



SISTEM PREDIKSI STATUS GIZI BALITA DENGAN MENGGUNAKAN *SUPPORT VECTOR REGRESSION*

RYAN HIDAYAT



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2013**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul Sistem Prediksi Status Gizi Balita dengan Menggunakan *Support Vector Regression* adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2013

Ryan Hidayat
NIM G64090093

Hak Cipta Dilindungi Undang-Undang

Hak Cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

ABSTRAK

RYAN HIDAYAT. Sistem Prediksi Status Gizi Balita dengan Menggunakan *Support Vector Regression*. Dibimbing oleh WISNU ANANTA KUSUMA dan HELDA KHUSUN.

Anak di bawah usia lima tahun (balita) adalah kelompok usia yang paling rentan dalam hal gizi dan kesehatan di suatu kelompok masyarakat. Masalah gizi sering kali terjadi pada usia balita, namun diagnosis masalah gizi saat ini masih dilakukan dengan mengukur langsung indikator gizi seperti berat badan, tinggi atau penanda biokimia dari beberapa nutrisi. Dengan kemajuan teknologi dalam komputasi dan program pengolahan data statistika, data non-nutrisional yang tersedia dapat digunakan untuk memprediksi status gizi balita. Penelitian ini bertujuan untuk meneliti penggunaan *support vector regression* (SVR) sebagai metode *machine-learning* untuk menemukan model yang dapat memprediksi status gizi balita serta mengembangkan sistem prediksi dari model SVR tersebut. Model terbaik dihasilkan dengan *kernel* RBF, yang memiliki tingkat korelasi tertinggi dan galat terendah di tiap jenis Z-score yang diprediksi. Dengan model SVR terbaik, dapat dikembangkan sistem yang dapat memprediksi Z-score, walaupun belum cukup akurat.

Kata kunci: balita, gizi, *machine-learning*, prediksi, SVR

ABSTRACT

RYAN HIDAYAT. Prediction System for Nutritional Status of Children Using Support Vector Regression. Supervised by WISNU ANANTA KUSUMA dan HELDA KHUSUN.

Children under five years are the most vulnerable age group in terms of nutrition and health in a community. Nutritional problems often occur at the age of five, but the diagnosis of malnutrition is still done by directly measuring nutrition indicators such as weight, height or biochemical markers of some nutrients. With the technological advances in computing and statistical data processing program, the available non-nutritional data can be used to predict the nutritional status of children. The objective of this study was to investigate the use of support vector regression (SVR) as a machine-learning method to find models that can predict the nutritional status of children and to develop prediction system from the SVR models. The best model was produced by RBF kernel, with the highest degree of correlation and the lowest error in each type of Z-score predicted. With the best SVR model, a system that can predict Z-score can be developed, although it is not quite accurate.

Keywords: children, machine-learning, nutrition, prediction, SVR



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

SISTEM PREDIKSI STATUS GIZI BALITA DENGAN MENGGUNAKAN *SUPPORT VECTOR REGRESSION*

RYAN HIDAYAT

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer
pada
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2013**



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Judul Skripsi: Sistem Prediksi Status Gizi Balita dengan Menggunakan *Support Vector Regression*

Nama : Ryan Hidayat

NIM : G64090093

Disetujui oleh

Dr Wisnu Ananta Kusuma, ST, MT
Pembimbing I

Ir Helda Khusun, MSc, PhD
Pembimbing II

Diketahui oleh

Dr Ir Agus Buono, MSi, MKom
Ketua Departemen

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah prediksi status gizi, yang merupakan bentuk kerja sama dengan SEAMEO RECFON pusat penelitian makanan dan nutrisi. Karya ilmiah ini berjudul Sistem Prediksi Status Gizi dengan Menggunakan *Support Vector Regression*.

Terima kasih penulis ucapkan kepada Bapak Dr Wisnu Ananta Kusuma, ST, MT dan Ibu Ir Helda Khusun, MSc, PhD selaku pembimbing. Di samping itu, penghargaan penulis sampaikan kepada lembaga SEAMEO RECFON, yang telah membantu selama pengumpulan data. Ungkapan terima kasih juga disampaikan kepada ayahanda Harianto, ibunda Nurjanah, kakak Bunga Oktora, adik Firmansyah dan seluruh keluarga, serta rekan-rekan atas segala doa dan kasih sayangnya.

Semoga karya ilmiah ini bermanfaat.

Bogor, Agustus 2013

Ryan Hidayat

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	2
METODE	3
Studi Pustaka	3
Pengumpulan Data	3
Pemilihan Data dan <i>Data Cleaning</i>	3
Support Vector Regression (SVR)	4
<i>Grid Search</i>	7
Status Gizi	7
Pemodelan Berdasarkan Fungsi <i>Kernel</i> SVR	8
Analisis dan Pemilihan Model Terbaik	8
HASIL DAN PEMBAHASAN	9
Pemilihan Data dan <i>Data Cleaning</i>	9
Kinerja Model Berdasarkan Fungsi <i>Kernel</i> SVR	11
Analisis dan Pemilihan Model Terbaik	20
SIMPULAN DAN SARAN	22
Simpulan	22
Saran	22
DAFTAR PUSTAKA	22
LAMPIRAN	24
RIWAYAT HIDUP	25

Hak Cipta Diliindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

DAFTAR TABEL

1	Peubah sebagai pengurutan	10
2	Peubah respon	10
3	Peubah penjelas	10
4	Perbandingan nilai R dan MSE dari tiap kernel pada model $Z_{TB/U}$	12
5	Perbandingan nilai R dan MSE dari tiap kernel pada model $Z_{BB/TB}$	15
6	Perbandingan nilai R dan MSE dari tiap kernel pada model $Z_{BB/U}$	18

DAFTAR GAMBAR

1	Diagram alir metode penelitian	4
2	Kategori kurang gizi (Atmarita dan Fallah 2004)	8
3	Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> RBF	12
4	Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> linear	12
5	Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> polinomial	13
6	<i>Scatter plot</i> $Z_{TB/U}$ aktual dan prediksi dengan <i>kernel</i> RBF	13
7	<i>Scatter plot</i> $Z_{TB/U}$ aktual dan prediksi dengan <i>kernel</i> linear	13
8	<i>Scatter plot</i> $Z_{TB/U}$ aktual dan prediksi dengan <i>kernel</i> polinomial	14
9	Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> RBF	15
10	Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> linear	15
11	Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> polinomial	16
12	<i>Scatter plot</i> $Z_{BB/TB}$ aktual dan prediksi dengan <i>kernel</i> RBF	16
13	<i>Scatter plot</i> $Z_{BB/TB}$ aktual dan prediksi dengan <i>kernel</i> linear	16
14	<i>Scatter plot</i> $Z_{BB/TB}$ aktual dan prediksi dengan <i>kernel</i> polinomial	17
15	Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> RBF	18
16	Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> linear	18
17	Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan <i>kernel</i> polinomial	19
18	<i>Scatter plot</i> $Z_{BB/U}$ aktual dan prediksi dengan <i>kernel</i> RBF	19
19	<i>Scatter plot</i> $Z_{BB/U}$ aktual dan prediksi dengan <i>kernel</i> linear	19
20	<i>Scatter plot</i> $Z_{BB/U}$ aktual dan prediksi dengan <i>kernel</i> polinomial	20
21	Perbandingan kinerja model SVR dengan <i>kernel</i> RBF untuk tiap jenis Z-score berdasarkan tingkat korelasi	21
22	Perbandingan kinerja model SVR dengan <i>kernel</i> RBF untuk tiap jenis Z-score berdasarkan galat	21

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

PENDAHULUAN

Latar Belakang

Anak di bawah usia lima tahun (balita) adalah kelompok usia yang paling rentan dalam hal gizi dan kesehatan di suatu kelompok masyarakat karena masa itu merupakan masa peralihan antara saat disapih dan mulai mengikuti pola makan orang dewasa. Suatu bangsa dianggap berhasil apabila tersedianya sumber daya manusia (SDM) yang berkualitas, yaitu SDM yang memiliki fisik yang tangguh, mental yang kuat dan kesehatan yang prima di samping penguasaan terhadap ilmu pengetahuan dan teknologi (Atmarita dan Fallah 2004). Proses pertumbuhan dan perkembangan balita sangat berpengaruh terhadap kualitas SDM di masa depan sehingga kesejahteraan dari kelompok usia ini merupakan indikator kesejahteraan masyarakat. Kelompok usia ini sangat tergantung pada penyediaan makanan dari orang tua mereka untuk memenuhi kebutuhan gizi mereka. Kemampuan orang tua untuk memenuhi kebutuhan gizi anak tergantung dari beberapa hal, antara lain pendapatan, ketersediaan makanan dan pengetahuan mereka (Atmarita dan Fallah 2004).

Selain itu, balita juga masih memiliki sistem kekebalan tubuh yang belum matang dan sangat rentan terhadap penyakit menular (Meliala dan Poerwanto 1991). Hal ini terbukti terkait dengan penurunan status gizi. Insiden penyakit menular sangat berhubungan dengan lingkungan dimana subjek hidup serta ketersediaan dan aksesibilitas pelayanan kesehatan di daerah tersebut. Dengan demikian, dapat disimpulkan bahwa status gizi anak-anak berhubungan dengan berbagai faktor yang saling terkait. Jika diurutkan secara sistematis, determinan yang berpengaruh pada masalah gizi yang dapat terjadi pada masyarakat terutama balita diawali dengan faktor penyebab langsungnya yaitu makan tidak seimbang dan penyakit infeksi. Hal tersebut secara tidak langsung disebabkan oleh penyediaan pangan tidak cukup, serta pola asuh anak dan pelayanan kesehatan yang tidak memadai. Pokok masalah yang terjadi pada masyarakat sehingga menyebabkan hal tersebut ialah kurangnya pendidikan, pengetahuan, dan keterampilan. Hal tersebut disebabkan oleh kurangnya pemberdayaan wanita dan keluarga, serta kurangnya pemanfaatan sumberdaya masyarakat. Sesungguhnya semua hal tersebut berakar dari permasalahan nasional seperti krisis ekonomi, sosial, dan politik, yang kemudian menimbulkan banyaknya pengangguran, inflasi, kurang pangan, dan kemiskinan (UNICEF 1998). Oleh karena itu, upaya perbaikan gizi akan lebih efektif dengan selalu mengkaji faktor penyebab tersebut.

