibu sampai anak berusia 2 tahun, yang mana dalam kurun waktu tersebut tubuh bekerja secara optimal untuk membentuk fisik dan kecerdasan seorang anak.

Semua asupan dan kecukupan gizi penting untuk pertumbuhan otak. Dampak dari kekurangan gizi sangatlah beragam[2]. Secara umum, kekurangan gizi berdampak besar terhadap perkembangan otak dan mengakibatkan dampak buruk jangka panjang seperti kelemahan berpikir dan kelemahan kemampuan kognitif [3]. Hal ini didukung pula dalam konsep

"*Nutritional programming*" pada 1000 HPK yang menjelaskan bahwa gizi berkaitan dengan kebiasaan pola makan dan lingkungan selama kehamilan hingga usia 2 tahun seorang anak mempengaruhi kondisi fisiknya kelak di masa yang akan datang [4].

Anak akan mengalami hambatan dalam pertumbuhan atau yang sering disebut dengan *stunting* atau balita pendek [3]. Upaya peningkatan status gizi masyarakat, termasuk penurunan prevalensi balita pendek menjadi salah satu prioritas pembangunan nasional yang tercantum dalam Rencana Pembangunan Jangka Menengah 2015-2019.

Gizi adalah substansi organik yang dibutuhkan organisme untuk fungsi normal dari sistem tubuh, pertumbuhan, pemeliharaan kesehatan, gizi merupakan komponen yang sangat dibutuhkan seorang anak untuk peningkatan pertumbuhan dan perkembangannya, terutama pada masa usia 6 – 24 bulan[5]. Penentuan gizi merupakan suatu upaya yang dilakukan dalam rangka meningkatkan kualitas kesehatan anak.

Dalam bidang gizi, antropometri telah diaplikasikan secara luar dan detail untuk menilai status gizi masyarakat[6]. Penilaian status gizi masyarakat dengan antropmetri pada dasarnya adalah mengukur perubahan pertumbuhan anak yang mencakup pengukuran berat badan dan panjang badan atau tinggi badan dengan membandingkan hasil pengukuran dengan baku sesuai indeks yang digunakan[7]. Dengan adanya standaritas untuk gizi anak yang memiliki kesesuaian antara *World Health Organization- National Center for Health Statistic* (WHO-NCHS), yang memiliki Keputusan Menteri Kesehatan Republik Indonesia Nomor 1995/Menkes/SK/XII/2010 tanggal 30 Desember 2010. Dengan adanya pernyataan di atas dimana setiap bayi terkadang memiliki status gizi yang berbeda, dirasa masyarakat perlu untuk mengetahui dari status gizi tersebut, agar memberikan kemudahan dalam penanganan yang tepat dan membantu dalam mencerdaskan anak bangsa. Dimana status gizi dimulai dari umur 6 bulan sampai 24 bulan atau usia 2 tahun, karena pada usia tersebut masih membutuhkan nutrisi otak yang memadai dan cukup dalam rangka perkembangan otak anak. Jadi yang dibutuhkan adalah unutk mengetahui status dari gizi pada anak, apa kurang gizi, atau cukup gizi, atau normal[8].

Tujuan dari penelitian ini adalah untuk mencari tingkat akurasi yang lebih tinggi antara menerapkan metode C.45 dan K-NN dalam mengklasifikasi penetuan gizi anak.

II. TINJAUAN PUSTAKA

A. Gizi

Pengertian gizi secara umum adalah substansi organik yang dibutuhkan organisme untuk fungsi normal dari sistem tubuh, pertumbuhan, pemeliharaan kesehatan. ilmu gizi (*nutrition science*) adalah "Ilmu yg mempelajari segala sesuatu tentang makanan dalam hubungannya dengan kesehatan optimal". Sedangkan yang disebut sebagai makanan adalah "Bahan selain obat yang mengandung zat-zat gizi dan atau unsur-unsur/ikatan kimia yang dapat diubah menjadi zat gizi oleh tubuh, yang berguna bila dimasukkan dalam tubuh. Pada zaman modern ini pengertian ilmu gizi tidak hanya sebatas fungsi dasar zat gizi dan kesehatan saja tetapi zat gizi sudah dihubungkan dengan kemampuan kerja, produktivitas, IQ, dan status ekonomi. Tujuan ilmu gizi sendiri adalah mencapai, memperbaiki, dan mempertahankan kesehatan tubuh melalui konsumsi makanan[9][10].

