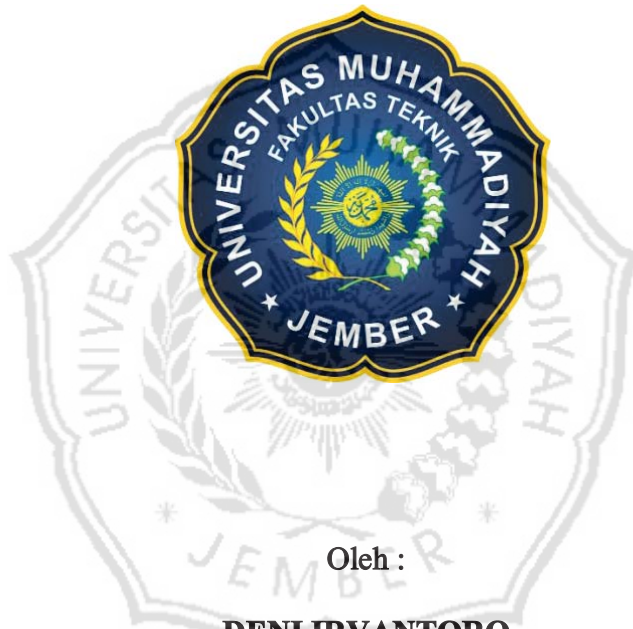


TUGAS AKHIR

***FEATURE SELECTION MENGGUNAKAN CHI-SQUARE
DAN N-GRAM DENGAN ALGORITMA NAÏVE BAYES
CLASSIFIER UNTUK ANALISIS SENTIMEN REVIEW
PRODUK ELEKTRONIK***



Oleh :

DENI IRVANTORO

1410651199

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH JEMBER

2019

TUGAS AKHIR

***FEATURE SELECTION MENGGUNAKAN CHI-SQUARE DAN
N-GRAM DENGAN NAIVE BAYES CLASSIFIER UNTUK
ANALISIS SENTIMEN REVIEW PRODUK ELEKTRONIK***

Diajukan Untuk Memenuhi Persyaratan Guna Meraih Gelar Sarjana Komputer

Teknik Informatika Universitas Muhammadiyah Jember



Oleh :

DENI IRVANTORO

1410651199

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH JEMBER

2019

HALAMAN PENGESAHAN

***FEATURE SELECTION MENGGUNAKAN CHI-SQUARE
DAN N-GRAM DENGAN ALGORITMA NAIVE BAYES
CLASSIFIER UNTUK ANALISIS SENTIMEN REVIEW
PRODUK ELEKTRONIK***

DENI IRVANTORO

NIM. 1410651199

Tugas Akhir Ini Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Komputer (S.Kom)

Di

Universitas Muhammadiyah Jember

Pembimbing I

Pembimbing II

Ilham Saifudin, S.Pd, M.Si.
NPK. 19891031 1 1703811

Reni Umilasari, S.Pd, M.Si.
NPK. 19910728 1 1703813

Penguji I

Penguji II

Wiwik Suharso, S.Kom, M.Kom.
NIP. 19760906 200501 1 003

Triawan Adi Cahyanto, S.Kom, M.Kom.
NPK. 12 03 719

Mengesahkan,
Dekan Fakultas Teknik

Mengetahui,
Ketua Program Studi Teknik Informatika

Ir. Suhartinah, MT.
NPK. 95 05 246

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

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini :

Nama : Deni Irvantoro

Nim : 1410651199

Jurusan : Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa karya ilmiah berupa skripsi yang berjudul : **“FEATURE SELECTION MENGGUNAKAN CHI-SQUARE DAN N-GRAM DENGAN NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN REVIEW PRODUK ELEKTRONIK”**, adalah hasil karya sendiri. Kecuali jika dalam beberapa kutipan substansi telah saya sebutkan sumbernya. Belum pernah diajukan pada institusi manapun, serta bukan karya plagiat atau jiplakan. Saya bertanggung jawab atas keaslian, keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah.

Demikian saya buat dengan sebenar-benarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta saya bersedia memperoleh sanksi akademik, jika ternyata ada pihak-pihak yang dirugikan dari pernyataan yang tidak benar tersebut.

