

TESIS

**KOMBINASI METODE K-NEAREST NEIGHBOR
DAN NAÏVE BAYES UNTUK KLASIFIKASI DATA**



MEGA KARTIKA SARI

No. Mhs : 135302022/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PROGRAM PASCASARJANA

UNIVERSITAS ATMA JAYA YOGYAKARTA

2015



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : MEGA KARTIKA SARI
Nomor Mahasiswa : 135302022/PS/MTF
Konsentrasi : Soft Computing
Judul Tesis : Kombinasi Metode K-Nearest Neighbor dan Naïve
Bayes untuk Klasifikasi Data

Nama Pembimbing

Tanggal

Tanda Tangan

Dra. Ernawati, M.T.

29/1/15

Dr. Pranowo, S.T., M.T.

29/01/2015



UNIVERSITAS ATMA JAYA YOGYAKARTA

PROGRAM PASCASARJANA

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : MEGA KARTIKA SARI
Nomor Mahasiswa : 135302022/PS/MTF
Konsentrasi : Soft Computing
Judul Tesis : Kombinasi Metode K-Nearest Neighbor dan Naïve
Bayes untuk Klasifikasi Data

Nama Penguji	Tanggal	Tanda Tangan
Dra. Ernawati, M.T. (Ketua)	29/1/15	
Dr. Pranowo, S.T., M.T. (Sekretaris)	29/01/2015	
Dr. Ir. Alb. Joko Santoso, M.T. (Anggota)	29/1/2015	

Ketua Program Studi

Prof. Ir. Suyoto, M.Sc., Ph.D.



UNIVERSITAS ATMA JAYA YOGYAKARTA

PROGRAM PASCASARJANA

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PERNYATAAN

Bersamaan dengan penelitian ini, maka saya yang bertanda tangan di bawah ini :

Nama : MEGA KARTIKA SARI
Nomor Mahasiswa : 135302022/PS/MTF
Konsentrasi : Soft Computing
Judul Tesis : Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes untuk Klasifikasi Data

menyatakan bahwa penelitian ini adalah hasil pemikiran sendiri dan bukan duplikasi dari karya tulis yang telah ada sebelumnya. Karya tulis yang telah ada sebelumnya dijadikan acuan oleh penulis guna melengkapi penelitian ini dan dinyatakan secara tertulis dalam penulisan acuan dan daftar pustaka.

Demikian pernyataan ini dibuat untuk digunakan sebagaimana mestinya.

Yogyakarta, 5 Januari 2015

Mega Kartika Sari

INTISARI

Penambangan data banyak digunakan untuk membantu menentukan keputusan dengan memprediksi tren data masa depan. Penambangan data juga dapat diimplementasikan pada berbagai bidang seperti pendidikan, medis, pemasaran, asuransi, dan lain sebagainya. Algoritma modifikasi juga sudah banyak dimunculkan oleh para peneliti dengan menggabungkan beberapa metode penambangan data. Penggabungan metode penambangan data digunakan untuk meningkatkan proses klasifikasi data dengan mengenali kelemahan dari sebuah metode penambangan data. Kelemahan dari metode tersebut kemudian diatasi dengan metode lain untuk dapat memperbaiki kelemahan tersebut.

Pada penelitian ini, akan dibahas penggabungan dua metode penambangan data untuk klasifikasi data. Metode klasifikasi yang digabungkan ialah KNN dan Naïve Bayes. Kedua metode ini termasuk dalam 10 peringkat metode penambangan data yang sering digunakan. KNN memiliki kelemahan pada tahap klasifikasi yang akan diatasi oleh tahapan yang dimiliki oleh Naïve Bayes.

Penerapan metode kombinasi dari KNN dan Naïve Bayes dapat meningkatkan kecepatan metode penambangan data dengan nilai keakuratan yang tinggi seperti KNN. Metode kombinasi diterapkan dengan bahasa pemrograman C++.

Kata Kunci : Klasifikasi, Penambangan data, Kombinasi Metode, KNN, Naïve Bayes.

ABSTRACT

Data mining is widely used to help determine the decision to predict the future trend of the data. Data mining can be implemented in various fields such as education, health, marketing, insurance, etc. Modification of the algorithm has been raised by many researchers to combine several data mining methods. Merging data mining methods are used to improve the process of data classification by recognizing the drawback of a data mining method. The drawback of the method then, can be solved by another method.

This research will discuss about merging two data mining methods to classify the data. The combined classification methods used in this research are KNN and Naïve Bayes. These methods are in the top 10 frequently used data mining. KNN has a drawback in the data classification phase which will be fixed by Naïve Bayes.

Application of this two combined method, KNN and Naïve Bayes, will accelerate data mining process with high accuracy values such as KNN. This combined method will be applied using C ++ programming language.

Keyword : Classification, Data mining, Combination Method, KNN, Naïve Bayes.

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas berkat-Nya penulis dapat menyelesaikan pembuatan tesis dengan baik. Tujuan penulisan tesis ini adalah untuk memenuhi sebagian persyaratan mencapai derajat sarjana Magister Teknik Informatika dari Program Pascasarjana Universitas Atma Jaya Yogyakarta..

Penulis menyadari bahwa selesainya pembuatan tesis ini tidak bisa terlepas dari bantuan yang diberikan berbagai pihak, baik yang bersifat langsung maupun tidak langsung. Untuk itu, penulis ingin mengucapkan terima kasih yang sebanyak-banyaknya kepada:

1. Profesor Ir. Suyoto, M.Sc., Ph.D. selaku Ketua Program Studi Magister Teknik Informatika yang telah memberikan kemudahan pada penulis dalam menyelesaikan studi Pascasarjana.
2. Ibu Dra. Ernawati, M.T. selaku Dosen Pembimbing I yang telah membimbing penulis dengan sabar selama proses penyelesaian tesis dan membagikan ilmu-ilmu yang bermanfaat selama penulis menyelesaikan studi Strata Dua.
3. Bapak. Dr. Pranowo, S.T., M.T. selaku Dosen Pembimbing II yang telah memberikan banyak sekali masukan kepada penulis dalam menyempurnakan tesis ini dan selalu memberikan semangat kepada penulis hingga tesis ini dapat diselesaikan.

4. Seluruh dosen Program Studi Magister Teknik Informatika dan karyawan Program Pascasarjana Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama menempuh studi.
5. Keluarga penulis yang selalu memberikan doa, dukungan serta semangat bagi penulis, hingga akhirnya penulis dapat menyelesaikan tesis ini.
6. Donny Avianto S.T., M.T. selaku teman penulis yang selalu membantu dan memberi semangat bagi penulis, hingga akhirnya penulis dapat menyelesaikan tesis ini.
7. Pihak yang tidak dapat penulis sebutkan satu persatu atas bantuannya dalam menyelesaikan menyelesaikan tesis ini.

Demikian penulisan tesis ini dibuat dengan sebaik-baiknya oleh penulis. Penulis menyadari bahwa bahwa penulisan tesis ini masih memiliki banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu saran dan kritik yang membangun demi penyempurnaan penulisan tesis ini akan selalu penulis nantikan.

Akhir kata, penulis mengharapkan semoga penulisan tesis ini dapat diambil berguna dan bermanfaat sehingga dapat memberikan inspirasi bagi pembacanya.

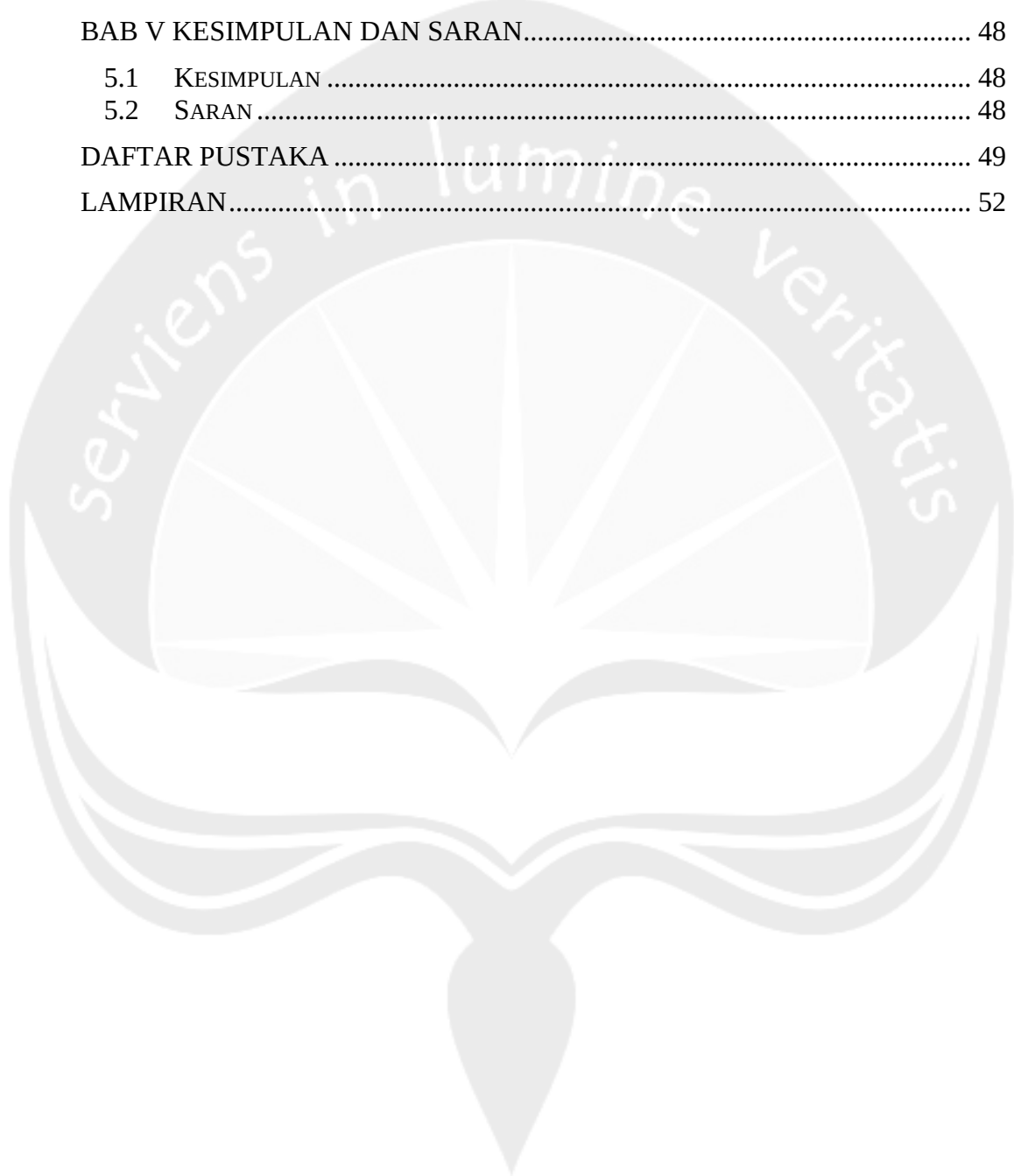
Yogyakarta, 19 Januari 2015

Penulis

DAFTAR ISI

INTISARI.....	V
ABSTRACT.....	VI
KATA PENGANTAR	VII
DAFTAR ISI.....	IX
DAFTAR GAMBAR	XI
DAFTAR TABEL.....	XII
BAB I PENDAHULUAN	13
1.1 LATAR BELAKANG	13
1.2 RUMUSAN MASALAH.....	16
1.3 BATASAN MASALAH.....	16
1.4 KEASLIAN PENELITIAN	16
1.5 MANFAAT PENELITIAN	18
1.6 TUJUAN PENELITIAN.....	18
1.7 SISTEMATIKA PENULISAN	18
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	20
2.1 TINJAUAN PUSTAKA	20
2.2 LANDASAN TEORI.....	22
2.2.1. Data, Informasi dan Pengetahuan	22
2.2.2. Klasifikasi Penambangan data	22
2.2.3. Teknik Klasifikasi	24
BAB III METODOLOGI PENELITIAN	27
3.1 METODE PENELITIAN KEPUSTAKAAN	27
3.2 METODE PEMBANGUNAN PERANGKAT LUNAK.....	27
3.2.1 Implementasi Perangkat Lunak.....	27
3.2.2 Pengujian Perangkat Lunak	27
3.3 LANGKAH PENELITIAN	28
3.4 ALUR METODE KOMBINASI KNN-NAÏVE BAYES	30
3.5 PENERAPAN ALGORITMA PADA C++	31
3.5.1 KNN.....	31
3.5.2 Naïve Bayes	32
3.5.3 Kombinasi KNN-Naïve Bayes.....	34
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	36
4.1 IMPLEMENTASI SISTEM	36
4.2 HASIL PENGUJIAN DAN ANALISIS SISTEM.....	38
4.2.1 Nursery Data Set	38

4.2.2	Car Evaluation Data Set	40
4.2.3	Balance Scale Data Set	41
4.2.4	Hasil Pengujian Metode Kombinasi	43
4.3	KELEBIHAN DAN KEKURANGAN ALGORITMA	46
BAB V KESIMPULAN DAN SARAN.....		48
5.1	KESIMPULAN	48
5.2	SARAN	48
DAFTAR PUSTAKA		49
LAMPIRAN.....		52



DAFTAR GAMBAR

Gambar 3. 1 Tahap pengolahan data.....	29
Gambar 3. 2 Alur metode kombinasi KNN-Naïve Bayes.....	31
Gambar 4. 1 Tahap implementasi pada sistem	37



DAFTAR TABEL

Tabel 1. 1 Perbandingan Penelitian Metode Kombinasi Penambahan data.....	17
Tabel 4. 1 Keterangan Atribut Nursery Data Set	39
Tabel 4. 2 Keterangan Kelas Nursery Data Set	40
Tabel 4. 3 Keterangan Atribut Car Evaluation Data Set.....	41
Tabel 4. 4 Keterangan Kelas Car Evaluation Data Set	41
Tabel 4. 5 Keterangan atribut Balance Scale Data Set	42
Tabel 4. 6 Keterangan Kelas Balance Scale Data Set.....	42
Tabel 4. 7 Jumlah Data yang Digunakan	43
Tabel 4. 8 Hasil Perbandingan Akurasi Metode Klasifikasi.....	44
Tabel 4. 9 Hasil Perbandingan Waktu Metode Klasifikasi.....	44