B. Data Mining

Data mining adalah sebuah pencarian pengetahuan dari data, yaitu dengan cara mengekstrak secara otomatis pola atau pengetahuan yang menarik (tidak sederhana, tersembunyi, tidak diketahui sebelumnya, berpotensi berguna) dari data dalam jumlah sangat besar. data mining memilik nama alternatif diantaranya *knowledge discovery (mining)* in *databases* (KDD), *knowledge extraction*, data/*pattern analysis*, data *archeology*, data *dredging*, *information harvesting*, *business intelligence*. Data mining pastinya yang akan dibahas pertama kali adalah sebuah data, jika data tidak valid dan juga data hanya sedikit hal ini akan membuat hasil dari penambangan pengetahuannya memiliki akurasi yang sangat kecil, sehingga hasil klasifikasi, prediksi, dan pengglompokkan dari data mining akan meleset dari kenyataan sesungguhnya. Dalam data mining masih ada jenis-jenisnya yang pertama adalah klasifikasi, *classte*, klasifikasi biasanya dilakukan untuk prediksi, pengelompokkan yang diketahui jenis-jenisnya, sedangkan *classter* digunakan untuk mengelompokkan yang tidak diketahui jumlah classnya. Algoritma dalam data mining banyak jenisnya salah satu algoritma dalam data mining adalah apriori, dan juga ada ID3, C4.5, K-NN, *Naïve bayes*, serta masih banyak yang lainnya, perlu diketahui bahwa setiap algoritma dalam data mining memiliki tingkat akurasi yang berbeda-beda tergantung dengan data yang digunakan, dan tidak akurasi yang mencapai 100% [11][12].

C. Metode C4.5

Metode C4.5 diperkenalkan Quinlan [10] sebagai versi terbaru dari ID3. Dalam induksi *tree* hanya bisa dilakukan pada fitur bertipe katagorikal (*nominal* atau *ordinal*), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe *numerik*, melakukan pemotongan *decision*

tree. Algoritma C4.5 juga menggunakan kriteria dalam menentukan fitur yang menjadi pemecah pada pohon yang diinduksi[13].

Algoritma C4.5 adalah pengembangan dari algoritma ID3. Oleh karena pengembangan tersebut algoritma C4.5 mempunyai prinsip dasar kerja yang sama dengan algoritma ID3. Hanya saja dalam algoritma C4.5 pemilihan atribut dilakukan dengan menggunakan *Gain Ratio* pada persaman (1):

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

Keterangan :

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah partisi atribut A
- |Si| : Jumlah Kasus pada partisi ke i
- |S| : Jumlah Kasus Dalam S

Sedangkan perhitungan nilai *Entropy* dapat dilihat pada persamaan (2):

$$\boldsymbol{entropy(A)} = -\sum_{i=1}^{n} \frac{|Si|}{|S|} \log_2 \frac{|Si|}{|S|}$$

Keterangan :

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah partisi atribut A
- |Si| : Jumlah Kasus pada partisi ke i
- |S| : Jumlah Kasus Dalam S (Putra 2013)

D. Metode KNN

K-*Nearest Neighbor* (K-NN) merupakan salah satu metode untuk mengambil keputusan menggunakan pembelajaran terawasi dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai[1]. Algoritma K-*Nearest Neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN merupakan algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikasi berdasarkan mayoritas dari kategori pada algoritma KNN[2]. Dimana kelas yang paling banyak muncul yang nantinya akan menjadi kelas hasil dari klasifikasi. Kedekatan didefinisikan dalam jarak metrik, seperti jarak *Euclidean*. Jarak *Euclidean* dapat dicari dengan menggunakan persamaan (3) berikut ini:

$$\boldsymbol{Dxy} = \sqrt{\sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{y_i})^2} \quad (3)$$

Keterangan :

D : jarak terdekat
x : data *training*
y : data *testing*
n : jumlah atribut individu antara 1 s.d. n
f : fungsi similitary atribut G antara kasus X dan kasus Y
i : Atribut individu antara 1 sampai dengan n

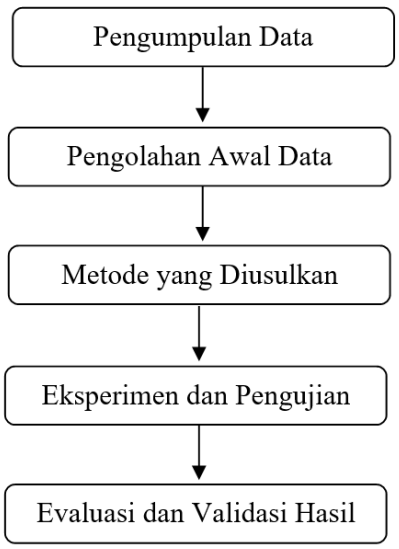
Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain :

1. Menentukan parameter K (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak *Euclid* (*queri instance*) masing-masing objek terhadap data sampel yang diberikan menggunakan persamaan 3.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *Euclid* terkecil.
4. Mengumpulkan kategori Y (Klasifikasi *Nearest Neighbor*)

Dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas maka dapat diprediksi nilai *query instance* yang telah dihitung[12][14].

III. METODE PENELITIAN

Pada penelitian ini, data gizi anak tersebut diolah menggunakan metode *data mining* sehingga diperoleh satu metode yang paling akurat dan dapat digunakan sebagai *rules* dalam mengklasifikasikan sesuai dengan data yang terdapat di Puskesmas Paiton. Dalam penelitian ini akan dilakukan beberapa langkah-langkah seperti gambar 1 menunjukkan sebuah tahapan penelitian.