Jember, 12 April 2019

Yang menyatakan

Deni Irvantoro

NIM. 141065119

MOTTO

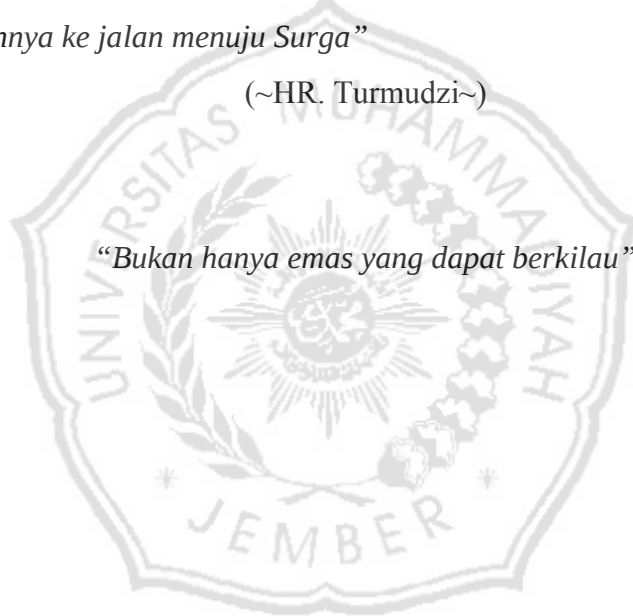
“Barang siapa yang mengkehendaki kehidupan dunia maka wajib baginya memiliki ilmu, dan barang siapa yang mengkehendaki kehidupan akhirat, maka wajib baginya memiliki ilmu, dan barang siapa yang mengkehendaki keduanya maka wajib baginya memiliki ilmu”

(~HR. Turmudzi~)

“Barang siapa yang menempuh jalan untuk mencari suatu ilmu. Niscaya Allah memudahkannya ke jalan menuju Surga”

(~HR. Turmudzi~)

“Bukan hanya emas yang dapat berkilau”



HALAMAN PERSEMBAHAN

Alhamdulillah atas 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 karunia-Nya dalam mempermudah segala urusan saya dalam rangka menyusun tugas akhir ini, tanpa bantuan - NYA saya tidak akan mampu menyelesaikan tugas akhir ini dengan sebaik-baiknya.
2. Kedua orang tua saya Bapak Samsul Arifin dan Ibu Hartatik dan juga saudara kandung saya Dedy Irawan terima kasih atas segala bantuan, doa, dan semangat dalam penyelesaian tugas akhir ini.
3. Kepada Uswatun Hasanah yang telah memberikan dukungan dan do'a untuk saya.
4. Kepada para sahabat-sahabat saya Risky Dwi P., M. Ali Khofi, Anas Abdillah M., Alin Janahiq, dan Cindi Armalian yang telah memberikan segenap nasehat, bantuan, dan tempat berkeluh kesah dari awal perkuliahan sampai terselesaikannya Tugas Akhir ini. Semoga Allah SWT membalasnya dengan penuh barokah.
5. Teman-teman keluarga besar KKN (Kuliah Kerja Nyata) 2017 Kelompok 5 Desa Pulo, Kec. Tempeh, Lumajang yang telah menjadi inspirasi dalam penulisan Tugas Akhir ini dan yang telah mengajarkan pentingnya bekerjasama dan bersosialisasi dalam kehidupan di tengah masyarakat.
6. Terakhir untuk almamaterku tercinta Universitas Muhammadiyah Jember dan Program Studi Teknik Informatika hingga saya mendapat 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, Adik dan Keluarga** yang telah memberikan Doa, Dukungan, dan dorongan moral maupun moril.
3. **Ilham Saifuddin S.Pd., M.Si. dan Reni Umilasari S.Pd., M.Si.** selaku dosen pembimbing yang banyak meluangkan waktu sampai tugas akhir ini selesai
4. Seluruh **Dosen penguji : Wiwik Suharso, S.Kom., M.Kom. dan Triawan Adi Cahyanto, M.Kom.** yang bersedia meluangkan waktu untuk menguji dan mengoreksi hasil tugas akhir ini.
5. Seluruh **teman-teman jurusan Teknik Informatika dan khususnya angkatan 2014**, dan seluruh teman-teman yang mengenalku di Universitas Muhammadiyah Jember.

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

Harapan sebagai penulis semoga dengan terselesaikannya Tugas Akhir ini. Dapat bermanfaat bagi pembaca khususnya untuk pengembang ilmu pengetahuan dan teknologi.

Jember, 12 April 2019

Penulis

KATA PENGANTAR



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

FEATURE SELECTION MENGGUNAKAN CHI-SQUARE DAN N-GRAM DENGAN ALGORITMA NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN REVIEW PRODUK ELEKTRONIK

Dalam menyelesaikan Tugas Akhir ini, penulis berpegang pada teori yang pernah di 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.

Penulis menyadari bahwa masih banyak kekurangan pada analisis dan pembuatan buku Tugas Akhir ini. Oleh karena itu, besar harapan 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, 12 April 2019

Penulis

DAFTAR ISI

HALAMAN PENGESAHAN	ii
PERNYATAAN KEASLIAN.....	iii
MOTTO	iv
HALAMAN PERSEMBAHAN	v
UNGKAPAN TERIMA KASIH	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
ABSTRAK	xv
ABSTRACT	xvi
BAB I Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II Tinjauan Pustaka.....	5
2.1 Latar Belakang	5

2.2	<i>Amazon.com, Inc</i>	7
2.3	<i>Analisis Sentimen</i>	7
2.4	<i>Text Mining</i>	8
2.4.1	<i>Case Folding</i>	9
2.4.2	<i>Tokenizing</i>	10
2.4.3	<i>Stopword Removal</i>	10
2.5	<i>Stemming</i>	11
2.5.1	<i>Porter Stemmer</i>	11
2.6	<i>N-Gram Tokenizer</i>	12
2.7	<i>Chi-Square</i>	13
2.8	<i>Naive Bayes Classifier</i>	14
2.8.1	<i>Teorema Bayes</i>	15
2.8.2	<i>Bentuk Klasifikasi Naive Bayes Classifier</i>	16
2.9	<i>Confussion Matrix</i>	17
BAB III METODOLOGI PENELITIAN		19
3.1	<i>Rancangan Penelitian</i>	19
3.2	<i>Preprocessing</i>	20
3.3	<i>Generate N-Gram Tokenizer</i>	21
3.4	<i>Feature Selection Menggunakan Chi-Square</i>	22
3.5	<i>Implementasi Algoritma Naive Bayes Classifier pada Uni-Gram</i>	24
3.5.1	<i>Proses Pelatihan Data</i>	24
3.5.2	<i>Proses Uji Data</i>	26
3.6	<i>Penerapan Algoritma Naive Bayes Classifier Pada Bi-Gram</i>	28
3.6.1	<i>Proses Pelatihan Data</i>	28

3.6.2	Proses Uji Data	30
3.7	Penerapan Algoritma <i>Naive Bayes Classifier</i> Pada <i>Uni+Bi-Gram</i> ..	32
3.7.1	Proses Pelatihan Data	32
3.7.2	Proses Uji Data	33
BAB IV HASIL DAN PEMBAHASAN		36
4.1	Pemrosesan Data	36
4.2	Hasil Dan Analisa	38
4.2.1	Hasil Pengujian Klasifikasi Menggunakan 25% Fitur.....	38
4.2.2	Hasil Pengujian Klasifikasi Menggunakan 50% Fitur.....	39
4.2.3	Hasil Pengujian Klasifikasi Menggunakan 75% Fitur.....	40
4.2.4	Hasil Pengujian Klasifikasi Menggunakan 100% Fitur.....	41
4.3	Komparasi Fitur <i>N-Gram</i>	42
4.3.1	Komparasi Akurasi.....	42
4.3.2	Komparasi Presisi	43
4.3.3	Komparasi <i>Recall</i>	44
BAB V KESIMPULAN DAN SARAN		45
5.1	Kesimpulan	45
5.2	Saran	46
DAFTAR PUSTAKA.....		47

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	5
Tabel 2. 2 Proses <i>case folding</i>	10
Tabel 2. 3 <i>Generate N-Gram</i> berdasarkan <i>string</i>	13
Tabel 3. 1 <i>Generate N-Gram</i>	21
Tabel 3. 2 <i>Observed Frequency</i>	22
Tabel 3. 3 <i>Expected Frequency</i>	23
Tabel 3. 4 Hasil <i>test statistic</i>	24
Tabel 3. 5 <i>Data Training</i>	25
Tabel 3. 6 Hasil Pengolahan Contoh <i>Data Training</i>	26
Tabel 3. 7 <i>Data Testing</i>	26
Tabel 3. 8 Uji Data Positif	27
Tabel 3. 9 Uji Data Negatif	28
Tabel 3. 10 <i>Training Data</i>	28
Tabel 3. 11 Hasil <i>Data Training</i>	30
Tabel 3. 12 <i>Data Testing</i>	30
Tabel 3. 13 Hasil Uji Data Positif	31
Tabel 3. 14 Hasil Uji Data Negatif	31
Tabel 3. 15 <i>Data Training</i>	32
Tabel 3. 16 Hasil <i>Data Training</i>	33
Tabel 3. 17 <i>Data Testing</i>	33
Tabel 3. 18 Hasil Uji Data Positif	34
Tabel 3. 19 Hasil Uji Data Negatif	35

Tabel 4. 1 Konversi Kategori Sentimen	36
Tabel 4. 2 Hasil Klasifikasi Menggunakan 25% Fitur	39
Tabel 4. 3 Hasil Klasifikasi Menggunakan 50% Fitur	40
Tabel 4. 4 Hasil Klasifikasi Menggunakan 75% Fitur	41
Tabel 4. 5 Hasil Klasifikasi Menggunakan 100% Fitur	41



DAFTAR GAMBAR

Gambar 2. 1 Proses <i>Text Mining</i>	9
Gambar 2. 2 Proses <i>Tokenizing</i>	10
Gambar 2. 3 Proses <i>Stopping</i>	10
Gambar 2. 4 Langkah-langkah Algoritma <i>Porter Stemmer</i>	11
Gambar 3. 1 Rancangan Penelitian	19
Gambar 4. 1 Performa Komparasi Akurasi Fitur <i>N-Gram</i>	42
Gambar 4. 2 Performa Komparasi Akurasi Fitur <i>N-Gram</i>	43
Gambar 4. 3 Performa Komparasi Akurasi Fitur <i>N-Gram</i>	44



DAFTAR LAMPIRAN

LAMPIRAN.....	50
Lampiran 1: Hasil Frekuensi Kemunculan <i>Term Uni-Gram</i>	50
Lampiran 2: Hasil Pembobotan Fitur <i>Uni-Gram</i>	51
Lampiran 3: Hasil Seleksi Fitur <i>Uni-Gram</i>.....	52
Lampiran 4: Hasil Klasifikasi <i>Uni-Gram</i>.....	56
Lampiran 5: Hasil Frekuensi Kemunculan <i>Term Bi-Gram</i>.....	57
Lampiran 6: Hasil Pembobotan Fitur <i>Bi-Gram</i>.....	58
Lampiran 7: Hasil Seleksi Fitur <i>Bi-Gram</i>	59
Lampiran 8: Hasil Klasifikasi <i>Bi-Gram</i>	63
Lampiran 9: Hasil Frekuensi Kemunculan <i>Term Uni+Bi-Gram</i>	64
Lampiran 10: Hasil Pembobotan Fitur <i>Uni+Bi-Gram</i>.....	65
Lampiran 11: Hasil Seleksi Fitur <i>Uni+Bi-Gram</i>	66
Lampiran 12: Hasil Klasifikasi <i>Uni+Bi-Gram</i>	70

DAFTAR PUSTAKA

- Ali, M. F., D. C. Utomo, B. D. Setiawan. *Automatic Essay Scoring System Using N-Gram and Cosine Similarity for Gamification Based E-Learning*. ICAIP 2017, Agustus 25-27, 2017, Bangkok, Thailand.
- Caroline, M. W., 2015. *Sistem Pendeteksi Pengirim Tweet dengan Metode Klasifikasi Naïve Bayes*. Jurnal Teknik Informatika dan Sistem Informasi. Vol.1, No.2, Agustus 2015.
- Chandani, Vinita., Romi Satria W., Purwanto., 2015. *Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada analisis Sentiment Review Film*. *Journal of Intelligent Systems*, Vol. 1, No.1, February 2015.
- Chen, Jingnian., et al., 2009. *Feature Selection for Text Classification with Naive Bayes*. *Expert System with Applications*, 36 (2009), 5432-5435.
- Hall, Mark., 2018. *Amazon.com American Company*. *Encyclopedia Britannica*. <https://www.britannica.com/topic/Amazoncom>. Di akses pada 10 Oktober 2018.
- Hemalata, I., G. P Saradhi Varma., A. Govardhan., 2012. *Preprocessing the Informal Text For Efficient Sentiment Analysis*. *Internasional Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Vol. 1, Issues 2, July-August 2012.

- Ipmawati, J., Kusriani., Emha T. L., 2017. *Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen*. Indonesian Journal on Networking and Security, Vol. 6, No. 1, Thn. 2017.
- Karaa, Wahiba. B. A., 2013. *A New Stemmer to Improve Information Retrieval*. International Journal of Network Security and Its Applications (IJNSA), Vol.5, No.4, July 2013.
- Korde, Vandana., C. Namrata Mahender. *Text Classification and Classifiers: Survey*. International Journal of Artificial Intelligence and Application (IJAIA), Vol.3, No.2, March 2012.
- Koshy, Susan., R. Padmavalli., 2015. *Text Categorization of Multi-Label Document Text Mining*. International Journal of Data Mining Techniques and Applications, Vol. 4, Issue: 2 Desember 2015 Page No.52-58. ISSN 2278-2419
- Ling, J., I. P. E. N. Kencana., 2014. T. Bagus Oka. *Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi-Square*. E-Jurnal Matematika. Vol.3, No.3, Agustus 2014, pp. 92-99.
- Ogadha, K. O., 2016. *N-Gram for Text Classification Using Supervised Machine Learning Algorithms. A thesis submitted in fulfillment for the degree of Doctor of Philosophy in Information Technology. University of Agriculture and Technology*.
- Prasanti, A. A., M. A. Fauzi., M. Tanzil F. *Klasifikasi teks Pengaduan Pada Sambat Online Menggunakan metode N-Gram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol.2, No.2, Februari 2018, hlm. 594-601.

- Pratama, N. D., Yuita A. S., Putra P. A., 2018. *Analisis Sentimen Pada Review Konsumen Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Chi-Square Untuk Rekomendasi Lokasi Makanan Tradisional*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol.2, No.9, September 2018, hlm. 2982-2988.
- Rachburee, Nachirat., Wattana Punlumjeak., 2015. *A Comparison of Feature Selection Approach Between Greedy, IG, Chi-Square, and mRMR in Educational Mining*. 2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand.
- Satapathy, R., Cambria E., Hussain A., 2017. *Sentiment Analysis in the Bio-Medical Domain*. Springer International Publishing AG, Gwerbestrasse 11, 6630 Cham, Switzerland. <https://doi.org/10.1007/978-3-319-68468-0>
- Yadav, S. H., Balu L. Parne., 2015. *A Survey on Different Text Categorization Techniques for Text Filtration*. IEEE Sponsored 9th International Conference on Intelligent System and Control (ISCO), 2013.
- Ye, Qiang., Ziqiong Zhang., Rob Law. *Sentiment Classification of Online Reviews To Travel Destinations By Supervised Machine Learning Approaches*. Expert System with Applications, 36 (2009), 6527-6535.
- Yusuf, 2010. *Chi-Square Independence Test*. <http://www.spss-tutorials.com/chi-square-independent-test/>. Di akses pada 10 Oktober 2018.