

Jon Idrison Molina
Hari Soetanto

Optimalisasi Kinerja Hasil Komparasi Algoritma C4.5 Dan Naive Bayes Menggunakan Particle Swarm Optimization Untuk Deteksi Dini Penentuan Status Gizi Pada Balita

Optimalisasi Kinerja Hasil Komparasi Algoritma C4.5 Dan Naive Bayes Menggunakan Particle Swarm Optimization
Untuk Deteksi Dini Penentuan Status Gizi Pada Balita

Jon Idrison Molina
Hari Soetanto



Jon Idrison Molina • Hari Soetanto

OPTIMALISASI KINERJA

Hasil Komparasi Algoritma C4.5 dan Naive Bayes
Menggunakan Particle Swarm Optimization
untuk Deteksi Dini Penentuan Status Gizi pada Balita

 **deepublish**
glorify and develop the intellectual of human's life

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Pelindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. Penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. Penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

**OPTIMALISASI KINERJA HASIL KOMPARASI ALGORITMA C4.5
DAN NAIVE BAYES MENGGUNAKAN PARTICLE SWARM OPTIMIZATION
UNTUK DETEKSI DINI PENENTUAN STATUS GIZI PADA BALITA**

Jon Idrison Molina & Hari Soetanto

Desain Cover :
Rulie Gunadi

Sumber :
<https://www.shutterstock.com/>

Tata Letak :
Gofur Dyah Ayu

Proofreader :
Mira Muarifah

Ukuran :
xvi, 120 hlm, Uk: 14x20 cm

ISBN :
978-623-02-3038-7

Cetakan Pertama :
Juni 2021

Hak Cipta 2021, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2021 by Deepublish Publisher
All Right Reserved

Hak cipta dilindungi undang-undang
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT DEEPUBLISH
(Grup Penerbitan CV BUDI UTAMA)

Anggota IKAPI (076/DIY/2012)

Jl. Rajawali, G. Elang 6, No 3, Drono, Sardonoharjo, Ngaglik, Sleman

Jl. Kaliurang Km. 9,3 – Yogyakarta 55581

Telp/Faks: (0274) 4533427

Website: www.deepublish.co.id

www.penerbitdeepublish.com

E-mail: cs@deepublish.co.id

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus, karena berkat dan hikmat-Nya sehingga penulis dapat menyelesaikan Tesis yang berjudul Optimalisasi Kinerja Hasil Komparasi Algoritma C4.5 dan Naïve Bayes Menggunakan Particle Swarm Optimization Untuk Deteksi Dini Penentuan Status Gizi Pada Balita. Tujuan dari penulisan tesis ini adalah sebagai salah satu syarat untuk menyusun tesis pada Program Studi Magister Ilmu Komputer, Universitas Budi Luhur Jakarta.

Rasa dan ucapan terima kasih penulis persembahkan kepada semua pihak yang telah membantu penulis dalam menyusun tesis ini:

1. Bapak Hari Soetanto, S.Kom, M.Sc, selaku dosen pembimbing tesis yang telah membimbing dan memotivasi penulis dalam mengerjakan tesis ini.
2. Bapak Kasih Hanggoro, MBA, Bapak Permenas L. Kolly, SE, Bapak Dr. Moedjiono, M.Sc yang sudah memberikan beasiswa untuk melanjutkan S2 di Universitas Budi Luhur.
3. Bapak Dr. Ir. Nazori AZ, M.T selaku Kaprodi MKom beserta Bapak dan Ibu Dosen pengampu mata kuliah di Program Studi Magister Ilmu Komputer Universitas Budi Luhur yang telah dengan sabar memberikan ilmu pengetahuan, pencerahan, dan bimbingan dalam belajar.
4. Keluargaku, Bapak, Mama dan Kakak adik tercinta yang selalu mendoakan, memotivasi ku untuk menyelesaikan studi S2.
5. Istri ku tercinta *Gilberti Sailana, S.Pd.* yang selalu setia mendoakan, memotivasi ku dalam menyelesaikan penulisan ini.
6. Kakak sekaligus Teman-teman seperjuangan, *Pa Romly, Pa Mesak, Pa Letto, Pa Jun dan Pa Lasarus* yang selalu bersama dalam susah dan senang mengejar cita-cita kami di UBL, yang juga selalu mendoakan dan saling memotivasi untuk meraih kesuksesan bersama.
7. *Mas Yudo Devianto, Pa Kisworo, Mas Endah Wijaya* sebagai rekan sekaligus motivator yang selalu setia bersama-sama menyelesaikan tugas-tugas.

8. Rekan-rekan mahasiswa MKOM Universitas Budi Luhur kelas XB, XL Semester 1 dan 2 dan MKOM Semester 3 konsentrasi Teknologi Sistem Informasi, terima kasih atas kebersamaan, kerja keras dan dukungan semangatnya.
9. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu penulis selama studi hingga tugas akhir ini.

Penulis menyadari masih banyak kekurangan dari tugas akhir tesis ini, oleh karena itu penulis mengharapkan kritik dan saran yang bersifat membangun demi kesempurnaan penelitian tesis nantinya. Semoga tesis ini masih dapat memberikan manfaat dari sisi keterbatasannya.

Kalabahi, 27 Januari 2021

DAFTAR ISI

	Halaman
KATA PENGANTAR	ii
DAFTAR ISI.....	v
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Masalah Penelitian	3
1.2.1 Identifikasi Masalah	3
1.2.2 Pembatasan Masalah.....	3
1.2.3 Rumusan Masalah.....	4
1.3. Tujuan dan Manfaat Penelitian	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian.....	4
1.4. Tata Urut Penulisan.....	5
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	7
2.1. Tinjauan Pustaka	7
2.1.1 Data Mining	7
2.1.2 Algoritma Klasifikasi Data Mining	11
2.1.3 Teknik - Teknik Data Mining.....	14
2.1.4 Algoritma Optimasi	19
2.1.5 Algoritma C4.5	20
2.1.6 Algoritma Naïve Bayes.....	26
2.1.7 Algoritma Particle Swarm Optimization (PSO)	30
2.1.8 Evaluasi dan Validasi Metode Klasifikasi Data Mining.....	32
2.1.9 Penentuan Status Gizi Pada Balita.....	34
2.1.10 RapidMiner	34
2.2. Tinjauan Studi.....	35
2.3. Tinjauan Obyek Penelitian	42
2.3.1 Profil Singkat Organisasi.....	42
2.3.2 Visi Misi Organisasi	42