Saat ini, diagnosis masalah gizi dilakukan dengan mengukur langsung indikator gizi seperti berat badan, tinggi atau penanda biokimia dari beberapa nutrisi. Proses ini sangat penting, tetapi sampai batas tertentu hal ini akan memakan biaya yang cukup besar dan banyak tantangan yang dihadapi setelah tugas ini sekarang didelegasikan ke tingkat kabupaten. Dalam beberapa kasus, hal ini berjalan sesuai rencana tetapi dengan beberapa defisiensi. Dalam banyak kasus, *monitoring* (pemantauan) sering kali diabaikan. Di sisi lain, sebenarnya banyak lembaga melakukan beberapa survei atau pemantauan variabel yang berhubungan dengan status gizi. Dengan kemajuan teknologi dalam komputasi dan program pengolahan data statistika, ada kesempatan menggunakan data *non-nutritional*

yang tersedia untuk memprediksi status gizi balita, dan begitu pula dengan kesejahteraan masyarakat. Oleh karena itu, penelitian ini akan mencoba menemukan model prediksi status gizi balita dengan menggunakan *support vector regression* (SVR) sebagai metode *machine-learning* serta mengembangkan sistem prediksi dari model SVR tersebut. Alasan pemilihan metode SVR karena metode ini dapat mengatasi kasus non linier dan *overfitting* yang terdapat pada prediksi status gizi ini. Penggunaan SVR sebagai sistem prediksi sebelumnya sudah dilakukan di berbagai bidang. Wu *et al.* (2004) menggunakan SVR untuk memprediksi waktu perjalanan sebagai ukuran dasar dalam transportasi. Agmalaro (2011) juga menggunakan SVR untuk memprediksi curah hujan bulanan Indramayu.

Perumusan Masalah

Permasalahan yang mendasari penelitian ini ialah bagaimana cara mendapatkan pola atau model yang dapat digunakan untuk memprediksi status gizi balita dengan *support vector regression*.

Tujuan Penelitian

Tujuan umum dari penelitian ini ialah untuk meneliti penggunaan *support vector regression* sebagai metode *machine-learning* untuk menemukan model yang dapat memprediksi status gizi balita serta mengembangkan sistem prediksi dari model SVR tersebut.

Manfaat Penelitian

Manfaat dari penelitian ini adalah menghasilkan sistem yang akurat dalam memprediksi status gizi di Indonesia beberapa tahun ke depan, sehingga dapat memberikan rekomendasi pada kebijakan program perbaikan gizi dan kesehatan masyarakat terutama balita di masa yang akan datang.

Ruang Lingkup Penelitian

Ruang lingkup pada penelitian ini, antara lain

- 1 data yang digunakan ialah data survei skala besar yang dihimpun oleh program Higher Education Network Ring Initiative (HENRI), yang terdiri atas hasil survei dari beberapa lembaga survei gizi, dan hanya dipilih data yang berasal dari survei Nutrition and Health Surveillance System (NSS) pada tahun 2000–2003 untuk anak berusia di bawah 2 tahun, dan
- 2 Z-score yang diprediksi hanya yang berhubungan dengan status gizi balita dengan kategori kurang gizi (*stunting*, *wasting*, dan *underweight*).

METODE

Penelitian ini dilakukan dengan melalui beberapa tahapan proses, proses tersebut dijabarkan dengan diagram alir yang dapat dilihat pada Gambar 1.

Studi Pustaka

Tahap awal penelitian ini ialah melakukan studi pustaka. Tinjauan pustaka yang dibutuhkan berasal dari buku, berkala ilmiah, artikel ilmiah, dan skripsi ataupun tesis yang berhubungan dengan penelitian. Studi pustaka bertujuan untuk memahami inti dan domain masalah serta langkah-langkah penyelesaian masalah tersebut dengan metode atau teknik yang digunakan dalam penelitian ini.

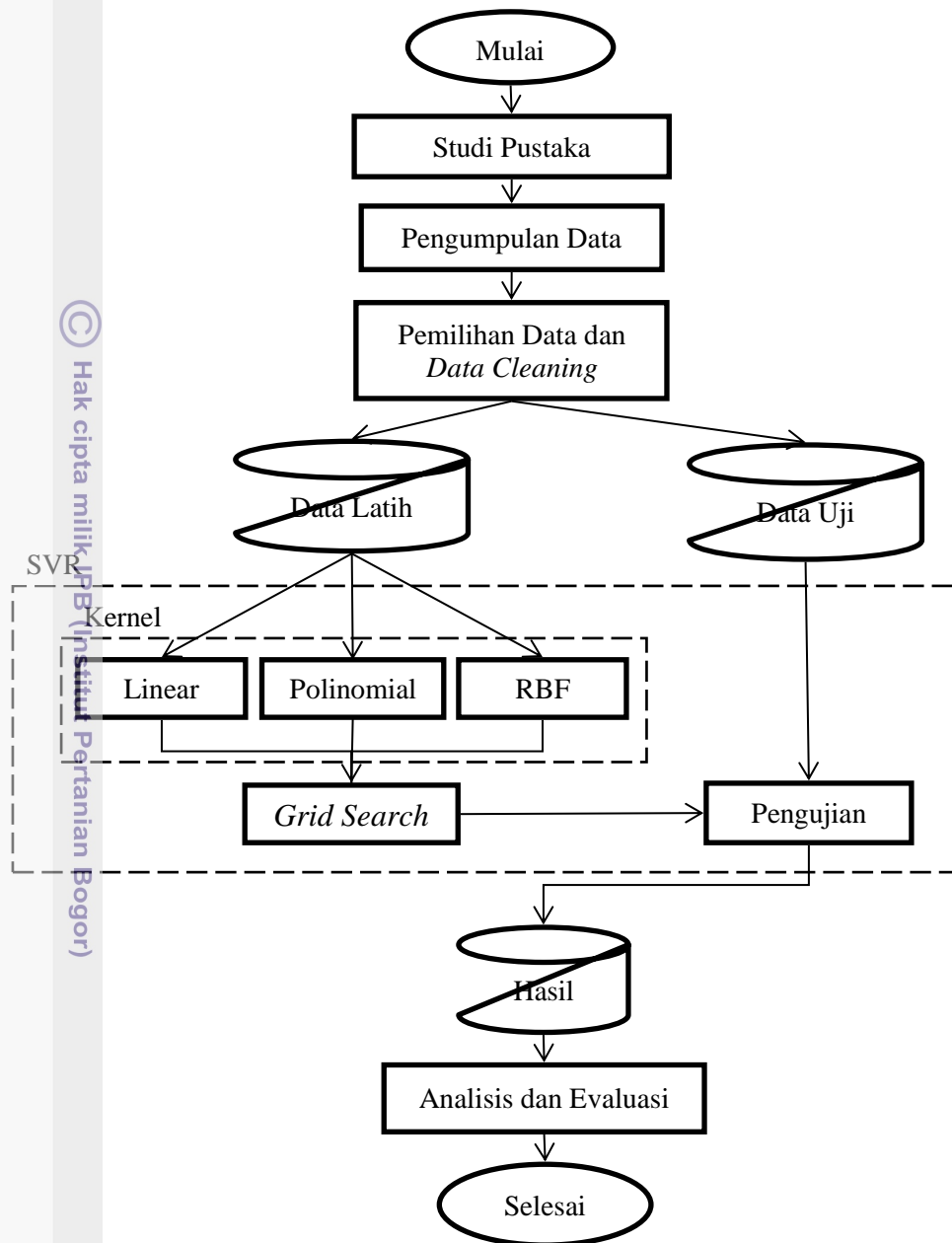
Pengumpulan Data

Data yang digunakan dalam penelitian ialah data survei skala besar gizi yang dihimpun oleh program HENRI yang berasal dari beberapa lembaga yang melakukan survei. Data tersebut terdiri atas survei Indonesia Family Life Survey (IFLS), Demographic and Health Surveys (DHS), dan NSS. IFLS merupakan survei yang diadakan oleh RAND Corporation, sedangkan DHS diadakan oleh Macro International Inc, dan NSS diadakan oleh Helen Keller International. Setelah dilakukan analisis dan pembersihan data, data yang sesuai dan tepat untuk digunakan pada penelitian ini ialah data survei NSS pada tahun 2000–2003 untuk anak berusia di bawah 2 tahun. Hasil survei terdiri atas variabel gizi dan faktor-faktor yang berhubungan dengan gizi, dari faktor rumah tangga hingga lingkungan. Data ini dapat digunakan sebagai sumber pembelajaran data untuk mengembangkan model prediksi pada metode *machine-learning* yang akan digunakan.

Pemilihan Data dan *Data Cleaning*

Tahapan pemilihan data bermaksud untuk memilih parameter-parameter yang digunakan untuk pengolahan data pada tahap berikutnya. Pemilihan parameter didasarkan pada tingkat pengaruh parameter tersebut terhadap status gizi balita. Parameter-parameter yang terpilih akan digunakan sebagai prediktor status gizi balita. *Data cleaning* atau pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan serta menghilangkan data yang memiliki *missing value*.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 1 Diagram alir metode penelitian

Support Vector Regression (SVR)

SVR merupakan penerapan *support vector machine* (SVM) untuk kasus regresi. Dalam kasus regresi output berupa bilangan riil atau kontinu (Smola dan Scholkopf 2004). Lebih lanjut Smola dan Scholkopf (2004) menjelaskan bahwa SVR mengimplementasikan prinsip *structural risk minimization* yang bertujuan untuk meminimalisasi sebuah hubungan dari kesalahan umum (*upper bound of generalization error*) daripada meminimalisasi kesalahan pelatihan (*training error*). SVR telah memperlihatkan sebagai metode yang bisa mengatasi masalah

overfitting. Hal ini bisa menghasilkan performansi yang lebih baik dibandingkan metode lainnya seperti *neural network* yang mengimplementasikan prinsip *empirical risk minimization*.

Ide Dasar SVR

Lebih lanjut Smola dan Scholkopf (2004) menjelaskan mengenai SVR dengan memisalkan suatu kondisi yang terdapat λ set data *training*, (x_j, y_j) dengan $j = 1, 2, \dots, \lambda$ dengan *input* $x = \{x_1, x_2, x_3\} \dots \subseteq \mathbb{R}^N$ dan *output* yang bersangkutan $y = \{y_1, \dots, y_\lambda\} \subseteq \mathbb{R}$. Dengan SVR, akan ditemukan suatu fungsi $f(x)$ yang mempunyai deviasi paling besar ε dari target aktual y_i untuk semua data *training*. Maka dengan SVR, ketika nilai ε sama dengan 0 akan didapatkan regresi yang sempurna.