Gambar 1. Tahapan Penelitian

Pada metode penelitian *eksperimen*, terdapat beberapa tahapan penelitian yang dilakukan seperti berikut:

a. Pengumpulan Data

Pada tahap ini dijelaskan bagaimana dan dari mana data dalam penelitian ini didapatkan, meliputi data *se-*

kunder dan data primer. Data *sekunder* berisi tentang sumber pemerolehan data untuk keperluan penelitian, sedangkan data *primer* berisi tentang data yang dihasilkan dari penelitian.

b. Pengolahan awal data

Pada bagian ini dijelaskan tentang tahap awal data mining. Pengolahan awal data meliputi proses *input* data ke format yang dibutuhkan, pengelompokan dan penentuan atribut data, serta pemecahan data untuk digunakan dalam proses pembelajaran *(training)* dan pengujian *(testing)*.

c. Model yang diusulkan

Pada tahap ini data dianalisis, dikelompokan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data kedalam data latihan (*training* data) dan data uji (*testing* data) juga diperlukan untuk pembuatan model. Metode yang diusulkan adalah C.45 dan *K-Nearest Neighbor* (K-NN), dimana akan dicari akurasi terbaik, dalam mengklasifikasi status gizi anak[6].

d. Eksperimen dan pengujian model

Pada bagian ini dijelaskan tentang langkah-langkah eksperimen meliputi cara pemilihan arsitektur yang tepat dari model atau metode yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat.

e. Evaluasi dan validasi hasil

Pada bagian ini dijelaskan tentang evaluasi dan validasi hasil penerapan metode pada penelitian yang dilakukan.

IV. PEMBAHASAN

Data gizi anak diambil pada tahun 2019, jumlah dataset 133 record. Dari jumlah tersebut akan dibagi 2 yaitu 3 atribut atau variabel X dan 1 label atau variabel Y, klasifikasi gizi anak terdiri dari gizi buruk, gizi kurang, gizi baik dan gizi lebih. Lebih detailnya perhatikan tabel 1 tentang karakteristik dataset gizi anak di Puskesmas Paiton.

Tabel 1. Karakteristik Dataset Gizi Anak Puskesmas Paiton

No	Nama Atribut	Variabel X	Keterangan
1	Umur	X1	*Numeric*
2	Berat Badan	X2	*Numeric*
3	Tinggi Badan	X3	*Numeric*
4	Label atau Y	Gizi Buruk, Gizi Kurang, Gizi Baik, Gizi Lebih	*Nominal*

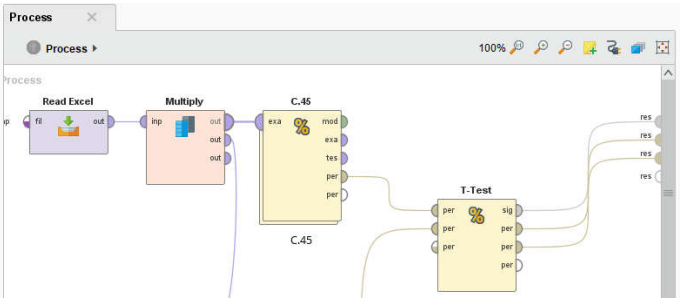
Dari dataset diatas, penelitian ini akan dilakukan pengujian dengan menggunakan metode C.45 dan K-NN setelah itu, akan dibandingkan antara dua metode seperti yang sudah dibahas sebelumnya. Tujuan dari perbandingan metode tersebut, adalah akan dicari satu metode terbaik dengan nilai akurasi tertinggi dalam menentukan status gizi anak. Perbandingan akan terjadi jika dari dua metode tersebut memiliki akurasi, pencarian akurasi sesuai dengan apa yang dilakukan. Maka dari itu, akan merinci secara detail dari dua metode yang dibandingkan, untuk lebih jelasnya perhatikan nilai hasil, dari setiap pembanding.

Dari dataset tersebut akan dikelola dan diuji ke beberapa metode yang telah ditentukan, algoritma tersebut adalah C4.5 dan K-NN. Penelitian ini akan merinci secara detail dari setiap metode yang terdapat di atas. Dibawah ini akan dirinci secara detail dari setiap algoritma tersebut.

A. Implementasi Metode C4.5

Algoritma C4.5 adalah bagian dari kelompak *Decision Tree*. Pada algoritma ini memilki *input* yang berupa *training sample* dan *sample*. *Training sample* merupakan data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah di uji kebenaranya. Sedang sampel adalah *field-field* data yang nantikan akan dipakai sebagai parameter atau pengukur dalam melakukan klasifikasi data. Dari *sample* dan *training sample* sudah dimiliki oleh dataset pada penelitian saat ini.

