

PENERAPAN ALGORITMA RANDOM FOREST DAN CROSS VALIDATION UNTUK PREDIKSI DATA STUNTING

Fadellia Azzahra ¹, Nana Suarna ², Yudhistira Arie Wijaya ³

¹²³ Program Studi Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon, Indonesia
Email: ¹fadelliaazzahra3@gmail.com, ²st_nana@yahoo.com, ³yudhistira010471@gmail.com
Email Penulis Korespondensi : fadelliaazzahra3@gmail.com

Abstract- The problem that often occurs in children whose growth does not match their age is chronic malnutrition or often called *stunting*. The problem of *stunting* children also occurs a lot in one of the Kelurahan in Cirebon City, namely at UPT Puskesmas XYZ. The high number of *stunting* cases makes the Puskesmas hold routine Posyandu activities every month. From Posyandu activities, the Puskesmas will obtain *stunting* reports from manual recording which will then be inputted using Microsoft Excel. Identifying and inputting data manually has the risk of experiencing data inaccuracies and nutritional status, which is an indicator for determining various kinds of *stunting*, can take a long time. To overcome these problems it is necessary to apply data mining using the Random Forest and Cross Validation algorithms for predicting *stunting* data. The research that will be carried out aims to find out how the Random Forest and Cross Validation algorithms work and to get the best accuracy, precision, recall, and RMSE values. The results of the research that has been carried out by applying the Random Forest algorithm method obtained the results of a confusion matrix test with a 90:10 data division in predicting weight/height showing an excellent score for an accuracy of 77.55%, a recall of 32.88%, a precision of 21.97%, and an RMSE of 0.402 with model 40 trees. Evaluation with Cross Validation produces the best value at 10-fold by obtaining an accuracy of 78.33%, precision of 39.13%, recall of 25.01%, and RMSE of 0.428 by making a model of 100 trees.

Keywords : Random Forest, Cross Validation, Prediction, Stunting.

Abstrak- Permasalahan yang seringkali terjadi pada anak dengan pertumbuhan yang tidak sesuai usianya merupakan kekurangan gizi kronis atau seringkali disebut *stunting*. Permasalahan anak *stunting* juga banyak terjadi di salah satu Kelurahan yang ada di Kota Cirebon yaitu di UPT Puskesmas XYZ, dari tingginya kasus *stunting* ini membuat Puskesmas menggelar kegiatan Posyandu rutin di setiap bulannya. Dari kegiatan Posyandu, Puskesmas akan memperoleh hasil laporan *stunting* dari pencatatan manual yang kemudian akan dilakukan penginputan data menggunakan Microsoft Excel. Pengidentifikasian dan penginputan data secara manual beresiko mengalami ketidaktepatan data serta status gizi yang menjadi indikator penentuan *stunting* yang beranekaragam dapat menghabiskan waktu lama. Untuk mengatasi permasalahan tersebut perlu melakukan penerapan data mining dengan algoritma Random Forest dan Cross Validation untuk prediksi data *stunting*. Pada riset yang akan dilakukan bertujuan untuk mengetahui cara kerja algoritma Random Forest dan Cross Validation serta untuk mendapatkan nilai akurasi, presisi, recall, dan RMSE terbaik. Hasil riset yang telah dilakukan dengan penerapan metode algoritma Random Forest memperoleh hasil pengujian confusion matrix dengan pembagian data 90:10 dalam memprediksi BB/TB menampilkan perolehan nilai yang sangat baik untuk akurasi 77.55%, recall 32.88%, presisi 21.97%, serta RMSE 0.402 dengan membuat model 40 pohon. Evaluasi dengan

Cross Validation menghasilkan nilai terbaik pada 10-fold dengan mendapatkan nilai akurasi 78.33%, presisi 39.13%, recall 25.01%, serta RMSE 0.428 dengan membuat model 100 pohon.

Kata Kunci : Random Forest, Cross Validation, Prediksi, Stunting.

I. PENDAHULUAN

Permasalahan yang seringkali terjadi pada anak dengan pertumbuhan yang tidak sesuai usianya merupakan kekurangan gizi kronis atau seringkali disebut *stunting* [1]. Di Indonesia pada tahun 2021 angka *stunting* mengalami penurunan menjadi sebesar 24.4 % dan di Jawa Barat angka *stunting* mencapai 24.5%, namun angka *stunting* yang cukup tinggi terjadi juga di Kota Cirebon sebesar 30.6% [2]. Status gizi anak ditentukan berdasarkan standar antropometri indikator yang paling umum digunakan untuk menghitung angka *stunting* adalah indeks tinggi badan pada umur (TB/U), berat badan pada umur (BB/U), dan berat badan pada tinggi badan (BB/TB) [3]. Permasalahan anak *stunting* juga banyak terjadi di salah satu Kelurahan yang ada di Kota Cirebon yaitu di UPT Puskesmas XYZ, dari tingginya kasus *stunting* ini membuat Puskesmas menggelar kegiatan Posyandu rutin di setiap bulannya. Posyandu yang diselenggarakan ini merupakan bentuk pemantauan terhadap pertumbuhan anak dengan cara melakukan pengukuran berat badan dan tinggi badan anak. Dari kegiatan Posyandu, Puskesmas akan memperoleh hasil laporan *stunting* dari pencatatan manual yang kemudian akan dilakukan penginputan data menggunakan Microsoft Excel. Pengidentifikasian dan penginputan data secara manual beresiko mengalami ketidaktepatan data serta status gizi yang menjadi indikator penentuan *stunting* yang beranekaragam dapat menghabiskan waktu lama. Dengan adanya kegiatan riset yang dilakukan menggunakan metode prediksi pada data *stunting* dengan penerapan data mining ini diharapkan dapat membantu proses pengolahan data secara tepat dan efisien.

Penggunaan algoritma Random Forest dalam menyelidiki permasalahan sindrom koroner kronis, dalam meninjau fungsi machine learning untuk penilaian menggunakan metode yang mengacu pada IBM dalam program python. Model pembagian data 70:30 mendapatkan hasil terbaik dengan akurasi 83,45%, recall 92,4%, serta presisi 85% [4].

Algoritma Random Forest untuk prediksi probabilitas gagal jantung sukses mendapatkan nilai untuk akurasi sebesar 82,6087% sebaliknya dengan menggunakan K-Fold serta dioptimasi menggunakan GridSearchCV menjadi 85,058% [5].

Untuk prediksi mungkin terjadinya diabetes menggunakan 3 perbandingan algoritma *Naive Bayes*, *Support Vector Machine*, dan *Random Forest* dengan tools Weka. Menampilkan hasil yang diperoleh *Random Forest* mempunyai nilai yang tinggi pada akurasi sebesar 97,88% dibanding dengan menggunakan algoritma lain [6].

Klasifikasi menggunakan algoritma *K-Nearest Neighbors* untuk menentukan permasalahan *stunting* bersumber pada parameter yang digunakan. Pengujian yang sangat baik dengan *confusion matrix* 90:10 mempunyai akurasi 97%, *recall* 98,46%, *presisi* 96,25%, serta *f1-score* 97,34%. Uji *5-fold cross validation* menghasilkan nilai rata-rata untuk akurasi sebesar 95,26% [7].

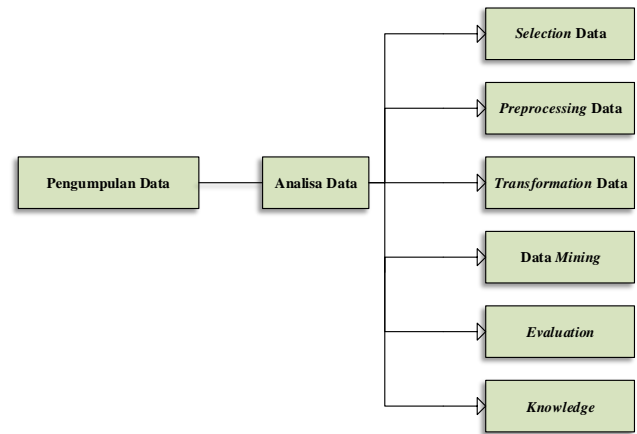
Klasifikasi status *stunting* anak dengan algoritma *K-Nearest Neighbor* menggunakan eliminasi *Backward*, akan mendapatkan hasil yang tepat dan cepat. Rata-rata akurasi perolehan dengan $k = 5$ sebesar 91,90% menggunakan 9 atribut, serta akurasi yang dihasilkan oleh algoritma *K-Nearest-Neighbor* dengan eliminasi *Backward* tambahan rata-rata memiliki nilai 92,20 % dengan menggunakan 8 atribut [8].

Riset dengan menggunakan algoritma *Random Forest* seringkali digunakan pada metode prediksi, karena dapat memperkirakan nilai secara cermat dan cepat [9]. Tidak hanya itu penggunaan algoritma *Random Forest* pula dapat menghindari *overfitting* [10]. *Overfitting* terjadi dikala model dilatih pada fitur yang lebih banyak bervariasi serta *underfitting* terjadi dikala model dilatih pada fitur yang lebih sedikit [11]. Dalam menghindari terbentuknya *overfitting* maupun *underfitting* terhadap machine learning dengan pengujian yang sudah dicoba, hingga validasi menggunakan metode *cross validation* baik untuk dicoba [12].

Pada riset yang akan dilakukan bertujuan untuk mengetahui cara kerja algoritma *Random Forest* dan *Cross Validation* serta mendapatkan nilai *akurasi*, *presisi*, *recall*, dan RMSE terbaik dari penelitian sebelumnya. Hasil yang diperoleh merupakan prediksi dari data *stunting* di UPT Puskesmas XYZ menggunakan tools RapidMiner dengan melewati proses *Knowledge Discovery in Database*. Pengujian dalam penelitian ini dilakukan menggunakan *confusion matrix* dengan pembagian 0.9 data latih 0.1 data uji dan dievaluasi menggunakan *cross validation* pada 3,5,7, dan 10-fold. Kombinasi *number of trees* pada algoritma *Random Forest* yang akan diuji dengan nilai *k-fold* adalah 20,40,60,80, dan 100 *trees* untuk prediksi berat badan pada tinggi badan data *stunting*. Penelitian ini diharapkan dapat memberikan solusi pada masalah yang terjadi, dengan melakukan metode prediksi dapat memberikan hasil analisis yang lebih baik dan detail, sehingga dapat membantu meningkatkan efisiensi dan efektivitas dalam mengelola data *stunting*.

II. METODOLOGI PENELITIAN

Metode yang digunakan dalam riset ini adalah deskriptif kuantitatif dengan prediksi menggunakan algoritma *Random Forest* dan *Cross Validation*. Tahapan dari penelitian yang dilakukan melalui proses sebagai berikut :



Gambar 1 Alur Penelitian

Pada metode penelitian yang digunakan dengan penerapan algoritma *Random Forest* dan *Cross Validation* dalam prediksi data *stunting* di UPT Puskesmas XYZ ini melalui tahapan proses *Knowledge Discovery in Databases* sebagai berikut :

A. Pengumpulan Data

Pengumpulan data dilakukan untuk mencari dan menentukan *objek* riset yang akan digunakan.

B. Analisa Data

1. Selection Data

Selection data dilakukan untuk memilih data yang digunakan dan berkaitan dengan riset, pemilihan ini dilakukan pada data *stunting*.

2. Preprocessing Data

Preprocessing data dilakukan untuk mengetahui data yang kosong. Jika terdapat data yang kosong pada data *stunting* perlu penghapusan data.

3. Transformation Data

Transformation data dilakukan untuk mengubah suatu tipe dari atribut data agar tepat dengan tipe data yang digunakan oleh algoritma *Random Forest*.

4. Data Mining

Data mining dilakukan untuk menerapkan algoritma *Random Forest* dalam proses prediksi berat badan pada tinggi badan data *stunting*.

5. Evaluation

Evaluation dilakukan untuk memperoleh suatu pola dari hasil proses evaluasi yang berguna dalam perkiraan tujuan yang dicapai. Di tahap ini dilakukan untuk menentukan nilai prediksi dari algoritma *Random Forest* yang akan menguji evaluasi nilai menggunakan *Confusion Matrix* dan *Cross Validation*.

6. Knowledge

Tahapan terakhir yaitu *Knowledge* penyampaian informasi hasil penelitian dari prediksi data *stunting* menggunakan tahapan KDD.

A. Algoritma Random Forest

Random forest ialah *classifier* yang memakai konsep *Bagging* serta *Bootstrapping*. Algoritma *Random Forest* memastikan nilai N awal selaku jumlah pohon yang

dibentuk ataupun buat membootstrap tiap pohon yang digunakan buat memilah informasi dari informasi pelatihan [13].

B. Confusion Matrix

Confusion Matrix merupakan perlengkapan ukur yang digunakan dalam menghitung performa ataupun akurasi dari proses klasifikasi. *Confusion Matrix* bisa digunakan buat menganalisa seberapa baik klasifikasi mengidentifikasi catatan dari bermacam jenis yang berbeda [14].

C. Prediksi

Secara universal prediksi sama semacam klasifikasi. Kerap digunakan di informasi *mining* ialah prediksi. Nilai hasil prediksi digunakan pada waktu mendatang bersumber pada informasi yang sudah terdapat lebih dahulu [15].

D. KDD atau Knowledge Discovery in Database

Knowledge Discovery in Database atau yang dikenal dengan tahapan KDD ini memiliki tujuan dalam eksplorasi juga analisis keseluruhan data dengan jumlah banyak serta ekstraksi informasi yang bermanfaat [16].

E. Cross Validation

Cross Validation merupakan teknik partisi data kedalam set *training* untuk latih model dan set *testing* untuk evaluasi model [17].

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan adalah proses penjelasan dari penelitian yang telah dilakukan, berikut merupakan hasil analisa data pada penerapan algoritma *Random Forest* dan *Cross Validation* untuk prediksi data *stunting*.

A. Hasil Pengumpulan Data

Pada pengumpulan data dilakukan observasi terlebih dahulu untuk memperoleh data yang ada di UPT Puskesmas XYZ. Dari hasil observasi tersebut data *stunting* balita yang akan dijadikan sebagai objek dari riset berhasil dikumpulkan sebanyak 503 data dan 14 atribut pada bulan Januari hingga September tahun 2022. Berikut ini merupakan tabel 3.1 dari data atribut *stunting* :

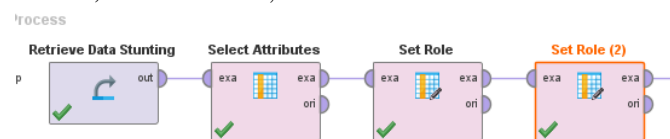
Tabel 1 Data Atribut

No	Data Atribut
1	No
2	Nama
3	Jenis Kelamin
4	Tanggal Lahir
5	Nama Orang Tua
6	Puskesmas
7	Posyandu
8	Umur Saat Ukur
9	Tanggal Pengukuran
10	Berat
11	Tinggi
12	BB/U
13	TB/U

B. Hasil Analisa Data

1. Selection Data

Pada tahapan ini dilakukan *selection* data terhadap data *stunting* menggunakan *tools* RapidMiner, atribut yang akan digunakan adalah atribut yang berpengaruh untuk menentukan perhitungan dalam riset yang akan dilakukan. Langkah pertama yang dilakukan adalah memilih operator *retrieve* untuk melihat keseluruhan data kemudian dilakukan seleksi menggunakan operator *select attribute*. Dari ke-14 atribut yang ada, setelah dilakukan seleksi kini menjadi hanya 8 atribut yang akan digunakan dalam riset antara lain Jenis Kelamin, Umur Saat Ukur, No, Tinggi, Berat, TB/U, BB/U dan BB/TB. Setelah itu, untuk menentukan suatu atribut yang bekerja sebagai label atau id dapat menggunakan operator *set role*. Berikut gambar 2 merupakan model proses dari operator *retrieve*, *select attribute*, dan *set role*.



Gambar 2 Model Selection Data

Pada gambar 3 menunjukan hasil dari operator *retrieve*, *select attribute*, dan *set role*. Untuk atribut yang dijadikan sebagai label adalah BB/TB yakni status penentu gizi pada data *stunting* yang akan dijadikan sebagai prediksi dan atribut No sebagai ID nilai pengidentifikasi yang unik.

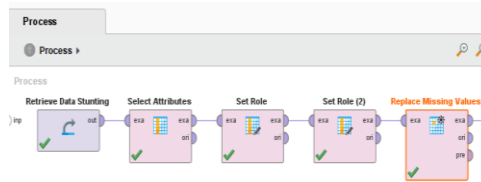
Row No.	No	BB/TB	JK	Umur Saat U...	Berat	Tinggi	BB/U	TB/U
1	1	Gizi Baik	L	58 Bulan	14.200	97.400	Berat Badan ...	Pendek
2	2	Gizi Baik	L	28 Bulan	10.200	83.600	Kurang	Pendek
3	3	Gizi Baik	L	25 Bulan	10.100	80.500	Berat Badan ...	Pendek
4	4	Gizi Lebih	P	30 Bulan	13	81.100	Berat Badan ...	Pendek
5	5	Risiko Gizi L...	L	20 Bulan	11	78	Berat Badan ...	Pendek
6	6	Gizi Baik	L	14 Bulan	8.100	72.800	Berat Badan ...	Pendek
7	7	Gizi Baik	P	0 Hari	3	47	Berat Badan ...	Pendek
8	8	Risiko Gizi L...	L	10 Bulan	9.500	69.500	Berat Badan ...	Pendek
9	9	Gizi Buruk	P	16 Bulan	6.300	71	Sangat Kurang	Pendek
10	10	Gizi Baik	P	16 Bulan	8.500	72.800	Berat Badan ...	Pendek
11	11	Gizi Lebih	P	6 Bulan	7.500	60	Berat Badan ...	Pendek
12	12	Gizi Kurang	L	0 Hari	2.800	50	Sangat Kurang	Pendek
13	13	Gizi Baik	P	20 Bulan	8	75.500	Kurang	Pendek
14	14	Gizi Baik	L	9 Bulan	7.100	67	Kurang	Pendek
15	15	Gizi Baik	P	49 Bulan	12.200	83.300	Kurang	Pendek

ExampleSet (503 examples, 2 special attributes, 6 regular attributes)

Gambar 3 Hasil Selection Data

2. Preprocessing Data

Preprocessing data dilakukan untuk pembersihan data yang teridentifikasi kosong atau tidak memiliki nilai. Pada proses ini dapat dilakukan menggunakan operator *missing value* di RapidMiner untuk meninjau apakah ada data yang kosong atau tidak konsisten. Dibawah ini gambar 4 merupakan model proses dari operator *missing value*.

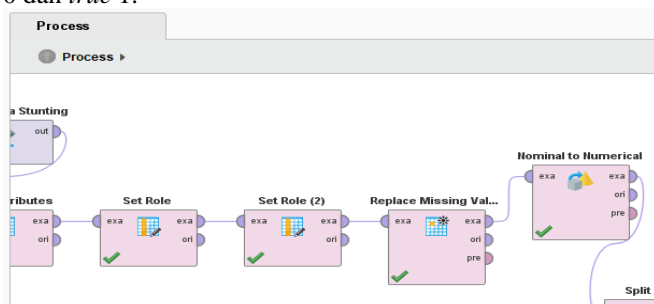


Gambar 4 Model Preprocessing Data

Berdasarkan hasil penerapan operator *missing value* pada data *stunting* tidak ada nilai atribut yang memiliki nilai kosong. Dengan seperti itu tidak perlu dilakukan *cleansing* data, karena data tetap lengkap sebesar 503 data tidak ada nilai yang kosong.

3. Transformation Data

Transformation data dilakukan untuk mengubah tipe atribut pada data *stunting* yang semula adalah data nominal, setelah dilakukan *transformation* akan berubah menjadi data numerik sesuai dengan ketentuan dari algoritma yang akan digunakan. Di RapidMiner untuk mengubah suatu tipe atribut dapat menggunakan operator *nominal to numerical*. Berikut gambar 5 merupakan model proses dari operator *nominal to numerical* dan gambar 6 menunjukkan hasil dari *transformation* tipe data pada atribut Jenis Kelamin, BB/U, dan TB/U pada data *stunting* yang berubah menjadi nilai *false* 0 dan *true* 1.



Gambar 5 Model Transformasi Data

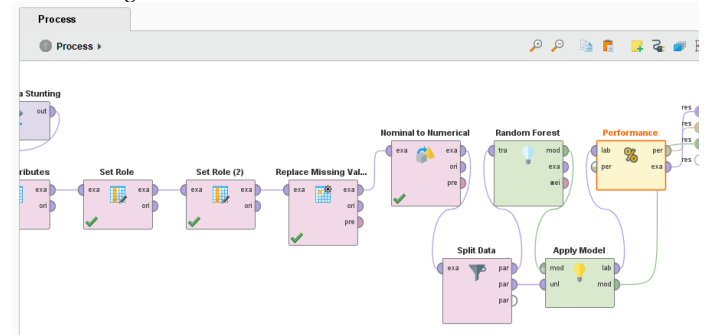
No	BB/TB	JK = L	JK = P	Usia Saat U...	Usia Saat U...	Usia Saat U...	Usia Saat U...	Usia Saat U...	Usia Saat U...
1	Gila Baik	1	0	1	0	0	0	0	0
2	Gila Baik	1	0	0	1	0	0	0	0
3	Gila Baik	1	0	0	0	1	0	0	0
4	Gila Lelah	0	1	0	0	0	1	0	0
5	Risiko Gila L...	1	0	0	0	0	0	1	0
6	Gila Baik	1	0	0	0	0	0	0	1
7	Gila Baik	0	1	0	0	0	0	0	0
8	Risiko Gila L...	1	0	0	0	0	0	0	0
9	Gila Buruk	0	1	0	0	0	0	0	0
10	Gila Baik	0	1	0	0	0	0	0	0
11	Gila Lelah	0	1	0	0	0	0	0	0
12	Gila Kurang	1	0	0	0	0	0	0	0
13	Gila Baik	0	1	0	0	0	0	1	0
14	Gila Baik	1	0	0	0	0	0	0	0

Gambar 6 Hasil Transformasi Data

4. Data Mining

Pada tahapan ini dilakukan penerapan data *mining* untuk prediksi data *stunting* yaitu dengan metode algoritma *Random Forest*. Riset yang dilakukan ini akan membuat sebuah model kombinasi dari nomor pohon pada algoritma *Random Forest* sekitar 20,40,60,80 dan 100 pohon. Pemodelan data *mining* di RapidMiner dapat menggunakan operator *Random Forest*, selanjutnya pembagian data sebelum dilakukan uji

menggunakan *confusion matrix* dengan split data sebesar 90 : 10 untuk data *training* juga data *testing*. Selanjutnya untuk melihat model dari data yang telah dilatih ini dapat menggunakan operator *apply* model pada RapidMiner dan hasil akhir untuk menampilkan nilai akurasi, presisi, *recall* serta RMSE dapat menggunakan operator *performance*. Berikut gambar 7 merupakan tampilan model proses untuk data *mining*.



Gambar 7 Model Data Mining

Pada tahapan data *mining* menggunakan algoritma *Random Forest* ini dari hasil uji *Confusion Matrix* 90:10 memperoleh nilai yang dapat dilihat pada tabel 3.2

Tabel 2 Hasil Uji Data Split 90:10

Pohon	Akurasi	Recall	Presisi	RMSE
20	77.55%	22.88%	20.30%	0.404
40	77.55%	32.88%	21.97%	0.402
60	77.55%	32.88%	21.97%	0.403
80	77.55%	25.00%	19.26%	0.406
100	77.55%	32.88%	21.97%	0.403

5. Evaluation

Tahapan *evaluation* adalah hasil dari proses evaluasi untuk mendapatkan pola baru dan perolehan nilai hasil prediksi dengan algoritma *Random Forest* untuk pengujian menggunakan *confusion matrix* dan *cross validation*.

a. Confusion Matrix

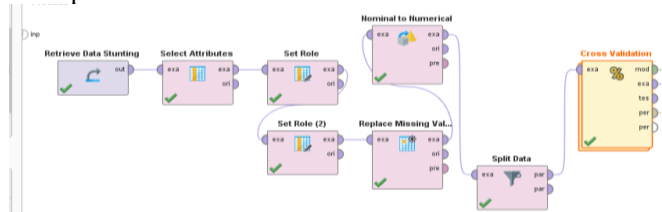
Pada tahapan ini perolehan nilai uji yang dihasilkan dengan *confusion matrix* menunjukkan bahwa dengan membangun model dari sebuah pohon akan menghasilkan nilai yang berbeda untuk setiap pohonnya. Berdasarkan pada tabel 3.2 menunjukkan jika perolehan nilai hasil uji menggunakan *confusion matrix* 90:10 dengan algoritma *Random Forest* nilai terbaik diperoleh untuk membuat model 40 pohon pada prediksi BB/TB menghasilkan nilai akurasi sebesar 77.55%, presisi 21.97%, *recall* 32.88% serta RMSE 0.402. Gambar 8 merupakan model evaluasi untuk prediksi data *stunting* BB/TB menggunakan perhitungan *confusion matrix*.

	true Gila Baik	true Gila Lelah	true Risiko Gila Lelah	true Gila Buruk	true Gila Kurang	true Obesitas	class precision
pred Gila Baik	36	1	4	0	3	0	81.82%
pred Gila Lelah	0	0	0	0	0	0	0.00%
pred Risiko Gila L...	0	0	0	0	0	0	0.00%
pred Gila Buruk	0	0	0	2	2	0	50.00%
pred Gila Kurang	1	0	0	0	0	0	0.00%
pred Obesitas	0	0	0	0	0	0	0.00%
class recall	97.30%	0.00%	0.00%	100.00%	0.00%	0.00%	

Gambar 8 Nilai Confusion Matrix

b. Cross Validation

Analisa hasil prediksi untuk data *stunting* BB/TB dievaluasi menggunakan *cross validation* dengan nilai *k-fold* yang akan diuji adalah 3,5,7,dan 10-fold. Pada RapidMiner untuk menguji nilai evaluasi dapat menggunakan operator *cross validation*, berikut ini gambar 3.8 adalah model proses dari operator *cross validation*.



Gambar 9 Model Evaluasi Cross Validation

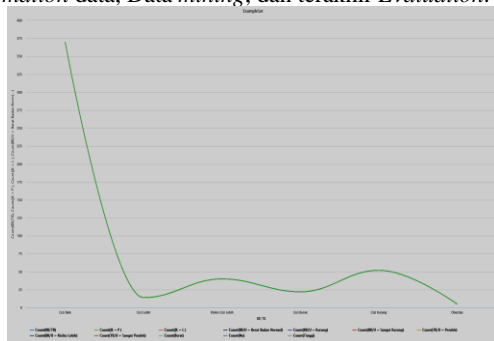
Untuk menguji nilai *k-fold cross validation* juga akan membuat model proses dengan nomor pohon 20,40,60,80,dan 100 pohon. Pada tabel merupakan hasil uji evaluasi dari setiap nilai *k-fold* yang paling optimal :

Tabel 3. 1 Prediksi Dengan Nilai K-Fold

Fold	Pohon	Akurasi	Recall	Presisi	RMSE
3	100	76.94%	26.12%	28.88%	0.435
5	100	77.73%	27.67%	43.56%	0.427
7	80	77.73%	25.97%	29.31%	0.430
10	100	78.33%	25.01%	39.13%	0.428

6. Knowledge

Pada tahapan ini dapat diketahui melalui gambar 10 merupakan grafik hasil penerapan algoritma *Random Forest* dan *Cross Validation* untuk prediksi data *stunting* melewati proses yang sesuai dengan tahapan KDD yaitu *Selection* data, *Preprocessing* data, *Transformation* data, *Data mining*, dan terakhir *Evaluation*.



Gambar 10 Visualisasi Grafik Prediksi BB/TB Stunting

Berdasarkan riset yang telah dilakukan dengan penerapan algoritma *Random Forest* dan *Cross Validation* untuk prediksi data *stunting* menghasilkan nilai prediksi BB/TB untuk Gizi baik, Gizi Lebih, Risiko Gizi Lebih, Gizi Kurang, Gizi Buruk, dan Obesitas. Perolehan nilai prediksi dengan algoritma *Random Forest* dengan pembagian data 90:10 memiliki nilai akurasi 77.55%, presisi 21.97%, *recall* 32.88% serta RMSE 0.402 pada model 40 pohon. Evaluasi dengan *Cross*

Validation menghasilkan nilai terbaik pada 10-fold dengan mendapatkan nilai akurasi 78.33%, presisi 39.13%, *recall* 25.01% serta RMSE 0.428 pada model 100 pohon.

IV. KESIMPULAN

Berdasarkan hasil riset yang telah dilakukan untuk mengetahui cara kerja algoritma *Random Forest* dan *Cross Validation* dengan membangun suatu model proses prediksi untuk data *stunting* dapat menggunakan *tools* RapidMiner dengan cara membuat operator serta mengelola parameter yang digunakan untuk prediksi BB/TB pada data *stunting*. Hasil riset yang telah dilakukan dengan penerapan metode algoritma *Random Forest* memperoleh hasil pengujian *confusion matrix* dengan pembagian data 90:10 dalam memprediksi BB/TB menampilkan perolehan nilai yang sangat baik untuk akurasi 77.55%, *recall* 32.88%, presisi 21.97%, serta RMSE 0.402 dengan membuat model 40 pohon. Evaluasi dengan *Cross Validation* menghasilkan nilai terbaik pada 10-fold dengan mendapatkan nilai akurasi 78.33%, presisi 39.13%, *recall* 25.01% serta RMSE 0.428 dengan membuat model 100 pohon.

V. REFERENCES

- [1] Kemenkes RI, "Laporan Prediksi Stunting Tahun 2020," 2020.
- [2] Kemenkes RI, *Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi dan Kabupaten/Kota Tahun 2021*, vol. 2, no. 1. 2021. doi: 10.36805/bi.v2i1.301.
- [3] "Peraturan Kemenkes RI," vol. 2, no. 1, pp. 1–12, 2020, [Online]. Available: <http://clik.dva.gov.au/rehabilitation-library/1-introduction-rehabilitation%0Ahttp://www.scrip.org/journal/doi.aspx?DOI=10.4236/as.2017.81005%0Ahttp://www.scrip.org/journal/PaperDownload.aspx?DOI=10.4236/as.2012.34066%0Ahttp://dx.doi.org/10.1016/j.pbi.201>
- [4] E. P. Cynthia, M. Afif Rizky A., A. Nazir, and F. Syafria, "Random Forest Algorithm to Investigate the Case of Acute Coronary Syndrome," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 369–378, 2021, doi: 10.29207/resti.v5i2.3000.
- [5] T. S. Parsaoran and Edric, "Prediksi Penyakit Gagal Jantung Dengan Menggunakan Random Forest," *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 176–181, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2445.
- [6] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021, doi: 10.32520/stmsi.v10i1.1129.
- [7] H. H. Sutarno, R. Latuconsina2, and A. Dinimaharawati3, "Prediksi Stunting Pada Balita Dengan Menggunakan Algoritma Klasifikasi K-Nearest Neighbors Stunting Prediction in Children Using K-Nearest Neighbors Classification

- Algorithm,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 6657–6661, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16456/16167>
- [8] S. Lonang and D. Normawati, “Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination,” *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 49, 2022, doi: 10.30865/mib.v6i1.3312.
- [9] M. Rianto and R. Yunis, “Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest,” *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 1, 2021, doi: 10.31294/p.v23i1.9781.
- [10] E. N. Candra, I. Cholissodin, and R. C. Wihandika, “Klasifikasi Status Gizi Balita menggunakan Metode Optimasi Random Forest dengan Algoritme Genetika (Studi Kasus: Puskesmas Cakru),” vol. 6, no. 5, pp. 2188–2197, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] J. Kolluri, V. K. Kotte, M. S. B. Phridviraj, and S. Razia, “Reducing Overfitting Problem in Machine Learning Using Novel L1/4 Regularization Method,” *Proc. 4th Int. Conf. Trends Electron. Informatics, ICOEI 2020*, no. September, pp. 934–938, 2020, doi: 10.1109/ICOEI48184.2020.9142992.
- [12] H. Gunawan and G. S. Budhi, “Penerapan Machine Learning dalam mendeteksi Fake Account pada Instagram,” no. 121, 2022.
- [13] Rakes, Jondri, and K. M. Lhaksamana, “Prediksi retweet berdasarkan feature user-based menggunakan metode klasifikasi Support Vector Machine,” vol. 8, no. 5, pp. 11183–11191, 2021.
- [14] G. Balita, S. Dengan, and P. K. Cross, “Aplikasi naïve bayes classifier (nbc) pada klasifikasi status gizi balita stunting dengan pengujian k-fold cross validation,” *J. Gaussian*, vol. 11, pp. 130–139, 2022.
- [15] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [16] S. K. M. P. Fitri Marisa, S. T. M. S. M. M. T. Anastasia Lidya Maukar, and S. S. M. M. S. I. Dr. Tubagus Mohammad Akhriza, *Data Mining Konsep Dan Penerapannya*. Deepublish, 2021.
- [17] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITO Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.