2.3.3 Struktur Organisasi Dinas Kesehatan Kabupaten Alor	43
2.3.4 Sarana dan Prasarana SI Pendukung Penelitian	45
2.4. Kerangka Konsep	46
2.5. Hipotesis.....	47
BAB III METODOLOGI DAN RANCANGAN PENELITIAN	48
3.1. Jenis Penelitian.....	48
3.2. Populasi dan Sampel	48
3.2.1 Populasi	48
3.2.2 Sampel	48
3.3. Variabel Penelitian	49
3.4. Metode Pengumpulan Data	50
3.4.1 Studi Kepustakaan	50
3.4.2 Data Sekunder.....	50
3.5. Teknik Analisis Data.....	50
3.6. Eksperimen dan Pengujian Model	55
3.7. Evaluasi dan Validasi Hasil	57
3.8. Langkah - Langkah Penelitian	58
3.9. Jadwal Penelitian.....	60
<u>BAB</u> IV PEMBAHASAN HASIL PENELITIAN.....	62
4.1. Analisis Sistem.....	62
4.1.1 Analisis Proses Bisnis Sistem Berjalan	62
4.1.2 Analisis Kebutuhan Fungsional, Nonfungsional, dan Penggun	63
4.2. Analisis Perilaku Sistem	67
4.2.1 Activity Diagram	67
4.2.2 Sequence Diagram	69
4.3. Perancangan Sistem	71
4.3.1 Deployment Diagram	71
4.3.2 Perancangan Infrastruktur Arsitektur	72
4.4. Konstruksi Sistem	74
4.4.1 Lingkungan Konstruksi	74
4.4.2 Konstruksi Database	74

4.5. Pengujian Sistem.....	78
4.5.1. Lingkungan Pengujian.....	78
4.5.2. Pengujian Validasi.....	79
4.5.3. Klasifikasi Algoritma <i>Naïve Bayes</i>	79
4.5.4. Klasifikasi Algoritma C4.5.....	82
4.6. Evaluasi dan Validasi.....	85
4.6.1. <i>Confusion Matrix</i> Algoritma <i>Naïve Bayes</i>	85
4.6.2. <i>Confusion Matrix</i> Algoritma C4.5.....	86
4.7. Analisa Hasil Komparasi	88
4.8. Optimasi Algoritma PSO dari Algoritma Terpilih.....	88
4.9. Penerapan Algoritma PSO	90
4.10. Implikasi Penelitian.....	91
4.10.1. Aspek Manajerial.....	91
4.10.2. Aspek Sistem	92
4.10.3. Rencana Implementasi.....	92
BAB V PENUTUP.....	94
5.1. Kesimpulan	94
5.2. Saran.....	94
DAFTAR PUSTAKA	96
<u>LAMPIRAN-LAMPIRAN</u>	99

BAB I PENDAHULUAN

1.1 Latar Belakang

Riset kesehatan dasar atau Riskedas 2013 yang diselenggarakan Badan Penelitian dan Pengembangan Kesehatan dari Kementerian Kesehatan Republik Indonesia melaporkan prevalensi gizi kurang pada balita di Indonesia sekitar 19,6 persen. Data Riskedas 2013 tersebut sebagai acuan yang bisa dilihat tentang masalah gizi yang dibagi dalam kategori gizi kurang (*underweight*), kurus (*wasting*), pendek (*stunting*), dan kegemukan (*obese*). Data ini menunjukkan beban ganda yang terjadi dalam masalah gizi di Indonesia yaitu masih ada kasus gizi kurang dan juga kasus gizi berlebih. Secara nasional pada tahun 2013, prevalensi kurus dan sangat kurus masih cukup tinggi yaitu masing-masing 12,1 persen dan 5,3 persen. Adapun masalah tubuh pendek atau stunting dalam istilah gizi menerangkan kondisi dimana tinggi badan anak tidak sesuai dengan umurnya pada balita di Indonesia saat ini masih cukup serius sekitar 37,2 persen. Pemenuhan gizi pada anak usia dibawah lima tahun (balita) merupakan faktor yang perlu diperhatikan dalam menjaga kesehatan, karena masa balita merupakan periode perkembangan yang rentan gizi. Kasus kematian yang terjadi pada balita merupakan salah satu akibat dari gizi buruk. Gizi buruk dimulai dari penurunan berat badan ideal seorang anak sampai akhirnya terlihat sangat buruk (Kemenkes RI, 2013).

Pertumbuhan dan perkembangan mengalami peningkatan yang pesat pada usia dini, yaitu dari 0 sampai 5 tahun. Masa ini sering juga disebut sebagai fase "Golden Age". *Golden age* merupakan masa yang sangat penting untuk memperhatikan tumbuh kembang anak secara cermat agar sedini mungkin dapat terdeteksi apabila terjadi kelainan. Selain itu, penanganan kelainan yang sesuai pada masa *golden age* dapat meminimalisir kelainan pertumbuhan dan perkembangan anak sehingga kelainan yang bersifat permanen dapat dicegah (Chamidah A. N. 2009). Tujuan utama pembangunan nasional adalah peningkatan kualitas Sumber Daya Manusia (SDM) yang dilakukan secara

berkelanjutan. Upaya peningkatan kualitas SDM dimulai pada proses tumbuh kembang anak sejak pembuahan sampai mencapai dewasa utama. Pada masa tumbuh kembang anak, pemenuhan kebutuhan dasar seperti perawatan dan makanan bergizi yang diberikan dengan penuh kasih sayang dapat membentuk SDM yang sehat, cerdas dan produktif ^(Octaviani dkk 2008). Kualitas seorang anak dapat dinilai dari proses tumbuh kembang. Proses tumbuh kembang merupakan hasil interaksi faktor genetik dan faktor lingkungan. Faktor genetik/keturunan adalah faktor yang berhubungan dengan gen yang berasal dari ayah dan ibu, sedangkan faktor lingkungan meliputi lingkungan biologis, fisik, psikologis, dan sosial ^(Chamidah A. N. 2009).

Salah satu upaya cukup penting terhadap peningkatan kualitas sumber daya manusia adalah upaya peningkatan status gizi. Status gizi merupakan salah satu faktor yang menentukan kualitas hidup dan produktifitas kerja ^(Octaviani dkk 2008). Peningkatan status gizi sangat diperlukan apalagi pada balita yang masih berada dalam masa *golden age*. Penilaian status gizi bisa menjadi salah satu metode untuk mencegah terjadinya permasalahan gizi. Status gizi balita akan diketahui dengan adanya penilaian status gizi pada balita ^(Supariasa dkk 2001). Penilaian status gizi sangat berperan dalam pemantauan gizi anak. Apabila penilaian status gizi dilakukan secara tepat dan akurat maka adanya tanda atau gejala gangguan pertumbuhan maupun perkembangan anak dapat diketahui secara dini sehingga bisa dilakukan penanggulangan dan pencegahan dalam upaya peningkatan status gizi pada balita. Serta terjadinya permasalahan gizi dapat diminimalisir. Untuk itu maka penentuan status gizi pada balita harus dilakukan secara cepat tepat dan akurat.

Puskesmas sebagai pelaksana teknis Dinas Kesehatan, mempunyai tugas pokok melakukan pendataan dan penilaian status gizi balita serta menyerahkan hasil penilaian ke Dinas Kesehatan. Penentuan status gizi yang selama ini dilakukan adalah dengan cara melihat dan menghitung pada tabel standar antropometri penilaian status gizi balita sesuai SK. Menteri Kesehatan Nomor:1995/Menkes/SK/XII/2010. Namun Puskesmas sering lambat dalam menangani pasien dalam jumlah yang banyak dan juga keterbatasan sumber daya terutama di wilayah yang kekurangan tenaga medis. Sehingga ini berpengaruh juga

terhadap hasil penilaian penentuan status gizi balita. Untuk itu Dinas Kesehatan perlu menyediakan sebuah metode yang optimal dalam penentuan status gizi balita untuk dipakai oleh Puskesmas dalam penentuan status gizi balita secara cepat, tepat dan akurat.

Pada penelitian ini, akan di bahas tentang komparasi penerapan algoritma C4.5 dan *Naïve Bayes* sehingga jika salah satu algoritma memiliki akurasi yang lebih tinggi maka akan dioptimalkan dengan menggunakan algoritma *Particle Swarm Optimization* (PSO) untuk menentukan status gizi balita.

1.2 Masalah Penelitian

1.2.1 Identifikasi Masalah

Berdasarkan uraian-uraian pada latar belakang masalah, maka identifikasi permasalahan yang muncul yaitu :

1. Keterbatasan sumber daya manusia sehingga hasil dari penentuan status gizi balita kurang akurat, terutama dalam membaca tabel antropometri untuk menentukan status gizi balita.
2. Dinas Kesehatan Kabupaten belum mempunyai metode yang lebih optimal untuk dipakai Puskesmas dalam menentukan status gizi balita secara cepat, tepat dan akurat.
3. Perlu adanya optimasi dari komparasi penerapan algoritma C4.5 dan *Naïve Bayes* menggunakan algoritma *Particle Swarm Optimization* (PSO).

1.2.2 Batasan Masalah

Membahas mengenai status gizi pada balita merupakan suatu hal yang kompleks sehingga membutuhkan waktu yang cukup lama serta kekurangan tentang pemahaman terhadap ilmu kesehatan maka penelitian ini dibatasi hanya pada :

1. Dalam penelitian ini, penentuan status gizi pada balita hanya didasarkan pada indeks Berat Badan menurut Umur (BB/U).

2. Komparasi penerapan algoritma C4.5 dan *Naïve Bayes* sehingga algoritma yang memiliki akurasi lebih tinggi akan dioptimalkan menggunakan algoritma *Particle Swarm Optimization* (PSO) untuk mendeteksi secara dini status gizi pada balita secara cepat, tepat dan akurat.
3. *Tools* yang akan digunakan adalah *software RapidMiner*

1.2.3 Rumusan Masalah

Untuk mengatasi berbagai hal yang diuraikan di dalam latar belakang dan masalah penelitian di atas, maka dapat dirumuskan rumusan masalah yang akan dibahas di dalam penulisan ini adalah seberapa akurat dari komparasi penggunaan algoritma C4.5 dan *Naïve Bayes* dibandingkan dengan penerapan algoritma *Particle Swarm Optimization* (PSO) diterapkan pada sistem deteksi dini status gizi pada balita ?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Adapun tujuan penelitian ini adalah untuk :

1. Meningkatkan hasil keakuratan menggunakan algoritma PSO dari hasil komparasi algoritma C4.5 dan *naïve bayes*.
2. Membuktikan bahwa algoritma PSO dapat meningkatkan presentase keakuratan algoritma yang terpilih sehingga lebih optimal dalam mendeteksi secara dini status gizi pada balita secara cepat, tepat dan akurat.

1.3.2 Manfaat Penelitian

Manfaat dari penelitian dan penulisan ini adalah sebagai berikut:

1. Dengan adanya penerapan metode yang optimal dalam mendeteksi secara dini status gizi pada balita maka bisa dilakukan penanggulangan dan pencegahan secara dini dalam upaya peningkatan status gizi pada balita.

2. Dapat memberikan edukasi kepada individu atau lembaga-lembaga yang berperan dalam menangani kesehatan masyarakat tentang pentingnya penentuan status gizi pada balita secara dini.
3. Menemukan algoritma mana yang memiliki akurasi yang lebih tinggi diantara algoritma C4.5 dan *Naïve Bayes* dalam mendeteksi secara dini status gizi pada balita secara cepat, tepat dan akurat.

1.4 Tata Urut Penulisan

Secara ringkas, penelitian ini dituangkan ke dalam 5 (lima) bab penulisan dengan struktur sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan dibahas mengenai latar belakang, permasalahan, perumusan masalah, tujuan dan manfaat penulisan serta sistematika penulisan.

BAB II LANDASAN PEMIKIRAN

Pada bab ini akan dibahas beberapa teori yang akan digunakan dalam penelitian ini diantaranya adalah teori mengenai konsep Algoritma C4.5 dan *Naïve Bayes* secara umum, tinjauan umum perancangan serta pengembangan sistem dengan pendekatan berorientasi obyek, tinjauan studi terhadap beberapa penelitian terdahulu, uraian pola pikir yang digunakan di dalam penelitian ini serta hipotesis.

BAB III METODOLOGI DAN RANCANGAN PENELITIAN

Pada bab ini akan dibahas mengenai metode penelitian yang digunakan, metode pemilihan sampel, metode pengumpulan data beserta teknik analisis, teknik perancangan sistem prototipe, teknik pengujian, langkah penelitian dan jadwal penelitian yang telah dilakukan.

BAB IV PEMBAHASAN HASIL PENELITIAN

Pada bab ini akan dibahas mengenai analisis kebutuhan sistem, model fungsional yang ada, model fungsional yang dikembangkan,

perancangan basisdata, perancangan arsitektur, perancangan sistem, implementasi sistem serta pengujian hipotesis.

BAB V

PENUTUP

Bab ini terdiri dari kesimpulan dan saran dan merupakan penutup dari tesis, dimana penulis menjabarkan hasil yang diperoleh dari penelitian ini berdasarkan perumusan masalah, hipotesis dan hasil pengujian hipotesis dan saran dirangkum dari implikasi penelitian.

BAB II
LANDASAN TEORI DAN KERANGKA KONSEP

BAB III

METODOLOGI DAN RANCANGAN PENELITIAN

Metodologi dan tahapan penelitian diperlukan sebagai panduan proses penelitian, sehingga proses penelitian dapat dilakukan secara sistematis.

3.1. Jenis Penelitian

Adapun jenis penelitian ini terbagi menjadi :

1. Penelitian Experimental, merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variabel atau atribut.
2. Penelitian perbandingan atau studi komparasi yakni dengan membandingkan algoritma C4.5 dengan *Naïve Bayes*.
3. Penelitian pengaruh optimasi menggunakan algoritma PSO dari algoritma yang terpilih.

3.2. Populasi dan Sampel

3.2.1. Populasi

Populasi dalam penelitian ini adalah pemilihan sampel dengan mengidentifikasi populasi target yaitu populasi yang relevan dengan tujuan masalah dalam penelitian. Populasi dalam penelitian ini adalah seluruh data penilaian status gizi balita yang terdapat pada dua Puskesmas yang ada di Kabupaten Alor.

3.2.2. Sampel

Penentuan sampel dilakukan dengan teknik *proposive* (teknik penarikan sampel bertujuan) yaitu penarikan sampel dengan memiliki target dan tujuan tertentu. Data seluruh jumlah populasi diambil data penilaian status gizi balita yang ada di 2 Puskesmas yang ada di Kabupaten Alor pada Bulan Januari dan Februari tahun 2015 sebanyak 206 data status gizi balita.

3.3. Variabel Penelitian

Dalam melakukan klasifikasi data status gizi balita pada penelitian ini akan dilakukan komparasi algoritma C4.5 dan *Naïve Bayes* untuk mengetahui algoritma yang paling akurat dalam mendeteksi status gizi balita.

Atribut dan nilai atribut diperoleh dari laporan data penilaian status gizi balita. Adapun atribut yang digunakan dalam penelitian ini antara lain :

1. Jenis Kelamin

Merupakan atribut yang berisi data jenis kelamin balita

2. Umur

Merupakan atribut yang berisi data umur balita

3. Tinggi Badan

Merupakan atribut yang berisi data tinggi badan balita

4. Berat Badan

Merupakan atribut yang berisi data berat badan balita

5. Remark

Merupakan atribut tambahan yang berisi penentuan status gizi balita.

3.4. Metode Pengumpulan data

3.4.1. Studi Kepustakaan

Pengumpulan data bersifat teoritis yang berhubungan dengan penelitian ini. Pengambilan data tersebut dilakukan dengan cara mempelajari literatur-literatur, jurnal-jurnal penelitian serta sumber-sumber lain yang berkaitan dalam penelitian ini.

3.4.2. Data Sekunder

Pengumpulan data dalam penelitian ini menggunakan penelitian lapangan, dilakukan dengan mengambil data dari 2 Puskesmas melalui Dinas Kesehatan Kabupaten Alor yaitu data gizi balita, berbentuk foto copy.

3.5. Teknik Analisis Data

Proses *Knowledge discovery in database* (KDD) adalah interaktif dan berulang, yang melibatkan banyak langkah dengan banyak keputusan yang dibuat oleh pengguna. ([Brachman & Anand 1996]) Memberikan pandangan praktis dari proses KDD, menekankan sifat interaktif dari proses, yang terdiri dari sembilan langkah.

Berikut ini adalah Sembilan langkah dalam KDD yaitu :

3.5.1. *Developing and understanding of application domain*

Pada tahap ini dilakukan pengembangan pemahaman domain aplikasi untuk persiapan langkah awal untuk memahami apa yang dilakukan dengan banyak keputusan (tentang transformasi, algoritma, representasi).

3.5.2. *Selection and creating a data set on which discovery will be performed*

Memilih dan menciptakan satu set data yang akan digunakan untuk penelitian harus ditentukan, termasuk mencari tahu data apa yang tersedia, memperoleh data tambahan yang diperlukan, dan kemudian mengintegrasikan semua data untuk penemuan *knowledge* ke dalam satu set data termasuk atribut yang akan dipakai untuk proses tersebut.

3.5.3. *Preprocessing and cleansing*

Pada tahap ini kehandalan data ditingkatkan dengan membersihkan data yang tidak lengkap (*missing value*) dan data tidak benar (*noise*). Data yang dipakai untuk klasifikasi sejumlah 165 data.

3.5.4. *Data Transformation*

Pada tahap ini disusun dan dikembangkan generasi data yang lebih baik untuk data mining. Tahap ini juga merupakan proses transformasi pada data yang telah dipilih sehingga data tersebut sesuai dengan proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

3.5.5. Choosing the appropriate Data Mining task

Pada tahap ini memilih jenis data mining yang digunakan. Pada penelitian ini untuk deteksi status gizi balita dipilih jenis data mining yang akan digunakan adalah klasifikasi.

3.5.6. Choosing the Data Mining Algorithm

Setelah pemilihan jenis data mining yang akan digunakan yaitu klasifikasi, maka selanjutnya menentukan algoritma klasifikasi yang akan digunakan. Pada penelitian ini algoritma yang dipilih untuk komparasi adalah C4.5 dengan *Naïve Bayes*.

3.5.7. Employing the Data Mining Algorithm

Tahap ini dilakukan untuk pengolahan data dengan algoritma yang telah dipilih untuk mendapatkan algoritma terbaik dengan tingkat akurasi yang tinggi dalam klasifikasi status gizi balita.

3.5.8. Evaluation

Dalam tahap ini dilakukan evaluasi dan menafsirkan pola yang didapatkan dari hasil algoritma yang dipakai untuk mengetahui aturan, kehandalan, dan lain-lain. Evaluasi dilakukan dengan menerapkan pola yang didapat dari proses sebelumnya terhadap data testing yang disediakan.

3.5.9. Using the discovered knowledge

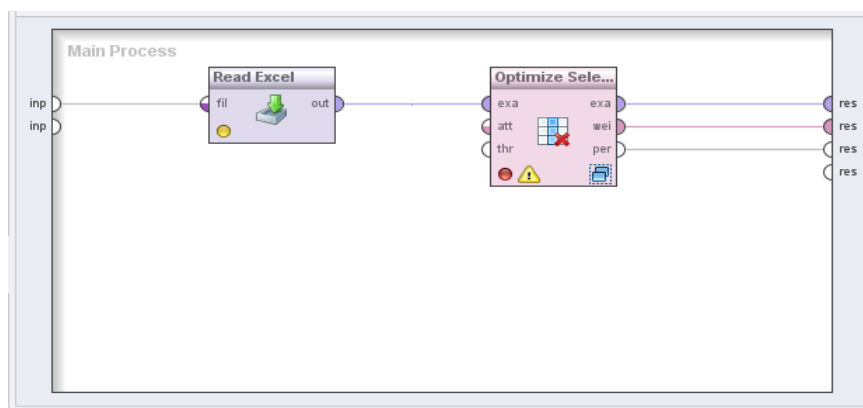
Pada tahap ini menggunakan pengetahuan yang diperoleh dari proses data mining untuk penerapan pada aplikasi atau lainnya. Pengetahuan klasifikasi status gizi balita diterapkan pada data baru untuk membuat klasifikasi status gizi balita.

3.6. Eksperimen dan Pengujian Model

Pengujian model menggunakan empat variable yang akan secara acak mengambil 10 % dari data training untuk sebagai data testing, proses ini diulang

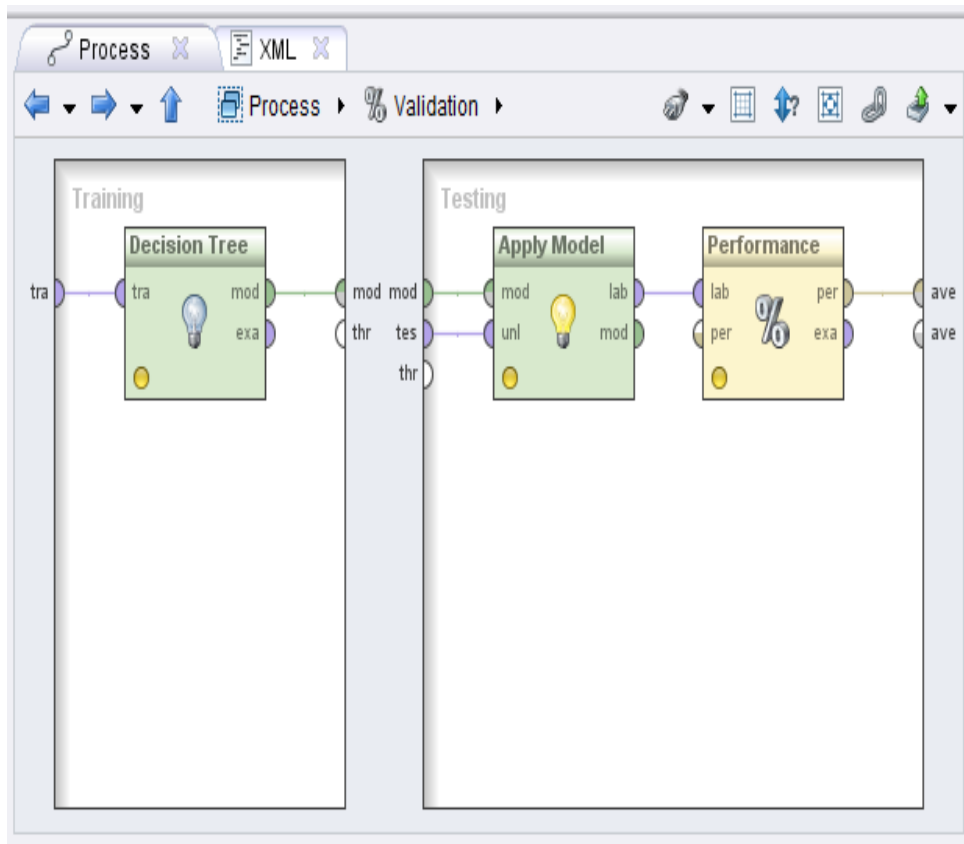
sebanyak 10 kali dan hasil pengujian model berupa *accuracy*, *precision*, dan *recall* di rata-ratakan. Proses pengujian ini dilakukan dengan rapid miner dalam building block yang digunakan untuk prediksi.

Dalam melakukan penelitian ini diperlukan eksperimen dan proses pengujian model yang diusulkan. Proses eksperimen dan pengujian model menggunakan bagian dari dataset yang ada. Semua dataset kemudian diuji dengan metode yang diusulkan pada aplikasi Rapid Miner 5. Berikut ini model yang diimplementasikan ke dalam aplikasi Rapid Miner 5, yaitu :



Gambar III-1 Model yang diusulkan pada Rapid Miner 5

Pada gambar III-1 merupakan hubungan model yang diusulkan yaitu read excel dari dataset. Kemudian menghubungkan antara kedua proses training dalam hal validation, hal ini dilakukan untuk mengestimasi performa dari operator. Setelah itu, hasil dari validation dalam training dihubungkan ke bagian optimize selection example sebagai set in.



Gambar III-2 Metode cross validation pada Rapid Miner 5

Dataset ditraining pada pada gambar III-2 yang menggunakan decision tree agar dapat dihasilkan performa dari algoritma yang digunakan. Dimana information gain diisikan dengan parameter seperti gambar III-3 dibawah ini:

Parameter	Value
criterion	information_gain
minimal size for split	4
minimal leaf size	2
minimal gain	0.1
maximal depth	20
confidence	0.25

Gambar III-3 Parameter untuk proses *decision tree*

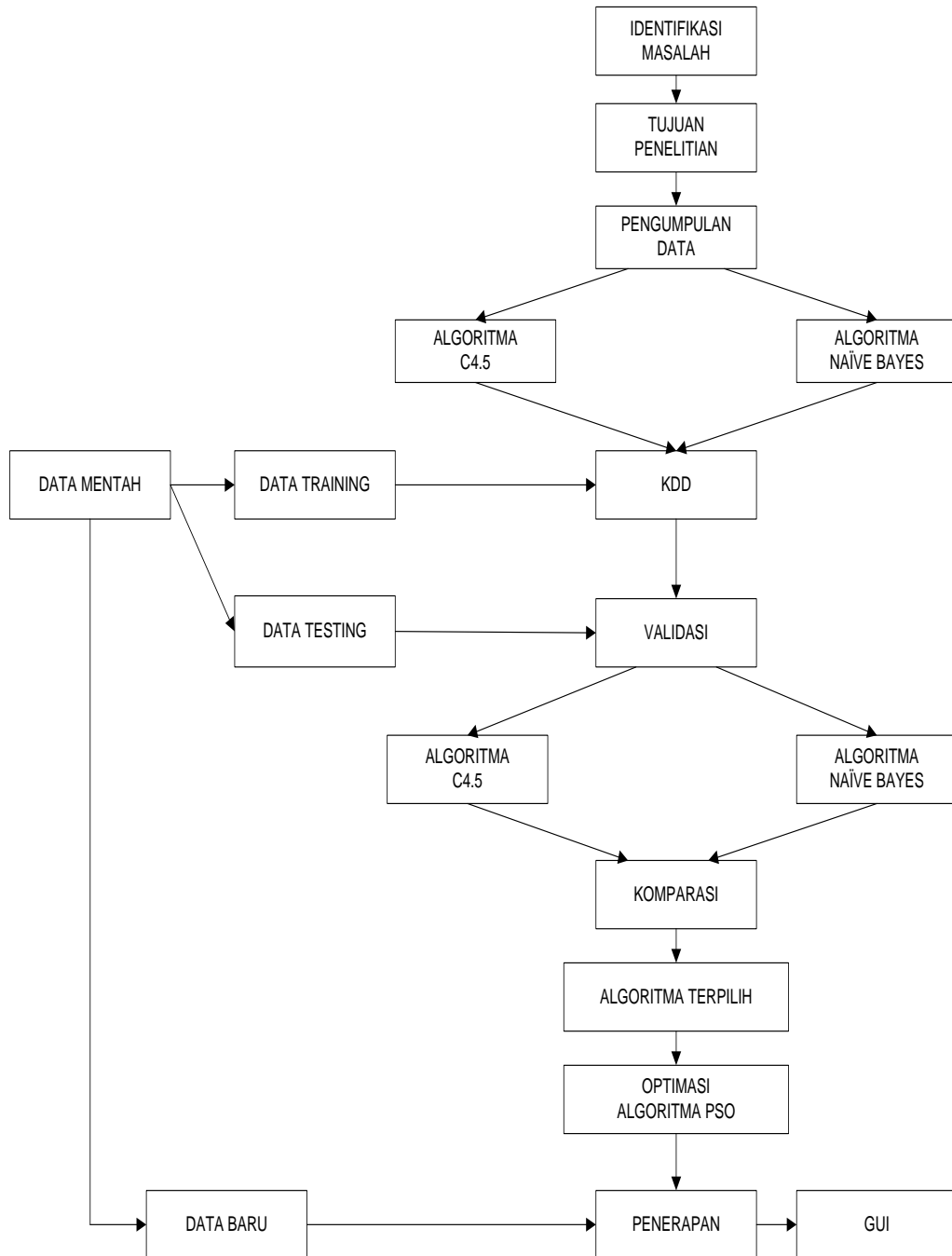
Hasil yang diperoleh kemudian dibandingkan untuk mengetahui metode mana yang paling akurat pada model yang diusulkan. Hasil yang dibandingkan antara lain atribut-atribut yang digunakan pada model, nilai AUC, nilai akurasi, performa model (*f-measure, precision, recall*).

3.7. Evaluasi dan Validasi Hasil

Evaluasi terhadap model yang terbentuk akan dilakukan dengan pengukuran akurasi. Akurasi diukur dengan menggunakan tabel *confusion matrix*. *Confusion Matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negative yang benar, dan prediksi negative yang salah. Sehingga model yang terbentuk dapat langsung diuji dengan data yang secara acak dipisahkan dengan empat variable.

Dengan *confusion matrix*, nilai akurasi dari model akan dibandingkan antara model yang terbentuk dengan algoritma C4.5 dan Naïve Bayes yang memiliki akurasi lebih tinggi dengan salah satu algoritma yang sudah dioptimasi. Untuk melihat kualitas model yang dihasilkan.

3.8. Langkah-Langkah Penelitian



Gambar III-4. Langkah-langkah penelitian

Permasalahan dalam penelitian ini adalah bagaimana cara mengkompaarsi hasil algoritma C4.5 dan *Naïve Bayes* agar keakurasiannya lebih dari 80-90%, dan

apakah dengan teknik optimasi algoritma PSO dapat lebih meningkatkan hasil akurasi dari algoritma yang terpilih. Tujuan penelitian ini untuk meningkatkan hasil keakuratan algoritma yang terpilih menggunakan algoritma PSO dari hasil komparasi algoritma C4.5 dengan *Naïve Bayes*, serta ingin membuktikan bahwa algoritma PSO dapat meningkatkan presentase keakuratan algoritma yang terpilih. Untuk mencapai tujuan tersebut maka dilakukan klasifikasi dengan dua algoritma yaitu C4.5 dan *Naïve Bayes*. Data gizi balita yang diambil untuk klasifikasi sejumlah 206 data. Kemudian data mentah tersebut dibagi menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 80% dan 20%.

Data *training* diolah dengan Sembilan langkah di *Knowledge Discovery in Databases* (KDD). Hasilnya berupa *rule* klasifikasi gizi balita dari algoritma C4.5 dan *Naïve Bayes*. setelah itu *rule* klasifikasi gizi balita dari dua algoritma tersebut divalidasi dengan menggunakan data *testing* dari validasi *rule* algoritma C4.5 dan *Naïve Bayes*. Dari hasil tersebut dipilih algoritma dengan tingkat akurasi tertinggi, setelah itu dilakukan optimasi menggunakan algoritma PSO agar hasil keakuratannya meningkat. Untuk penerapan digunakan data baru dengan menggunakan algoritma optimasi PSO, kemudian ditampilkan dalam bentuk GUI.

BAB IV
PEMBAHASAN HASIL PENELITIAN

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil pengolahan data dan analisa, maka dapat diambil kesimpulan sebagai berikut :

1. Bahwa untuk meningkatkan keakurasian teknik komparasi algoritma C4.5 dan *Naïve Bayes* dengan teknik optimasi adalah dengan menggunakan algoritma *Particle Swarm Optimization* (PSO).
2. Teknik optimasi menggunakan algoritma *Particle Swarm Optimization* dapat lebih meningkatkan hasil akurasi dari algoritma yang terpilih dari hasil komparasi algoritma C4.5 dan *Naïve Bayes*, algoritma yang terpilih adalah algoritma yang paling tinggi nilai *accuracy* pada model klasifikasi C4.5 dengan tingkat *accuracy* 72.13%, setelah dioptimasi dengan algoritma *Particle Swarm Optimization* tingkat *accuracy* meningkat menjadi 80.04%.

5.2. Saran

Berdasarkan pembahasan dan kesimpulan di atas maka penulis dapat menyarankan beberapa hal dari hasil penelitian ini sebagai berikut :

1. Algoritma C4.5 dapat memberikan solusi dalam deteksi status gizi balita dan algoritma *Particle Swarm Optimization* dapat meningkatkan presentase keakuratan algoritma C4.5, sehingga dapat lebih baik lagi dalam memberikan pemecahan masalah klasifikasi deteksi status gizi balita.
2. Aspek sistem pengambilan keputusan dan pengembangan sistem, model ini dapat diterapkan pada Puskesmas-Puskesmas menggunakan *software* RapidMiner atau aplikasi *interface* yang sudah dirancang untuk melakukan deteksi status gizi balita dapat dipergunakan secara langsung oleh tenaga medis yang berada di Puskesmas. Rancangan interface dibuat

sederhana sehingga mudah dalam penggunaannya. Sistem deteksi yang dirancang dapat digunakan oleh pihak Puskesmas untuk mengetahui status gizi balita sehingga jika balita berada dalam kondisi selain gizi baik maka dapat dilakukan tindakan–tindakan medis untuk meningkatkan status gizi balita.

3. Peneliti selanjutnya dapat melakukan penelitian lanjutan dengan membandingkan lebih dari dua algoritma menggunakan algoritma data mining yang lain seperti *K-Nearest Neighbor*, *Random Forest*, *Linear Discriminant Analysis* dan *Neural Network* untuk meningkatkan keakurasian dalam deteksi status gizi balita di atas 80% sehingga lebih akurat lagi dalam melakukan deteksi dini status gizi balita.

DAFTAR PUSTAKA

- [Chamidah A. N. 2009] Chamidah, A N. 2009, Deteksi Dini Gangguan Pertumbuhan dan Perkembangan Anak. *Jurnal Pendidikan Khusus* Vol 1.
- [Octaviani dkk 2008] Octaviani, Ulfa., N. Juniarti., dan A. Mardiyah, 2008, Hubungan Keaktifan Keluarga Dalam Kegiatan Posyandu Dengan Status Gizi Balita Di Desa Rancaekek Kulon Kecamatan Rancaekek. Universitas Padjajaran. Bandung.
- [Supariasa dkk 2001] Supariasa I. D. N., B. Bakri., dan I. Fajar, 2001, *Penilaian Status Gizi*, EGC. Jakarta.
- [Turban dkk 2005] Turban, E., Aronson, J.E., & Liang, T., 2005, *Decision Support Sistem and Intelligent Sistem*.
- [Larose 2005] Larose, Daniel, T., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey : Jhon Willey & Son Inc.
- [Mohammed & Wagner 2014] Mohammed J, Zaki., & Wagner, Meira, Jr., 2014, *Data Mining And Analysis : Fundamental Concepts and Algorithms*. Rensselaer Polytechnic Institute, Troy, New York & Universidade Federal de Minas Gerais, Brazil
- [Sani & Dedy 2010] Sani, S., Dedy, S., Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data, 2010.
- [Usama dkk 1996] Usama, Fayyad., Gregory, P.S., & Padhraic, Smyth., 1996, *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence (AAAI).
- [Brachman & Anand 1996] Brachman, R., and Anand, T., 1996, *The Process of Knowledge Discovery in Databases: A Human-Centered Approach*. In *Advances in Knowledge Discovery and Data Mining*. Uthurusamy. Menlo Park, Calif.: AAAI Press.
- [Quinlan 1993] Quinlan, J. R., 1993, "C4.5: Programs for Machine Learning", California: Morgan Kaufmann.
- [Wu & Kumar 2009] Wu., & Kumar, V., 2009, *The Top Ten Algorithms in Data Mining*. Boca Raton : Chapman & Hall.

- [Han & Kamber 2006] Han, J., & Kamber, M., 2006, *Data Mining Concept and Techniques*. San Francisco: Morgan Kauffman.
- [Kusrini 2009] Kusrini, E.T. Luthfi., 2009, *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [Bramer 2007] Bramer, Max., 2007, *Principles of Data Mining*. London: Springer
- [Haupt & Haupt 2004] Haupt, R. L., Haupt, S.E., 2004, *Partical Genetic Algorithm* (second edition). New Jersey: Wiley.
- [Shukla 2010] Shukla, 2010, *Real Life Application of Soft Computing*. Taylor and Francis Group, LLC.
- [Sucahyo 2003] Sucahyo, G. Y., Data Mining Menggali Informasi yang Terpendam, 2003, <http://wsilfi.staff.gunadarma.ac.id/Downloads/files/4413/yudho-datamining.pdf> (Diakses 16 Maret 2015).
- [Vercellis 2009] Vercellis, C., 2009, *Business Intelligence: Data Mining and Optimization for Decision Making*. John Wiley & Sons, Inc.
- [Liao 2007] Liao, T.Warren, and Evangelos Triantaphyllou., 2007, *Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications*. 5 Toh Tuck Link: World Scientific Publishing Co. Pte. Ltd.
- [Gorunescu 2011] Gorunescu, Florin, 2011, *Data Mining Concepts, Model and Techniques*. Springer; 1st Edition. edition (June 24, 2011).
- [Alpaydin 2010] Alpaydin, Ethem., 2010, *Introduction to Machine Learning 2nd*. London: The MIT Press.
- [Maimon 2010] Maimon, Oded, and Lior Rokach., 2010, *Data Mining and Knowledge Discovery Handbook*. Israel: Springer Science and Business Media.
- [Craig & David, 1986] Craig, Stanfill., and David Waltz., 1986, *Toward Memory-Based Reasoning*. Thinking Machines Corporation, Cambridge. Computer Dept, Brandeis Univ, Waltham.
- [Dennis dkk 2013] Dennis., dkk, Belajar *Data Mining* dengan RapidMiner, 2013.
- [Jin dkk 2003] Jin, Huang., Jingjing, Lu., and Charles X. Ling., 2003, *Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy*.

Department of Computer Science, The University of Western Ontario London.

[Hartana 2014] Hartana, Wijaya., 2014, Optimasi Penerapan Algoritma Genetika Dengan Menggunakan Metode C4.5 Untuk Memprediksi Penyakit Kanker Payudara. Tesis, Jakarta: Universitas Budi Luhur.

[Sevi 2012] Sevi, Andriasari., 2012, Kajian Penerapan Algoritma C4.5, *Naïve Bayes*, dan *Neural Network* dalam deteksi nasabah penjualan kredit motor bermasalah: Studi kasus pada PT. MACF Pringsewu Lampung. Tesis, Jakarta: Universitas Budi Luhur.

[Renny, 2013] Renny, Aisyah., 2013, Kajian Komparasi Penerapan Algoritma C4.5, *Neural Network* dan *K-Nearest Neighbor* dalam Penentuan Kelayakan Kredit Kendaraan Bermotor : Studi Kasus PT. RAZ. Tesis, Jakarta: Universitas Budi Luhur.

[Sugiyono, 2005] Sugiyono, 2005, Memahami Penelitian Kualitatif, Bandung: Alfabeta.

[Kemenkes RI, 2013] Kemenkes RI, 2013, Penyajian Pokok-Pokok Hasil Riset Kesehatan Dasar 2013, Kementerian Kesehatan Republik Indonesia, Jakarta.

[Setiyaji, 2017] Setiyaji, B. (2017). SISTEM PAKAR DIAGNOSA KERUSAKAN MOTOR 4 TAK HONDA VARIO DENGAN ALGORITMA C4. 5 (Doctoral dissertation, STMIK Sinar Nusantara Surakarta).

Bustami, B. (2013). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2).

[Dwiasnati & Devianto, 2019] Dwiasnati, S. (2019). Optimasi Prediksi Keputusan Calon Nasabah Potensial menggunakan Algoritma C 4.5 berbasis Particle Swarm Optimization. *Jurnal Informatika*, 6(2), 286-292.

[Septikasari, 2018] Septikasari, M. (2018). Status gizi anak dan faktor yang mempengaruhi. UNY Press.



Jon Idrison Molina adalah dosen pada Program Studi Teknik Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tribuana Kalabahi. Ia lahir di Masmur, Desa Kallesa, Kecamatan Pureman, Alor-NTT, 16 Januari 1987. Ia tamat SD Negeri Kolomana pada tahun 1999; SMP Negeri 2 Kalabahi pada tahun 2002; dan SMK Negeri 1 Kalabahi pada tahun 2005. Gelar Sarjana Komputer diperolehnya dari Program Studi Sistem Informasi STIKOM Uyelindo Kupang pada tahun 2010 dan pendidikan S-2 diperolehnya dari Program Studi Magister Ilmu Komputer Universitas Budi Luhur Jakarta pada tahun 2015 dalam bidang Teknologi Sistem Informasi.

Mata kuliah yang sedang dan pernah diampu terdiri dari Bahasa Rakitan, Komunikasi Data, Rekayasa Perangkat Lunak, Keamanan Sistem dan Data, Kecerdasan Buatan, Desain Grafis, Pengantar Ilmu dan Teknologi Maritim, Perancangan dan Implementasi Perangkat Lunak, Pemrograman Jaringan, dan Sistem Multimedia. Ia telah menerbitkan beberapa artikel ilmiah di jurnal lokal dan jurnal nasional, serta pernah menjadi presenter di Seminar Nasional Pengaplikasian Telematika (SINAPTIKA) yang diselenggarakan oleh Fakultas Ilmu Komputer Universitas Mercu Buana pada tanggal 12 Desember 2020.



Hari Soetanto merupakan staf pengajar atau dosen Universitas Budi Luhur, yang berhasil menyelesaikan pendidikan S-1 pada tahun 1992 di STMIK Budi Luhur atau Universitas Budi Luhur, dan menyelesaikan pendidikan S-2, pada tahun 1997 di Asian Institute Technology, dan berhasil menyelesaikan pendidikan S-3 pada tahun 2018 di Universitas Gadjah Mada. Saat ini ia menjabat sebagai Deputy Digitalisasi dan Pengembangan Pembelajaran di Universitas Budi Luhur, dan Pengkaji Bidang Teknologi Informasi Universitas Budi Luhur. Beberapa riwayat pekerjaan di

antaranya yaitu: Pada tahun 1999–2006 menjadi Ketua jurusan Teknik Informatika STMIK Budi Luhur; pada tahun 1999–2004 menjabat sebagai Ka. PPM STMIK Budi Luhur; pada tahun 1996–1999 menjabat sebagai Kepala Seksi Pengabdian Masyarakat pada Pusat PPM STMIK Budi Luhur; pada tahun 1995–1996 menjabat sebagai Kepala Sub Bagian Humas dan Percetakan Yayasan Pendidikan Budi Luhur; pada tahun 1995–1996 menjabat sebagai Kepala Seksi Pengabdian pada Masyarakat pada Bid. PPM STMIK Budi Luhur; dan pada tahun 1994–1999 menjabat sebagai Kepala Laboratorium Komputer Bidang PPM STMIK Budi Luhur.



Penerbit Deepublish (CV BUDI UTAMA)
Jl. Rajawali, Gang Elang 6 No.3, Drono, Sardonoharjo, Ngaglik, Sieman
Jl. Kaliurang Km 9,3 Yogyakarta 55581
Telp/Fax : (0274) 4533427
Anggota IKAPI (076/DIY/2012)
cs@deepublish.co.id @penerbitbuku_deepublish
Penerbit Deepublish www.penerbitbukudeepublish.com

Kategori : Ilmu Komputer

ISBN 978-623-02-3038-7



9 786230 230387