Pada gambar 2 dipaparkan desain implementasi menggunakan aplikasi *rapidminer* 8.2, dimana dataset yang digunakan pada *read excel* ditunjukkan pada gambar 4.2 untuk *multiply* berfungsi sebagai multi input yang digunakan untuk mengkomparasi metode antara C.45 dan K-NN. *T-Test* digunakan untuk menguji kebenaran atau kepalsuan hipotesis yang menyatakan bahwa diantara dua buah *mean sample* yang diambil secara *random* dari populasi yang sama.

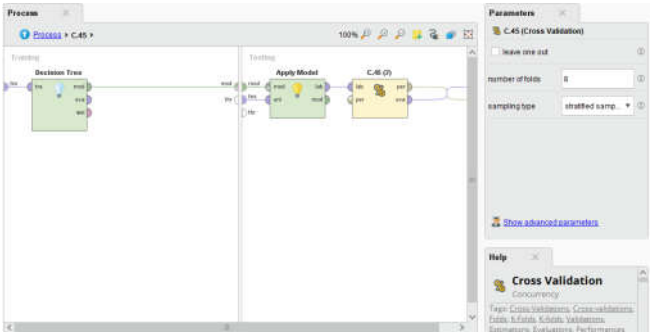


Gambar 2. Desain Implementasi Metode C.45

X1	X2	X3	Y
numeric	numeric	numeric	nominal
attribute	attribute	attribute	label
6	5.700	61.200	Gizi Buruk
6	6.400	63.300	Gizi Kura...
6	7.100	65.500	Gizi Kura...
6	7.900	67.600	Gizi Baik
6	8.800	69.800	Gizi Baik
6	9.800	71.900	Gizi Baik
6	10.900	74	Gizi Lebih

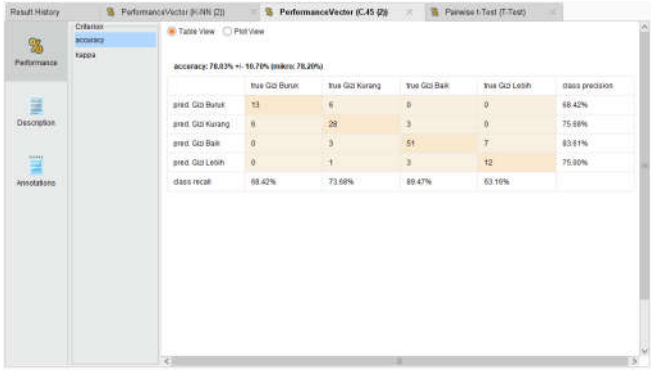
Gambar 3. Dataset yang digunakan dalam *rapidminer* 8.2

Dalam implementasi *rapidminer* metode C.45 ada di dalam *cross validation* seperti yang ditunjukkan gambar 4, dan untuk *number of folds* pada parameters berisi 8 maka validasi yang digunakan berada pada validasi ke – 8.



Gambar 4. *Cross validation* metode C.45

Hasil hitung ditunjukkan gambar 5 dari dataset gizi anak dengan aplikasi *tools rapidminer* 8.2 menggunakan metode C.45. Hasil yang di dapatkan merupakan hasil dari metode C.45 yang menggunakan validasi ke – 8 dengan akurasi : 78.03 %, dimana akurasi tersebut akurasi tertinggi dari metode C.45.



Gambar 5. Hasil akurasi metode C.45

Number of Validation ini digunakan untuk mencari parameter yang paling baik dari sebuah algoritma klasifikasi, sedangkan akurasi menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya. Dengan demikian hasil dari validasi ke – 2 sampai ke – 10 ditunjukkan pada tabel 2 hasil akurasi dari metode C.45.

Tabel 2. Hasil Akurasi dari Metode C.4.5

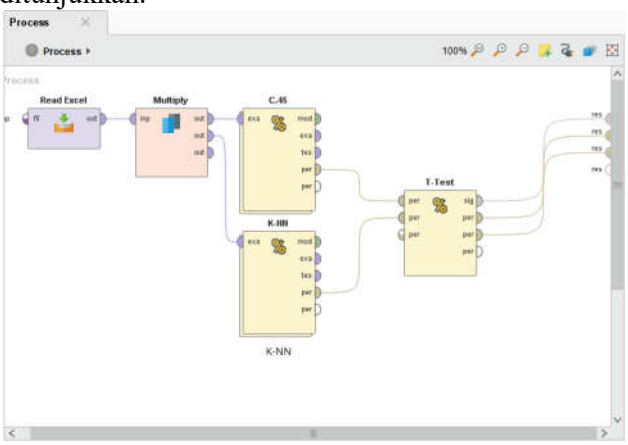
Number of Validation	**Akurasi**
Ke-2	65,40%
Ke-3	72,91%
Ke-4	65,46%
Ke-5	72,99%
Ke-6	76,75%
Ke-7	72,93%
Ke-8	78,03%
Ke-9	76,89%
Ke-10	78,02%

Pada dataset gizi anak telah dilakukan pengujian untuk mencari tingkat akurasi, pengujian ini melakukan 10 validasi, dimana rata-rata nilai akurasi adalah 74,26%, dan nilai akurasi tertinggi adalah 78,03% di pada validasi ke-8. Jadi berdasarkan pengujian diatas metode C.45 memiliki nilai akurasi yang bisa di katakan melebihi 50%.