Model SVR dengan parameter w dan b dapat dinyatakan sebagai berikut:

$$f(x) = w^T \phi(x) + b \quad (1)$$

dengan $f(x)$ adalah *output* dari model dan *input* x dipetakan ke dalam ruang fitur dengan menggunakan fungsi *kernel* $\phi(x)$. Masalah regresi dari SVR dapat dinyatakan oleh masalah optimasi yang mengestimasi koefisien w dan b dengan cara meminimalkan fungsi risiko (*risk function*) yang didefinisikan dalam Persamaan 2:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} L_{\varepsilon}(y_i, f(x_i)) \quad (2)$$

yang memenuhi:

$$\begin{aligned} y_i - w\phi(x_i) - b &\leq \varepsilon \\ w\phi(x_i) - y_i + b &\leq \varepsilon, i = 1, 2, \dots, \lambda \end{aligned}$$

dengan

$$L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon & \text{if } |y_i - f(x_i)| \geq \varepsilon \\ 0 & \text{for other cases} \end{cases} \quad (3)$$

Faktor $\|w\|^2$ dinamakan *reguralisasi*. Meminimalkan $\|w\|^2$ akan membuat suatu fungsi setipis mungkin, sehingga bisa mengontrol kapasitas fungsi. Faktor kedua dalam fungsi tujuan adalah kesalahan empirik (*empirical error*) yang diukur dengan ε -insensitive *loss function*. Menggunakan ide ε -insensitive *loss function* harus meminimalkan norm dari w agar mendapatkan generalisasi yang baik untuk fungsi regresi f . Oleh karena itu, perlu dilakukan penyelesaian problem optimasi berikut:

$$\min \frac{1}{2} \|w\|^2 \quad (4)$$

yang memenuhi

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

$$y_i - w\phi(x_i) - b \leq \varepsilon$$

$$w\phi(x_i) - y_i + b \leq \varepsilon, i = 1, 2, \dots, \lambda$$

$$\sum e^2 = \sum (y - \hat{y})^2 = \sum (y - (b_0 + b_1 x))^2$$

Diasumsikan terdapat suatu fungsi f yang dapat mengaproksimasi semua titik (x_i, y_i) dengan presisi ε . Dalam kasus ini diasumsikan semua titik terdapat dalam rentang $\varepsilon \pm f$ (*feasible*). Dalam hal ketidaklayakan (*infeasible*), mungkin ada beberapa titik yang mungkin keluar dari rentang $\varepsilon \pm f$, dapat ditambahkan variabel ξ, ξ^* untuk mengatasi masalah pembatas yang tidak layak (*infeasible constraint*) dalam problem optimasi. Selanjutnya masalah optimasi di atas bisa diformulasikan sebagai berikut:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} (\xi, \xi^*) \quad (5)$$

yang memenuhi

$$y_i - w\phi(x_i) - b - \xi_i \leq \varepsilon, i = 1, 2, \dots, \lambda$$

$$w\phi(x_i) - y_i + b - \xi_i \leq \varepsilon, i = 1, 2, \dots, \lambda$$

$$\xi_i, \xi_i^*$$

Konstanta $C > 0$ menentukan tawar-menawar (*trade off*) antara ketipisan fungsi f dan batas atas deviasi lebih dari ε masih ditoleransi. Semua deviasi lebih besar daripada ε akan dikenakan penalti sebesar C . Dalam SVR, ε ekuivalen dengan akurasi dari aproksimasi terhadap data *training*. Nilai ε yang kecil terkait dengan nilai yang tinggi pada variabel *slack* ξ_i^* dan akurasi aproksimasi yang tinggi. Sebaliknya, nilai yang tinggi untuk ε berkaitan dengan nilai ξ_i^* yang kecil dan aproksimasi yang rendah. Menurut persamaan (5) nilai yang tinggi untuk variabel *slack* akan membuat kesalahan empirik mempunyai pengaruh yang besar terhadap faktor regulasi. Dalam SVR, *support vector* adalah data *training* yang terletak pada dan di luar batas f dari fungsi keputusan, karena itu jumlah *support vector* menurun dengan naiknya ε .

Dalam formulasi dual, problem optimisasi dari SVR adalah sebagai berikut:

$$\max_{\alpha} - \sum_{i=1}^{\lambda} \sum_{j=1}^{\lambda} (a_i - a_i^*) (a_j - a_j^*) K(x_i, x_j) + \sum_{i=1}^{\lambda} (a_i - a_i^*) y_i - \varepsilon \sum_{i=1}^{\lambda} (a_i + a_i^*) \quad (6)$$

yang memenuhi

$$\sum_{i=1}^{\lambda} (a_i + a_i^*)$$

$$0 \leq a_i \leq C, i = 1, 2, \dots, \lambda$$

$$0 \leq a_i^* \leq C, i = 1, 2, \dots, \lambda$$

dengan nilai C ditentukan sendiri.

$K\langle x_i, x_j \rangle$ ialah *dot-product kernel* yang didefinisikan sebagai $K\langle x_i, x_j \rangle = \phi^T(x_i)\phi(x_j)$. Dengan menggunakan *langrange multiplier* dan kondisi optimalitas, fungsi regresi secara eksplisit dirumuskan sebagai berikut:

$$f(x) = \sum_{i=1}^{\lambda} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (7)$$

Fungsi Kernel

Fungsi yang lebih ekspresif biasanya digunakan untuk mengolah data yang kompleks dibandingkan dengan fungsi linear biasa. SVR metransformasikan *input* yang non-linear ke dalam ruang fitur yang dimensinya lebih tinggi. Hal inilah yang dilakukan oleh *kernel*. Kemudian, SVR akan melakukan perhitungan linear untuk menemukan *hyperplane* yang optimal pada ruang fitur tersebut. *Kernel* akan memproyeksikan data ke dalam ruang fitur berdimensi tinggi untuk menaikkan kemampuan komputasi dari mesin pembelajaran linear.

Beberapa bentuk fungsi *kernel* yang digunakan penelitian ini, antara lain,

Fungsi Linear

Persamaan fungsi linear adalah $K(x, y) = x^T y + C$

Fungsi Polinomial

Persamaan fungsi polinomial adalah $K(x, y) = (\alpha x^T y + C)^d$

Fungsi Gaussian/*Radial Basis Function* (RBF)

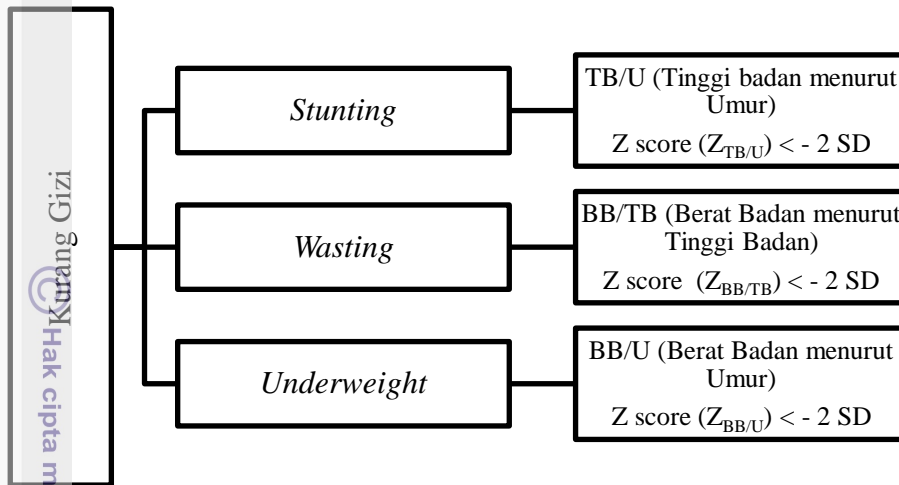
Persamaan fungsi RBF adalah $K(x, y) = \exp(-\gamma \|x - y\|^2)$

Grid Search

Metode *grid search* ialah salah satu metode umum yang sering digunakan untuk estimasi parameter agar mendapatkan hasil yang optimum. Metode ini melibatkan penyusunan *grid* yang cocok dalam suatu ruang dimensi, mengevaluasi fungsi objektif dari seluruh titik *grid*, dan menemukan titik *grid* yang sesuai dengan fungsi objektif yang memiliki nilai optimum (Rao 2009). Secara sederhana prinsip kerja *grid search* adalah dengan menentukan beberapa nilai parameter pada rentang tertentu, kemudian memilih parameter pada nilai terbaik pada rentang tersebut dan melakukan pencarian berulang pada *grid* (rentang nilai) yang lain.

Status Gizi

Status gizi anak adalah keadaan kesehatan anak yang ditentukan oleh derajat kebutuhan fisik energi dan zat-zat gizi lain yang diperoleh dari pangan dan makanan yang dampak fisiknya diukur secara antropometri (Suharjo 1996), dan dikategorikan berdasarkan standar baku WHO-NCHS dengan indeks BB/U (Berat Badan menurut Umur), TB/U (Tinggi Badan Menurut Umur) dan BB/TB (Berat Badan menurut Tinggi Badan).



Gambar 2 Kategori kurang gizi (Atmarita dan Fallah 2004)

Dalam menentukan klasifikasi status gizi digunakan satuan angka yang disebut dengan Z-score. Cara menghitung Z-score yaitu dengan membagi hasil pengurangan sebuah parameter dengan median nilai pada tabel baku rujukan yang digunakan dari parameter yang bersangkutan kemudian dibagi dengan standar deviasinya. Standar deviasi (SD) dihitung dari nilai median pada karakteristik pengukuran (jenis kelamin umur dan indeks) dikurangi dengan nilai -1 SD di dalam daftar baku rujukan pada karakteristik yang sama (Ali 2008). Pada penelitian ini difokuskan pada prediksi kategori kurang gizi seperti yang ditunjukkan pada Gambar 2.

Pemodelan Berdasarkan Fungsi *Kernel* SVR

Setelah dilakukan pembagian data (data *training* dan *testing*), langkah berikutnya ialah pemodelan dan pengujian. Pada percobaan ini, dilihat kinerja model berdasarkan fungsi *kernel* SVR. *Kernel* SVR yang akan digunakan adalah *kernel* linear, polinomial, dan RBF. Performa atau kinerja model berdasarkan fungsi *kernel* dapat diketahui melalui nilai koefisien korelasi (R) dan nilai galat *mean squared error* (MSE). Model yang terbaik adalah model dengan nilai koefisien korelasi (R) terbesar dan nilai galat MSE terkecil (paling mendekati 0).

Analisis dan Pemilihan Model Terbaik

Tahapan analisis dilakukan dengan melihat nilai akurasi (korelasi) dan galat dari setiap model yang dihasilkan. Pada penelitian ini digunakan koefisien korelasi (R) dan MSE untuk mengamati tingkat akurasi dan galat dari model. Model yang terbaik adalah model yang memiliki nilai koefisien korelasi (R) tertinggi (paling mendekati 1) dan nilai galat MSE terkecil (paling mendekati 0).

HASIL DAN PEMBAHASAN

Pemilihan Data dan *Data Cleaning*

Pemilihan data dilakukan dengan mempertimbangkan korelasi peubah penjelas (x) terhadap peubah respon (y). Peubah penjelas dalam penelitian ini ialah parameter-parameter yang memengaruhi nilai dari peubah respon dan yang bertindak sebagai peubah respon ialah nilai Z-score dari tiap status gizi balita (*stunting*, *wasting*, dan *underweight*). Untuk mengetahui korelasi suatu parameter terhadap status gizi, dilakukan peninjauan dari beberapa literatur yang menjelaskan mengenai faktor-faktor yang memengaruhi status gizi balita. Selain itu, untuk parameter-parameter yang tidak dijelaskan dalam literatur dilakukan uji *anova* (*analysis of varian*). Hal ini dimaksudkan untuk menguji korelasi atau pengaruh dua buah peubah (peubah penjelas dan peubah respon) dan mengukur kuatnya korelasi antara peubah yang satu dengan peubah yang lainnya.

