

**ANALISIS SENTIMEN *REVIEW* PRODUK *LAPTOP*
MENGUNAKAN ALGORITMA *NAIVE BAYES* DENGAN
SELEKSI FITUR *CHI SQUARE***

TESIS



**Oleh :
Suhendra
1311600199**

**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2015**

**ANALISIS SENTIMEN *REVIEW* PRODUK *LAPTOP*
MENGUNAKAN ALGORITMA *NAIVE BAYES* DENGAN
SELEKSI FITUR *CHI SQUARE***

TESIS

**Diajukan untuk memenuhi salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (MKOM)**



Oleh :

Suhendra

1311600199

**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2015**



PROGRAM STUDI : MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama Mahasiswa : Suhendra
Nomor Induk Mahasiswa : 1311600199
Konsentrasi : Teknologi Sistem Informasi
Judul Tesis : Analisis Sentimen *Review Produk Laptop*
Menggunakan Algoritma *Naive Bayes* Dengan
Seleksi Fitur *Chi Square*

Jakarta, 15 Agustus 2015

Tim Penguji :
Ketua,
(Prof. Dr. Moedjiono, M.Sc.)

Tanda Tangan:

Anggota,
(Dr. Ir. Nazori AZ, M.T.)

Pembimbing Utama,
(Ir. Dana Indra Sensuse, Ph.D)

Ketua Program Studi

(Dr. Ir. Nazori AZ, M.T)



**PROGRAM STUDI : MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

LEMBAR PERNYATAAN

Nama Mahasiswa : Suhendra
Nomor Induk Mahasiswa : 1311600199
Program Studi : Magister Ilmu Komputer
Konsentrasi : Teknologi Sistem Informasi
Fakultas / Program : Pasca Sarjana

Menyatakan bahwa Tesis yang berjudul :

**ANALISIS SENTIMEN *REVIEW* PRODUK *LAPTOP* MENGGUNAKAN
ALGORITMA *NAIVE BAYES* DENGAN SELEKSI FITUR *CHI SQUARE***

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain.
2. Saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, 15 Agustus 2015



(Suhendra)

ABSTRAK

Analisis sentimen yang akan dilakukan pada penelitian ini menggunakan *dataset review* produk *laptop* yang berasal dari situs web *amazon.com*. Karena saat ini hasil *review* produk *laptop* secara *online* semakin meningkat, sehingga bisa dengan mudah untuk mendapatkan informasi tersebut. Jumlah *dataset* yang diambil dari situs tersebut adalah 500 *review* yang dibagi menjadi 250 *review* positif dan 250 *review* negatif. Bahasa pemrograman dalam penelitian ini digunakan *Python* dan *plugin Natural Language Toolkit(NLTK)* untuk melakukan tahapan penelitian seperti *preprocessing*, *training*, *testing* dan juga evaluasi hasil. Penelitian ini memiliki tujuan untuk mengetahui seberapa besar pengaruh dari pemilihan fitur *Chi Square* dan *N-Gram* dalam menganalisa sentimen *review* produk *laptop* dengan teknik klasifikasi *Naïve Bayes*. Untuk evaluasi hasil menggunakan akurasi, *precision*, *recall* dan juga *f-measure*. Hasil dari penggunaan dari algoritma *Chi Square* dan *N-gram* yang digunakan sebagai pemilihan fitur terbukti dapat meningkatkan hasil akurasi algoritma *naïve bayes* yang semula 80,15% menjadi 94,44%.

Kata Kunci : *Naïve Bayes*, *N-Gram*, *Chi Square*, *Review* , *Laptop* .

ABSTRACT

Sentiment analysis that will be performed on this research using a laptop product review datasets derived from the amazon.com web site. Because the currently the results of the laptop product reviews are provided online, it is easy to obtain such information. The number of datasets taken from that site is 500 reviews which is divided into 250 positive reviews and negative reviews. Python programming language and the plug-in Natural Language Toolkit (NLTK) in this study are used in the preprocessing, training, testing and evaluation stages. This research aims to know the extent of the influence of the features selection of Chi Square and the N-Gram in analyzing sentiment laptop product review with Naïve Bayes classification techniques. Evaluation of the results employed accuracy, precision, recall and also f-measure. The result of the features selection stages are used to improve the accuracy of the Naïve Bayes Algorithm from 80,15% to 94,44%.

Keyword : Naïve Bayes, N-Gram, Chi Square, Review , Laptop.

KATA PENGANTAR

Puji syukur senantiasa penulis panjatkan ke hadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan pembuatan tesis yang berjudul “ANALISIS SENTIMEN *REVIEW* PRODUK *LAPTOP* MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DENGAN SELEKSI FITUR *CHI SQUARE*”

Dalam penyusunan tesis ini penulis tidak mampu berbuat apa-apa tanpa dukungan dan bimbingan dari semua pihak, oleh sebab itu saya mengucapkan terima kasih kepada :

1. Bapak Prof. Ir. Suryo Hapsoro Tri Utomo, Ph.D, selaku Rektor Universitas Budi Luhur Jakarta
2. Bapak Prof. Dr. Moedjiono, M.Sc., selaku Direktur Program Pascasarjana Universitas Budi Luhur Jakarta
3. Bapak Ir. Dana Indra Sensuse, Ph.D, selaku Pembimbing Utama.
4. Bapak Dr. Ir. Nazori AZ, M.T., selaku Ketua Program Bidang Studi Magister Ilmu Komputer Program Pascasarjana Universitas Budi Luhur Jakarta.
5. Seluruh Dosen dan Staff Universitas Budi Luhur Jakarta yang tidak bisa penulis sebutkan satu persatu.
6. Rekan-rekan yang telah membantu selama proses belajar mengajar di Universitas Budi Luhur Jakarta
7. Ayah, Ibu dan 3 Adikku (M.Supandi, Marlina, Anita L., Angga S. dan Adam K.) yang telah memberikan banyak inspirasi kepada penulis.

Dalam penelitian dan penyusunan tesis ini, penulis menyadari masih banyak kekurangan dan ketidaksempurnaan. Oleh karena itu penulis mengharapkan kritik, saran dan masukan yang bersifat membangun demi penyempurnaan penelitian di masa yang akan datang terutama bagi peneliti yang ingin melakukan penelitian dengan topik yang sama.

Jakarta, Agustus 2015

Penulis

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	2
1.2.1 Identifikasi Masalah	2
1.2.2 Batasan Masalah	3
1.2.3 Rumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	4
1.4 Tata Urut Penulisan	4
1.5 Daftar Pengertian	5
BAB II LANDASAN TEORI DAN KERANGKA KONSEP	6
2.1 Tinjauan Pustaka	6
2.1.1 <i>Pengertian Data Mining</i>	6
2.1.2 <i>Text mining</i>	11
2.1.3 Analisa Sentimen (<i>Sentiment Analysis</i>)	12
2.1.4 <i>Review Produk</i>	13
2.1.5 Pemilihan Fitur (<i>Feature Selection</i>)	14
2.1.6 Seleksi Fitur <i>Chi Square</i>	16
2.1.7 Algoritma <i>Naïve Bayes</i>	17
2.1.8 <i>N-Gram</i>	19
2.1.8.1 <i>N-Gram Based Text Categorization</i>	20
2.1.8.2 <i>Learning</i>	20

2.1.8.3 <i>Testing</i>	21
2.1.9 Validasi dan Evaluasi Algoritma <i>Data Mining</i>	23
2.1.10 <i>Focus Group Discussion</i>	24
2.1.11 Metode McCall.	27
2.1.12 <i>Python</i>	29
2.1.12 <i>Natural Language Toolkit</i> (NLTK).	30
2.2 Tinjauan Studi	31
2.3 Tinjauan Objek Penelitian	34
2.4 Kerangka Konsep	35
2.5 Hipotesis Penelitian	36
BAB III METODOLOGI DAN PERANCANGAN PENELITIAN	37
3.1 Metodologi Penelitian	37
3.2 Metode Pemilihan Sampel	37
3.3 Metode Pengumpulan Data	37
3.4 Instrumentasi	38
3.5 Teknik Analisis, Rancangan, dan Pengujian	38
3.5.1 Teknik Analisis.....	38
3.5.2 Teknik Rancangan	38
3.5.3 Teknik Pengujian.....	39
3.5.3.1 Teknik Pengujian Model.....	39
3.5.3.2 Pengujian Validasi	41
3.5.3.3 Pengujian Kualitas	41
3.5.3.4 Instrumen Pengujian	42
3.5.3.5 Pengolahan Data Pengujian	44
3.6 Langkah-Langkah Penelitian.....	45
3.7 Jadwal Penelitian	47
BAB IV PEMBAHASAN HASIL PENELITIAN	48
4.1 Analisis	48
4.2 Perancangan Model	48
4.2.1 Prototipe Model.....	50
4.2.2 Pengujian Model.....	52

4.2.2.1 <i>Preprocessing</i>	53
4.2.2.2 Akurasi Proses untuk Algoritma <i>Naïve Bayes</i>	59
4.2.3 Hasil Pengujian Kualitas	62
4.2.3.1 Hasil Pengujian Validasi	62
4.2.4 Hasil Pengujian Kualitas	63
4.3 Implikasi Penelitian	70
4.3.1 Aspek Sistem	70
4.3.2 Aspek Manajerial.....	70
4.3.3 Aspek Penelitian Lanjut	70
4.4 Rencana Implementasi.....	71
4.4.1 Tahapan Implementasi.....	71
4.4.2 Penjelasan Kegiatan Implementasi.....	71
BAB V KESIMPULAN DAN SARAN	73
5.1 Kesimpulan.....	73
5.2 Saran.....	73
DAFTAR PUSTAKA	74
LAMPIRAN-LAMPIRAN.....	78
RIWAYAT HIDUP.....	108

DAFTAR GAMBAR

Gambar	Halaman
II-1 Proses KDD	8
II-2 Langkah-langkah dalam Proses <i>Data Mining</i>	9
II-3 Prototipe Review Produk di amazon.com	13
II-4 <i>Decision Rules</i> pada Klasifikasi Review Produk di amazon.com	13
II-5 Taksonomi Masalah Pengurangan Dimensi	15
II-6 Contoh penggunaan tabel hash	20
II-7 Contoh penghitungan jarak dengan mekanisme <i>out-of-place measure</i>	22
II-8 Gambaran umum kategorisasi teks dengan menggunakan <i>N-gram</i>	23
II-9 Kerangka Konsep Penelitian	35
III-1 Pengujian Model	40
III-2 Langkah-langkah Penelitian	45
IV-1 Perancangan Model.	49
IV-2 Tampilan Awal Prototipe Sentimen Analisis	50
IV-3 Tampilan Pada Saat Klik Tombol <i>Browse</i> untuk <i>Preprocessing</i>	51
IV-4 Tampilan Pada Saat Klik Tombol <i>Browse</i> untuk akurasi	52
IV-5 Hasil Penggunaan Algoritma <i>Naïve Bayes</i>	61
IV-6 Hasil Setelah Penggunaan <i>Chi Square</i> dan <i>N-gram</i>	61

DAFTAR TABEL

Tabel	Halaman
II-1 Contoh pemotongan <i>N-gram</i> berbasis karakter.....	19
II-2 Contoh pemotongan <i>N-gram</i> berbasis kata.	20
II-3 Rangkuman Studi Yang Relevan	33
III-1 Spesifikasi Komputer yang Digunakan	39
III-2 Skala Pengukuran	42
III-3 Kisi-kisi Pengukuran Kualitas Perangkat Lunak dan Indikator	43
III-4 Kriteria Persentase Tanggapan Responden Terhadap Skor Ideal	45
III-5 Jadwal Penelitian	47
IV-1 Perbandingan teks sebelum dan sesudah dilakukan proses <i>Lower Case</i>	53
IV-2 Perbandingan teks sebelum dan sesudah dilakukan proses <i>Tokenization</i>	54
IV-3 Perbandingan teks sebelum dan sesudah dilakukan proses <i>Stopword Removal</i>	56
IV-4 Perbandingan teks sebelum dan sesudah dilakukan proses <i>Porter Stemmer</i>	58
IV-5 Perbandingan teks sebelum dan sesudah dilakukan proses <i>Lemmatization</i>	59
IV-6 Hasil Rata-Rata dari Evaluasi Hasil	61
IV-7 Evaluasi Hasil Dengan <i>Chi Square</i> dan <i>N-gram</i>	61
IV-8 Hasil Pengujian FGD.....	62
IV-9. Kriteria Persentase Tanggapan Responden Terhadap Skor Ideal	64

IV-10 Keterkaitan Kriteria dengan Faktor	65
IV-11 Tanggapan Responden Berdasarkan Faktor <i>Correctness</i>	66
IV-12 Tanggapan Responden Berdasarkan Faktor <i>Reliability</i>	67
IV-13 Tanggapan Responden Berdasarkan Faktor <i>Usability</i>	67
IV-14 Tanggapan Responden Berdasarkan Faktor <i>Maintainability</i>	68
IV-15 Hasil Pengujian Secara Keseluruhan	69
IV-16 Rencana Implementasi.....	71



DAFTAR LAMPIRAN

Lampiran	Halaman
1. Hasil <i>Focus Group Discussion</i>	78
2. Kuesioner	82
3. Dataset <i>ReviewLaptop</i> Negatif.....	84
4. Dataset <i>ReviewLaptop</i> Positif	87
5. Kode Program Aplikasi.....	95
6. Kartu Bimbingan Tesis	107
7. Daftar Riwayat Hidup	108

DAFTAR PUSTAKA

- [Arbuckle 2010] Arbuckle, A., *Python Testing. Beginner's Guide*, Packt Publishing, 2010.
- [Basari et al. 2013] Basari, A. S. H., Hussin, B., Pramudya Ananta, G. I., Zeniarja, J., "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization". *Procedia Engineering*, 2013.
- [Berry 2010] Berry, M.W. & Kogan, J.. *Text Mining Application and theory*. WILEY : United Kingdom. 2010.
- [BUM 2014] Manalu, Boy Utomo."Analisis Sentimen Pada *Twitter* Menggunakan *Text Mining*." Skripsi, Universitas Sumatera Utara, 2014.
- [Bustami 2013] Bustami. "Penerapan Algoritma *Naive Bayes* Untuk mengklasifikasi Data Nasabah Asuransi". Universitas Malikussaleh, 2013.
- [Bramer 2007] Bramer, Max, *Principles of Data Mining*, Springer, London, 2007.
- [Cavnar 1994] Cavnar, W.B., and Trenkle, J.M., *N-Gram-Based Text Categorization Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*, 2014.
- [Dehkharghani et al. 2014] Dehkharghani, R., Mercan, H., Javeed, A., & Saygin, Y., *Sentimental causal rule discovery from Twitter*. *Expert Systems with Applications*, 2014.
- [Downey et al. 2008] Downey, A. B., Elkner, J., Meyers, C., *How to Think Like a Computer Scientist Learning with Python*, O'Reilly Media, 2008.
- [Feldman 2007] Feldman, R., & Sanger, J., *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press : New York. 2007.
- [Gorunescu 2011] Gorunescu, F., *Data Mining: Concept Models and Technique*, Morgan Kaufmann Publishers, San Fransisco, 2006.
- [Haddi et al. 2013] Haddi, E., Liu, X., Shi, Yong, *The Role of Text Pre-processing in Sentiment Analysis*. London : Brunel University, 2013.
- [Han 2006] Han, J., and Kamber, M., *Data Mining: Concepts and Techniques Second Edition*. Morgan Kaufmann publisher : San Francisco. 2006.