

DAFTAR ISI

	Hal
JUDUL	i
HALAMAN PENGESAHAN	ii
PANITIA UJIAN DISERTASI	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	vii
DAFTAR TABEL	ix
DAFTAR SIMBOL	xi
KATA PENGANTAR	xiii
ABSTRAK	xv
ABSTRACT	xvi
MOTTO	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian	7
1.4 Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	8
2.1 <i>Naïve Bayes</i>	8
2.1.1 <i>Naïve Bayes</i> Untuk Klasifikasi	9
2.1.2 <i>Naïve Bayes</i> (NB) untuk Atribut Campuran	11
2.1.3 <i>Naïve Bayes Imputation</i> (NBI).....	13
2.2 Konsep <i>Self Organizing Map</i> (SOM)	16
2.2.1 <i>Self Organizing Map Imputation</i> (SOMI)	18
2.2.2 <i>Multiple Imputations</i>	20
2.2.3 <i>Self Organizing Map Single Imputation</i> (SOMSI).....	21
2.2.4 <i>Self-Organizing Map Multiple Imputation</i> (SOMMI)	24
2.3 Model <i>hybrid SOMI Naïve Bayes</i>	26
2.4 Algoritma Evolusi untuk Masalah Pemilihan Fitur	28
2.5 Algoritma Genetika	29

2.5.1 Representasi Kromosom	31
2.5.1.1 Pengkodean Biner.....	32
2.5.1.2 Pengkodean Permutasi (permutation encoding)	32
2.5.1.3 Pengkodean Nilai (Value Encoding).....	33
2.5.2 Fungsi <i>Obyektif Kebugaran (Fitness)</i>	33
2.5.3 <i>Selection</i>	33
2.5.3.1 <i>Roulette Wheel Selection</i>	34
2.5.3.2 Seleksi Turnamen	34
2.5.3.3 Seleksi Elit (<i>Elitism Selection</i>).....	35
2.5.4 <i>Crossover</i>	35
2.5.5 Mutasi	36
2.6 Pemilihan Atribut dengan Algoritma Genetika pada Klasifikasi Naïve Bayes.....	36
2.7 Optimasi <i>Missing value</i> pada metode SOM <i>Naïve Bayes</i> dengan menggunakan Algoritma Genetika	41
2.8 Algoritma Genetika Untuk Pengaturan Bobot SOM.....	44
BAB III KERANGKA KONSEP DAN HIPOTESIS	47
3.1. Kerangka Konsep.....	47
3.2. Hipotesis	49
BAB IV METODE PENELITIAN.....	51
4.1 Penggunaan Data	51
4.2 Tahapan Umum Kegiatan Penelitian	52
4.2.1 Preprocessing Data	53
4.2.2 Teknik penyiapan missing data.....	54
4.2.3 Pengembangan Model Imputasi Data menggunakan <i>Self Organizing Map Imputation (SOMI)</i> dan SOMMI pada data yang hilang.....	54
4.3 Proses <i>Hybrid</i> SOMINB.....	57
4.4 Proses Seleksi Fitur dengan Algoritma Genetika pada <i>Naïve Bayes</i>	60
4.5 Proses Pengembangan model SOMIGANB	61
4.6 Optimasi Bobot Algoritma SOM Dengan Algoritma Genetika	63
4.7 Pengukuran Kinerja Klasifikasi.....	65

4.8 Evaluasi Validasi Testing k-fold Cross Validation	66
4.9 Spesifikasi Kebutuhan	67
BAB V HASIL PENELITIAN DAN PEMBAHASAN.....	68
5.1 Pengolahan Data Awal (Preprocessing)	69
5.2 Proses Imputasi pada <i>Missing Value</i>	70
5.3 Hasil Evaluasi Kinerja <i>Self Organizing Map</i> (SOM)	72
5.3.1 Proses Imputasi menggunakan SOMSI pada <i>Missing Value</i>	73
5.3.2 Proses Imputasi menggunakan SOMMI pada <i>Missing Value</i>	74
5.3.3 Hasil Evaluasi Kinerja <i>Self Organizing Map Single Imputation</i> (SOMSI).	75
5.3.4 Hasil Evaluasi Kinerja <i>Self Organizing Map Multiple</i> <i>Imputation</i> (SOMMI)	80
5.3.5 Evaluasi Kinerja Imputasi.....	82
5.4 Evaluasi Kinerja Klasifikasi <i>Naïve Bayes</i> (NB)	84
5.4.1 Evaluasi Kinerja <i>Naïve Bayes Imputation</i> (NBI).....	85
5.4.2 Evaluasi Eksperimental SOMINB untuk Dataset Ucirepository	87
5.5 Kinerja Seleksi Fitur Algoritma Genetika pada Klasifikasi <i>Naïve Bayes</i> (GANB)	91
5.6 Kinerja Model hybrid SOMI, Algoritma Genetika Dalam Klasifikasi <i>Naïve Bayes</i>	95
5.7 Kinerja Optimalisasi bobot Awal untuk <i>Self Organizing</i> <i>Map</i> (SOM) Menggunakan Algoritma Genetika	99
BAB VI KESIMPULAN DAN SARAN.....	103
6.1 Kesimpulan	103
6.2 Saran	104
DAFTAR PUSTAKA	105
LAMPIRAN	114

DAFTAR GAMBAR

Gambar 2.1 Imputasi data yang hilang dengan SOMI 21

Gambar 2.2 Imputasi data yang hilang dengan SOMMI..... 24

Gambar 2.3 Ilustrasi kinerja algoritma GA 30

Gambar 2.4 Ilustrasi panjang kromosom GA 31

Gambar 2.5 Proses pemilihan kromosom dengan *Roda Rolet* 34

Gambar 2.6 lustrasi pada pertukaran kromosom dengan One-point, two-point, and uniform crossover methods 35

Gambar 2.7 Fitur terpilih 37

Gambar 2.8 Ilustrasi Kromosom bobot dan seleksi fitur..... 43

Gambar 2.9 Diagram alir GASOM..... 45

Gambar 2.10 Representasi kromosom pada bobot masing-masing fitur..... 47

Gambar 3.1 *State of the art* imputasi data dengan menggabungkan metode SOMI, GA dan *Naïve Bayes*..... 48

Gambar 4.1 Diagram tahapan utama penelitian 53

Gambar 4.2 Proses umum diagram algoritma SOMMI..... 56

Gambar 4.3 Flowchart *Self Organizing Map Imputation* (SOMI) dan NB..... 59

Gambar 4.4 Flowchart Preprosesing SOMI/GA/NB..... 62

Gambar 4.5 Proses uji coba *k-fold cross Validation* 66

Gambar 5.1 Alur proses pengujian 68

Gambar 5.2 Visualisasi fitur yang hilang data hepatitis dengan persentase tertentu 71

Gambar 5.3 *Quantization errors* pada SOM 72

Gambar 5.4 Hubungan *epoch* dengan *learning rate* pada SOM 72

Gambar 5.5 Visualisasi Data *Missing Value* pada segmentasi pelanggan 77

Gambar 5.6 Hasil Learning SOM Pengelompokan Data 5 *Cluster* Segmentasi Pelanggan 78

Gambar 5.7 Proses imputasi SOMMI Segmentasi Pelanggan berdasarkan *cluster* dan RMSEnya..... 79

Gambar 5.8 Hasil Kinerja imputasi SOMMI Segmentasi Pelanggan berdasarkan perbedaan *cluster*..... 81

Gambar 5.9 Hasil Uji Coba SOMMI dibandingkan metode lainnya 83

Gambar 5.10 Resume *Confusion matrix* untuk untuk data sampel data iris..... 84

Gambar 5.11 Persebaran data Iris	86
Gambar 5.12 Grafik perbandingan metode imputasi NBI dengan metode lainnya...	87
Gambar 5.13 Hubungan antara <i>error</i> terhadap pengelompokan partitif dari data set a. Heart dan b. Horse Colic	88
Gambar 5.14 Perbedaan Populasi berdasarkan fungsi kebugaran.....	91
Gambar 5.15 Hasil seleksi fitur pada GA.....	92
Gambar 5.16 Mekanisme operasi crossover.....	99
Gambar 5.17 Mekanisme operasi mutasi	99
Gambar 5.18 Hubungan antara perbedaan populasi dan fungsi kebugaran	99
Gambar 5.19 Hubungan pengaruh jumlah generasi terhadap fungsi kebugaran.....	100
Gambar 5.19 Perbedaan kinerja GASOM dan SOM.....	102

DAFTAR TABEL

Tabel 2.1 Sampel Data D keseluruhan yang Mengandung *Missing value* 14

Tabel 2.2 Sampel Data keseluruhan $X_D = X_C \cup X_M$ 14

Tabel 2.3 Sampel Data Lengkap dengan label kelas dianggap sebagai atribut..... 15

Tabel 2.4 Sampel Data dengan data yang hilang dianggap sebagai kelas..... 15

Tabel 2.5 lustrasi representasi kromosom biner 32

Tabel 2.6 Ilustrasi representasi kromosom riil..... 32

Tabel 2.7 Ilustrasi representasi kromosom kontinu, karakter dan kata 33

Tabel 2.8 Proses Mutasi 36

Tabel 2.9 Fitur terpilih..... 42

Tabel 4.1 Detail data set yang digunakan 51

Tabel 4.2 Parameter jaringan SOM 54

Tabel 4.3 Kriteria Terminasi GASOM..... 64

Tabel 4.4 Mekanisme Perhitungan 66

Tabel 4.5 Proses uji coba k-fold cross Validation..... 67

Tabel 5.1 Konversi Data Atribut Campuran pada Sampel Data Hepatitis 69

Tabel 5.2 Hasil transformasi z-score 70

Tabel 5.3 Identifikasi Data *Missing Value* 70

Tabel 5.4 Hasil Preimputasi dan Pascaimputasi dengan SOMSI 73

Tabel 5.5 Hasil Preimputasi dan Pascaimputasi dengan SOMMI..... 75

Tabel 5.6 Statistik Deskriptif Data Transformasi Segmentasi Pelanggan
berdasarkan Persentase *Missing* 76

Tabel 5.7 Hasil Imputasi pada 1 variabel *Missing* pada Segmentasi Pelanggan
pada kelompok *Missing* 25 Persen 76

Tabel 5. 8 Simulasi Hasil Imputasi 2 Variabel pada Segmentasi Pelanggan
Persentase *Missing* 25 Persen..... 77

Tabel 5. 9 Hasil Imputasi untuk Set Data dari Berbagai Metode (%)..... 79

Tabel 5.10 Nilai RMSE Proses imputasi Segmentasi Pelanggan yang menggunakan
bobot berulang pada perbedaan *cluster* 80

Tabel 5.11 Nilai RMSE Proses imputasi Segmentasi Pelanggan yang menggunakan
bobot pada variasi *missing* data 81

Tabel 5.12 Uji korelasi r dan uji t berpasangan data segmentasi pelanggan. 82

Tabel 5.13 Hasil Uji performa klasifikasi NB untuk sampel Iris	84
Tabel 5.14 Hasil pengujian korelasi data Iris	85
Tabel 5.15 Nilai rata-rata sebaran data dan Standar Deviasi untuk imputasi.....	86
Tabel 5.16 Hasil perbedaan kinerja metode imputasi nilai imputasi untuk atribut campuran.....	89
Tabel 5.17 Perbandingan persentase akurasi metode percobaan data set Heart dengan menggunakan beberapa metode.....	89
Tabel 5.18 Perbandingan persentase akurasi metode percobaan data set Diabetes dengan menggunakan semua metode.....	90
Tabel 5.19 Perbandingan persentase akurasi metode percobaan data set Australian Credit dengan menggunakan beberapa metode.....	90
Tabel 5.20 Perbandingan persentase akurasi metode percobaan data set Horse Colic dengan menggunakan beberapa metode	90
Tabel 5.21 Jumlah fitur terpilih pada data set multivariate Sonar.....	92
Tabel 5.22 Jumlah fitur terpilih pada data set Segmentasi Pelanggan	93
Tabel 5.23 Perubahan probabilitas Crossover P_c algoritma genetika pada klasifikasi NB data Segmentasi Pelanggan.....	93
Tabel 5.24 Perubahan probabilitas Crossover algoritma genetika pada klasifikasi NB data Sonar.....	94
Tabel 5.25 Perubahan probabilitas mutasi algoritma genetika pada klasifikasi NB data Segmentasi Pelanggan	94
Tabel 5.26 Perubahan probabilitas mutasi algoritma genetika pada klasifikasi NB data Sonar	95
Tabel 5.27 Perbandingan akurasi Klasifikasi (%) SOMIGANB data Segmentasi Pelanggan dengan beberapa metode lainnya.....	96
Tabel 5.28 Perbandingan akurasi Klasifikasi (%) SOMIGANB data Heart dengan beberapa metode lainnya.....	97
Tabel 5.29 Perbandingan akurasi Klasifikasi (%) SOMIGANB Data WDBC dengan beberapa metode lainnya Data WDBC	97
Tabel 5.30 Perbandingan <i>F-Measure</i> pada Dataset Segmentasi Pelanggan	98
Tabel 5.31 Perbandingan populasi terhadap waktu eksekusi	101

DAFTAR SIMBOL

X	: Data dengan <i>class</i> yang belum diketahui
H	: Hipotesis data merupakan suatu <i>class</i> spesifik
$P(H X)$: Probabilitas hipotesis H berdasar kondisi X (<i>posterior probabilitas</i>)
$P(H)$: Probabilitas hipotesis H (<i>prior probabilitas</i>)
$P(X H)$: Probabilitas hipotesis X berdasar kondisi pada hipotesis H
$P(X)$: Probabilitas hipotesis X
P	: Peluang
X_i	: Nilai atribut ke $-i$
x_i	: Nilai atribut ke $-i$
Y	: Kelas yang dicari
Y	: Nilai sub kelas Y yang dicari
π_y	: <i>Mean</i> yang menyatakan rata-rata dari seluruh atribut
σ_y	: Standar deviasi yang menyatakan varian dari seluruh atribut
v	: Nilai output hasil klasifikasi <i>Naïve Bayes</i> .
R_C	: Bilangan diskrit dinyatakan
R_d	: Atribut bilangan kontinyu adalah.
$\varepsilon(t)$: Parameter gain adaptasi yaitu $[0,1]$
$\pi_{(M_x)}(.)$: Bobot menggantikan nilai yang hilang pada M_x dari sampel x .
d	: Urutan atribut, atribut kelas dilambangkan sebagai
c_m	: 1, 2, ..., m adalah jumlah kelas.
$x_{i_{mis}}$: Nilai-nilai yang hilang
w	: Komponen vektor bobot
$x_{i,UB}$: Batas atas dan variabel desain ke- i ,
$x_{i,LB}$: Batas bawah variabel desain ke- i ,
$\Theta = (\theta_1, \theta_2, \dots, \theta_k)$: Nilai fitur model campuran yang dikonversi ke bentuk nilai diskrit
m	: Jumlah cluster
n_j	: Jumlah data dalam cluster j
n	: Jumlah dari semua data.
F	: Indeks fitur

nf	: Jumlah fitur
n_s	: Jumlah fitur yang dipilih
MSE	: Perhitungan kesalahan <i>Mean Square Error</i>
$RMSE$: Akar kesalahan <i>Mean Square Error</i>
Y_i	: Data sebenarnya (data awal),
\hat{Y}_i	: Data hasil estimasi (data akhir),
n	: Jumlah data.
NBI	: <i>Naive Bayes Imputation</i>
GAI	: <i>Genetic Algorithm Imputation</i>
SOMI	: <i>Self Organizing Map Imputation</i>
SOMMI	: <i>SOM Multiple Imputation</i>
PSOI	: <i>Particle Swarm Optimization Imputation</i>
ACOI	: <i>Ant Colony Optimization Imputation</i>
GA KNNI	: <i>Genetic Algorithm k-Nearest Neural Network Imputation</i>
GANBI	: <i>Genetic Algorithm Naive Bayes Imputation</i>
GAINB	: <i>Genetic Algorithm Imputation Naive Bayes</i>
FCMI	: <i>Fuzzy C-Means Imputation</i>
KMI	: <i>K-Means Imputation</i>
GASVM	: <i>Genetic Algorithm Support Vector Machine</i>
GAKNN	: <i>Genetic Algorithm k-Nearest Neural Network</i>
GARF	: <i>Genetic Algorithm Random Forest</i>
GADT	: <i>Genetic Algorithm Decision Tree</i>
GA KNNI	: <i>Genetic Algorithm k-Nearest Neural Network Imputation</i>
SOMINB	: <i>Self Organizing Map Imputation Naive Bayes</i>
GASOMINB	: <i>Genetic Algorithm Self Organizing Map Imputation Naive Bayes</i>
GASOM	: <i>Genetic Algorithm Self Organizing Map</i>

KATA PENGANTAR

Alhamdulillahirobbil'alamin, puji syukur atas segala limpahan nikmat dan karunia Allah SWT, yang telah memberikan petunjuk, rahmat dan ridho-Nya. Sehingga penulis dapat menyelesaikan disertasi ini dengan judul "**Optimasi Pemodelan Hybrid Naive Bayes Berbobot dengan Menggunakan Algoritma Genetika Pada Data Missing Value**".

Saya mengucapkan terima kasih yang sebesar-besarnya kepada yang terhormat kedua promotor Dr. Miswanto, M.Si. selaku promotor pertama dan Dr. Herry Suprajitno, M.Si. selaku Ko-promotor, yang dengan penuh perhatian dan kesabaran selalu meluangkan waktu, memberikan pengarahan dan motivasi dalam penulisan disertasi ini. Penulis dapat menyelesaikan disertasi ini, tidak terlepas dari bantuan dan kerjasama dari berbagai pihak, sehingga menyampaikan terima kasih yang tak terhingga kepada:

1. Prof. Dr. Mohammad Nasih SE., M.T., Ak., CMA selaku rektor universitas Airlangga yang telah memberikan kesempatan mendapatkan beasiswa BUDIDN Program doctoral S3 MIPA, Universitas Airlangga
2. Prof. Win Darmanto, M.Si., Ph.D. selaku Dekan Fakultas Sains dan Teknologi, yang telah memberikan fasilitas kepada saya untuk mengikuti dan menyelesaikan pendidikan pada Program Doctoral S3 MIPA, Universitas Airlangga
3. Dr. Miswanto, M.Si., selaku promotor pertama disertasi atas arahan, bimbingan dan motivasinya dalam menyelesaikan perkuliahan maupun penulisan disertasi ini.
4. Dr. Herry Suprajitno, M.Si., selaku Ko-promotor disertasi atas arahan, bimbingan dan motivasinya dalam menyelesaikan perkuliahan maupun penulisan disertasi ini.
5. Dr. Alfinda Novi Kristanti, DEA., selaku Koprodi S3 MIPA yang telah memberikan dukungan untuk melaksanakan disertasi ini.
6. Seluruh Pengajar dan staf Program Doktorat MIPA yang telah mentransfer ilmu pengetahuannya melalui kegiatan perkuliahan serta membantu kelancaran pengurusan administrasi perkuliahan dan penyelesaian disertasi ini.

7. Suamiku Syofyan Haryanto tersayang dan anak-anakku Mohammad Rafi Maulana Haryanto, Yuka Aqila Haryanto dan Naila Rifaya Haryanto yang dengan penuh kesabaran, mendukung dan mendoakan demi selesainya studi ini. Semoga kita semua selalu mendapat ridlo-Nya dalam keberkahan. Aamiin.
8. Orang tua Penulis Bapak Imam Suhadi dan Ibu Sutijah dan Mertuaku Bapak Buchari dan Ibu Uriza Latifah, terimakasih atas segala do'a dan dukungannya sehingga penulis dapat menyelesaikan disertasi ini tepat waktu.
9. Mahasiswa Program Studi S3 MIPA Angkatan 2016 yang selalu kompak dan saling mendukung, saling mendoakan baik dalam perkuliahan maupun dalam penyelesaian penulisan disertasi ini.

Semoga Allah SWT membalas kebaikan semua pihak yang telah memberi kesempatan, dukungan dan bantuan dalam menyelesaikan disertasi ini. Penulis menyadari disertasi ini merupakan dasar penelitian yang akan dikembangkan dalam waktu jangka panjang, oleh karena itu saran dan kritik yang membangun sangat diharapkan demi kesempurnaan tulisan ini, sehingga disertasi ini memberikan manfaat yang baik bagi agama, bangsa dan negara.

Surabaya, 24 Juni 2020

Penulis