Berdasarkan penggunaannya, peubah yang digunakan dalam penelitian ini dibagi menjadi dua, yaitu peubah sebagai pengurutan dan peubah sebagai data latih dan data uji yang akan digunakan dalam SVR. Peubah sebagai pengurutan, antara lain, tahun, provinsi, kabupaten, kecamatan, kelurahan/desa, dan jenis kelamin (Tabel 1), sedangkan peubah sebagai data latih dan data uji ialah peubah penjelas dan peubah respon yang telah dijelaskan sebelumnya. Setelah dilakukan proses pemilihan data dan *data cleaning*, ditentukan 3 buah peubah respons (Tabel 2) dan 13 buah peubah penjelas (Tabel 3). Sedikitnya jumlah peubah peubah penjelas yang digunakan dalam penelitian ini disebabkan karena beberapa peubah lainnya yang sesungguhnya berpengaruh terhadap status gizi tidak dapat diikutsertakan dan hal ini disebabkan karena data pada peubah-peubah tersebut terlalu banyak *missing value*. Mengacu pada UNICEF (1998), faktor-faktor yang memengaruhi status gizi yang digunakan pada penelitian ini antara lain:

- Karakteristik sosial, ekonomi dan demografi: masa pendidikan ibu, tingkat kekayaan, penggunaan yodium, dan ukuran rumah tangga.
- Faktor genetik dan status gizi ibu: tinggi badan ibu dan index massa tubuh (IMT) ibu.
- Penyakit infeksi: diare seminggu terakhir.
- Status gizi anak: berat badan lahir anak dan umur anak.
- Riwayat reproduksi ibu: urutan lahir anak.
- Praktek menyusui: menyusui (iya/tidak).

Pada dasarnya, masih banyak faktor-faktor yang belum diikutsertakan, seperti sanitasi lingkungan, fasilitas pelayanan kesehatan, durasi menyusui, imunisasi, penyakit-penyakit infeksi lainnya, perilaku merokok, dan lain-lain. Pada akhirnya, jumlah data yang digunakan pada penelitian ini ialah sebanyak 168 907. Data latih dan data uji yang digunakan adalah data yang berasal dari data survei NSS pada tahun 2000–2003. Data yang dipilih sebagai data latih ialah data pada tahun 2000–2002 yang berjumlah sebanyak 134 761, sedangkan yang digunakan sebagai data uji ialah data pada tahun 2003 yang berjumlah sebanyak 34 146.

Tabel 1 Peubah sebagai pengurutan

Peubah	Jenis data	Kualitatif/Kuantitatif
Sumber data	Nominal	Kualitatif
Tahun	Ordinal	Kualitatif
Provinsi	Nominal	Kualitatif
Kabupaten	Nominal	Kualitatif
Kecamatan	Nominal	Kualitatif
Kelurahan/Desa	Nominal	Kualitatif
Jenis kelamin	Nominal	Kualitatif

Tabel 2 Peubah respon

Peubah	Z-Score	Jenis data	Kualitatif/Kuantitatif
Tinggi badan menurut umur ($Z_{TB/U}$)		Rasio	Kuantitatif
Berat badan menurut tinggi Badan ($Z_{BB/TB}$)		Rasio	Kuantitatif
Berat badan menurut umur ($Z_{BB/U}$)		Rasio	Kuantitatif

Tabel 3 Peubah penjelas

Peubah	Jenis data	Kualitatif/Kuantitatif	Referensi
Jenis kelamin	Nominal	Kualitatif	Arisman (2004)
Tinggi badan ibu	Rasio	Kuantitatif	Uji <i>Anova</i>
Index massa tubuh (IMT) ibu	Rasio	Kuantitatif	Uji <i>Anova</i>
Berat badan lahir anak	Rasio	Kuantitatif	Adair dan Guilkey (1997)
Diare seminggu terakhir	Nominal	Kualitatif	Lima dan Guerrant (1992)
Penggunaan garam (yodium)	Nominal	Kualitatif	Depkes (1996)
Masa pendidikan ibu	Rasio	Kuantitatif	Uji <i>Anova</i>
Urutan lahir anak	Ordinal	Kualitatif	El Taguri <i>et al.</i> (2008)
Tingkat kekayaan	Rasio	Kuantitatif	Larrea dan Freire (2002)
Ukuran rumah tangga	Rasio	Kuantitatif	Larrea dan Freire (2002)
Menyusui	Nominal	Kualitatif	El Taguri <i>et al.</i> (2008)
Umur ibu ketika melahirkan	Rasio	Kualitatif	El Taguri <i>et al.</i> (2008)
Umur anak	Rasio	Kualitatif	El Taguri <i>et al.</i> (2008)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar IPB.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Kinerja Model Berdasarkan Fungsi *Kernel* SVR

Percobaan ini dilakukan dengan membandingkan kinerja dari ketiga fungsi *kernel* SVR yang telah ditentukan sebelumnya, yaitu *kernel* linear, polinomial, dan RBF. Model SVR yang akan dihasilkan bertujuan untuk memprediksi ketiga jenis Z-Score, yaitu $Z_{TB/U}$, $Z_{BB/TB}$, dan $Z_{BB/U}$. Hasil analisis kinerja fungsi *kernel* SVR dapat terlihat dari tingkat korelasi dan nilai galat estimasi terhadap data aktual masing-masing *kernel*.

Pelatihan dengan menggunakan SVR membutuhkan beberapa parameter sesuai dengan *kernel*-nya. Pada percobaan pertama, nilai parameter yang digunakan pada tiap *kernel* dipilih sesuai dengan nilai *default* dari tiap parameter tersebut (Hsu *et al.* 2010). Jika hasilnya masih belum optimum, langkah berikutnya ialah menggunakan metode *grid search* untuk menentukan parameter terbaik untuk menghasilkan fungsi *kernel* yang optimum. Untuk efisiensi waktu komputasi, data yang digunakan dalam proses *grid search* hanya sekitar 10% dari data latih yang dipilih secara acak, yaitu sebanyak 13 417. Proses pelatihan ini dilakukan pada perangkat lunak MATLAB R2010b dengan menggunakan *library* LibSVM 3.17.

Model SVR untuk Z-score Tinggi Badan Menurut Umur ($Z_{TB/U}$)

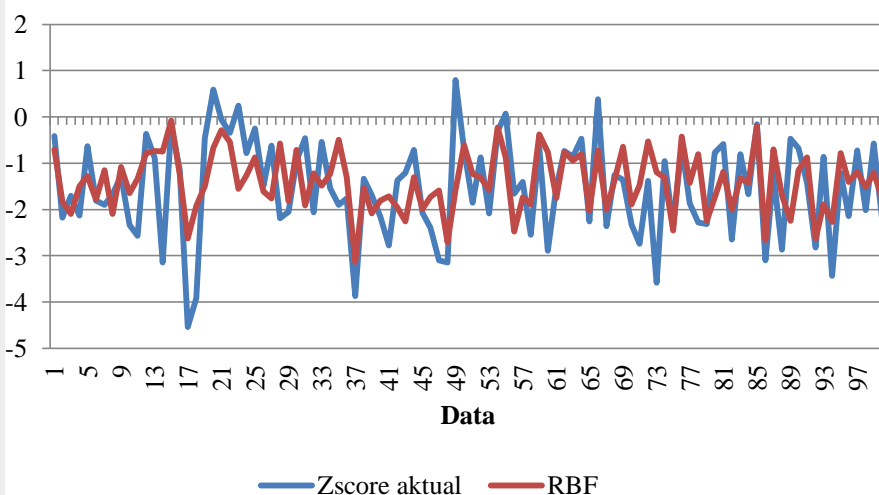
Percobaan pertama menggunakan *kernel* RBF dengan nilai parameter *default* dan menghasilkan nilai koefisien korelasi (R) sebesar 0.5212 dan nilai MSE sebesar 1.3251. Hasil ini belum cukup optimum sehingga langkah berikutnya perlu dilakukan metode *grid search* untuk mendapatkan nilai parameter terbaik. Parameter yang dihasilkan dari proses *grid search* ialah nilai C (*cost*) sebesar 64.0, nilai parameter G (*gamma*) sebesar 0.03125, dan nilai P (*epsilon*) sebesar 0.125. Kemudian, nilai parameter-parameter tersebut digunakan untuk setiap fungsi *kernel* sesuai dengan kebutuhannya. *Kernel* RBF menggunakan ketiga jenis parameter tersebut sehingga menghasilkan nilai koefisien korelasi (R) sebesar 0.5226 dan nilai MSE sebesar 1.32238. *Kernel* linear tidak perlu menggunakan parameter G, cukup menggunakan parameter C dan P sehingga menghasilkan nilai R sebesar 0.5142 dan nilai MSE sebesar 1.33772. *Kernel* polinomial menggunakan ketiga jenis parameter tersebut ditambah dengan parameter r (*coef0*) sebesar 0 dan D (*degree*) sebesar 2 sehingga menghasilkan nilai R sebesar 0.5175 dan nilai MSE sebesar 1.33186. Penjelasan lebih lanjut dari kinerja fungsi *kernel* pada model SVR dijelaskan dalam grafik perbandingan serta *scatter plot* pada 3–8. Grafik dan *scatter plot* tersebut menggambarkan hubungan antara Z-score aktual dan prediksi dari setiap fungsi *kernel*. Hubungan yang kuat antara nilai aktual dan prediksi menunjukkan korelasi yang semakin kuat dan semakin kecil pula ukuran galat antara nilai Z-score aktual dan prediksi.

Secara keseluruhan, tingkat korelasi tertinggi dan nilai galat terendah dihasilkan dengan menggunakan *kernel* RBF, yaitu nilai R sebesar 0.5226 dan nilai MSE sebesar 1.32238. Hal tersebut terlihat dari grafik maupun *scatter plot* yang dihasilkan dari *kernel* RBF pada 3 dan 6, gambar tersebut menunjukkan hubungan korelasi antara Z-score aktual dan prediksi yang dihasilkan dari *kernel* RBF. Sebaliknya *kernel* yang menghasilkan tingkat korelasi terendah dan nilai galat tertinggi ialah *kernel* linear, yaitu dengan nilai R sebesar 0.5142 dan nilai

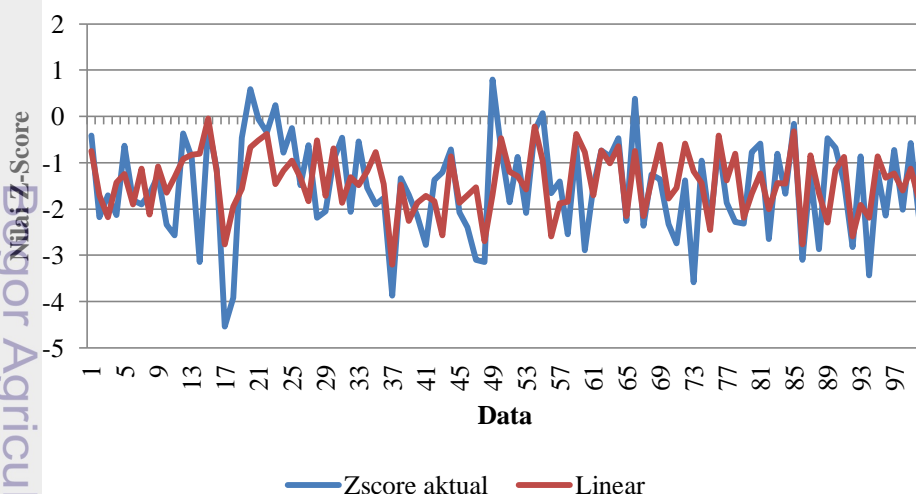
MSE sebesar 1.33772. Perbandingan nilai R dan MSE yang dihasilkan dari tiap *kernel* diperlihatkan pada Tabel 4. Dengan demikian, urutan model prediksi dengan memiliki kinerja terbaik ialah RBF, polinomial, dan linear yang memiliki kinerja paling rendah.