B. Implementasi K-NN

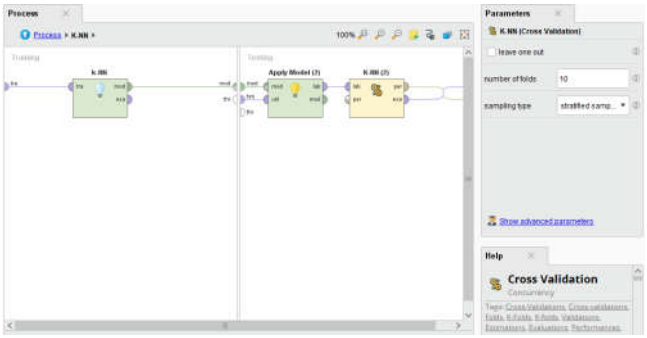
Metode K-NN atau *K-Nearest Neighbor* merupakan salah satu metode yang menggunakan algoritma *supervised* dimana hasil dari menghitungan dari *query instance* yang akan diklasifikasikan berdasarkan mayoritas pada kategori K-NN. Manfaat dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. Pada dataset yang sudah disiapkan pada data gizi anak di Puskesmas Paiton akan dilakukan pengujian dengan menggunakan metode K-NN, dan analisis dari akurasi yang telah didapat.

Pada gambar 6 dipaparkan desain implementasi komparasi metode C.45 dan K-NN menggunakan aplikasi *rapidminer* 8.2, dimana dataset yang digunakan pada *read excel* sama persis seperti yang ditunjukkan.



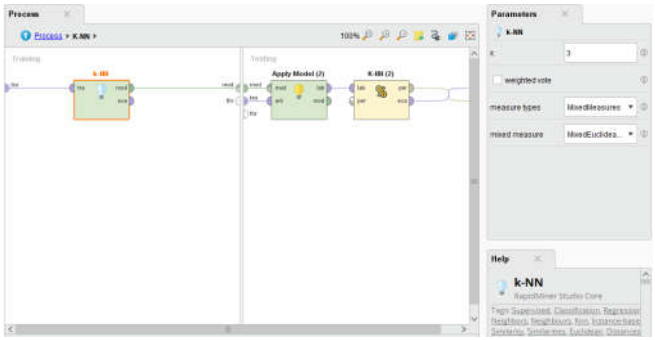
Gambar 6. Desian implementasi komparasi metode

Dalam Implementasi *rapidminer* metode K-NN ada di dalam *cross validation* seperti yang ditunjukkan gambar 7, dan untuk *number of folds* pada *parameters* berisi 10 maka validasi yang digunakan berada pada validasi ke – 10.



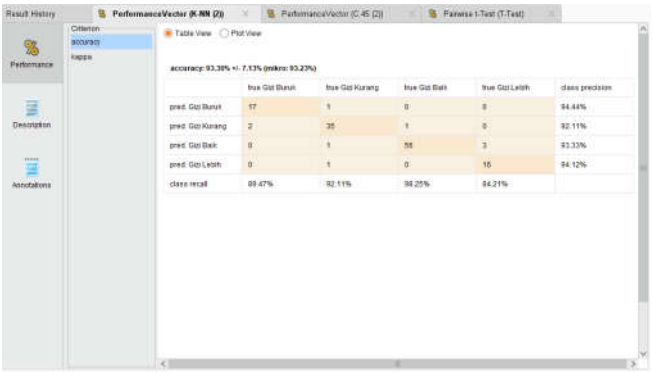
Gambar 7. *Cross Validation* metode K-NN

Selain dari *number of folds* metode K-NN memiliki nilai k harus ganjil ditunjukkan seperti gambar 8, karena mencari tetangga terdekat.



Gambar 8. Nilai k metode K-NN

Hasil hitung ditunjukkan gambar 9 dari dataset gizi anak dengan aplikasi *tools rapidminer* 9 menggunakan metode K-NN. hasil yang di dapatkan merupakan hasil dari metode K-NN yang menggunakan validasi ke – 10 dan nilai k : 3 dengan akurasi : 93.30 %, dimana akurasi tersebut akurasi tertinggi dari metode K-NN.



Gambar 9. Hasil Akurasi metode K-NN

Number of Validation ini digunakan untuk mencari parameter yang paling baik untuk sebuah algoritma klasifikasi, sedangkan nilai k yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, dengan menggunakan *cross validation*. Dengan demikian hasil dari validasi ke – 2 sampai ke – 10 dan k 1 samapai k 15 ditunjukkan pada tabel 3 hasil akurasi dari metode K-NN.

