

***Backward Elimination Untuk Meningkatkan Akurasi Kejadian Stunting Dengan Analisis
Algoritma Support Vector Machine***

Agus Byna^{1*}, Fadhiyah Noor Anisa¹

¹Akademi Kebidanan Sari Mulia Banjarmasin

Email: agusbyna37@akbidsarimulia.ac.id, fadhiyah_noor_anisa@akbidsarimulia.ac.id

Abstrak

Latar Belakang: Prevalensi stunting pada balita di Indonesia masih tinggi terutama pada usia 2-3 tahun. Faktor risiko stunting antara lain panjang badan lahir, asupan, penyakit dan infeksi, genetik, dan status sosial ekonomi keluarga. Stunting terutama pada anak usia diatas 2 tahun sulit diatasi, sehingga penelitian mengenai faktor risiko stunting pada anak usia diatas 2 tahun diperlukan. Penggunaan Data di bidang kesehatan di perlukan sebagai tolak ukur untuk mencari hubungan, analisis, dan faktor-faktor

Tujuan: Selain memberikan keilmuan di bidang Informatikan juga berguna bagi ilmu kesehatan dalam pengelolaan data dalam mengambil keputusan. Serta Dapat mengurangi dampak tentang kejadian stunting

Metode: Backward Elimination dengan algoritma Support Vector Machine

Hasil: nilai akurasi sebesar 81.62% dan nilai AUC sebesar 0.921 dengan tingkat diagnose Excellent Classification, namun setelah dilakukan penambahan yaitu Backward Elimination dengan Algoritma Support Vector Machine nilai akurasi sebesar 90.16% dan nilai AUC sebesar 0,962 dengan tingkat diagnosa Excelent Classification. Dari 13 atribut menjadi 10 atribut, Sehingga kedua metode tersebut memiliki perbedaan tingkar akurasi yaitu sebesar 8,54% dan perbedaan nilai AUC sebesar 0,041.

Simpulan: Penerapan metode *Backward Elimination* dapat meningkatkan nilai akurasi pada algoritma *SVM* dan juga menseleksi atribut/variable.

Kata Kunci: *Backward Elimination, Data Mining, Kejadian Stunting, SVM.*

Backward Elimination to Improve the Accuracy of Stunting Events with Support Vector Machine Analysis***Abstrak***

Background: *The prevalence of stunting in infants in Indonesia is still high, especially at 2-3 years of age. Risk factors for stunting include birth length, intake, disease and infection, genetic, and family socioeconomic status. Stunting, especially in children aged over 2 years is difficult to overcome, so research on risk factors for stunting in children aged over 2 years is needed*

Purpose: *In addition to providing knowledge in the field of Informatics, it is also useful for health science in managing data in making decisions. Can reduce the impact of stunting events. Backward Elimination.*

Method: *with Support Vector Machine algorithm*

Result: *of accuracy value is 81.62% and AUC value is 0.921 with the level of diagnosis Excellent Classification, but after adding the Backward Elimination with Support Vector Machine Algebraic accuracy value is 90.16% and AUC value is 0.962 with level Excelent Classification diagnosis. From the 13 attributes to 10 attributes, so the two methods have differences in accuracy of 8.54% and the difference in AUC value is 0.041.*

Conclusion: *The application of the Backward Elimination method can increase the accuracy value of the SVM algorithm and also select the attribute /variable.*

Keywords: *Backward Elimination, Data Mining, Stunting Events, SVM.*

Pendahuluan

Dalam Bidang kesehatan secara luas sudah mengenal *data mining* yang memberikan informasi akurat. *Data mining* selalu digunakan dalam bidang kesehatan sebagai inovasi dalam mengambil keputusan yang akurat sehingga bermanfaat untuk pengembangan informasi diberikan menjadi lebih akurat.

Khususnya dalam kejadian stunting yaitu mempunyai permasalahan gizi yang kronis sehingga membutuhkan penanganan yang tepat. Saat melakukan penanganan perlu diketahui faktor risiko stunting diantaranya yaitu frekuensi kunjungan posyandu dan kenaikan berat badan (Destiadi, Nindya, & Sumarmi, 2016).

Jumlah anak mengalami kejadian stunting sangat banyak yang dapat memberikan dampak indikasi di masyarakat, masalah ini berlangsung cukup lama. Kejadian stunting dipengaruhi oleh banyak faktor diantaranya pemberian ASI eksklusif.

