

KONTROL EKSPRESI WAJAH BERDASARKAN KLASIFIKASI TEKS MENGUNAKAN METODE NAIVE BAYES

Mochamad rizki¹, Akuwan saleh²,
Mahasiswa Jurusan Teknik Telekomunikasi¹, Dosen Pembimbing²
Politeknik Elektronika Negeri Surabaya
Institut Teknologi Sepuluh Nopember
Kampus PENS-ITS Keputih Sukolilo Surabaya 60111
Telp (+62)31-5947280, 5946114, Fax. (+62)31-5946114
Email : aleiytelkom06@yahoo.com

Makalah Proyek Akhir

ABSTRAK

Komunikasi dapat dilakukan dari informasi verbal dan non-verbal, verbal dapat berupa tulisan yang diperoleh dari kata, kalimat, paragraf dan sebagainya untuk penggalan informasi teksnya menggunakan klasifikasi teks. Pada proses klasifikasi itu akan digunakan data set yang telah diketahui kelas emosinya yaitu anger, joy dan sadness dengan menggunakan metode Naïve Bayes. Akan dilihat sejauh mana metode itu dapat mengklasifikasikan data emosi dalam bahasa Inggris. Hasil dari klasifikasi teks tersebut digunakan sebagai control ekspresi wajah dengan pemrograman VRML untuk memvisualisasikan ekspresi wajah tersebut.

Kata Kunci : *klasifikasi teks, naive bayes classifier, vrml, emosi..*

1. PENDAHULUAN

1.1 Latar Belakang

Kemunculan teknologi agen cerdas menyadarkan telah ada peluang untuk mengembangkan antarmuka sebagai perbaikan model interaksi antara manusia dan komputer; simulasi karakter virtual untuk aplikasi berbeda seperti hiburan, pendidikan dan sebagainya. Karakter yang hidup dipercaya dapat meningkatkan ketertarikan pada game komputer. Dewasa ini interaksi antara manusia dan komputer telah dilakukan melalui text, mouse atau keyboard bersamaan dengan cepatnya perkembangan komputer grafis dan teknologi pengenalan wicara membuat interaksi lebih adaptif, fleksibel dan berorientasi manusia [1]

Agen lama digagas dan lebih dari dua dekade, agen adalah segala tanggapan terhadap lingkungan melalui sensor dan bertindak sesuai lingkungan tersebut sebagai efeknya. Agen membutuhkan input dan output, input dapat berupa kamera, microphone sedangkan output hampir sama dengan manusia berupa suara atau gerakan isyarat [2].

Sebuah agen dapat dikatakan cerdas jika dilengkapi dengan emosi [3], sehingga agen perlu ditambahkan emosi. Komunikasi dapat dilakukan

dari informasi verbal dan non-verbal, verbal dapat berupa tulisan yang diperoleh dari kata, kalimat, paragraf dan sebagainya. Nonverbal dapat sebuah isyarat tubuh [4] dalam beberapa studi tentang emosi dan interaksi manusia dan komputer di dasarkan pada analisis ekspresi wajah, meskipun demikian hampir semua fokus pada analisis data psikologi atau pengenalan wajah, penghilangan aspek komunikasi non-verbal [5]

Pengambil ekspresi wajah untuk mewakili suatu emosi sebagai contoh senyuman dapat diartikan senang atau gembira, sedangkan agar tafsiran emosi menjadi lebih baik harus tetap memperhatikan informasi verbal. Dalam penelitian ini untuk mengenali komunikasi verbalnya dengan berbasis klasifikasi teks.

1.2 Rumusan Permasalahan

Perumusan masalah pada penelitian ini didasarkan pada bagaimana mengasilkan ekspresi wajah pada model tiga dimensi yang mengandung lebih dari satu jenis emosi. Dengan demikian perumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut :

1. Bagaimana mengenali jenis emosi dari sebuah teks berbahasa Inggris ?

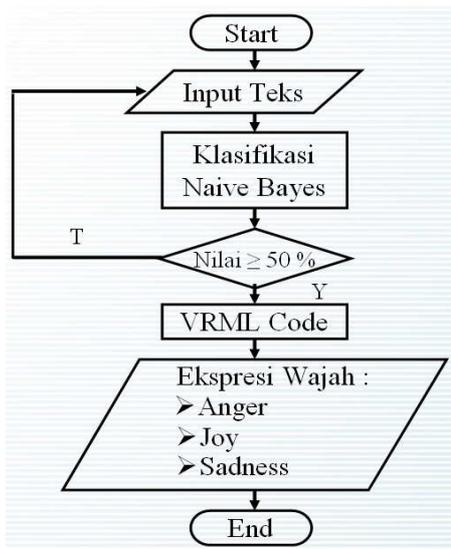
2. Bagaimana menampilkan emosi berupa ekspresi wajah yang dipengaruhi oleh lebih dari satu emosi ?

