

TUGAS AKHIR

Analisis Perbandingan Algoritma Naïve Bayes dan Algoritma C4.5 Untuk Klasifikasi Multi Data

Disusun Untuk Melengkapi dan Memenuhi Syarat Kelulusan
Guna Meraih Gelar Sarjana Komputer
Program Studi Teknik Informatika Universitas Muhammadiyah Jember



DWI LUKI ARISANDY

1310651061

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2017

HALAMAN PENGESAHAN

**ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN
ALGORITMA C4.5 UNTUK KLASIFIKASI MULTI DATA**

Oleh :

DWI LUKI ARISANDY
1310651061

Telah Mempertanggung Jawabkan Tugas Akhirnya Pada Seminar Hasil Tugas
Akhir Hari Jumat Tanggal 27 Juli 2017 Sebagai Salah Satu Syarat Kelulusan Dan
Mendapatkan Gelar Sarjana Komputer (S.Kom.)
di Program Studi Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember

Disetujui Oleh :

Dosen Penguji :
Penguji I

Dosen Pembimbing :
Pembimbing I

Taufiq Timur Warisaji, M.Kom
NPK. 08 04 486
Penguji II

Victor Wahanggara, S. Kom., M.Kom
NPK. 12 09 739
Pembimbing II

Deni Arifianto M.Kom
NPK. 11 03 588
Mengesahkan,
Dekan Fakultas Teknik

Yeni Dwi Rahayu S.ST, M.Kom
NPK. 11 03 590
Mengetahui,
Ketua Program Studi Teknik Informatika

Ir. Suhartinah, MT
NPK. 95 05 246

Yeni Dwi Rahayu, S. ST., M. Kom.
NPK. 11 03 590

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini, saya:

Nama : Dwi Luki Arisandy

NIM : 1310651061

Jurusan : Teknik Informatika

Menyatakan dengan sesungguhnya bahwa karya ilmiah berupa Tugas Akhir yang berjudul : “*Analisis Perbandingan Algoritma Naive Bayes dan Algoritma C4.5 Untuk Klasifikasi Multi Data*”, adalah benar-benar karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya dan belum pernah diajukan pada institusi manapun serta bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapatkan sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, 27 Juli 2017

Hormat Kami,

Dwi Luki Arisandy

Nim. 1310651061

HALAMAN PERSEMBAHAN

Alhamdulillah kehadiran Allah SWT yang telah memberikan jalan-Nya sehingga tugas akhir ini berhasil diselesaikan. dalam penelitian yang dituliskan ini penulis mempersembahkan karya ini untuk orang-orang yang ikut membantu dalam usaha, pemikiran, inspirasi dan semangat dalam proses penyelesaian penelitian ini, antara lain adalah :

1. Allah SWT atas segala rahmat dan karuniaNYA dalam mempermudah segala urusan saya dalam rangka menyusun tugas akhir ini tanpa bantuanNYA saya tidak akan mampu menyelesaikan tugas akhir ini dengan sebaik-baiknya.
2. Kedua orang tua saya bapak Tonaji dan Ibu Nuryamah, terimakasih atas segala bantuan doa dan semangat dalam rangka penyelesaian tugas akhir ini.
3. Kepada para sahabat-sahabat terimakasih atas doa, bantuan dan segenap nasehat yang diberikan selama ini, semoga Allah membalasnya dengan penuh barokah.
4. Teman-teman seperjuangan mahasiswa Universitas Muhammadiyah Jember dan Progam Studi Informatika hingga saya mendapatkan gelar Sarjana Komputer ini.
5. Terakhir, almamaterku tercinta Universitas Muhammadiyah Jember dan Program Studi Teknik Informatika hingga saya mendapatkan gelar Sarjana Komputer ini.

Akhirnya, dengan segala kerendahan hati penulis menyadari masih banyak terdapat kekurangan – kekurangan, sehingga penulis mengharapkan adanya saran dan kritik yang bersifat membangun demi kesempurnaan tugas akhir ini.

UNGKAPAN TERIMA KASIH

Alhamdulillah, atas berkah dan karunia Allah SWT penulis dapat menyelesaikan Tugas Akhir ini dan dalam pelaksanaan pembuatannya penulis banyak mendapatkan bimbingan dan dukungan dari berbagai pihak. Oleh karena itu pada kesempatan kali ini Penulis menyampaikan terima kasih yang tulus dan ikhlas kepada :

1. **Allah SWT**, karena Nikmat, Perlindungan, Pertolongan, dan Ridho-Nya saya mampu menyelesaikan Tugas Akhir ini serta hambanya yang termulia Nabi Besar **Muhammad SAW**.
2. **Bapak, Ibu, Kakak, Adik dan Keluarga** yang telah memberikan Doa, cinta dan kasih sayang baik itu berupa dorongan moril maupun spirituil.
3. **Ibu Yeni Dwi Rahayu, S. ST., M. Kom.** selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember dan dosen pembimbing kedua saya.
4. **Victor Wahanggara, S. Kom., M. Kom** Selaku dosen pembimbing yang banyak meluangkan waktu sampai tugas akhir ini selesai.
5. Seluruh **Dosen Penguji : Taufiq Timur Warisaji, M. Kom dan Deni Arifianto M. Kom** yang bersedia meluangkan waktu untuk menguji dan mengoreksi hasil tugas akhir kami ini.
6. **Dosen** dan semua **Staf** di fakultas teknik, yang telah memberikan ilmunya yang akan sangat berguna untuk penulis dan masyarakat umum.
7. Seluruh **teman-teman Jurusan Teknik Informatika dan khususnya angkatan 2013**.
8. Seluruh **Mahasiswa Universitas Muhammadiyah Jember** dari segala jurusan yang pernah kukenal dan mengenalku. Serta semua pihak yang telah membantu terselesaikannya Tugas Akhir ini.

Semoga Allah SWT senantiasa memberikan perlindungan dan memberikan balasan yang lebih di kemudian hari.

Harapan kami sebagai penulis semoga dengan terselesainya Tugas Akhir ini. Dapat bermanfaat bagi kami khususnya untuk pengembangan ilmu pengetahuan dan teknologi.

Jember, 01 Maret 2017

Penulis

KATA PENGANTAR



Alhamdulillah kami panjatkan rasa syukur kepada Allah SWT. karena hanya dengan rahmat, hidayah dan Inayah-Nya kami dapat menyelesaikan Tugas Akhir ini dengan judul :

ANALISIS PERBANDINGAN ALGORITMA NAIVE BAYES DAN ALGORITMA C4.5 UNTUK KLASIFIKASI MULTI DATA

Dalam menyelesaikan Tugas Akhir ini, kami berpegang pada teori yang pernah kami dapatkan dan bimbingan dari dosen pembimbing Tugas Akhir. Dan pihak – pihak lain yang sangat membantu hingga sampai terselesaikannya Tugas Akhir ini. Tugas Akhir ini merupakan salah satu syarat akademis untuk memperoleh gelar Sarjana Komputer (S.Kom.) di Fakultas Teknik, Universitas Muhammadiyah Jember.

Kami menyadari bahwa masih banyak kekurangan pada analisis dan pembuatan buku Tugas Akhir ini. Oleh karena itu, besar harapan kami untuk menerima saran dan kritik dari para pembaca. Semoga buku ini dapat memberikan manfaat bagi para Mahasiswa Universitas Muhammadiyah Jember pada umumnya dan dapat memberikan nilai lebih untuk para pembaca pada khususnya.

Jember, 27 Juli 2017

Penulis

MOTTO

“Dan janganlah kamu berjalan di muka bumi ini dengan sombong, karena sesungguhnya kamu sekali-kali tidak dapat menembus bumi dan sekali-kali kamu tidak akan sampai setinggi gunung”

(QS. Al-Israa': 37)

Usaha menghasilkan Harga.

(~ Luki Arie ~)

Hargailah usahamu, hargailah dirimu. Harga diri memunculkan disiplin diri.

Ketika anda memiliki keduanya, itulah kekuatan sesungguhnya.

(~ Clint Eastwood ~)

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
PERNYATAAN.....	ii
ABSTRAK	iii
PERSEMBAHAN.....	v
UNGKAPAN TERIMAKASIH.....	vi
KATA PENGANTAR	viii
MOTTO	ix
DAFTARI ISI	x
DAFTARI TABEL	xiii
DAFTARI GAMBAR	xv
BAB I. PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan	2
1.4. Batasan Masalah	3
1.5. Manfaat	3
BAB II. TINJAUAN PUSTAKA	5
2.1. Penelitian Terdahulu	5
2.2. Klasifikasi	6
2.2.1. Konsep Klasifikasi	6
2.2.2. Model	6
2.2.3. Pengukuran Kinerja Klasifikasi.....	9
2.3. Pengertian Analisis	11
2.4. Naive Bayes	11
2.4.1. Teorema Bayes	11
2.4.2. Naive Bayes untuk Klasifikasi	14
2.5. Algoritma C4.5	19
2.5.1. Pengertian Algoritma C4.5	19
2.5.2. Kelebihan dan Kekurangan Algoritma C4.5	20

2.5.3. Contoh Perhitungan Algoritma C4.5	21
2.6. Set Data	28
2.6.1. Jenis Data dalam Set Data	28
2.6.2. Jenis Set Data	31
2.6.3. Karakteristik Set Data.....	31
2.6.4. Data Rekord.....	32
2.7. K-fold Cross Validation	35
2.8. Pengertian Microsoft Excel.....	36
2.9. Kelebihan dan Kekurangan Microsoft Excel.....	37
2.10. Tool Weka.....	38
2.10.1 Pengertian Tool Weka.....	38
2.10.2 Kelebihan dan Kekurangan Tool Weka.....	39
2.10.3 Contoh Pengertian Tool Weka.....	40
BAB III. METODOLOGI PENELITIAN	48
3.1. Studi Literatur	49
3.2. Pengumpulan Data	49
3.3. Preprocessing data	50
3.3.1. Pembersihan Data	50
3.3.2. Perubahan Format Data	51
3.3.3. Analisis Data	54
3.4. Analisis Spesifikasi.....	55
3.5. Implementasi.....	56
3.6. Pengujian.....	57
3.7. Analisis	59
3.8. Kesimpulan	59
BAB IV. HASIL DAN PEMBAHASAN	60
4.1. Implementasi dan Pengujian Algoritma Naive Bayes	49
4.2. Implementasi dan Pengujian Algoritma C4.5	64
4.3. Analisis	68
4.4. Kesimpulan	70
BAB V. KESIMPULAN DAN SARAN	72

5.1. Kesimpulan	72
5.2. Saran	72
DAFTAR PUSTAKA	79

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	5
Tabel 2.2 Matriks konfusi untuk klasifikasi dua kelas	9
Tabel 2.3 Data latih klasifikasi hewan	16
Tabel 2.4 Probabilitas fitur dan kelas	17
Tabel 2.5 Contoh Data Training C4.5	22
Tabel 2.6 Contoh Perhitungan C4.5	23
Tabel 2.7 Contoh Perhitungan Total Entropy	24
Tabel 2.8 Contoh Hasil Perhitungan node 1 (Langkah 2)	25
Tabel 2.9 Contoh Hasil Perhitungan node 1 (Langkah 3)	27
Tabel 2.10 Tipe atribut	30
Tabel 2.11 Data rekord	33
Tabel 2.12 Data transaksi	34
Tabel 2.13 Data matriks	34
Tabel 2.14 Matriks istilah dokumen	34
Tabel 2.15 Contoh data set	36
Tabel 2.16 Contoh eksperimen	36
Tabel 2.12 Widget Orange	37
Tabel 3.1 Data set Penelitian	42
Tabel 3.2 Jumlah Data Hasil Pembersihan	42
Tabel 3.3 Kategori Jumlah Record	54
Tabel 3.4 Kategori Jumlah Variabel	55
Tabel 3.5 Kategori data set Penelitian	55
Tabel 3.6 Spesifikasi Hardware & Software	56
Tabel 4.1 Hasil Pengujian Data 1	60
Tabel 4.2 Hasil Pengujian Data 2	60
Tabel 4.3 Hasil Pengujian Data 3	61
Tabel 4.4 Hasil Pengujian Data 4	61
Tabel 4.5 Hasil Pengujian Data 5	62
Tabel 4.6 Hasil Pengujian Data 6	62

Tabel 4.7 Hasil Pengujian Data 7	62
Tabel 4.8 Hasil Pengujian Data 8	63
Tabel 4.9 Hasil Pengujian Data 9	63
Tabel 4.10 Hasil Pengujian Data 10	64
Tabel 4.11 Hasil Pengujian Data 1	64
Tabel 4.12 Hasil Pengujian Data 2	65
Tabel 4.13 Hasil Pengujian Data 3	65
Tabel 4.14 Hasil Pengujian Data 4	65
Tabel 4.15 Hasil Pengujian Data 5	66
Tabel 4.16 Hasil Pengujian Data 6	66
Tabel 4.17 Hasil Pengujian Data 7	67
Tabel 4.18 Hasil Pengujian Data 8	67
Tabel 4.19 Hasil Pengujian Data 9	67
Tabel 4.20 Hasil Pengujian Data 10	68
Tabel 4.21 Analisa Akurasi Pada Teknik Cross Validation K=5	69
Tabel 4.22 Analisa Akurasi Pada Teknik Cross Validation K=10	69
Tabel 4.23 Analisa Akurasi Pada Teknik Use Training Set	70
Tabel 4.24 Perbandingan dua dataset	70

DAFTAR GAMBAR

Gambar 2.1 Proses pekerjaan klasifikasi	7
Gambar 2.2 Pohon Keputusan Node 1	25
Gambar 2.3 Pohon Keputusan Node 1.1	26
Gambar 2.3 Pohon Keputusan Node 1.1.2	27
Gambar 2.2 Ilustrasi proses klasifikasi (decision tree).....	31
Gambar 2.3 Ilustrasi proses <i>Hierarchical Clustering</i>	31
Gambar 2.4 Ilustrasi <i>Cross Validation</i> beberapa metodologi	32
Gambar 2.5 Tampilan awal Weka	41
Gambar 2.6 Tampilan <i>preprocess</i> pada WEKA.....	41
Gambar 2.7 <i>file</i> hasil <i>capturing paket</i> yang akan diolah	42
Gambar 2.8 Hasil <i>preprocess</i> dari laporan jaringan	42
Gambar 2.9 Memilih metode <i>Classify</i> yang akan digunakan.....	43
Gambar 2.10 Memilih algoritma yang digunakan <i>Classifier</i>	44
Gambar 2.11 Test <i>Option</i> yang digunakan pada penelitian.....	46
Gambar 2.12 <i>More Option</i> yang dapat diatur untuk pengujian	41
Gambar 2.13 <i>Classifier evaluation options</i>	41
Gambar 2.14 Atribut kelas yang digunakan dalam klasifikasi.....	41
Gambar 3.1 Tahap penelitian	48
Gambar 3.2 Alur proses konversi format data.....	51
Gambar 3.3 Tahap pertama import data set.....	52
Gambar 3.4 Tahap pemilihan karakter pemisah	53
Gambar 3.5 Tahap pemberian format data kolom	53
Gambar 3.6 Tahap pemberian nama variabel	54
Gambar 3.7 Alur Pengujian dengan <i>Weka</i>	56
Gambar 3.8 Skenario Pengujian	58
Gambar 4.1 Vizualisasi dataset <i>Adult</i>	75
Gambar 4.1 Vizualisasi dataset <i>Poker</i>	76

DAFTAR PUSTAKA

- Berry, I. H. and Browne, M. 2006. *Lecture Notes in DATA MINING*. USA: World Scientific.
- Jawei Han, Micheline Kamber (2006). *Concept and techniques*. San Fransisco: Elsevier Inc.
- Prasetyo, Eko. (2012). *Data Mining konsep dan aplikasi menggunakan matlab*. Yogyakarta: Andi.
- Arrasyid, R.A. (2013). “Efektivitas Media *Software Excel* Dalam Model Pembelajaran *Contextual Teaching And Learning* Sebagai Upaya Meningkatkan Hasil Belajar Pokok Bahasan Kertas Kerja Mata Pelajaran Ekonomi Pada Siswa Kelas XI IPS Sma Negeri 2 Magelang Tahun Ajaran 2012/2013”. *Skripsi*. Semarang : Pendidikan Ekonomi Universitas Negeri Semarang.
- Wicaksana, P.D. (2015). “Perbandingan Algoritma K-Nearest Neighbors dan Naive Bayes Untuk Studi Data “*Wisconsin Diagnosis Breast Cancer*”. *Skripsi*. Yogyakarta : Teknik Informatika Universitas Sanata Dharma.
- Astuti, P. (2016). “*Komparasi Penerapan Algoritma C45, Knn Dan Neural Network Dalam Proses Kelayakan Penerimaan Kredit Kendaraan Bermotor*”.
- Kamus Besar Bahasa Indonesia 2012-2015. “Arti Kata Analisa”, (Online), (Kamus Besar Bahasa Indonesia (KBBI) Online - arti kata analisa, diakses 8 Januari 2017)
- Metode C4.5, (2015) “[Contoh](http://mardani-g4ul.blogspot.co.id/2011/12/contoh-soal-algoritma-c45.html) Soal algoritma C4.5 dan Perhitungannya”. di akses [<http://mardani-g4ul.blogspot.co.id/2011/12/contoh-soal-algoritma-c45.html>] (diakses 7 Januari 2017).

Nugraha, Erfan. (2010). “Perbedaan Mekanisme, Proses, Tinjauan, Analisis, dan Evaluasi”. Diakses dari [https://drummerfan.wordpress.com/2010/03/25/perbedaan-mekanisme-proses-tinjauan-analisis-dan-evaluasi/] (diakses 8 Januari 2017).

Wihardi yaya. (2013). “K-Folds Cross Validation”. Diakses dari [http://blog.yayaw.web.id/riset/k-folds-cross-validation] (diakses 16 Januari 2017).

Website ILMUSKRIPSI. <http://www.ilmuskripsi.com>

Website WEKA. <http://www.cs.waikato.ac.nz/ml/weka/>.