Tabel 4 Perbandingan nilai R dan MSE dari tiap *kernel* pada model $Z_{TB/U}$

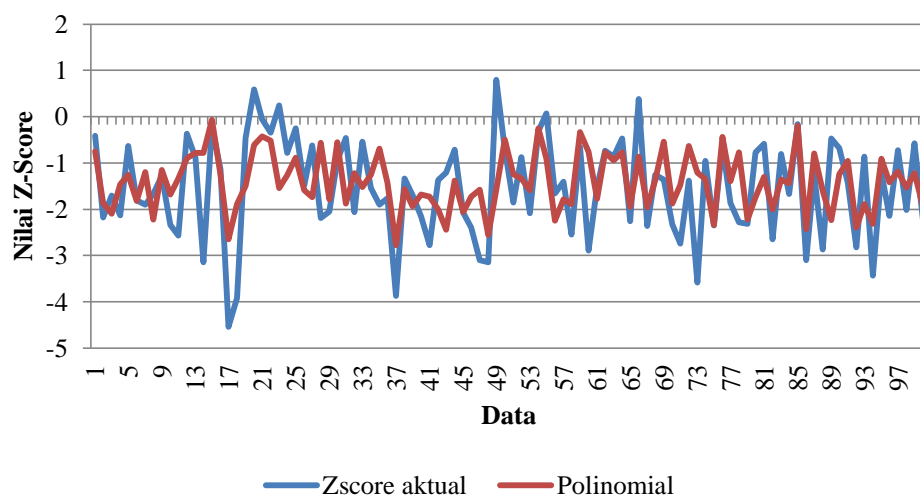
Kernel	R	MSE
RBF	0.5226	1.32238
Polinomial	0.5175	1.33186
Linear	0.5142	1.33772



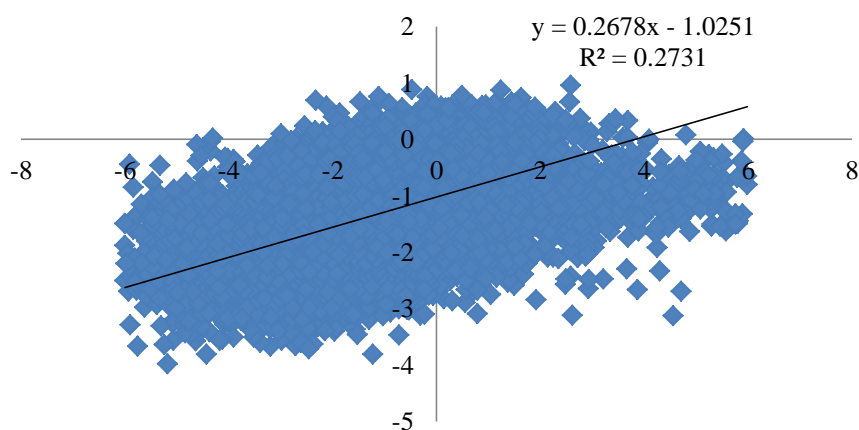
Gambar 3 Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* RBF



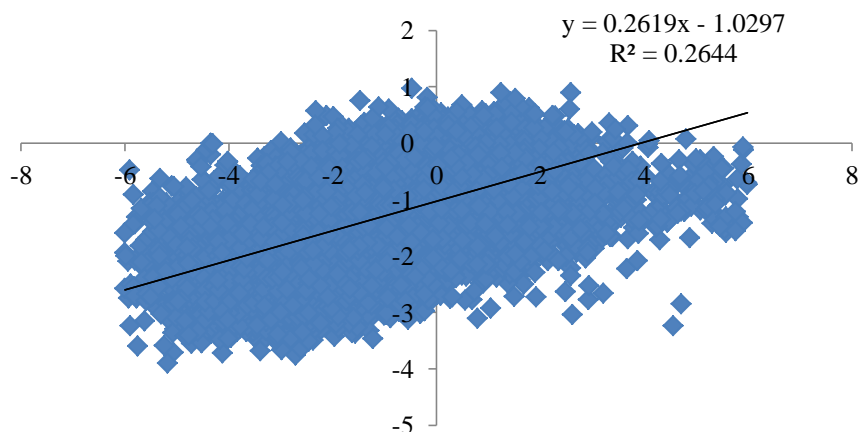
Gambar 4 Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* linear



Gambar 5 Perbandingan $Z_{TB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* polinomial



Gambar 6 *Scatter plot* $Z_{TB/U}$ aktual dan prediksi dengan *kernel* RBF



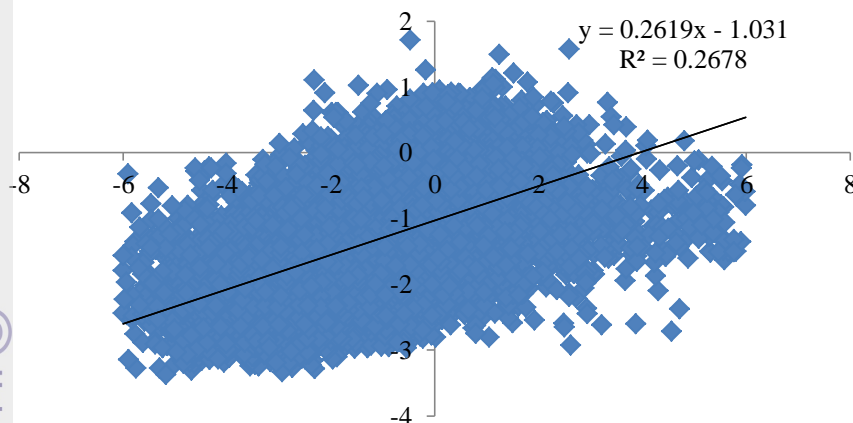
Gambar 7 *Scatter plot* $Z_{TB/U}$ aktual dan prediksi dengan *kernel* linear

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 8 *Scatter plot* $Z_{TB/U}$ aktual dan prediksi dengan *kernel* polinomial

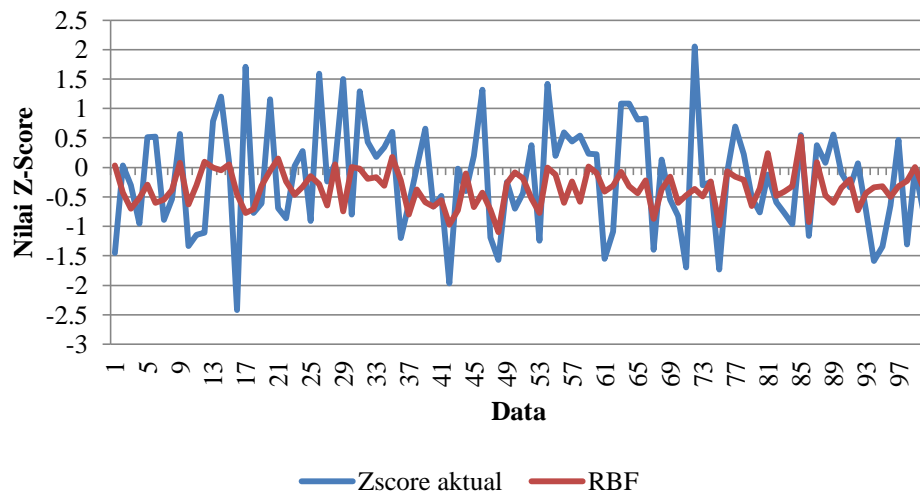
Model SVR untuk Z-score Berat Badan Menurut Tinggi Badan ($Z_{BB/TB}$)

Percobaan pertama menggunakan *kernel* RBF dengan nilai parameter *default* dan menghasilkan nilai koefisien korelasi (R) sebesar 0.3087 dan nilai MSE sebesar 1.15419. Hasil ini belum cukup optimum sehingga langkah berikutnya perlu dilakukan metode *grid search* untuk mendapatkan nilai parameter terbaik, parameter yang dihasilkan dari proses *grid search* ialah nilai C (*cost*) sebesar 0.5, nilai parameter G (*gamma*) sebesar 0.5, dan nilai P (*epsilon*) sebesar 0.5. Kemudian, nilai parameter-parameter tersebut digunakan untuk setiap fungsi *kernel* sesuai dengan kebutuhannya. *Kernel* RBF menggunakan ketiga jenis parameter tersebut sehingga menghasilkan nilai koefisien korelasi (R) sebesar 0.3159 dan nilai MSE sebesar 1.14815. *Kernel* linear tidak perlu menggunakan parameter G, cukup menggunakan parameter C dan P sehingga menghasilkan nilai R sebesar 0.2841 dan nilai MSE sebesar 1.17241. *Kernel* polinomial menggunakan ketiga jenis parameter tersebut ditambah dengan parameter r (*coef0*) sebesar 0 dan D (*degree*) sebesar 2 sehingga menghasilkan nilai R sebesar 0.3016 dan nilai MSE sebesar 1.15919. Penjelasan lebih lanjut dari kinerja fungsi *kernel* pada model SVR dijelaskan dalam grafik perbandingan serta *scatter plot* pada Gambar 9–14. Grafik dan *scatter plot* tersebut menggambarkan hubungan antara Z-score aktual dan prediksi dari setiap fungsi *kernel*. Hubungan yang kuat antara nilai aktual dan prediksi menunjukkan korelasi yang semakin kuat dan semakin kecil pula ukuran galat antara nilai Z-score aktual dan prediksi.

