

**KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN
LAYANAN SMS “SUARA KONSUMEN” TERHADAP
PRODUK MENGGUNAKAN METODE K-NN**

Skripsi



oleh
DAVID ADDIWIJAYA HANZ
22094686

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2014

**KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN
LAYANAN SMS “SUARA KONSUMEN” TERHADAP
PRODUK MENGGUNAKAN METODE K-NN**

Skripsi



Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

DAVID ADDIWIJAYA HANZ
22094686

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2014

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN LAYANAN SMS “SUARA KONSUMEN” TERHADAP PRODUK MENGGUNAKAN METODE K-NN

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 18 Juni 2014



DAVID ADDIWIJAYA HANZ

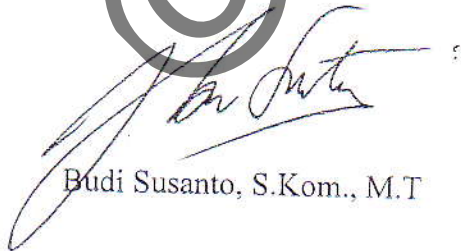
22094686

HALAMAN PERSETUJUAN

Judul Skripsi : Klasifikasi Sentimen Pembeli berdasarkan Layanan SMS “Suara
Konsumen” terhadap Produk menggunakan Metode K-NN
Nama Mahasiswa : DAVID ADDIWIJAYA HANZ
N I M : 22094686
Matakuliah : Tugas Akhir
Kode : TIW276
Semester : Genap
Tahun Akademik : 2013/2014

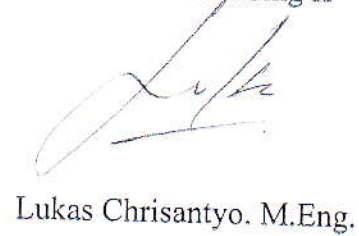
Telah diperiksa dan disetujui
di Yogyakarta,
Pada tanggal 28 - Mei - 2014

Dosen Pembimbing I



Budi Susanto, S.Kom., M.T

Dosen Pembimbing II



Lukas Chrisantyo. M.Eng.

HALAMAN PENGESAHAN

KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN LAYANAN SMS "SUARA KONSUMEN" TERHADAP PRODUK MENGGUNAKAN METODE K-NN

Oleh: DAVID ADDIWIJAYA HANZ / 22094686

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 16 Juni 2014

Yogyakarta, 18 Juni 2014
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, S.Kom., M.T.
2. Lukas Chrisantyo, M.Eng.
3. Kristian Adi Nugraha, S.Kom., M.T.
4. Antonius Rachmat C., S.Kom., M.Cs



Dekan

(Drs. Wimmie Handiwidjojo, M.T.)

Ketua Program Studi



(Nugroho Agus Haryono, M.Si)

INTISARI

KLASIFIKASI SENTIMEN PEMBELI BERDASARKAN LAYANAN SMS “SUARA KONSUMEN” TERHADAP PRODUK MENGUNAKAN METODE K-NN

Layanan “SUARA KONSUMEN” dibuat produsen dengan tujuan untuk dapat mendapatkan saran, kritik, maupun pengaduan terhadap produk dari produsen. Layanan ini juga adalah salah satu cara mempermudah produsen tersebut dapat mengetahui secara langsung bagaimana sentimen pelanggan terhadap produk yang dikeluarkan oleh produsen dengan mengetahui jumlah komentar negatif dan komentar positif, namun terbatas jumlahnya karakter dalam SMS menjadi masalah yang utama karena akan menyebabkan harus penggunaan kata-kata dan struktur yang ringkas namun padat dan jelas dalam menyampaikan isi dan maksud dari pengirim SMS.

Penelitian ini akan diteliti bagaimana mengklasifikasikan SMS yang menggunakan struktur dan kata-kata yang terbatas untuk mendapatkan sentimen dari sebuah layanan SMS “Suara Konsumen” melalui pendekatan text mining. Penelitian ini menggunakan Algoritma K-NN dalam melakukan klasifikasi, selain Algoritma K-NN penelitian ini akan membobot kata menggunakan metode TF-IDF, Feature Selection dalam pemilihan kata-kata juga Cosine Similarity dalam mengukur tingkat kedekatan antar dokumen.

Dalam menentukan keberhasilan, diukur tingkat akurasi dari dokumen yang dilakukan percobaan dan hasilnya sangat akurat dengan nilai rata-rata akurasi sebesar 89% pada variasi nilai K dan Feature Selection. Nilai akurasi pada sistem klasifikasi sentimen SMS optimal pada $K = 3$ dan $K = 7$ yaitu 90% dan dengan error rate dibawah 10%.

Kata Kunci : *K-Nearest Neighbor*, *Sentiment Analyst*, klasifikasi dokumen, komentar positif, komentar negatif, SMS, *Cosine Similarity*

DAFTAR ISI

HALAMAN JUDUL	
PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
UCAPAN TERIMA KASIH	vi
INTISARI	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Metodologi Penelitian	4
1.6 Sistematika Penulisan	5
BAB II LANDASAN TEORI	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori	7
2.2.1 Sentimen Analisis	7
2.2.2 Text Mining	8
2.2.3 Algoritma K-Nearest Neighbor	9
2.2.3 Metode TF-IDF	10

2.2.4 Cosine Similiarity	12
BAB III ANALISIS DAN PERANCANGAN SISTEM.....	13
3.1 Kebutuhan Sistem.....	13
3.1.1 Kebutuhan Fungsional.....	13
3.1.2 Non Fungsional.....	14
3.1.3 Pengembangan Sistem.....	15
3.2 Gambaran Kerja Sistem.....	17
3.2.1 Topologi Jaringan.....	17
3.2.2 Flowchart Kerja Sistem.....	18
3.3 Perancangan Database.....	21
3.4 Perancangan Desain.....	26
3.5 Rancangan Pengujian Sistem.....	30
3.2.1 Environment Pengujian.....	30
3.2.2 Komentar Training.....	31
3.2.2 Komentar Testing.....	32
BAB IV IMPLEMENTASI DAN ANALISIS SISTEM.....	34
4.1 Implementasi Sistem.....	34
4.2 Analisis Sistem.....	40
4.2.1 Hasil Evaluasi dengan variasi Nilai K dan Feature Selection.....	40
4.2.2 Perbandingan Evaluasi variasi nilai K dan Nilai Feature Selection.....	57
BAB V KESIMPULAN DAN SARAN.....	59
5.1 Kesimpulan.....	59
5.2 Saran.....	59
DAFTAR PUSTAKA.....	60
LAMPIRAN.....	

DAFTAR TABEL

Tabel 3.1 Tabel dokumen_training	21
Tabel 3.2 Tabel singkatan.....	22
Tabel 3.3 Tabel inbox.....	22
Tabel 3.4 Tabel token_training.....	23
Tabel 3.5 Tabel dokumen_join.....	23
Tabel 3.6 Tabel stopwords.....	24
Tabel 3.7 Tabel feature_selection.....	24
Tabel 3.8 Tabel Komentar Training	31
Tabel 3.9 Tabel Komentar Testing.....	32
Tabel 4.1 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 3 dan Feature Selection 10%.....	41
Tabel 4.2 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 3 dan Feature Selection 30%.....	42
Tabel 4.3 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 3 dan Feature Selection 50%.....	43
Tabel 4.4 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 3 dan Feature Selection 100%.....	44
Tabel 4.5 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 5 dan Feature Selection 10%.....	45
Tabel 4.6 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 5 dan Feature Selection 30%.....	46
Tabel 4.7 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 5 dan Feature Selection 50%.....	47
Tabel 4.8 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 5 dan Feature Selection 100%.....	48
Tabel 4.9 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 7 dan Feature Selection 10%.....	49
Tabel 4.10 Confusion Matrix untuk Kelas Positif dan Negatif pada K = 7 dan Feature Selection 30%.....	50

Tabel 4.11 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 7$ dan Feature Selection 50%	51
Tabel 4.12 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 7$ dan Feature Selection 100%	52
Tabel 4.13 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 9$ dan Feature Selection 10%	53
Tabel 4.14 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 9$ dan Feature Selection 30%	54
Tabel 4.15 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 9$ dan Feature Selection 50%	55
Tabel 4.16 Confusion Matrix untuk Kelas Positif dan Negatif pada $K = 9$ dan Feature Selection 100%	56
Tabel 4.17 Tabel Perbandingan Nilai Akurasi pada tiap Nilai K dan Feature Selection	57

©UKYD

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Klasifikasi Algoritma K-Nearest Neighbor.....	10
Gambar 3.1 Blok Diagram Klasifikasi	15
Gambar 3.2 Topologi Jaringan	17
Gambar 3.3 Flowchart Sistem Klasifikasi SMS	18
Gambar 3.4 Flowchart Proses Pengelompokan Komentar dan SPAM.....	18
Gambar 3.5 Skema Database	18
Gambar 3.6 Skema Situs	26
Gambar 3.7 Skema Pengaturan File Di Sistem Server	26
Gambar 3.8 Tampilan Login	27
Gambar 3.9 Tampilan Dashboard Sistem	27
Gambar 3.10 Tampilan Halaman Training	28
Gambar 3.11 Tampilan Komentar Positif	28
Gambar 3.12 Tampilan Komentar Negatif	28
Gambar 4.1 Implementasi Preprocessing 1	44
Gambar 4.2 Implementasi Preprocessing 2	44
Gambar 4.3 Implementasi Preprocessing 3	45
Gambar 4.4 Implementasi Preprocessing 4.....	45
Gambar 4.5 Implementasi Pembobotan TF_IDF 1	46
Gambar 4.6 Implementasi Pembobotan TF_IDF 2	47
Gambar 4.7 Implementasi Pembobotan TF_IDF 3	47
Gambar 4.8 Implementasi Pembobotan TF_IDF 4	48
Gambar 4.9 Implementasi Cosine Similarity	48
Gambar 4.10 Implementasi Metode K-NN	49
Gambar 4.11 Grafik Perbandingan Nilai Akurasi	58

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kepuasan konsumen menjadi salah satu faktor utama untuk dapat meningkatkan mutu produk tersebut. Beberapa produsen telah memberi sebuah layanan SMS yang disebut “SUARA KONSUMEN”. Layanan ini bertujuan untuk dapat mendapatkan saran, kritik, maupun pengaduan terhadap produk dari produsen tersebut. Layanan ini juga adalah salah satu cara mempermudah produsen tersebut dapat mengetahui secara langsung bagaimana sentimen pelanggan terhadap produk yang dikeluarkan oleh produsen dengan mengetahui jumlah komentar negatif dan komentar positif. Untuk mengetahui apakah komentar yang masuk ke dalam layanan “SUARA KONSUMEN” maka diperlukan biasanya perlu membaca satu per satu SMS yang masuk dan baru diklasifikasikan.

Untuk mempermudah operator dari produsen untuk mengetahui sentimen dari pelanggan terhadap produk yang telah dipasarkan, maka diperlukan sebuah sistem yang mampu mengklasifikasikan setiap komentar yang masuk untuk dikelompokkan kedalam sentimen positif dan sentimen negatif. Banyak cara dalam proses mengklasifikasikan suatu dokumen kedalam berbagai kelas namun dalam penelitian ini penulis akan menggunakan metode K-Nearest Neighbor (K-NN).

Metode K-Nearest Neighbor telah banyak digunakan dalam mengklasifikasikan dokumen, namun dalam penelitian ini akan menggunakan juga metode kamus yang akan digunakan dalam membangun komentar yang sebelumnya dalam bentuk tidak baku menjadi bentuk baku. Setelah menjadi baku baru komentar tersebut akan diklasifikasikan dengan metode K-Nearest Neighbor

yang metodenya akan dibantu dengan penggunaan *feature selection*, menggunakan pembobotan dari metode TF-IDF dan Cosinne Similiarity.

1.2 Rumusan Masalah

Dari latar belakang diatas, masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut :

1. Dapatkah algoritma *K-Nearest Neighbour* diimplementasikan untuk mengklasifikasikan pesan SMS kedalam kelas sentimen positif dan sentimen negatif?
2. Seberapa akurat sistem dalam mengklasifikasikan sentimen pelanggan yang masuk ke dalam layanan SMS “Suara Konsumen”?

1.3 Batasan Masalah

Batasan masalah yang membatasi ruang lingkup penelitian ini meliputi :

1. Komentar yang dipakai berbahasa Indonesia, karena dalam database kamus kata akan digunakan database dari kata-kata singkatan berbahasa Indonesia.
2. Komentar Training hanya akan diambil dari layanan Suara Konsumen sebanyak 200 SMS. Dimana 100 SMS merupakan komentar bernada positif dan 100 SMS merupakan komentar bernada negatif.
3. Nilai K pada pencarian nilai yang terdekat menggunakan angka ganjil dan menggunakan 4 macam nilai K yaitu $K = 1$, $K = 3$, $K = 5$, dan $K = 6$.
4. Nilai Feture Selection akan menggunakan 4 macam nilai yaitu 10%, 30%, 50%, dan 100%.
5. Komentar yang masuk akan diklasifikan menjadi 2 kategori yaitu Sentimen Positif yang merupakan komentar yang isinya bisa memuji, memberi semangat, atau merasa puas, contoh : “enak, memuaskan, senang”. Sentimen negatif adalah komentar yang berisikan menghina, puas, ataupun marah, contoh : “jijik, mahal, jorok”.

6. Pengujian terhadap data dilakukan hanya pada tingkat kalimat. Data training dan data yang akan uji adalah data yang mempunyai lebih dari 1 kata.
7. Sistem ini tidak memproses kalimat yang kompleks yaitu:
 - a. kata sentimen positif atau negatif memiliki orientasi berlawanan dalam domain namun digunakan dengan pengaplikasian yang berbeda . Misalnya, "bisa" biasanya menunjukkan sentimen negatif, misalnya, "bisa ular itu menjalar keseluruh tubuh" tetapi juga dapat menyiratkan sentimen positif , misalnya , "saya bisa mengerjakan soal".
 - b. Sarkasme, yaitu kalimat yang mempunyai maksud yang berbeda namun digunakan sebagai sentimen sebaliknya, seperti "hebat sekali mobil ini, sampai berhenti tiap 10 meter".
 - c. Banyak kalimat tanpa kata-kata sentimen juga dapat menyiratkan pendapat. Contoh : "mesin cuci ini menggunakan banyak air" menyiratkan sentimen negatif tentang mesin cuci karena menggunakan banyak sumber daya (air).
8. Sebuah SMS akan dikatakan sebagai SPAM jika:
 - a. SMS yang memiliki content yang sama atau nilai yang sama maksimal 3 kali dalam 1 kali 24 jam.
 - b. Sistem akan diberi database yang berisi kata-kata sentimen positif maupun kata-kata sentimen negatif yang berhubungan dengan komentar terhadap produk tersebut, dan jika SMS yang masuk berisi sebanyak kurang 15% dari kata-kata tersebut maka SMS tersebut akan dimasukkan kedalam SPAM.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk melihat sentimen positif dan sentimen negatif komentar-komentar yang terdapat pada layanan SMS Suara Konsumen, dan komentar tersebut diharapkan dapat diklasifikasikan dengan benar menggunakan metode K-Nearest Neighbour.

Adapun Hasil dari penelitian ini akan membantu produsen produk dalam melihat seberapa memuaskan produk yang telah mereka pasarkan dan memudahkan operator dalam mengelompokkan komentar yang masuk ke dalam

layanan Suara Konsumen dan melihat seberapa akurat penggunaan metode K-Nearest Neighbor dalam mengklasifikasikan sebuah komentar dari SMS yang masuk kedalam layanan Suara Konsumen.

1.5 Metode Penelitian

Dalam penelitian ini diperlukan beberapa data-data yang menunjang proses penelitian.

1. Studi Kepustakaan dalam tahapan studi pustaka, penulis akan mempelajari teori-teori dari buku , *e-book* maupun jurnal yang dapat mendukung dalam penulisan tugas akhir.
2. Melakukan analisis terhadap masalah yang ada, batasan yang dimiliki dan keperluan yang diperlukan.
3. Melakukan analisis dan mencermati kata-kata dalam komentar yang dapat berhubungan dengan data yang akan diolah. Bagaimana kata-kata tersebut dapat dikategorikan dalam komentar atau data yang menjadi SPAM.
4. Fokus Grup dalam mencari nilai data training dalam menguji sitem maka akan dilakukan fokus group dalam menentukan nilai yang dari data training yang telah didapat dari komentar yang masuk kedalam layanan suara konsumen. Fokus group ini akan diisi dari 10 untuk menentukan apakah data yang masuk merupakan sentimen positif dan sentimen negatif.
5. Metodologi pengembangan aplikasi akan menggunakan model waterfall dalam menganalisis masalah, mendesain sistem, mengkoding, dan melakukan testing.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini secara garis besar akan dibagi menjadi beberapa bab yaitu :

Bab 1, Pendahuluan. Bab ini memberikan gambaran umum mengenai penelitian yang akan dilakukan mencakup latar belakang, perumusan masalah, batasan, tujuan, metode dan sistematika penulisan tugas akhir ini.

Bab 2, Landasan teori. Bab ini berisi tinjauan pustaka dan landasan teori yang dipakai dalam penelitian ini. Pada bab ini akan dijabarkan konsep dan teori dari berbagai sumber yang terkait dengan penelitian ini.

Bab 3, Perancangan sistem. Bab ini berisi mengenai rancangan sistem yang akan dibangun dalam penelitian ini. Rancangan yang ada berupa : rancangan spesifikasi sistem, rancangan diagram sistem, rancangan antarmuka sistem berupa input dan output. Dan tahapan yang dilakukan dalam pembuatan sistem.

Bab 4, Implementasi dan Analisis Sistem. Dalam bab ini berisi gambar capture dari sistem yang didalamnya terdapat penjelasan dan pembahasan mengenai implementasi sistem yang telah dirancang pada bab sebelumnya. Bab ini juga akan menjelaskan Analisis dari sistem yang telah dibuat.

BAB 5, Kesimpulan dan Saran. Berisi mengenai kesimpulan yang didapat dari penelitian yang telah dilakukan serta saran-saran dari penulis terhadap penelitian ini sehingga dapat referensi bagi penelitian selanjutnya.

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan Analisis dan implementasi sistem yang telah dibuat maka diperoleh kesimpulan sebagai berikut:

1. Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* dapat diimplementasikan ke dalam sistem untuk digunakan dalam mengklasifikasi sentimen positif atau sentimen negatif konsumen terhadap suatu produk yang berbentuk pesan SMS dalam layanan Suara Konsumen.
2. Sistem klasifikasi komentar SMS dengan metode *K-Nearest Neighbor* memberikan nilai akurasi yang cukup tinggi dengan nilai rata-rata 89% serta nilai keakuratan akan optimal ketika menggunakan $K = 3$ dan $K = 7$ yang nilai keakuratannya mencapai 90%.

5.2. Saran

Saran untuk pengembangan dan perbaikan sistem adalah:

1. Perlu ditambahkan komentar yang memiliki lebih banyak kosakata sehingga dapat menambah variasi kata sehingga sistem menjadi semakin akurat.
2. Ditambahkan kosakata singkatan dan mengurangi jumlah *stopword* yang dapat mengurangi tingkat kesalahan sistem dalam mengklasifikasi komentar yang masuk kedalam sistem karena perkembangan kosakata bahasa akan semakin bervariasi.
3. Pada program bantu SMS *Gateway* yaitu Gammu memiliki tingkat kestabilan rendah pada modem yang menggunakan koneksi USB oleh karena itu diperlukan Modem yang menggunakan port khusus agar *services* Gammu tidak perlu berpindah-pindah port.

DAFTAR PUSTAKA

- El-Halees, A. (2011). Arabic Opinion Mining using Combined Classification Approach.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concepts and Techniques*. Morgan Kaufmann.
- Handoko, B. Y. (2008). Perbandingan Metode Naive Bayes Classifier (NBC) dengan K-Nearest Neighbor (K-NN) untuk Klasifikasi Dokumen.
- Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews.
- Kowalski, G. (2011). *Information Retrieval Architecture and Algorithms*. Ashburn: Spinger.
- Kusumo, G. D. (2013). Klasifikasi Sentimen Pembaca berdasar Komentar terhadap Berita menggunakan Metode Naive Bayes Classifier.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining* . Morgan & Claypool Publishers.
- Liu, B. (2007). *Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data*. Chicago: Springer.
- Pang, B., Lee, L., & Vaithyanathan, S. (2003). Thumbs up? Sentiment Classification using Machine Learning.
- Wiebe, J., Bruce, R., & O'Hara, T. (1999). Development and use of a gold standard data set for. *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)* , 246–253.