

Penerapan Algoritma Support Vector Machine dan K-Nearest Neighbor Menggunakan *Feature Selection Backward Elimination* Untuk Prediksi Status Penderita Stunting Pada Balita

Abdul Yunus Labolo¹, Sarlis Mooduto², Andi Bode³, Ivo Colanus Rally Drajana⁴

Teknik Informatika, Universitas Ichsan Gorontalo, Universitas Pohuwato

Email: abdulyunuslabolo@gmail.com¹, sarlis mooduto@gmail.com², andibode22@gmail.com³, ivocolanusrally@gmail.com⁴

Abstrak

Stunting adalah malnutrisi yang ditandai dengan tinggi badan, diukur dengan standar deviasi dari WHO. Dinas Kesehatan Provinsi Gorontalo khususnya dibidang Gizi mengenai stunting, selama ini melakukan kegiatan pemantauan tiap-tiap puskesmas dan posyandu. Pemantauan dan pendataan terkait stunting di berbagai puskesmas di wilayah Gorontalo merupakan faktor penting dalam menentukan faktor tumbuh kembang baik dalam kandungan maupun bayi yang dilahirkan. Masalah yang sering muncul adalah data yang dikumpulkan untuk underestimasi selalu tidak akurat setiap bulannya, karena hanya perkiraan yang dihitung berdasarkan kasus Puskesmas. Prediksi yang akurat diperlukan untuk mengatasi permasalahan yang ada. Data mining didefinisikan sebagai ekstraksi informasi berharga atau berguna dari industri pertambangan atau database yang sangat besar. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM) menggunakan *feature selection backward elimination*. Berdasarkan hasil eksperimen, diprediksi jumlah penderita stunting menggunakan algoritma *Support Vector Machine* (SVM), dan *k-Nearest Neighbor* (K-NN) menggunakan *Backward Elimination* (BE). Tingkat *error* terkecil hasil RMSE 2,476 pada algoritma *k-nearest neighbor*. Adapun perbandingan antara hasil prediksi jumlah penderita stunting dibulan januari yaitu 23 orang dengan data aktual jumlah penderita stunting yakni 26 orang. Hasil prediksi menghasilkan nilai keakuratan 88,46%.

Kata Kunci: prediksi; penderita stunting; SVM; K-NN; *backward elimination*

ABSTRACT

Stunting is a nutritional deficiency characterized by height, measured by the WHO standard deviation. The Gorontalo Provincial Health Office, especially in the field of nutrition related to stunting, has so far carried out monitoring activities at every puskesmas and posyandu. Monitoring and data collection related to stunting at various health centers in the Gorontalo area is an important factor in determining growth and development factors both in the womb and in babies born. The problem that often arises is that the data collected for underestimation is always inaccurate every month, because only estimates are calculated based on Puskesmas cases. Accurate predictions are needed to overcome the existing problems. Data mining is defined as the extraction of valuable or useful information from the mining industry or very large databases. This study uses the K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) algorithms by using the backward elimination feature selection. Based on the experimental results, it is predicted that the number of stunting sufferers uses the Support Vector Machine (SVM) algorithm, and k-Nearest Neighbor (K-NN) uses Backward Elimination (BE). The smallest error rate is the result of RMSE 2,476 on the k-nearest neighbor algorithm. The comparison between the predicted number of stunting sufferers in January is 23 people with the actual data on the number of stunting sufferers, which is 26 people. Prediction results produce an accuracy value of 88.46%.

Keywords: prediction; stunting; SVM; K-NN; backward elimination

A. PENDAHULUAN

Stunting adalah bentuk malnutrisi yang ditandai dengan tinggi badan yang bergantung pada usia, diukur dengan standar deviasi WHO. *Tuberkulosis* (U) menunjukkan masalah gizi kronis yang disebabkan oleh penderitaan jangka panjang, terutama kemiskinan, gaya hidup yang buruk, pola asuh yang buruk / kekurangan gizi sejak lahir, yang mengarah ke anak kecil (Setyawati. 2018: 834).

Pembangunan kesehatan saat ini sedang berlangsung di Provinsi Gorontalo untuk mencapai derajat kesehatan masyarakat. Salah satu indikatornya dapat diambil dari Indeks Pembangunan Manusia (IPM) Provinsi Gorontalo tahun 2013, yaitu 71,77. Kinerja ini mengalami peningkatan dibandingkan tahun 2012 sebesar 71,31. Kinerja IPM tidak terlepas dari keterlibatan lintas sektoral pemerintah dalam perencanaan

dan pelaksanaan program kesehatan, khususnya kesehatan ibu dan anak.. Dinas kesehatan Provinsi Gorontalo khususnya dibidang Gizi mengenai stunting, selama ini melakukan kegiatan pemantauan tiap tiap puskesmas dan posyandu.

Pemantauan dan pendataan terkait stunting di berbagai puskesmas di wilayah Gorontalo merupakan faktor penting dalam menentukan faktor tumbuh kembang baik dalam kandungan maupun bayi yang dilahirkan. Masalah yang sering muncul di Dinas Kesehatan Provinsi Gorontalo adalah ketika data dikumpulkan untuk evaluasi stunting selalu tidak akurat dan tidak pasti setiap bulannya, karena hanya perkiraan yang diperhitungkan berdasarkan kasus yang dilakukan oleh puskesmas.

Data mining didefinisikan sebagai ekstraksi data atau upaya untuk mengekstrak informasi yang berharga dan berguna dari database yang sangat besar dan merupakan pengetahuan yang tersembunyi di dalam database yang di proses untuk menemukan pola dan teknik statistik matematika, kecerdasan buatan, dan machine learning (Utomo. 2020: 437). Ada beberapa metode peramalan dalam data mining, antara lain *regresi linier*, *C-45*, *Naive Bayes*, *K-Nearest Neighbor*, *Support Vector Machine*, dan lainnya. Kemajuan ilmu pengetahuan dan teknologi telah meningkatkan pemahaman kita tentang berbagai bidang. lingkungan dan mengarah ke peristiwa yang lebih dapat diprediksi. Kecenderungan untuk dapat memprediksi peristiwa dengan lebih akurat, terutama dalam perekonomian, merupakan dasar perencanaan yang baik (Irfan. 2017: 9).

Pada penelitian ini, algoritma *K-Nearest Neighbor* (K-NN) *Support Vector Machine* (SVM) dipilih dengan memilih fitur eliminasi terbalik. Algoritma SVM digunakan dalam penelitian ini karena Algoritma SVM memiliki keunggulan menunjukkan kinerja yang sangat baik dalam

peramalan time series (Drajana. 2017: 116). Seperti halnya algoritma K-NN, algoritma ini merupakan teknik klasifikasi data dengan konsistensi yang kuat, dan K-NN memiliki keuntungan mengumpulkan data pelatihan yang ketat yang efektif terhadap sejumlah besar data pelatihan. (Bode. 2017: 188). Pemilihan atribut adalah proses memilih subset dari atribut awal dengan menghilangkan atribut fitur asing. Metode eliminasi mundur berkinerja lebih baik daripada metode statistik dalam memilih fitur. Performa terbaik dapat dicapai melalui sensitivitas, spesifisitas, dan akurasi (Bode. 2019: 104).

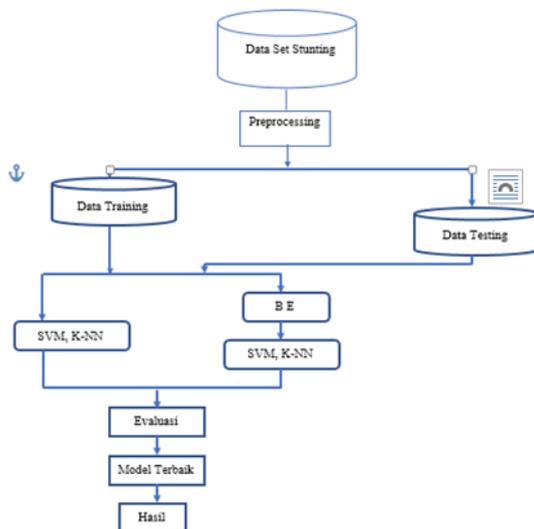
Pada penelitian ini, kedua metode tersebut diuji dengan menggunakan metode eliminasi terbalik untuk memprediksi status retardasi pertumbuhan kepada anak dibawah usia 5 tahun.

B. METODE

2.1. Jenis dan Desain Penelitian

Metode yang digunakan pada penelitian ini yakni analisis deskriptif menggunakan pendekatan kuantitatif dengan menekankan analisa terhadap data-data bersifat angka (numerik).

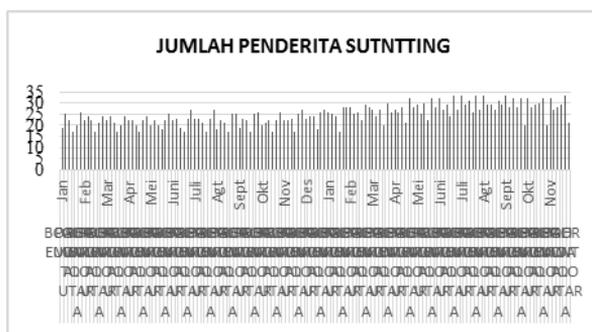
Masalah yang ada dalam penelitian ini berhubungan dengan datamining berupa data penderitanya stunting. Selanjutnya dilakukan proses pengolahan data set, data yang diolah berupa data bulanan yang akan diubah menjadi data bulanan *time series*. Kemudian data *time series* univariate akan diubah menjadi data *time series* multivariate. Model metode desain dilihat pada gambar 1 dibawah ini:



Gambar 1. Model Desain Metode Ekperimen

2.2. Data

Data merupakan data penderita stunting dari Dinas Kesehatan Provinsi Gorontalo. Data tersebut berupa jumlah penderita stunting tahun 2017 dan 2018. Bentuk data dapat dilihat pada 2 gambar dibawah ini:



Gambar 2. Data Jumlah Penderita Stunting seprovinsi Gorontalo 2018

2.3. Preprocessing

Pada tahap ini adalah pengolahan data awal sebelum di lakukan prediksi. Tahap awal menjelaskan bahwa data di ubah menjadi data *time*

series dan telah melaluitahap perubahan data dari *ascending* ke data *descending* sesuai dengan Tabel 1 dibawah ini:

Tabel 1. Bentuk Data Time Series

NAMA KAB/KOTA	BULAN	JUMLAH
BOALEMO	Jan	19
GORONTALO	Jan	25
POHUWATO	Jan	22
BONE BOLANGO	Jan	17
GORONTALO UTARA	Jan	20
KOTA GORONTALO	Jan	26
BOALEMO	Feb	22
GORONTALO	Feb	24
POHUWATO	Feb	22
BONE BOLANGO	Feb	17
GORONTALO UTARA	Feb	21
KOTA GORONTALO	Feb	24

Setelah data menjadi data *time series* kemudia data diubah dari data *univariate* menjadi *multivariate*. Data jumlah penderita sutntting diubah menjadi 5 periode dan memiliki variabel $xt-1$, $xt-2$, $xt-3$ $xt-4$ dan $xt-5$ Ini adalah variabel bebas dan xt adalah variabel terikat. Data *multivariate* ditunjukkan pada Tabel 2 di bawah ini:

Tabel 2. Multivariate 5 Periode

$xt-5$	$xt-4$	$xt-3$	$xt-2$	$xt-1$	xt
32	27	28	29	33	21
20	32	27	28	29	33
32	20	32	27	28	29
30	32	20	32	27	28
29	30	32	20	32	27
28	29	30	32	20	32
32	28	29	30	32	20
20	32	28	29	30	32
32	20	32	28	29	30
28	32	20	32	28	29
32	28	32	20	32	28
28	32	28	32	20	32
33	28	32	28	32	20
29	33	28	32	28	32
31	29	33	28	32	28

C. HASIL DAN PEMBAHASAN

Data yang dikumpulkan adalah data jumlah penderita sutntting dengan 138 *record* data. Tujuan dari penelitian ini adalah untuk mengetahui efektivitas algoritma *Support Vector Machine* (SVM) *K-Nearest Neighbor* (K-NN) dengan menggunakan opsi Back Elimination yang bertujuan untuk memberikan model prediksi terbaik.

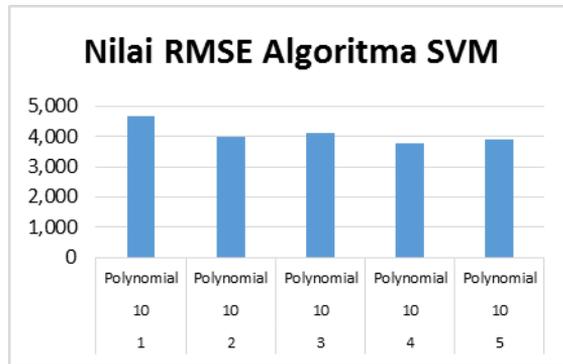
3.1. Parameter Support Vector Machine

Pada fase ini dilakukan pengujian untuk mendapatkan model terbaik dengan algoritma *support vector machine* (SVM) pengujian pada 10 *number of validation* pada *type kernel polynomial*. Nilai terkecil dari *root mean squared error* (RMSE) yang dihasilkan dipilih sebagai model terbaik yang dapat dilihat pada Tabel 3 sebagai berikut:

Tabel 3. Nilai RMSE Algoritma Support Vector Machine

	Periode Of Validation	Type Kernel	RMSE
1	10	Polynomial	4,660
2	10	Polynomial	4,007
3	10	Polynomial	4,121
4	10	Polynomial	3,752
5	10	Polynomial	3,896

Berdasarkan Tabel 3, tabel ringkasan hasil eksperimen pemilihan model menggunakan algoritma *Support Vector Machine* (SVM) dengan jumlah validasi 10 dan tipe kernel polinomial berdasarkan least square mean error (RMSE). Kemudian ditemukan model terbaik berdasarkan tingkat kesalahan terkecil pada variabel periode 4 dengan nilai RMSE sebesar 3,752. Grafiknya dapat dilihat pada Gambar 3 di bawah ini:



Gambar 3. Grafik Perbandingan Nilai RMSE Algoritma SVM

Sesuai dengan Gambar 3 menjelaskan bahwa dapat dilihat nilai error terkecil terdapat pada variabel periode 4 dengan nilai RMSE 3,752.

3.2. Parameter SMV Menggunakan Backward Elimination

Pada tahap ini dilakukan upaya untuk mendapatkan model terbaik dengan menguji algoritma *Support Vector Machine* menggunakan fungsi *backward elimination* pick pada 10 titik validasi untuk tipe kernel polinomial. Hasil RMSE ditunjukkan pada Tabel 4 di bawah ini.:

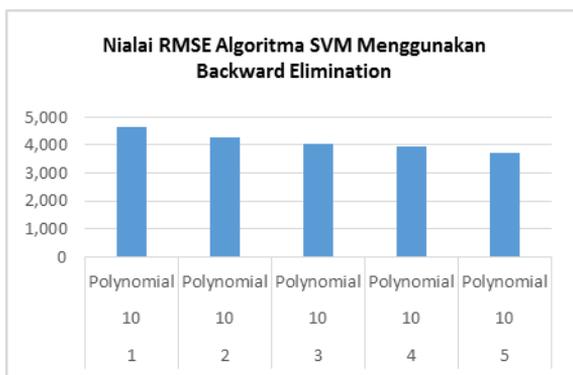
Tabel 4. Nilai RMSE Algoritma SVM Menggunakan *Backward Elimination*

Periode Of Validation	Type Kernel	RMSE
1	Polynomial	4,660
2	Polynomial	4,254
3	Polynomial	4,035
4	Polynomial	3,932
5	Polynomial	3,731

Berdasarkan Tabel 4 yang merupakan tabel ringkasan hasil pemilihan model *eksperimen* menggunakan algoritma *Support Vector Machine* (SVM), menggunakan fungsi seleksi eliminasi terbalik, rangkaian 10 jenis validasi *multicore* berdasarkan nilai kuadrat terkecil. (RMSE): Kemudian

model terbaik ditemukan berdasarkan kesalahan terkecil dari variabel periode ke-5 - 3731 Rs. Dengan demikian, menambahkan fitur seleksi dapat meningkatkan kinerja algoritma. Grafiknya dapat dilihat pada Gambar 4 di bawah ini:

Gambar 4. Grafik Perbandingan Algoritma SMV Menggunakan *Backward Elimination*



Sesuai dengan gambar 4 menjelaskan bahwa dapat dilihat nilai *error* terkecil terdapat pada variabel periode 5 dengan nilai RMSE 3,731.

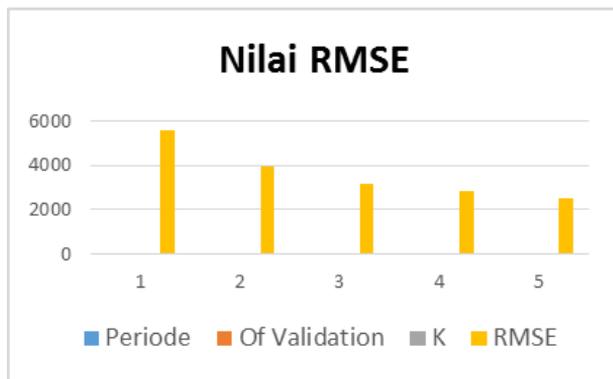
3.3. Parameter K-Nearest Neighbor

Pada tahapan selanjutnya untuk menentukan jarak atau jumlah tetangga terdekat, maka kemudian mengujikan dengan menggunakan metode KNN. Nilai RMSE uji coba KNN dapat dilihat pada Tabel 5 berikut:

Tabel 5. Nilai RMSE Algoritma K-Nearest Neighbor

Periode Of Validation	K	RMSE	
1	10	3	5,548
2	10	3	3,136
3	10	3	3,138
4	10	3	2,620
5	10	3	2,476

Tabel 5 merupakan hasil percobaan penentuan model menggunakan algoritma *k-nearest neighbor* dengan *cross-validation* 10 pada nilai 3 k3 berdasarkan *root mean square error* (RMSE). Model terbaik dari hasil eksperimen didasarkan pada nilai error terkecil untuk periode ke-5 yaitu nilai RMSE sebesar 2,476. Grafiknya dapat dilihat pada Gambar 5 di bawah ini:



Gambar 5. Grafik Perbandingan Niali RMSE Algoritma K-NN

Sesuai dengan Gambar 5 menjelaskan bahwa dapat dilihat nilai *error* terkecil terdapat pada variabel periode 5 dengan nilai RMSE 2,476.

3.4. Parameter K-NN Menggunakan Backward Elimination

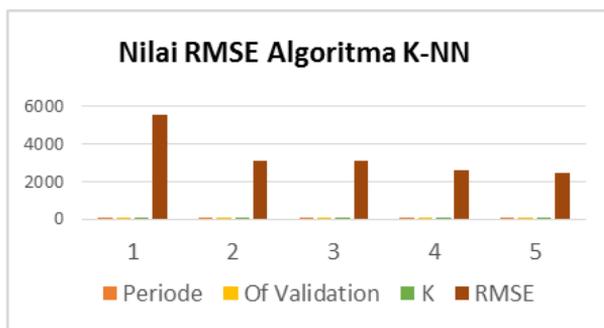
Dengan menggunakan opsi seleksi yaitu algoritma *K-Nearest Neighbor* dengan menggunakan *eliminasi* terbalik. Nilai RMSE dapat dilihat di bawah ini pada Tabel 6:

Tabel 6. Nilai RMSE Algoritma K-NN Menggunakan *Backward Elimination*

Periode	Of Validation	K	RMSE
1	10	3	5,548
2	10	3	3,953
3	10	3	3,175

4	10	3	2,862
5	10	3	2,485

Tabel 6 merupakan hasil penentuan model eksperimen menggunakan algoritma persamaan *k-nearest* menggunakan eliminasi regresi dengan validasi silang 10 berdasarkan 3 nilai berdasarkan RMSE. Model terbaik hasil eksperimen didasarkan pada nilai *error* terkecil terutama pada variabel periode ke-5 yaitu nilai RMSE 2,485. Dengan demikian, menambahkan fitur seleksi tidak memberikan algoritma K-NN nilai kesalahan minimum. Grafiknya dapat dilihat pada Gambar 6 di bawah ini:



Gambar 6. Grafik Perbandingan Algoritma K-NN Menggunakan *Backward Elimination*

Sesuai dengan Gambar 6 menjelaskan bahwa dapat dilihat nilai error terkecil terdapat pada variabel periode 5 dengan nilai RMSE 2,485.

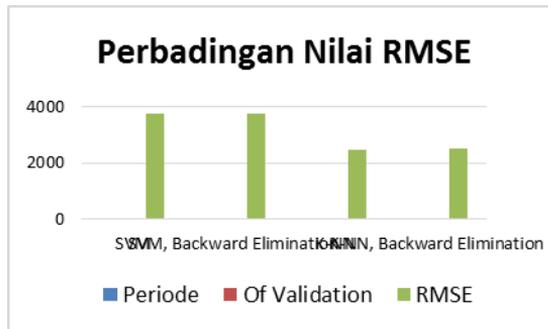
3.5. Evaluasi

Model dipilih berdasarkan *root mean square error* (RMSE) yang diperoleh. Berdasarkan hasil eksperimen, dimungkinkan untuk menentukan model mana yang digunakan untuk memprediksi jumlah penyandang tunagrahita, khususnya model dengan tingkat kesalahan paling rendah, lihat Tabel 7 di bawah ini.:

Tabel 7. Perbandingan Support Vectot Machine, *K-Nearest Neighbor* Menggunakan Fitur Seleksi

Model	Periode	Of Validation	RMSE
SVM	4	10	3,752
SMV, BE	5	10	3,731
K-NN	5	10	2,476
K-NN, BE	5	10	2,485

Tabel 7 yang membandingkan hasil algoritma *K-Nearest-Neighbor* dengan fungsi selektor eliminasi terbalik *Support Vector Machine*, menunjukkan bahwa tiga model teratas yang dihasilkan oleh algoritma *K-Nearest Neighbor* adalah tiga. hasil dari kesalahan terkecil. secara khusus nilai RMSE pada variabel periode ke-5 adalah 2,476.



Gambar 7. Grafik Perbandingan Model Terbaik

Sesuai dengan Gambar 7 yaitu grafik perbandingan model *Support Vector Machine* (SVM), *K-Nearest Neighbor* (K-NN) menggunakan fitur seleksi *Backward Elimination*. Dari grafik dapat dilihat model terbaik untuk data jumlah penderita sutntting adalah model algoritma *K-Nearest Neighbor*.

3.6. Implementasi

Pada tahapan implementasi ini adalah penerapan model terbaik algoritma *k-nearest neighbor* dari hasil eksperimen uji coba yang dilakukan yaitu dengan sampel data set 1 *record* untuk prediksi jumlah penderita stunting, hasil prediksi dapat dilihat dari Tabel 8 sebagai berikut:

Tabel 8. Hasil Prediksi Bulan Kedepan

Bulan	Jumlah
Januari	23

Sesuai dengan Tabel 8 yang berupa hasil prediksi satu bulan kedepan bulan januari 2019. Perbandingan pesentasi jumlah aktual dan hasil prediksi dapat dilihat pada Tabel 9 berikut:

Tabel 9. Perbandingan Persentasi Bulan Januari 2019

Bulan	Hasil Predisksi	Data Aktual	Persentasi
Januari	23	26	88,46%

Sesuai dengan Tabel 9 yang berupa hasil perbandingan antara hasil prediksi jumlah penderita stunting dengan data jumlah penderita stunting Dinas Kesehatan Provinsi Gorontalo Januari 2019. Hasil prediksi menghasilkan nilai keakuratan 88,46%, dengan demikian hasil prediksi dinyatakan berhasil karena keakuratan yang dihasilkan keakuratan baik diatas 50% mendekati nilai kesempurnaan.

Sampel perhitungan

$$\frac{\text{Hasil Prediksi}}{\text{Jumlah (Nilai Aktual)}} \times 100$$

$$\frac{23}{26} \times 100 = 88,46\%$$

D. PENUTUP

Simpulan dan Saran

Melihat hasil eksperimen yang telah diproses, pada prediksi jumlah penderita stunting dengan penerapan algoritma *support vector machine* (SVM) dan *k-nearest neighbor* (K-NN) menggunakan *backward elimination* (BE) telah berhasil dilakukan. Maka yakni nilai hasil dari prediksi tersebut dapat diimplementasikan untuk bahan pertimbangan atau kebijakan didalam pengambilan keputusan. Tingkat *error* terkecil hasil RMSE 2,476 pada algoritma *k-nearest neighbor* dibandingkan algoritma *support vector machine*. Adapun perbandingan antara hasil prediksi jumlah penderita stunting dibulan januari yaitu 23 orang dengan data aktual jumlah penderita stunting yakni 26 orang. Hasil prediksi menghasilkan nilai keakuratan 88,46%.

Saran kepada peneliti selanjutnya untuk menggunakan *feature selection* yang lain yang lebih tepat pada algoritma K-NN dengan tujuan menghasilkan nilai *error* lebih kecil dalam melakukan prediksi.

DAFTAR PUSTAKA

- Bode, A. 2017. K-nearest neighbor dengan feature selection menggunakan backward elimination untuk prediksi harga komoditi kopi arabika. *ILKOM Jurnal Ilmiah*, 9(2), 188-195.
- Bode, A. 2019. Perbandingan Metode Prediksi Support Vector Machine Dan Linear Regression Menggunakan Backward Elimination Pada Produksi Minyak Kelapa. *Simtek: jurnal sistem informasi dan teknik komputer*, 4(2), 104-107.
- Drajana, I. C. R. 2017. Metode support vector machine dan forward selection prediksi pembayaran pembelian bahan baku kopra. *ILKOM Jurnal Ilmiah*, 9(2), 116-123.

- Irfan, M., Ayuningtias, L. P., & Jumadi, J. 2017. Analisa Perbandingan Logic Fuzzy Metode Tsukamoto, Sugeno, Dan Mamdani (Studi Kasus: Prediksi Jumlah Pendaftar Mahasiswa Baru Fakultas Sains Dan Teknologi Uin Sunan Gunung Djati Bandung). *Jurnal Teknik Informatika*, 10(1), 9-16.
- Setyawati, V. A. V. 2018. Kajian stunting berdasarkan umur dan jenis kelamin di Kota Semarang. *Proceeding of The URECOL*, 834-838.
- Utomo, D. P., & Mesran, M. 2020. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437-444.