1.3 Batasan Permasalahan

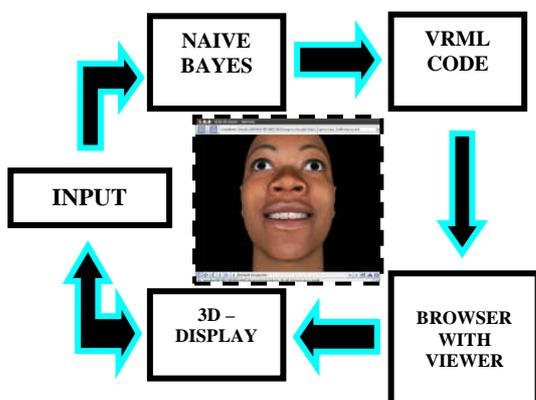
Batasan masalah dalam pembuatan proyek akhir kali ini adalah klasifikasi teks yang sudah diketahui nilainya dan dibatasi 3 kategori emosi yaitu joy, anger dan sadness dalam hal ini kita dapat mengkategorikan suatu ekspresi dengan menggunakan metode naive bayes. Dan dalam aplikasinya teknik yang digunakan adalah menggunakan program rainbow dengan tampilan GUI dan software yang digunakan adalah menggunakan bahasa pemrograman Shell.

2.1. Perencanaan Sistem

Perancangan dan implementasi sistem akan digunakan dengan mendesain metode-metode yang ada dan diimplementasikan dengan menggabungkan metode-metode tersebut sehingga didapatkan hasil yang diinginkan.



Gambar 2.1 Flowchart System



Gambar 2.2 Blok Diagram Sistem

Proyek tugas akhir kontrol ekspresi wajah berdasarkan klasifikasi teks menggunakan naive bayes ini bekerja berdasarkan Klasifikasi teks. Klasifikasi teks adalah proses pengelompokan dokumen kedalam kelas berbeda, dalam tahapannya tiap dokumen d menunjuk pada satu kelas tertentu maka dibutuhkan proses untuk menggali informasi dari dokumen tersebut. Sehingga dokumen tersebut harus dapat merepresentasikan dari kelasnya sehingga tiap kata yang muncul dalam dokumen mempunyai nilai.

2.2. Teori Penunjang

2.2.1 Rainbow

Rainbow merupakan sebuah *program* untuk mengklasifikasikan dokumen yang didalamnya terdapat beberapa jenis metode untuk klasifikasi. Rainbow didasari dengan Library Bag of Words (libbow). Program ini dibuat oleh Prof. Andrew McCallum beserta beberapa anak muridnya. Program Rainbow ini diciptakan pada tahun 1996 dan telah mengalami beberapa perubahan. Perubahan terakhir terjadi pada tahun 1998.

Program Rainbow biasanya mengklasifikasi data menurut atribut yang dimiliki data tersebut dan dikategorikan ke dalam kelas yang sesuai dengan data tersebut. Dengan adanya program ini maka tidak perlu lagi membaca dokumen satu per satu dan mengklasifikasinya. Program rainbow dapat mengklasifikasi ribuan dokumen hanya dalam hitungan detik. Program ini bertujuan untuk mendukung perkembangan di bidang Information Retrieval dengan menyajikan sarana untuk mengklasifikasikan dokumen agar kecepatan pada proses retrieval dapat ditingkatkan lagi.

2.2.2 GUI (Graphical User Interface)

Zenity mempunyai peran sebagai pembuat GUI pada shell programing atau yang disebut *Grafik User Interface* dimana mempunyai fungsi mempermudah user untuk mengoperasikan programnya melalui grafis secara interaktif. Zenity mempunyai banyak opsi seperti:

--entry	meminta input dari keyboard
--info	menampilkan text yang di mana berfungsi sebagai sebuah info.
--list	membuat list berdasarkan column dan row secara tertabel dan lain-lain.

Dengan opsi seperti diatas zenity dapat digunakan untuk membuat sebuah question dialog box. Disamping itu zenity juga dapat digunakan untuk aplikasi yang lain seperti calendar, entry, error, info, file selection, list, notification, progress,

warning, scale, dan text info. Pada bab ini akan di ilustrasikan bagaimana membuat aplikasi zenity dialog.