Tabel 3. Hasil Akurasi dari Metode K-NN

	Metode K-NN							
	K1	**K3**	**K5**	**K7**	**K9**	**K11**	**K13**	**K15**
Ke-2	77,44%	70,68%	72,98%	67,70%	63,91%	62,40%	60,91%	60,14%
Ke-3	79,76%	64,66%	65,47%	65,41%	66,93%	67,72%	67,69%	67,04%
Ke-4	84,25%	86,52%	81,22%	78,21%	69.90%	68,38%	65,42%	63,17%
Ke-5	84,27%	91,03%	79,80%	75,98%	69,20%	68,55%	66,24%	65,47%
Ke-6	87,25%	86,43%	80,47%	78,95%	72,92%	68,41%	68,45%	66,21%
Ke-7	83,46%	90,23%	77,44%	75,94%	69,17%	68,42%	63,16%	65,41%
Ke-8	86,35%	89,38%	78,22%	77,44%	73,76%	67,05%	67,00%	67,00%
Ke-9	85,13%	90,95%	82,06%	78,25%	73,02%	72,17%	68,47%	66,93%
Ke-10	84,89%	93,30%	80,44%	81,15%	70,66%	69,23%	67,75%	64,73%

Pada metode K-NN, pengujian ini mencari akurasinya dari validasi ke-10, dimana nilai k masing-masing validasi adalah 16, dan nilai k harus ganjil, karena mencari tetangga terdekat. Akurasi tertinggi adalah 93,30%, ada di validasi ke 10 dan nilai k 3. Metode K-NN sudah memiliki nilai akurasi yang tinggi yang bisa di katakan melebihi 90%.

C. Evaluasi dan Validasi Metode

Pada tahapan evaluasi dilakukan dengan memperoleh hasil *Accuracy*, kemudian evaluasi terhadap model mengukur akurasi dengan c*onfusion matrik* yang menitikberatkan pada *class* secara umum dan proses dengan menggunakan 10-fold *cross validation*.

Tabel 4. *Confusion Matrik*

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	TrueNegative (TN)

Formulasi perhitungan yang digunakan seperti dibawah ini:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (4.1)$$

$$Sensitivity = TP\ rate = \frac{TP}{TP+FN} \quad (4.2)$$

$$Specificity = TN\ rate \frac{TN}{TN+FP} \quad (4.3)$$

$$\text{FP}\ rate = \frac{FP}{FP+TN} \quad (4.4)$$

$$Precision = \frac{TP}{TP+FN} \quad (4.5)$$

$$Recall = \frac{TP}{TP+FP} \quad (4.6)$$

$$F - Measure = \frac{2RP}{R+P} \quad (4.7)$$

$$G - Mean = \sqrt{Sensitivity * Specificity} \quad (4.8)$$

Evaluasi dengan menggunakan *F-Measure*, rata-rata harmonil dari dua angka cenderung lebih dekat dengan lebih kecil dari dua, maka dari itu, nilai *F-Measure* yang tinggi dapat memastikan bahwa kedua data tersebut adalah *sensitivity* dan *presisi* yang cukup tinggi. Jika hanya kinerja kelas positif dianggap sebagai dua langkah penting maka itu disebut dengan *TP rate* dan *Possitive Predective Value* (*PP Value*). Pada dasarnya penghitungan (4.1), (4.5) dan (4.6) di atas akan digunakan sebagai evaluasi nilai yang didapatkan dari hasil dari akurasi dari algoritma di atas.

Di bawah ini terdapat beberapa *Confusion Matrik* dari beberapa metode, seperti apa yang sudah dipaparkan di atas tentang algoritmanya. *Confusion Matrik* yang akan diambil adalah nilai tertinggi dari dua metode yang sudah diuji coba. Akan tetapi yang akan diperdetail adalah metode C.45 dan K-NN, karena algoritma tersebut adalah pembahasan dalam penelitian ini.

Tabel 5. *Confusion Matrik* dari Metode C4.5

	true Gizi buruk	true Gizi Kurang	true Gizi Baik	true Gizi Lebih	class precision
Pred. Gizi Buruk	13	6	0	0	68,42%
Pred. Gizi Kurang	6	28	3	0	75,68%
Pred. Gizi Baik	0	3	51	7	83,61%
Pred. Gizi Lebih	0	1	3	12	75,00%
class recall	68,42%	73,68%	89,47%	63,16%	

Diketahui cara menghitung rumus untuk menghitung *accuracy* dapat di lihat pada rumus (4.1). Maka jika diimplementasikan akan seperti berikut:

$$Accuracy = \frac{13 + 28 + 51 + 12}{Jumlah\ Dataset}$$

$$Accuracy = \frac{104}{133}$$

$$Accuracy = 0{,}781955 = \mathbf{78,19}\%$$

Pada bagian di atas adalah cara untuk menghitung *accuracy*, dibagian ini akan dipaparkan bagaimana cara menghitung *class recall* dan *class precision* dapat di lihat pada rumus (4.5) dan (4.6). Konsepnya nilai dari acuan ditambah dengan semua nilai prediksi. Lebih jelasnya sebagai berikut:

$$Class\ recall = \frac{13}{13 + 6}$$

$$Class\ recall = \frac{13}{19}$$

$$Class\ recall = 0{,}6842105 = \mathbf{68,42}\%$$

Nilainya sama dengan tabel *confusion matrik* pada *class recall*. Dan untuk menghitung *class precision* sebagai berikut:

$$Class\ precision = \frac{13}{13 + 6}$$

$$Class\ precision = \frac{13}{19}$$

$$Class\ precision = 0{,}6842105 = \mathbf{68,42}\%$$

Nilainya sama dengan tabel *confusion matrik* pada *class precision.*

Tabel 6. *Confusion Matrik* dari Metode K-NN

	true Gizi buruk	true Gizi Kurang	true Gizi Baik	true Gizi Lebih	class precision
Pred. Gizi Buruk	17	1	0	0	94,44%
Pred. Gizi Kurang	2	35	1	0	92,11%
Pred. Gizi Baik	0	1	56	3	93,33%
Pred. Gizi Lebih	0	1	0	16	94,12%
class recall	89,47 %	92,11%	98,2 5%	84,21 %	

Pada *confusion matrik* pada metode K-NN dari nilai akurasi yang tertinggi, dari penghitungan dengan menggunakan validasi ke -10 dan K 3.

Diketahui cara menghitung rumus untuk menghitung *accuracy* dapat di lihat pada rumus (4.1). Maka jika diimplementasikan akan seperti berikut:

$$Accuracy = \frac{17 + 35 + 56 + 16}{Jumlah\ Dataset}$$

$$Accuracy = \frac{124}{133}$$

$$Accuracy = 0{,}9323308 = \mathbf{93,23}\%$$

Pada bagian di atas adalah cara untuk menghitung *accuracy*, dibagian ini akan dipaparkan bagaimana cara menghitung *class recall* dan *class precision*dapat di lihat pada rumus (4.5) dan (4.6). Konsepnya nilai dari acuan ditambah dengan semua nilai prediksi. Lebih jelasnya sebagai berikut:

$$Class\ recall = \frac{17}{17 + 2}$$

$$Class\ recall = \frac{17}{19}$$

$$Class\ recall = 0{,}8947368 = \mathbf{89,47}\%$$

Nilainya sama dengan tabel *confusion matrik* pada *class recall*. Dan untuk menghitung *class precision* sebagai berikut:

$$Class\ precision = \frac{17}{17 + 1}$$

$$Class\ precision = \frac{17}{18}$$

$$Class\ precision = 0{,}9444444 = \mathbf{94,44}\%$$

Nilainya sama dengan tabel *confusion matrik* pada

class precision.

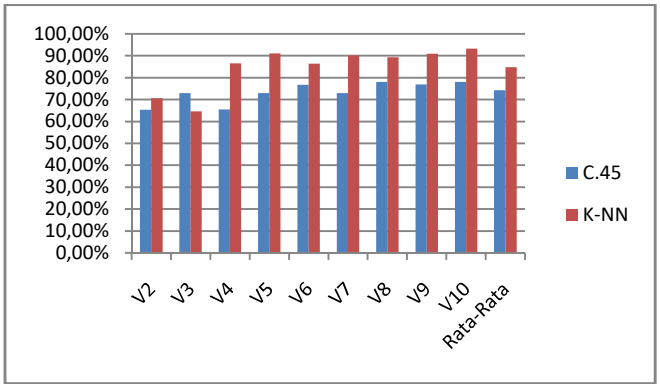
D. Analisis Hasil Komparasi

Setelah nilai akurasi sudah didapatkan dari dua metode C.45 dan K-NN, maka dari penelitian ini akan dibandingkan akurasinya dan hasil dari *T-Test*, hasil akurasi tertinggi pada dataset gizi anak dengan *T-Test* A : 0.933 dan C : 0.933 dengan demikian akurasi tertinggi yang dihasilkan *T-Test* sama dengan metode K-NN validasi ke – 10 dan k 3 akurasi: 93,30%. Untuk lebih detailnya perhatikan tabel 7 Perbandingan kedua metode yang sudah dipaparkan di atas, sebagai berikut:

Tabel 7. Perbandingan Metode Sesuai Akurasi

Eksperiment	**Metode**	
	C4.5	**K-NN dan pada K3**
Ke-2	65,40%	70,68%
Ke-3	72,91%	64,66%
Ke-4	65,46%	86,52%
Ke-5	72,99%	91,03%
Ke-6	76,75%	86,43%
Ke-7	72,93%	90,23%
Ke-8	78,03%	89,38%
Ke-9	76,89%	90,95%
Ke-10	78,02%	93,30%
Rata-rata	**74,34%**	**84,80%**

Dari hasil di atas akan dibuat model grafik dengan tujuan lebih diperjelas perbandingan dari metode C.45 dan K-NN yang digunakan dalam penelitian ini. kronologi gambar 10 Perbandingan grafik dari kedua metode yang sudah dipaparkan di atas sebagai berikut :



Gambar 10. Perbandingan Akurasi Grafik

Dari gambar grafik menunjukkan warna merah memiliki nilai yang lebih unggul dari pada warna biru, warna merah adalah metode K-NN.