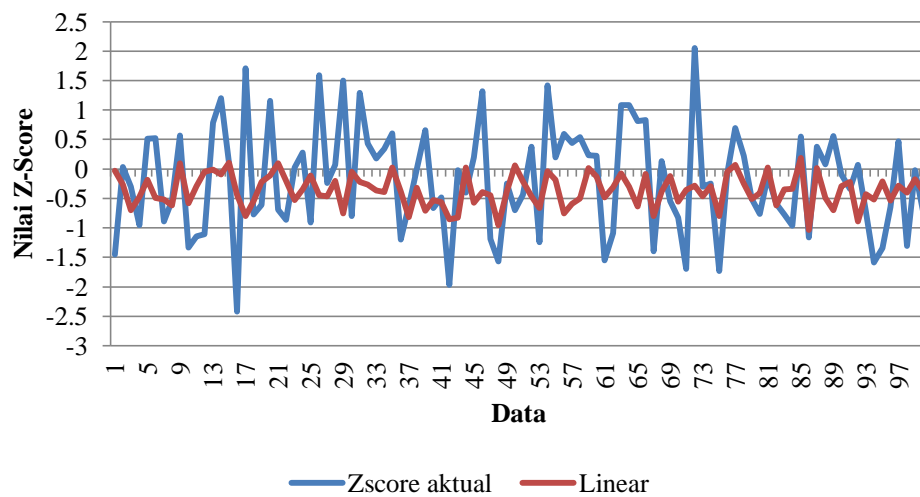
Secara keseluruhan, tingkat korelasi tertinggi dan nilai galat terendah dihasilkan dengan menggunakan *kernel* RBF, yaitu nilai R sebesar 0.3159 dan nilai MSE sebesar 1.14815. Hal tersebut terlihat dari grafik maupun *scatter plot* yang dihasilkan dari *kernel* RBF pada Gambar 9 dan 12, gambar tersebut menunjukkan hubungan korelasi antara Z-score aktual dan prediksi yang dihasilkan dari *kernel* RBF. Sebaliknya *kernel* yang menghasilkan tingkat korelasi terendah dan nilai galat tertinggi ialah *kernel* linear, yaitu dengan nilai R sebesar 0.2841 dan nilai MSE sebesar 1.17241. Perbandingan nilai R dan MSE yang dihasilkan dari tiap *kernel* diperlihatkan pada Tabel 5. Dengan demikian, urutan model prediksi dengan memiliki kinerja terbaik ialah RBF, polinomial, dan linear yang memiliki kinerja paling rendah.

Tabel 5 Perbandingan nilai R dan MSE dari tiap kernel pada model $Z_{BB/TB}$

Kernel	R	MSE
RBF	0.3159	1.14815
Polinomial	0.3016	1.15919
Linear	0.2841	1.17241

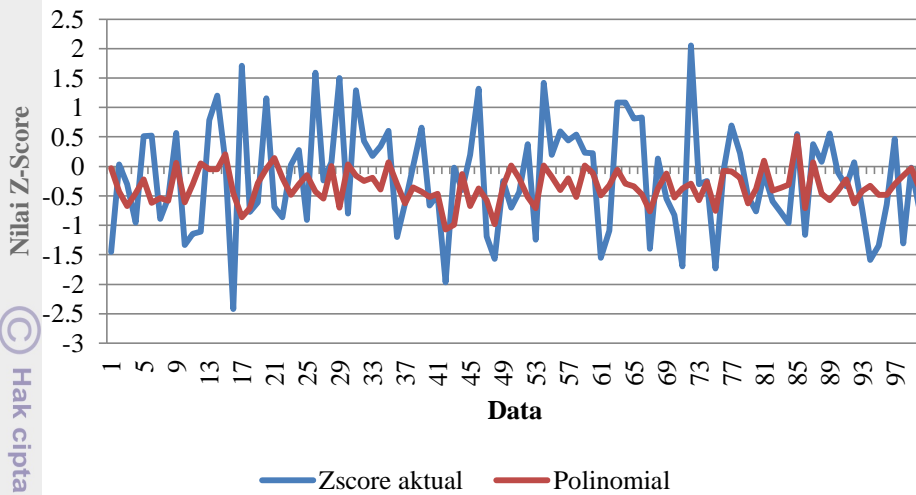


Gambar 9 Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* RBF

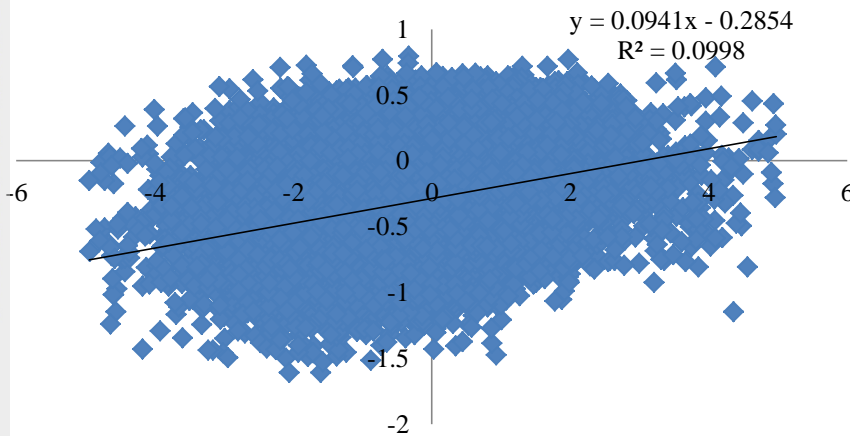


Gambar 10 Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* linear

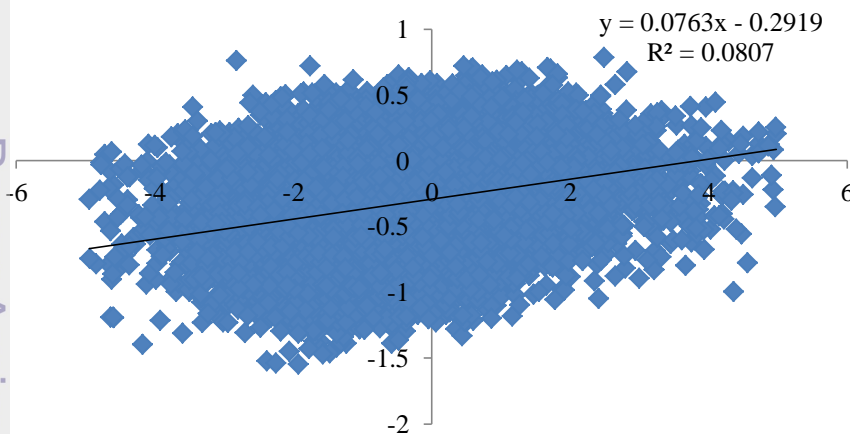
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



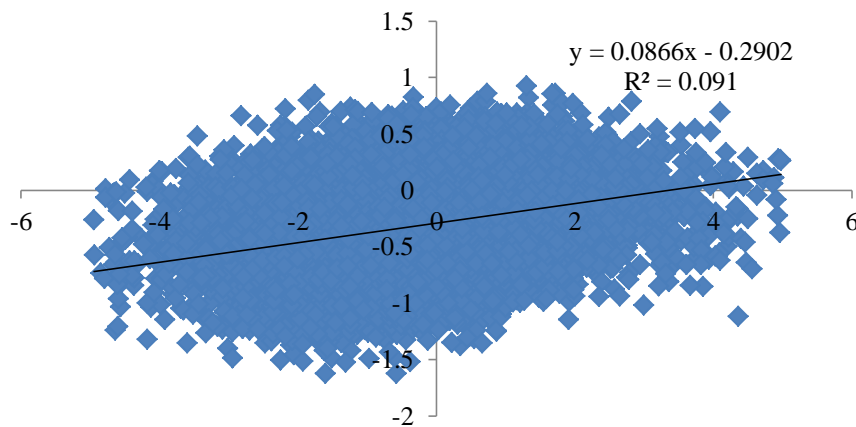
Gambar 11 Perbandingan $Z_{BB/TB}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* polinomial



Gambar 12 *Scatter plot* $Z_{BB/TB}$ aktual dan prediksi dengan *kernel* RBF



Gambar 13 *Scatter plot* $Z_{BB/TB}$ aktual dan prediksi dengan *kernel* linear



Gambar 14 *Scatter plot* $Z_{BB/TB}$ aktual dan prediksi dengan *kernel* polinomial

Model SVR untuk Z-score Berat Badan Menurut Umur ($Z_{BB/U}$)

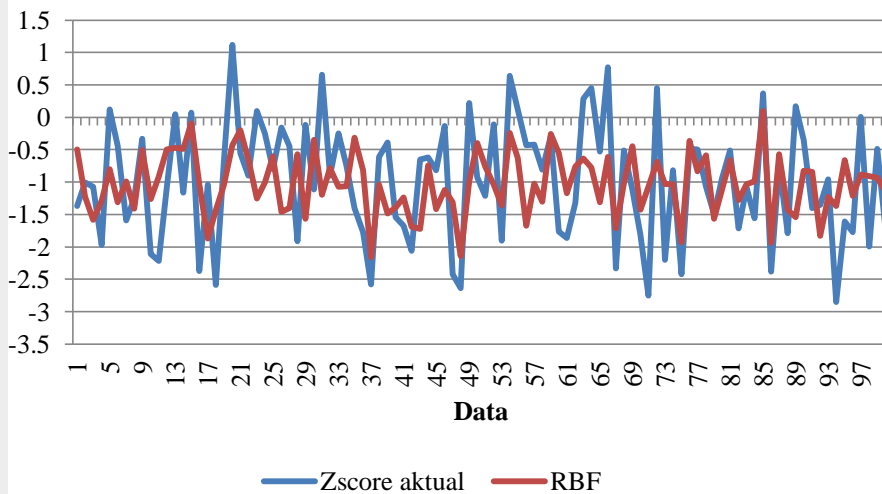
Percobaan pertama menggunakan *kernel* RBF dengan nilai parameter *default* dan menghasilkan nilai koefisien korelasi (R) sebesar 0.4877 dan nilai MSE sebesar 0.93367. Hasil ini belum cukup optimum sehingga langkah berikutnya perlu dilakukan metode *grid search* untuk mendapatkan nilai parameter terbaik. Parameter yang dihasilkan dari proses *grid search* ialah nilai C (*cost*) sebesar 64.0, nilai parameter G (*gamma*) sebesar 0.03125, dan nilai P (*epsilon*) sebesar 0.0625. Kemudian, nilai parameter-parameter tersebut digunakan untuk setiap fungsi *kernel* sesuai dengan kebutuhannya. *Kernel* RBF menggunakan ketiga jenis parameter tersebut sehingga menghasilkan nilai koefisien korelasi (R) sebesar 0.490 dan nilai MSE sebesar 0.9262. *Kernel* linear tidak perlu menggunakan parameter G , cukup menggunakan parameter C dan P sehingga menghasilkan nilai R sebesar 0.476 dan nilai MSE sebesar 0.9430. Pada *kernel* polinomial menggunakan ketiga jenis parameter tersebut ditambah dengan parameter r (*coef0*) sebesar 0 dan D (*degree*) sebesar 2 sehingga menghasilkan nilai R sebesar 0.484 dan nilai MSE sebesar 0.9337. Penjelasan lebih lanjut dari kinerja fungsi *kernel* pada model SVR dijelaskan dalam grafik perbandingan serta *scatter plot* pada Gambar 15–20. Grafik dan *scatter plot* tersebut menggambarkan hubungan antara Z-score aktual dan prediksi dari setiap fungsi *kernel*. Hubungan yang kuat antara nilai aktual dan prediksi menunjukkan korelasi yang semakin kuat dan semakin kecil pula ukuran galat antara nilai Z-score aktual dan prediksi.