2.2.3 AWK

Statement Control

Script awk telah dibahas pada praktikum sebelumnya dengan beberapa kelebihan untuk manipulasi file teks. Seperti bahasa pemrograman lainnya awk juga memiliki statement control diantaranya untuk pengkondisian menggunakan perintah *if*, perulangan dengan perintah *while* dan *for*.

Selain statement diatas, perintah *break* dan *continue* pada proses loop di awk juga bisa dipergunakan. Ketika sampai pada perintah *break*, proses loop menggunakan perintah *for* atau *while* segera dihentikan selanjutnya melakukan control kembali untuk mengikuti statement loop berikutnya. Ketika ketemu perintah *continue*, segera memulai iterasi dari awk yaitu mengikuti alur control dari perintah *next*. Ketika *next* ditemukan, maka baris input dibaca dan dieksekusi setelah itu kembali lagi melakukan.

3.1 Pembuatan Sistem dan Hasil pengujian

Percobaan pada tugas akhir ini menggunakan dua jenis data yang berbeda. Data pertama adalah data-emosi-en. Data tersebut terbagi atas tiga kategori, yaitu anger, joy dan sadness. Data kedua adalah 20Newsgroups dataset yang merupakan kumpulan e-mail yang berjumlah 18828 dokumen. 20Newsgroups dataset memiliki 20 kategori dan dapat diunduh pada <http://people.csail.mit.edu/jrennie/20Newsgroups/>. Dokumen e-mail yang terdapat pada 20Newsgroups dataset yang digunakan pada percobaan ini merupakan dokumen-dokumen yang telah dihilangkan tag header-nya.

Contoh dokumen yang digunakan pada tugas akhir ini dapat dilihat pada tabel di bawah:

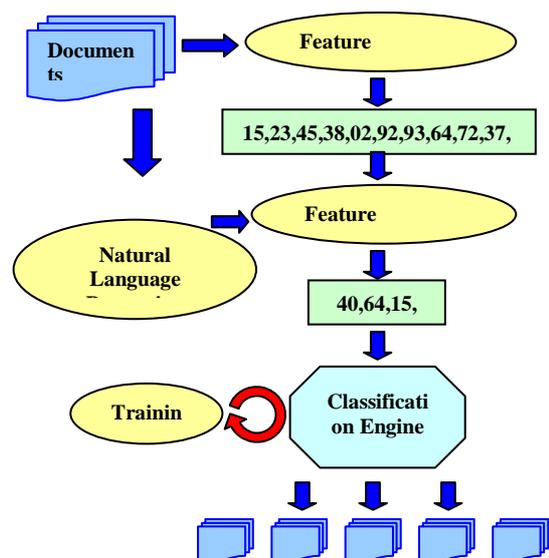
Kategori	Jumlah Teks
data-emosi-en	
Anger	1088
Joy	1090
Sadness	1083
20Newsgroups Dataset	
alt.atheism	799
comp.graphics	973
comp.os.ms-windows.misc	985
comp.os.ms-windows.misc	982
comp.sys.mac.hardware	961
comp.windows.x	980
misc.forsale	972
rec.autos	990
rec.motorcycles	994

rec.sport.baseball	994
rec.sport.hockey	999
sci.crypt	991
sci.electronics	981
sci.med	990
sci.space	987
soc.religion.christian	997
talk.politics.guns	910
talk.politics.mideast	940
talk.politics.misc	775
talk.religion.misc	628

3.1.1 Klasifikasi Teks

Klasifikasi atau kategorisasi teks adalah proses penempatan suatu teks ke suatu kategori atau kelas sesuai dengan karakteristik dari teks tersebut. Dalam *text mining*, klasifikasi mengacu kepada aktifitas menganalisis atau mempelajari himpunan teks *pre-classified* untuk memperoleh suatu model atau fungsi yang dapat digunakan untuk mengelompokkan teks lain yang belum diketahui kelasnya ke dalam satu atau lebih kelas *predefined* tersebut.

Klasifikasi teks adalah proses pengelompokan dokumen kedalam kelas berbeda, dalam tahapannya tiap dokumen *d* menunjuk pada satu kelas tertentu maka dibutuhkan proses untuk menggali informasi dari dokumen tersebut. Sehingga dokumen tersebut harus dapat merepresentasikan dari kelasnya sehingga tiap kata yang muncul dalam dokumen mempunyai nilai.



Gambar 2.3 Diagram Proses Klasifikasi Teks

3.1.2 Emosi

Emosi dapat digambarkan sebagai keadaan yang pada umumnya disebabkan oleh suatu

kejadian penting sebuah subyek yang meliputi (a) keadaan mental sadar yang dinyatakan dengan kemampuan mengenali, kualitas perasaan dan diarah untuk beberapa subyek, (b) gangguan jasmani pada beberapa organ tubuh, (c) pengenalan ekspresi pada wajah, suara dan isyarat tubuh, (d) kesiapan untuk melakukan tindakan tertentu. Karenanya emosi dalam sosio-biologi adalah kecenderungan mental (conative dan kognitif), keadaan, proses dan model komputasi harus spesifikasi semirip mungkin [6].

Sejumlah penelitian tentang emosi manusia telah dilakukan sehingga ada kesepakatan tentang emosi dasar [7]

1. Marah sebagai ganjalan atau frustrasi dari peran atau tujuan yang di rasakan orang lain
2. Sedih digambarkan sebagai kegagalan atau kerugian tentang peran dan tujuan.
3. Senang digambarkan sebagai berhasil atau bergerak menuju selesainya peran yang bernilai atau tujuan.

3.1.3 Perbedaan Perlakuan Preprocessing

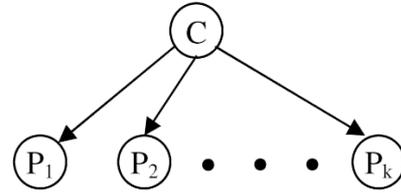
Di dalam klasifikasi teks untuk kategori artikel kata-kata seperti “tidak”, “bukan”, “tanpa” dianggap tidak penting sehingga di masukan dalam daftar kata yang bias dihilangkan (stopword), namun dalam hal ini (klasifikasi teks emosi) kata-kata “tidak”, ”bukan”, “tanpa” menjadi sangat berarti sehingga tidak boleh dihilangkan.

Perlakuan proses sebelum klasifikasi menjadi sangat penting supaya data yang akan diolah benar-benar mewakili maksud dari dokumen yang ditulis seringkali penggunaan kata “bukan”, “tanpa” dan “tidak” menjadi sangat penting, perbedaan arti “tanpa cinta” dan “tidak senang” dapat menempatkan dokumen dalam kelas berbeda sehingga tidak menutup kemungkinan dalam preprocessing ada modifikasi kata terlebih dahulu [8]

3.1.4 Naive Bayes

Klasifikasi-klasifikasi Bayes adalah klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Untuk klasifikasi Bayes sederhana yang lebih dikenal sebagai naïve Bayesian Classifier dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Asumsi ini disebut *class conditional independence* yang dibuat untuk memudahkan perhitungan-perhitungan pengertian ini dianggap “naive”, dalam bahasa lebih sederhana naive itu mengasumsikan bahwa kemunculan suatu term kata dalam suatu kalimat tidak dipengaruhi kemungkinan kata-kata yang lain dalam kalimat padahal dalam kenyataannya bahwa kemungkinan

kata dalam kalimat sangat dipengaruhi kemungkinan keberadaan kata-kata yang dalam kalimat. Dalam Naïve Bayes di asumsikan prediksi atribut adalah tidak tergantung pada kelas atau tidak dipengaruhi atribut laten



Gambar 2.4 Klasifikasi Naive Bayes sebagai aringan bayes dengan atribut prediksi (P1, P2, Pk) dan kelas (C)

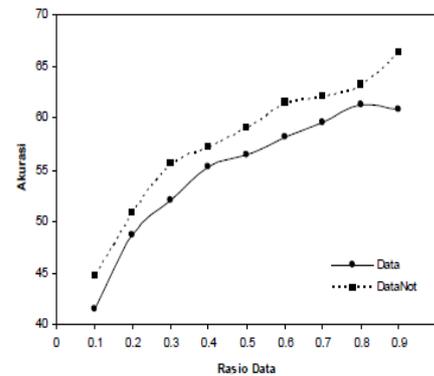
C adalah adalah anggota kelas dan X adalah variabel acak sebuah vektor sebagai atribut nilai yang diamati. c mewakili nilai label kelas dan x mewakili nilai atribut vector yang diamati. Jika diberikan sejumlah x tes untuk klasifikasi maka probabilitas tiap kelas untuk atribut prediksi vektor yang diamati adalah

$$p(C = c | X = x) = \frac{p(C = c) p(X = x | C = c)}{p(C = c)}$$

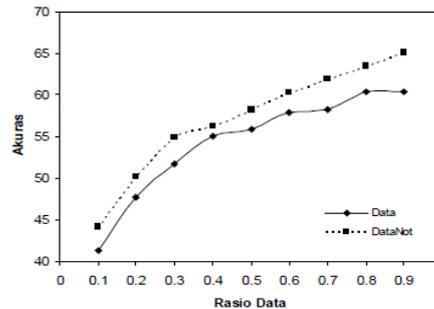
X = x adalah mewakili kejadian $X_1 = x_1 \wedge X_2 = x_2 \wedge \dots \wedge X_k = x_k$

Jumlah dari $p(C = c | X = x)$

Untuk semua kelas adalah 1



Gambar 2.5 Akurasi Penggunaan Data dan DataNot Dengan Metode Multinomial Naive Baves



Gambar 3.1.6 Akurasi Penggunaan Data dan DataNot Dengan Naive Bayes

3. KESIMPULAN DAN SARAN

3.1 KESIMPULAN

Dari hasil pengujian dan analisa pada bab sebelumnya maka dapat diambil kesimpulan

1. Pemilihan metode terbaik adalah yang mempunyai tingkat akurasi yang tinggi dan juga dipastikan simpangan bakunya yang cenderung lebih kecil. Dari data pengujian terangkum dibawah ini dapat disimpulkan bahwa metode klasifikasi teks yang terbaik adalah metode *Support Vector Machine* dengan tingkat akurasi 52.71% » 53% dan simpangan baku 0.07.
2. Hasil klasifikasi menggunakan metode *Naive Bayes* tidak memberikan perbaikan yang signifikan saat rasio data 0.4 untuk percobaan DataNot (data modifikasi) dan Data (data asli).
3. Metode *Support Vector Machine* lebih baik meskipun sangat mudah terjadi error dalam data training dan metode *Naive Bayes* sangat cocok untuk klasifikasi teks berbahasa Inggris.

3.2 Saran

Saran-saran yang bisa disampaikan adalah sebagai berikut:

1. Dengan menggunakan *Support Vector Machine* data yang di dapat mempunyai tingkat akurasi yang tinggi dan dipastikan simpangan bakunya cenderung kecil dibandingkan metode lain.
2. Perlu penelitian yang lebih mendalam dalam klasifikasi tanggapan emosi dengan menggunakan metode lain.
3. Perlu adanya pengembangan dalam penggunaannya untuk mengklasifikasikan sebuah teks dengan ekspresi wajah.
4. Penelitian dapat dikembangkan menjadi tanggapan dari hasil klasifikasi menggunakan animasi sederhana

4. DAFTAR PUSTAKA

- [1] Gregorius S. Budhi, Ibnu Gunawan, Ferry Yuwono, “*Algoritma Porter Stemmer For Bahasa Indonesia Untuk Pre-Processing Text Mining Berbasis Metode Market Basket analysis*”, UK Petra Jurusan Teknik Informatika.
- [2] Hearst, M. A. 1997. “*Text data mining: Issues, techniques, and the relationship to information access*”. Presentation notes for UW/MS workshop on data mining, July 1997.
- [3] Yudi Wibisono, “*Klasifikasi Berita Berbahasa Indonesia menggunakan Naive Bayes Classifier*”, Jurusan Pendidikan Matematika FPMIPA UPI, Seminar Nasional Matematika 2005 di Universitas Pendidikan Indonesia, Bandung, tanggal 20 Agustus 2005.
- [4] Kusriani, Emha Taufiq Luthfi, “*Algoritma Data Mining*”, Andi 2009
- [5] Danisman Taner, Alpkocak Adil. (2008). *Feeler: Emotion Classification of Text Using Vector Space Model*.
- [6] Guo, Qinglin (2008). *The Similarity Computing of Documents Based on VSM*, Springer Verlag Berlin
- [7] Dunham, Margareth H (2003) *Data Mining Introductory and Advanced Topics*, New Jersey: Prentice Hall
- [8] Chenometh, Megan, Song, Min (2009) *Text Categorization*, dalam *Encyclopedia of Data Warehouse & Data Mining*, IGI Global, hal. 1936-1941
- [9] Mahinovs, Aigars, TiwariText, Ashutosh (2007) *Classification Method Review. Decision Engineering Report Series*, Cranfield University
- [10] Asirvatham, Arul Prakash, Ravi, Kranthi Kumar (200x) *Web Page Categorization Based on Document Structure*
- [11] Nigam, Kamal Et.al (1998) *Learning to Classify Text from Labeled and Unlabeled Documents*, 15th *National Conference on Artificial Intelligence (AAAI-98)*
- [12] Sebe Nicu, Lew Michael S. (2002). *Emotion Recognition Using a Cauchy Naive Bayes Classifier*.
- [13] Chuang Zee-Jing, Chung-Hsien Wu.2004. *Multi modal Emotion Recognition from Speech and Text*.