V. KESIMPULAN

Berdasarkan pada pembahasan sebelumnya, tentang klasifikasi gizi anak di Puskesmas Paiton berdasarkan 3 faktor yaitu *variable* umur, berat badan dan tinggi badan. Dimana penelitian ini menggunakan komparasi dua metode, metode tersebut adalah C.45 dan K-NN, yang tujuannya mencari akurasi dan akurasi yang tertinggi akan digunakan dalam penelitian ini sebagai acuan dari penentuan gizi di Puskesmas Paiton Probolinggo. Penelitian ini menggunakan *k-fold* validasi dari ke-2 sampai 10. Nilai akurasi yang paling tinggi adalah metode KNN dengan akurasi **93,30%** di *k-fold* validasi ke-10 dan nilai k 3, sedangkan nilai tertinggi dari metode C.45 dengan akurasi **78,03** di *k-fold* validasi ke-8. Dan nilai rata-rata, dari kedua metode tersebut sebagai berikut: Nilai rata-rata C.45 adalah **74,34%** dan K-NN adalah **84,80%**. Jadi akurasi tertinggi pada metode K-NN.

DAFTAR PUSTAKA

[1] H. Saleh, M. Faisal, and R. I. Musa, "Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor," Simtek J. Sist. Inf. dan Tek. Komput., vol. 4, no. 2, pp. 120–126, 2019, doi: 10.51876/simtek.v4i2.60.

[2] M. A. Rahman, N. Hidayat, and A. A. Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," J. Pengemb. Teknol. Inf. dan Ilmu Komput. Vol. 2, No. 12, Desember 2018, hlm. 6346-6353 e-ISSN, vol. 2, no. 12, pp. 925–928, 2018.

[3] Kementrian Kesehatan RI, Infodatin : Situasi Balita Pendek di Indonesia. 2016.

[4] N. Hermaduanti and S. Kusumadewi, "Sistem Pendukung Keputusan Berbasis Sms untuk Menentukan Status Gizi dengan Metode K-Nearest Neighbor," Semin. Nas. Apl. Teknol. Inf., pp. 49–56, 2008.

[5] Y. D. Iriani, "Sistem Pendukung Keputusan Status Gizi Balita Menggunakan K-Nearest Neighbor," Universitas Jember, 2015.

[6] Y. R. Kaesmetan, "Klasifikasi Status Gizi Balita di Kelurahan Oesapa barat Menggunakan K-Nearest Neighbor," Multitek, vol. 11, no. 1, pp. 42–50, 2017.

[7] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, "Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 9, pp. 925–932, 2017.

[8] A. A. Afandie, M. N., Cholissodin, I., Supianto, "Implementasi Metode K-Nearest Neighbor Untuk Pendukung Keputusan Pemilihan Menu Makanan Sehat dan Bergizi," DORO Repos. J. Mhs. FILKOM Univ. Brawijaya, vol. 3, no. 1, 2014.

[9] Ida Mardelana, Dasar-Dasar Ilmu Gizi Konsep Dan Penerapan Pada Asuhan Keperawatan. Yogyakarta: Pustaka Baru Press, 2017.

[10] S. Supangat, A. R. Amna, and T. Rahmawati, "Implementasi Decision Tree C4.5 Untuk

Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun," Teknika, vol. 7, no. 2, pp. 73–78, 2018, doi: 10.34148/teknika.v7i2.90.

[11] M. Faid and M. Jasri, Teori dan praktek data mining dengan weka dan netbeans. Bali: Istiqlal Publishing Group, 2017.

[12] A. S. Hutasoit, P. Tarigan, and E. R. Siagian, "Implementasi Data Mining Klasifikasi Status Gizi Balita Pada Posyandu Medan Timur Dengan Menggunakan Metode C4.5," J. Pelita Inform., vol. 17, no. 2, pp. 120–125, 2018.

[13] R. Wajhillah and E. Mutiara, "Penerapan Algoritma C4.5 Untuk Diagnosa Status Gizi Pada Anak Balita Berbasis Website," Swabumi, vol. 4, no. 2, pp. 178–185, 2016.

[14] R. I. Ndaumanu and M. R. Arief, Kusrini, "Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor," Jatisi, vol. 1, no. 1, pp. 1–15, 2014, [Online]. Available: http://www.mdp.ac.id/jatisi/vol-1-no-1/JATISI_Vol_1_No_1_September_2014_1.pdf.