Secara keseluruhan, tingkat korelasi tertinggi dan nilai galat terendah dihasilkan dengan menggunakan *kernel* RBF, yaitu nilai R sebesar 0.490 dan nilai MSE sebesar 0.9262. Hal tersebut terlihat dari grafik maupun *scatter plot* yang dihasilkan dari *kernel* RBF pada Gambar 15 dan 18, gambar tersebut menunjukkan hubungan korelasi antara Z-score aktual dan prediksi yang dihasilkan dari *kernel* RBF. Sebaliknya *kernel* yang menghasilkan tingkat korelasi terendah dan nilai galat tertinggi ialah *kernel* polinomial, yaitu dengan nilai R sebesar 0.476 dan nilai MSE sebesar 0.9430. Perbandingan nilai R dan MSE yang dihasilkan dari tiap *kernel* diperlihatkan pada Tabel 6. Dengan demikian, urutan

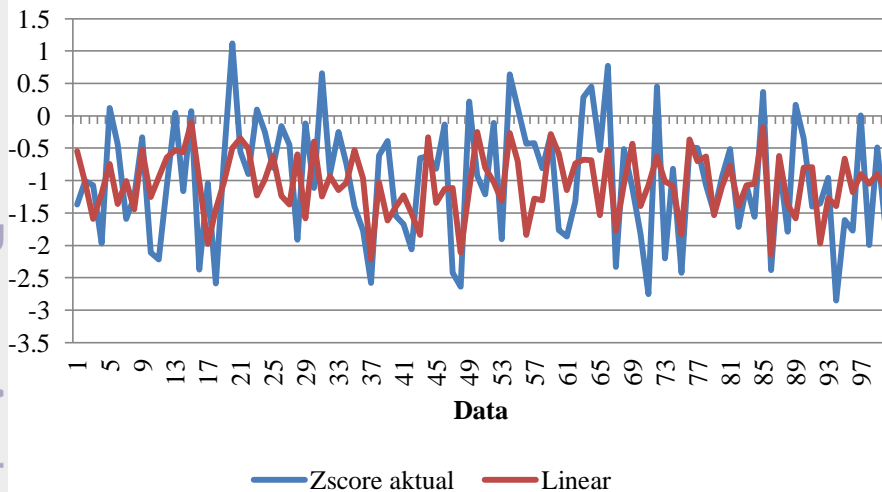
model prediksi dengan memiliki kinerja terbaik ialah RBF, polinomial, dan linear yang memiliki kinerja paling rendah.

Tabel 6 Perbandingan nilai R dan MSE dari tiap kernel pada model $Z_{BB/U}$

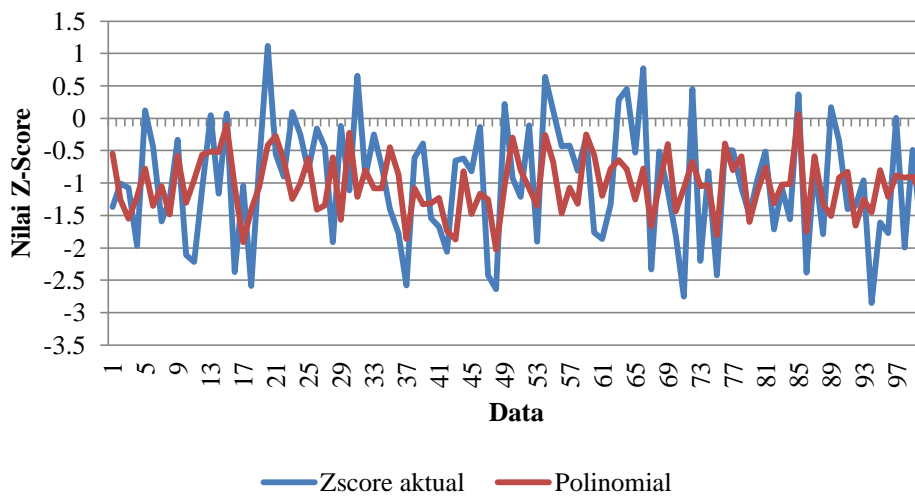
Kernel	R	MSE
RBF	0.490	0.9262
Polinomial	0.484	0.9337
Linear	0.476	0.9430



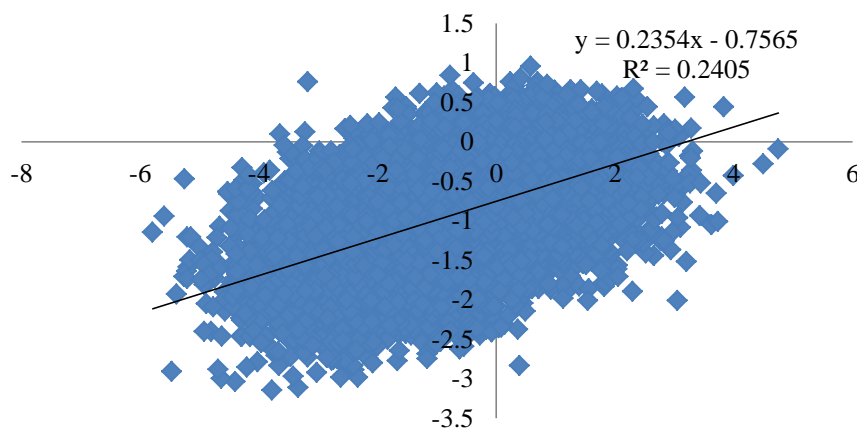
Gambar 15 Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* RBF



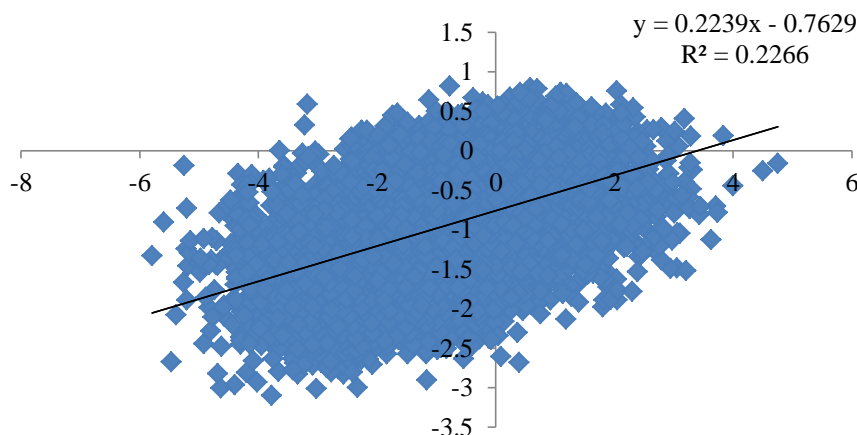
Gambar 16 Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* linear



Gambar 17 Perbandingan $Z_{BB/U}$ aktual dan prediksi pada 100 data awal dengan menggunakan *kernel* polinomial



Gambar 18 *Scatter plot* $Z_{BB/U}$ aktual dan prediksi dengan *kernel* RBF



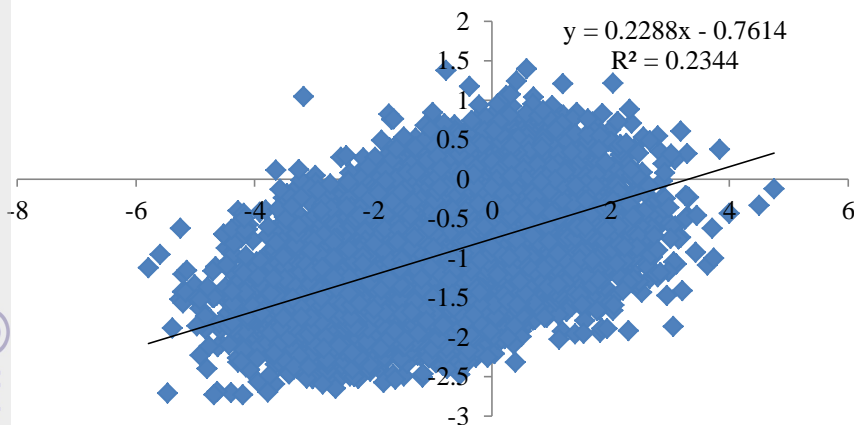
Gambar 19 *Scatter plot* $Z_{BB/U}$ aktual dan prediksi dengan *kernel* linear

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik IPB (Institut Pertanian Bogor)

Bogor Agricultural University

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.



Gambar 20 Scatter plot $Z_{BB/U}$ aktual dan prediksi dengan *kernel* polinomial

Analisis dan Pemilihan Model Terbaik

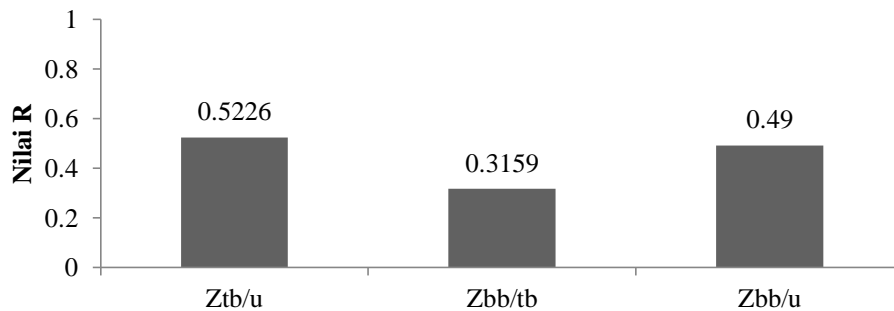
Terlihat pada setiap grafik perbandingan nilai Z-score aktual dan prediksi menunjukkan bahwa rata-rata hasil prediksi tidak dapat mengestimasi nilai-nilai ekstrim, maksud ekstrim disini ialah Z-score yang memiliki nilai sangat rendah ataupun sangat tinggi dibandingkan dengan nilai Z-score lainnya. Contoh nilai ekstrim dapat terlihat pada Gambar 3, gambar tersebut menampilkan perbandingan nilai $Z_{TB/U}$ aktual dan prediksi. Nilai Z-score sangat rendah ditunjukkan pada data ke-17 yang memiliki nilai Z-score aktual sebesar -4.54, sedangkan nilai Z-score prediksinya sebesar -2.63. Selisih nilai aktual dan prediksi dapat dikatakan cukup besar dan hal ini sangat memengaruhi tingkat akurasi secara keseluruhan.