Masalah kesehatan masyarakat dianggap berat bila prevalensi pendek sebesar

30-39 persen dan serius bila prevalensi pendek ≥ 40 persen (WHO, 2013).

Banyak sekali para peneliti terutama pakar kesehatan menggunakan *data mining* sebagai pemanfaatan dalam mencari solusi yang terbaik dalam penanganan suatu penyakit. penelitian yang dilakukan pada data pasien penderita penyakit hepatitis menggunakan metode klasifikasi *data mining* melakukan perbandingan antara Algoritma C4.5 menghasilkan akurasi 77,29% dan nilai AUC 0,846 dan Algoritma Naive Bayes menghasilkan akurasi 83,71% dan nilai AUC 0,812 sehingga kedua metode ini akurat dalam melakukan prediksi untuk penyakit hepatitis (Septiani, 2017).

Penggunaan algoritma Naive Bayes dalam memprediksi terdapat hasil nilai akurasi Naive Bayes sebesar 89,08%, setelah dilakukan pemilihan fitur dengan menggunakan Backward Elimination didapatkan nilai akurasi yang lebih tinggi yaitu sebesar 90,69%, dengan melihat nilai akurasi maka algoritma Naive Bayes berbasis Backward Elimination meningkatkan akurasi

dalam memprediksi keputusan klien Telemarketing (Sulaehani,2016).

Penggunaan *SVM* dengan *PSO* yang menggunakan data dari instansi Pengukuran kinerja dilakukan dengan menghitung rata-rata error yang terjadi melalui besaran *Root Mean Square Error (RMSE)*. Semakin kecil nilai dari masing masing parameter kinerja ini menyatakan semakin dekat nilai prediksi dengan nilai sebenarnya. Dengan demikian dapat diketahui algoritma yang lebih akurat. Hasil *RMSE Support Vector Machines* Berbasis *PSO* adalah 37.685 (Setiawan,2016)

Dalam hal ini penggunaan metode *Support Vector Machine* yang mempunyai kelebihan *SVM* dapat menyelesaikan masalah *decision* khususnya jika sampel data yang ada kecil, tetapi *SVM* memiliki kelemahan pada sulitnya pemilihan fitur yang sesuai dan optimal pada bobot atribut yang digunakan sehingga menyebabkan tingkat akurasi prediksi menjadi rendah. Sehingga penulis memilih seleksi fitur variable yang akurat dengan menggunakan *Backward Elimination* untuk memaksimalkan tingkat akurasi dalam memprediksi kejadian stunting.

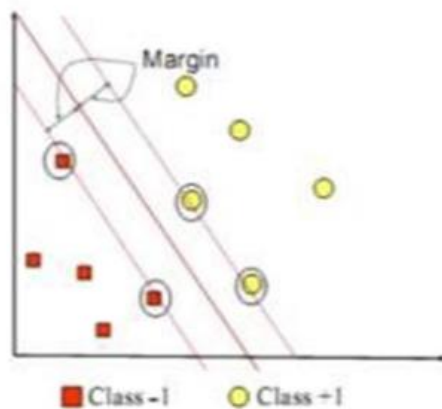
Bahan dan Metode

Stunting didefinisikan sebagai indikator status gizi TB/U sama dengan atau kurang dari minus dua standar deviasi (-2 SD) di bawah rata-rata dari standar. Stunting merupakan keadaan tubuh yang pendek dan sangat pendek sehingga melampaui defisit -2 SD dibawah median panjang atau tinggi badan (Manary & Solomons, 2009).

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data kemudian proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar(Iarose,2014).

Support Vector Machine adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. Gambar 1 memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class : +1 dan -1. Pattern yang tergabung pada class

- 1 disimbolkan dengan warna merah (kotak), sedangkan pattern pada class +1, disimbolkan dengan warna kuning (lingkaran). Problem klasifikasi dapat diterjemahkan dengan usaha menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut (Andina, Pham, 2007).

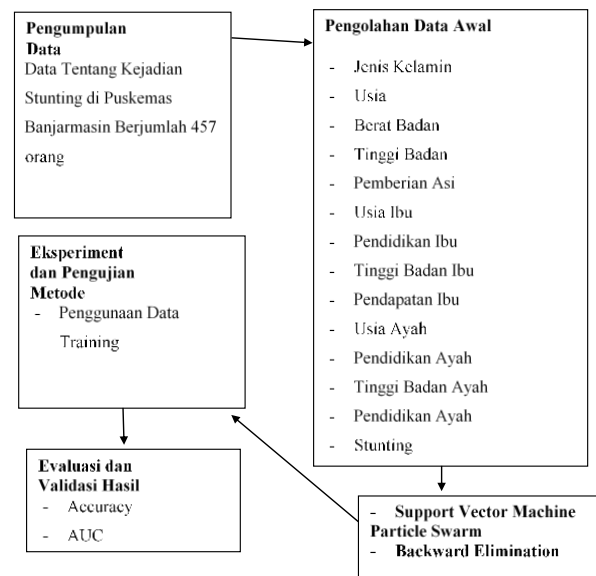


Gambar 1 SVM berusaha untuk menemukan hyperplane terbaik yang memisahkan

Backward Elimination merupakan metode yang dapat menghilangkan atribut yang tidak signifikan dari model (Vu, Muttaqi, & Agalgaonkar, 2015). Untuk prosedur Backward Elimination, model dimulai dengan semua atribut yang ada di dalamnya, dan atribut dengan persial statistik terkecil dihapus. Atribut yang digunakan akan diimplementasikan dengan Backward Elimination, sebelumnya data yang digunakan dijadikan data numerik dan diregresikan,

sebelum mendapatkan nilai F terlebih dahulu mencari nilai prediksi, rata-rata prediksi, MSR dan MSE.

Data yang digunakan yang berhubungan dengan tentang kejadian Stunting (bayi/anak yang bertubuh pendek) tahun 2018 di beberapa puskesmas Banjarmasin. Jumlah data sample yang di ambil adalah 457 data. Dari data tersebut menunjukkan 43% adalah pasien yang mengalami tentang kejadian Stunting dan 57% adalah pasien yang tidak mengalami tentang kejadian stunting.



Gambar 2 Metode Tahapan Penelitian

Pengelolaan data yang mempunyai atribut yang relevan sesuai dengan format input Algoritma Support Vector Machine, di gunakan 14 atribut/variable adalah Jenis

Kelamin,Usia,Berat Badan,Tinggi Badan, Pemberian Asi, Usia Ibu, Pendidikan Ibu,Tinggi Badan Ibu,Pendapatan Ibu,Usia Ayah,Pendidikan Ayah,Tinggi Badan Ayah,Pendidikan Ayah,Stunting.

Hasil

Uji coba Algoritma Support Vector Machine, dengan menampilkan nilai *weight* pada setiap attribute dengan menggunakan kernel type: *Dot*, *Kernel Gama* : 1.0, *C* : 0.0. maka hasilnya akan terlihat seperti di bawah :

Table 1 Tabel Atribut Nilai Weight SVM berdasarkan yang paling tinggi

Urutan	Atribut	Weight
1	Usia	28.49095
2	TB Ayah	12.62433
3	Pendidikan Ayah	10.10149
4	PD Ayah	9.601811
5	Pendidikan IBU	9.254653
6	PD IBU	2.372363
7	Usia Ibu	2.142191
8	TB IBU	0.389527
9	ASI	0.076472
10	JK	-1.67513
11	Usia Ayah	-5.57718
12	BBL/BB	-9.35166
13	PBL/TB	-50.0537

Pengujian Kembali di lakukan menggunakan *Cross X Validation* model confusion matrix akan membentuk matrix yang terdiri dari true positif atau tupel positif dan true negatif atau true sehingga didapatkan hasil accuracy dan AUC (Area Under Curve). berdasarkan nilai

C maka terlihat hasil pada table seperti berikut :

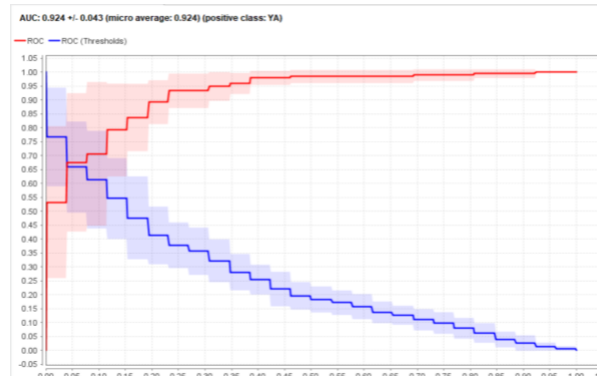
Tabel 2 Table perbandingan Accuracy dan AUC berdasarkan Nilai C

C	Accuracy	AUC
0.0	85.32%	0.92
1.0	85.32%	0.92
2.0	84.88%	0.92
3.0	85.76%	0.92
4.0	85.32%	0.92
5.0	85.54%	0.92
6.0	85.33%	0.917
7.0	86.20%	0.921
8.0	86.41%	0.924

Hasil terbaik pada eksperimen SVM di atas adalah dengan C=8.0 dan dihasilkan accuracy 86.41 % dan AUC nya 0.924. di bawah ini adalah table hasil akurasi.

Tabel 3 Hasil Akurasi Algoritma SVM C =8.0

	TRUE TIDAK	TRUE YA	CLASS PRECISION
PRED. TIDAK	229	31	88.08%
PRED. YA	31	166	84.26%
CLASS RECALL	88.08%	84.26%	



Gambar 3 Nilai AUC dalam grafik ROC algoritma SVM

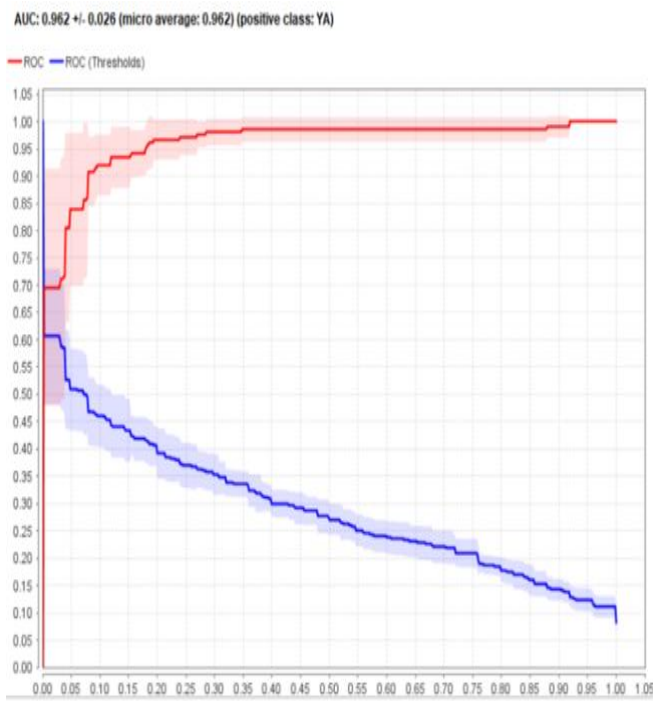
Hasil Backward Elimination dengan Algoritma SVM (Support Vector Machine)

ini, dari 13 atribut yang digunakan pada data data set kejadian stunting di seleksi menggunakan Backward Elimination menjadi 3 atribut yaitu jenis kelamin, berat badan, tinggi badan kemudian, di dalam testing ada fungsi apply model dan performance yang akan menampilkan nilai accuracy dan nilai AUC beserta grafik ROC pada saat di jalankan seperti table di bawah ini .

Tabel 4 Hasil Akurasi Metode Backward Elimination dengan Algoritma SVM

	true TIDAK	true YA	class precision
pred. TIDAK	247	32	88.53%
pred. YA	13	165	92.70%
class recall	95.00%	83.76%	

Berdasarkan data yang terdapat pada *confusion matrix* diatas maka nilai accuracy menunjukan bahwa Backward Elimination dengan algoritma SVM adalah sebesar 90.16%.



Gambar 4 Nilai AUC dalam grafik ROC Metode Backward Elimination dengan Algoritma SVM

Hasil nilai AUC (Area Under Curve) yang didapat dari pengolahan ROC untuk metode Backward Elimination dengan algoritma SVM dengan menggunakan data training adalah 0.962 dapat dilihat pada gambar 4 dengan tingkat diagnosa excellent classification.

Pembahasan

Metode Backward Elimination dengan algoritma SVM memiliki nilai akurasi yang lebih tinggi dibanding hanya menggunakan algoritma SVM. Selisih akurasi sebesar 3.75%, dapat dilihat pada Tabel 5 dibawah ini.

Table 5 Hasil Analisis Pengujian Algoritma SVM dan Backward Elimination dengan Algoritma SVM

	Akurasi	AUC
Algoritma SVM	86.41%	0.924
Backward Elimination Algoritma SVM	90.16%	0.962
Analisis Perbandingan	3.75%	0.038

Metode Backward Elimination dengan Algoritma SVM menghasilkan nilai 0.962 dengan nilai diagnosa *Excelent Classification*, yang artinya mempunyai nilai diagnosa *Excelent Classification* dan selisih nilai keduanya sebesar 0.38. Dengan demikian metode Backward Elimination dengan algoritma klasifikasi SVM selain dapat meningkatkan nilai akurasi dalam memprediksi klasifikasi juga dapat mengeliminasi atribut agar menjadi lebih mudah dengan hasil yang lebih akurat sehingga dapat memberikan solusi permasalahan bagi pakar kesehatan didalam menentukan tentang kejadian stunting.

Hasil penelitian untuk nilai akurasi algoritma SVM senilai 81.62%, sedangkan untuk nilai akurasi Backward Elimination dengan Algoritma SVM senilai 90.16% sehingga selisih untuk nilai akurasi sebesar 8.54%. Dan dievaluasi menggunakan ROC

curve untuk keduanya yaitu, nilai AUC berdasarkan ROC curve untuk algoritma SVM bernilai 0.921 dengan tingkat diagnosa *Excellent classification*, sedangkan untuk Backward Elimination dengan Algoritma SVM bernilai 0.962 dengan tingkat diagnosa *Excellent classification* yang sama-sama mempunyai nilai klasifikasi yang memuaskan, sehingga didapatkan selisih nilai AUC sebesar 0.041, tidak hanya meningkatkan akurasi Backward Elimination juga menseleksi attribute/variable yang sebelumnya berjumlah 13 menjadi 10 variabel/attribute, dari seleksi ini lah akurasi yang di hasilkan meningkat. Penerapan metode *Backward Elimination* dapat meningkatkan nilai akurasi pada algoritma SVM dan juga menseleksi atribut/variable.

Daftar Pustaka

- Andina, D. and Pham, D.T., 2007. *Computational intelligence*. Berlin Heidelberg New York: Springer.
- Gorunescu, F. 2011. *Data Mining Concepts, Model and Techniques*. Berlin: Springer.
- Ina, W.T., 2013. *Klasifikasi Data Rekam Medis Berdasarkan Kode Penyakit Internasional*

- Kusuma, K. E., & NURYANTO, N. (2013). Faktor risiko kejadian stunting pada anak usia 2-3 tahun (Studi di Kecamatan Semarang Timur) (Doctoral dissertation, Diponegoro University).
- Kusrini dan Luthfi, E. T., 2009. *Algoritma Data Mining*, Andi Offset, Yogyakarta
- Larose, D.T., 2014. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.
- Menggunakan Algoritma C4. 5. *Jurnal Media Elektro*, 2(1), p.105.
- Manary, M.J. & Solomons, N.W. 2009. *Gizi Kesehatan Masyarakat Gizi dan Perkembangan Anak*. Penerbit Buku Kedokteran EGC.
- Marmi dan Kukuh Rahardjo. 2012. *Asuhan Neonatus, Bayi, Balita, dan Anak Prasekolah*. Yogyakarta : Pustaka Pelajar.
- Narin, A., Isler, Y., & Ozer, M. 2014. Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance. *Computers in biology*.
- Rudolf, M. & Levene, M. 2006. *Paediatrics and Child Health*. Blackwell Publishing.
- Roesli. 2000. *Mengenal ASI Eksklusif*. Jakarta: Pustaka Pengembangan Swadaya Nusantara.
- Septiani, W. D. 2017. Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76-84.
- Soetjiningsih. 2004. *Tumbuh Kembang Anak*. Penerbit Buku Kedokteran (EGC), Jakarta.
- Vu, D. H., Muttaqi, K. M., & Agalgaonkar, A. P. 2015. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, 140, 385-394.
- Yu, L., Chen, G., Koronios, a., Zhu, S., & Guo, X. 2007. *Application and Comparison of Classification Techniques in Controlling Credit Risk*. World Scientific, 111.s