Secara statistik, nilai korelasi dan galat yang dihasilkan dari setiap pelatihan menunjukkan hasil yang kurang baik, yang dimaksud dengan hasil kurang baik ialah nilai koefisien korelasi (R) yang kurang mendekati nilai 1 dan nilai MSE yang kurang mendekati nilai 0. Rendahnya nilai korelasi dan nilai galat serta kemampuan mengestimasi yang kurang baik pada Z-score ekstrim ini dapat disebabkan oleh jumlah peubah penjelas (parameter-parameter penentu status gizi) yang masih tergolong sedikit dan belum cukup mewakili semua faktor-faktor penyebab kurang gizi (UNICEF 1998), sehingga model prediksi yang terbentuk akan memiliki nilai akurasi yang kurang baik. Selain itu, hal tersebut juga dapat disebabkan oleh kurang banyaknya penggunaan data yang digunakan untuk proses *grid search* dalam menentukan parameter fungsi *kernel*, namun untuk penggunaan jumlah data yang besar pada *grid search*, konsekuensinya ialah memakan waktu komputasi yang sangat lama, sehingga hal ini akan tidak efisien. Meskipun demikian, jika dilihat dari sisi positifnya ialah dengan parameter yang terbatas pun ternyata dapat memprediksi nilai suatu Z-score untuk menentukan status gizi balita, walaupun dengan nilai akurasi yang kurang baik. Jika ditelaah lebih mendalam, model prediksi yang menunjukkan tingkat korelasi tertinggi dan nilai galat terendah dari semua jenis Z-score ialah dengan menggunakan *kernel* RBF. Hal sesuai ini dengan panduan SVM yang menjelaskan bahwa *kernel* RBF memang lebih unggul dalam kasus-kasus *machine-learning* pada umumnya (Hsu et al. 2010).

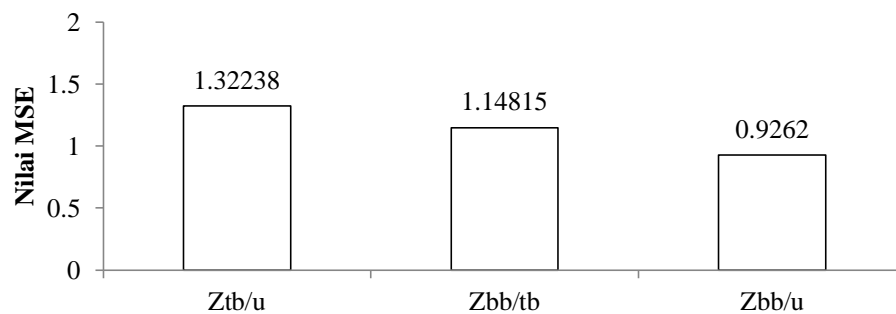
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

Selain itu, jika dilihat dari segi waktu komputasi atau *running time*, waktu yang dibutuhkan dari pelatihan dan pengujian masing-masing kernel juga memiliki perbedaan yang signifikan. Saat dilakukan *running* model dengan kernel linear dan polinomial waktu yang dibutuhkan hingga mendapatkan hasil prediksi bisa mencapai 1.5 jam hingga 2 jam. Sedangkan saat *running* dengan menggunakan kernel RBF, waktu yang dibutuhkan hingga mendapatkan hasil prediksi hanya sekitar 30 menit, selisih 1 jam hingga 1.5 jam dibandingkan dengan *kernel* lainnya. Dari sini dapat terlihat bahwa baik dari segi hasil prediksi maupun waktu komputasi, *kernel* RBF lebih unggul dibanding *kernel* lainnya.

Jika diperhatikan secara seksama dari ketiga model SVR dengan *kernel* RBF yang dihasilkan untuk tiap Z-score, ternyata model yang menghasilkan tingkat korelasi tertinggi ialah model SVR untuk Z-score Tinggi Badan menurut umur ($Z_{TB/U}$), yaitu dengan nilai R sebesar 0.5226 dan model yang menghasilkan galat terendah ialah model SVR untuk Z-score Berat Badan menurut Umur ($Z_{BB/U}$), yaitu dengan nilai MSE sebesar 0.9262. Penjelasan lebih lanjut mengenai perbandingan kinerja model SVR dengan *kernel* RBF untuk tiap jenis Z-score dapat dilihat pada diagram batang yang terdapat pada Gambar 21 dan 22. Diagram tersebut menunjukkan perbedaan yang tidak terlalu signifikan tingkat korelasi maupun galat dari tiap model SVR tersebut dan terlihat bahwa hasil dari ketiga model tersebut menunjukkan hasil yang kurang baik.



Gambar 21 Perbandingan kinerja model SVR dengan *kernel* RBF untuk tiap jenis Z-score berdasarkan tingkat korelasi



Gambar 22 Perbandingan kinerja model SVR dengan *kernel* RBF untuk tiap jenis Z-score berdasarkan galat

SIMPULAN DAN SARAN

Simpulan

Berdasarkan serangkaian percobaan dengan meneliti dan menganalisis sumber data, kemudian mengidentifikasi parameter-parameter (peubah penjelas) yang digunakan, dan mengembangkan model *machine-learning* dengan menggunakan metode *support vector regression* sehingga mendapatkan hasil penelitian berupa model yang dapat memprediksi Z-score yang berhubungan dengan status gizi balita. Pada validasi model prediksi, dilakukan pengujian dengan memprediksi data uji pada tahun 2003 berdasarkan model prediksi yang telah dihasilkan. Model terbaik dihasilkan dengan menggunakan *kernel* RBF, model tersebut menghasilkan tingkat korelasi tertinggi dan galat terendah di tiap jenis Z-score yang diprediksi. Secara keseluruhan dengan menggunakan *kernel* RBF, tingkat korelasi tertinggi dihasilkan oleh model SVR untuk Z-score Tinggi Badan menurut Umur ($Z_{TB/U}$), yaitu dengan nilai R sebesar 0.5226 dan galat terendah dihasilkan oleh model SVR untuk Z-score Berat Badan menurut Umur ($Z_{BB/U}$), yaitu dengan nilai MSE sebesar 0.9262. Dapat disimpulkan, dengan menggunakan model prediksi terbaik dari tiap Z-score tersebut dapat dikembangkan sistem yang dapat memprediksi Z-score, walaupun belum cukup akurat.

Saran

Pencarian parameter-parameter untuk fungsi *kernel* SVR dapat menggunakan algoritma lain seperti algoritma genetika atau *particle swarm optimization* agar mendapatkan nilai parameter yang lebih optimum. Selain itu, untuk memperoleh hasil dengan akurasi terbaik perlu ditambahkan parameter-parameter yang mempengaruhi status gizi balita yang belum terdapat dalam penelitian ini, seperti status imunisasi, pemberian vitamin A, durasi pemberian ASI, sanitasi lingkungan, dan lain-lain. Untuk pengembangan sistem disarankan untuk mengembangkan sistem yang dapat digunakan untuk *mutiplatform* dan dengan *user interface* yang sederhana dan interaktif, sehingga mempermudah pengguna dalam memanfaatkan sistem prediksi yang akan dikembangkan.

DAFTAR PUSTAKA

- Adair LS, Guilkey DK. 1997. Age-specific determinants of stunting in Filipino children. *The Journal of Nutrition*. 127(2):314–320.
- Agmalaro MA. 2011. Pemodelan statistical downscaling data GCM menggunakan support vector regression untuk memprediksi curah hujan bulanan Indramayu [tesis]. Bogor (ID): Institut Pertanian Bogor.
- Ali AR. 2008. Penilaian status gizi anak [Internet]. [diunduh 2013 Apr 8]. Tersedia pada: <http://arali2008.files.wordpress.com/2008/08/penilaian-status-gizi-anak.doc>.
- Arisman. 2004. *Gizi dalam Daur Kehidupan*. Jakarta (ID): Buku Kedokteran EGC.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

- Atmarita, Fallah TS. 2004. Analisis situasi gizi dan kesehatan masyarakat. Di dalam: *Widyakarya Nasional Pangan dan Gizi VIII*; 2004 Mei 14–19; Jakarta, Indonesia. Jakarta (ID): Departemen Kesehatan.
- [Depkes] Departemen Kesehatan RI. 1996. *Gangguan Akibat Kekurangan Yodium*. Jakarta (ID): Depkes RI.
- El Taguri A, Betimal I, Mahmud SM, Ahmed AM, Goulet O, Galan P, Hercberg S. 2009. Risk factors for stunting among under-fives in Libya. *Public Health Nutrition*. 12(8):1141–1149.
- Hsu CW, Chang CC, Lin CJ. 2010. A practical guide to support vector classification [Internet]. [diunduh 2013 Apr 2]. Tersedia pada: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- Parrea C, Freire W. 2002. Social inequality and child malnutrition in four Andean countries. *Rev Panam Salud Publica*. 11(5–6):356–364.
- Pima AAM, Guerrant RL. 1992. Persistent diarrhea in children: epidemiology, risk factors, pathophysiology, nutritional impact, and management. *Epidemiologic Reviews*. 14(1):222–242.
- Reliala AM, Poerwanto S. 1991. Masalah kesehatan di Indonesia bagian timur. *Media Penelitian dan Pengembangan Kesehatan*. 1(4):9–20.
- Rao SS. 2009. *Engineering Optimization: Theory and Practice*. New Jersey (US): J Wiley.
- Smola AJ, Scholkopf B. 2004. A tutorial on support vector regression. *Statistics and Computing*. 14(3):199–222.
- Suharjo. 1996. *Gizi dan Pangan*. Yogyakarta (ID): Kanisius.
- [UNICEF] United Nations Children's Fund. 1998. *The State of the World's Children 1998*. New York (US): Oxford Univ Pr.
- Wu CH, Ho JM, Lee DT. 2004. Travel-time prediction with support vector regression. *IEEE Transactions on Intelligent Transportation Systems*. 5(4):276–281.

Lampiran 1 Tampilan antarmuka sistem prediksi status gizi balita



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.

RIWAYAT HIDUP

Penulis dilahirkan di Bogor pada tanggal 14 Agustus 1991 sebagai anak kedua dari tiga bersaudara pasangan Ir Harianto, MSc dan Prof Dr Ir Nurjanah, MS. Penulis mengenyam pendidikan menengahnya di SMP Negeri 4 Bogor (2003–2006). Pada tahun 2009, penulis menyelesaikan pendidikan di SMA Negeri 1 Bogor. Penulis lulus seleksi masuk Institut Pertanian Bogor (IPB) pada tahun yang sama melalui jalur Ujian Talenta Mandiri (UTM) dan diterima sebagai mahasiswa di Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam.

Penulis aktif di berbagai organisasi seperti Himpunan Mahasiswa Ilmu Komputer (Himalkom), Forum Komunikasi Alumni Muslim SMAN 1 Bogor (Forkom Alims), dan di beberapa kepanitiaan seperti OMI (2010 dan 2011), IT Today (2011), dan GENUS (2012). Penulis juga menjadi asisten praktikum pada Mata Kuliah Penerapan Komputer (2011 dan 2012) dan Basis Data (2013). Selain itu, pada tahun 2013 penulis juga pernah mengikuti Program Kreativitas Mahasiswa (PKM) kategori PKM Karsa Cipta yang didanai DIKTI dengan judul Palwi: Sistem Pakar Penentuan Tanaman Palawija Berdasarkan Iklim Dan Kondisi Lahan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB.