ANALISIS PENGARUH FRASA PADA DETEKSI EMOSI DARI TEKS MENGGUNAKAN *VECTOR SPACE MODEL*

Ranap Sitorus, Harry Soekotjo Dachlan, dan Wijono

***Abstract*— Text communication is one way of conveying information about one's attitudes and emotional state. Emotion is also very important role in taking a decision such as doing emotion detection from the text of the questionnaire. Text emotions are detected at various stages to be recognizable by the computer. The steps taken are preprocessing the case folding, tokenizing, filtering and stemming. In this emotional detection research, For grouping words based on word class, using POS-Tagging where approach based rule rule based containing combination of word class which most likely when combined will form a phrase in order to facilitate the computer in understanding the characteristics of a phrase. The phrases-based detection-based tokenisation process uses the Hidden Markov Model (HMM) POS-Tagger for easy classification using the Tf-Idf and VSM methods. This study aims to classify text communication in Indonesia language into emotional expression classes into 3 basic emotional classes and serve as training documents and test documents obtained from the results of student questionnaires as many as 264 documents. The computer is able to detect emotion with the corpus data by performing sentences into words or phrases using Chunk that can classify emotions happy, disappointed and afraid. From the test results for 90% training data and 10% test data in detecting emotions using Phrase got 92.59%**

Keyword— tf-idf, Vector Space model, emotion, phrase.

***Abstrak*– Komunikasi teks adalah salah satu cara menyampaikan informasi tentang sikap dan keadaan emosional seseorang. Emosi juga sangat berperan penting dalam mengambil sebuah keputusan seperti melakukan deteksi emosi dari teks kuesioner. Emosi teks didedeksi dengan berbagai tahap agar dapat dikenali oleh komputer. Tahapan yang dilakukan adalah melakukan *preprosesing* yaitu *case folding, tokenizing, filtering dan stemming*. Dalam penelitian deteksi emosi ini, Untuk pengelompokan kata berdasarkan kelas kata, yaitu dengan menggunakan POS-Tagging dimana pendekatan berdasar aturan rule-based yang berisi kombinasi kelas kata yang kemungkinan besar bila digabungkan akan membentuk frasa agar memudahkan komputer dalam memahami ciri-ciri sebuah frasa. Proses tokenisasi berbasis deteksi frasa menggunakan Hidden Markov Model (HMM) POS-Tagger untuk memudahkan klasifikasi menggunakan metode Tf-Idf dan VSM. Penelitian ini bertujuan untuk mengklasifikasikan komunikasi teks dalam Bahasa Indonesia menjadi kelas ekspresi emosi kedalam 3 kelas emosi dasar dan dijadikan sebagai dokumen pelatihan dan dokumen uji yang diperoleh dari hasil kuesioner mahasiswa sebanyak 264 dokumen. Komputer mampu mendeteksi emosi dengan data corpus dengan melakukan pemenggalan kalimat menjadi kata/frase menggunakan Chunk yang dapat mengklasifikasikan emosi senang, kecewa dan Takut. Dari hasil uji coba untuk data *training* 90 % dan data uji 10 % dalam mendeteksi emosi dengan menggunakan Frasa mendapat hasil 92,59%.**

***Kata Kunci— tf-idf, Vector Space model, emosi, frasa*.**

I. PENDAHULUAN

A. Latar Belakang

KOMUNIKASI teks adalah salah satu yang digunakan untuk menyampaikan informasi dan juga berisi informasi tentang sikap dan keadaan emosional seseorang [1]. Emosi bersifat umum sangat penting didalam semua aspek kehidupan, dimana emosi merupakan salah satu faktor yang akan mempengaruhi keputusan hubungan manusia dengan lingkungan sosial dengan membentuk perilaku keseharian seseorang dalam berkomunikasi. Komunikasi yang terjadi pada era teknologi saat ini bersifat multi fungsi, baik ditinjau dari segi perangkat maupun prosesnya. Teknologi perangkat yang digunakan untuk berkomunikasi sangat berkembang cepat, mulai dari telepon menggunakan jaringan internet, *text messaging* dan *video call* yang kesemuanya dilakukan secara langsung antar pengguna. Akan tetapi komunikasi secara tidak langsung juga dapat dilakukan seperti untuk memberikan *review* terhadap kualitas sebuah produk. Salah satu contoh yaitu memberikan Saran dan kritik terhadap sumber daya manusia (SDM) menggunakan metode kuesioner dan beberapa pertanyaan tambahan karena dengan metode kuesioner teks ini bebas menuliskan apa yang akan disampaikan.

Kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan secara tertulis kepada responden untuk dijawab [2]. PT. X adalah perguruan tinggi swasta yang menerapkan pemberian saran dan kritik terhadap kualitas dosen mengajar melalui pengisian kuesioner. Pengisian kuesioner tersebut dilakukan untuk memberikan kritikan dan saran terhadap pelayanan

Ranap Sitorus, Harry Soekotjo Dachlan and Wijono are with the Electrical Engineering Department of Brawijaya University, Malang, Indonesia (corresponding author provide phone 0341-554166; (*ranap.natalinas@gmail.com*).

dosen pada mahasiswa.

Permasalahan pada makalah ini adalah sering munculnya kesalahan persepsi seseorang dalam menginterpretasikan beragam komunikasi teks mengakibatkan permasalahan baru dalam ruang lingkup sosial manusia. Komputer adalah sebagai media untuk melakukan komunikasi teks, yang mana masih sangat susah untuk mengetahui kondisi emosi seseorang karena interaksi dilakukan secara tidak langsung, tidak ada nada dan intonasi dalam media teks.

Berdasarkan permasalahan tersebut, maka muncullah sebuah gagasan bagaimana mengolah dan menganalisa data rekapitulasi hasil kuesioner dengan cara mengelompokkan atau klasifikasi opini mahasiswa yang tertuang dalam kuesioner pada beberapa kategori atau kelas. Proses pengelompokan berdasarkan deteksi tingkat emosi mahasiswa menggunakan sumber data hasil kuesioner mahasiswa berupa teks. Oleh karena itu kategori yang digunakan untuk pengelompokkan opini mahasiswa adalah kategori emosi senang, kecewa dan takut. Pada kategori senang disini nantinya mencerminkan sebuah rasa bahagia atau gembira dimana mahasiswa terhadap proses pembelajaran yang dilakukan oleh dosen. Pada kategori kecewa adalah kurangnya pelayanan atau memberikan pernyataan ketidakpuasan terhadap dosen sedangkan takut adalah kebalikan dari rasa puas tersebut. Ketiga kategori emosi tersebut telah mempresentasikan emosi mahasiswa terhadap Dosen. Pengkategorian emosi ini didasarkan seringnya dijumpai hasil kuesioner yang berupa ekspresi senang ketika pengguna senang, kecewa , dan takut.

Adapun penelitian tentang emosi dan sentiment yang sudah dilakukan pada layanan media sosial seperti deteksi emosi terhadap isi blog yang diperoleh dari *lexicon* emosi dan mampu meningkatkan nilai *Accuracy*nya [4],[5], deteksi emosi terhadap isi email dan dibuat sebagai pengembangan lexicon emosi [7]. dan micro-blog [7]. Dengan metode yang sudah digunakan untuk mendeteksi emosi seperti keyword-spotting, rule- based dan statistical untuk pembelajaran mesin pencarian dokumen. akan tetapi deteksi emosi yang dilakukan tersebut hanya terbatas pada bahasa Inggris. Untuk Deteksi emosi dengan bahasa Indonesia hanya pengembangan leksikon emosi yaitu pemilihan seed words dan perluasan leksikon [8].

Berdasarkan uraian tersebut, maka dalam penelitian yang dilakukan kali ini menggunakan dua buah metode yaitu TF-IDF dan VSM (Vector Space Model). Dalam Fitur TF-IDF dengan *discounted-cumulative* digunakan untuk menangani karakter topik yang muncul pada sumber data, yang berkelanjutan dengan *discounted cumulative* untuk mengekstrak topik agar mampu meningkatkan jumlah topik terekstrak yang sesuai. Sedangkan *Vector Space Model* (VSM) merupakan salah satu metode dalam pengklasifikasian teks yang digunakan untuk menentukan jenis emosi yang dihasilkan karena cara kerja model ini efisien, mudah dalam merepresentasikan dan dapat diimplementasikan pada *document- matching* terhadap bahasa Indonesia.

Dalam penelitian deteksi emosi ini, Untuk pengelompokan kata berdasarkan kelas kata, yaitu dengan menggunakan POS-Tagging dimana pendekatan berdasar aturan rule-based yang berisi kombinasi kelas kata yang kemungkinan besar bila digabungkan akan membentuk frasa agar memudahkan komputer dalam memahami ciri-ciri sebuah frasa. Selanjutnya dilakukan proses penghitungan bobot kata yang terdapat dalam sumber data (corpus) menggunakan metode Term Frequency-Inverse Documen Frequency (TF-IDF). Sedangkan menggunakan metode Vector Space Model (VSM) untuk menentukan relevansi antara Fitur - fitur fase dalam menemukan korelasi antara yang sangat penting yaitu untuk mengurangi dimensi data multidimensi yang dapat digunakan untuk menilai emosi pengguna dengan fitur ekstraksi, VSM digunakan untuk mengetahui hubungan tiap-tiap kata yang membentuk sebuah kalimat. Emosi di setiap pengguna akan disajikan dengan nilai persentase menggunakan pendekatan VSM dengan mengambil kesamaan antara permintaan dan data corpus.

B. *Preproccessing*

Preprocessing dilakukan untuk menghindari data yang tidak sempurna, interupsi data, dan data yang tidak konsisten [9]. Dokumen teks tidak dapat diproses langsung oleh algoritma pencarian, sehingga diperlukan proses untuk menghasilkan data numerik yang akan digunakan dalam perhitungan dengan menggunaan text preproccesing [10].

Contoh penggunaan model ruang *vector* dengan data yang digunakan sebagai berikut:

Kata kunci (KK) = Dosen ini menyenangkan dan enak diajak berdiskusi.

Senang = Beliau baik dan menyenangkan jg enak mengajarnya.

Kecewa = Dosen ini kurang menyenangkan.

Takut = Hmm.. Galak susah diajak berdiskusi.

Tahapan *preprosessing* yang akan dilakukan meliputi:

a. *Case folding* adalah proses untuk mengubah huruf besar menjadi huruf kecil pada teks.

Kata kunci (KK) = dosen ini menyenangkan dan enak diajak berdiskusi.

Senang = beliau baik dan menyenangkan jg enak mengajarnya.

Kecewa = dosen ini kurang menyenangkan.

Takut = hmm..galak susah diajak berdiskusi.

b. *Tokenizing* merupakan proses pemilahan tiap kata pada teks berdasarkan spasi [11].

Kata kunci (KK) =| dosen | ini | menyenangkan | dan |enak | diajak| berdiskusi|.

Senang = | beliau | baik | dan | menyenangkan | jg | enak | mengajarnya |.

Kecewa = | dosen | ini | kurang | menyenangkan |.

Takut = | hh.. | galak| susah | diajak | berdiskusi | .

c. *Filtering* karakter merupakan proses *filtering* karakter untuk menghilangkan karakter - karakter bawaan serta karakter lainnya seperti tanda baca dan angka yang tidak termasuk dalam pemrosesan teks [12].

Kata kunci (KK) =| beliau | baik |dan | menyenangkan | dan |enak | diajak| berdiskusi|

Senang = | beliau | baik | dan | menyenangkan | jg | enak| mengajarnya |
Kecewa = | dosen | ini | kurang | menyenangkan |
Takut = | hh | galak | susah | diajak | berdiskusi |

d. Merubah singkatan adalah proses mengubah singkatan menjadi kata baku melibatkan *table database* yang berisi daftar singkatan dirubah menjadi kata baku yang sesuai.
Kata kunci (KK) =| dosen | ini | menyenangkan | dan |enak | diajak| berdiskusi|
Senang = | beliau | baik | dan | menyenangkan | juga | enak | mengajarnya |
Kecewa = | dosen | ini | kurang | menyenangkan |
Takut = | hh | galak | susah | diajak | berdiskusi |

e. *Stemming* merupakan proses tahap terakhir pemprosesing teks, untuk menghilangkan semua imbuhan baik terdiri dari awalan maupun akhiran[13].
Kata kunci (KK) =| dosen | ini | senang | dan |enak | ajak| diskusi|
Senang = | beliau | baik | dan | senang | juga | enak | ngajar |
Kecewa = | dosen | ini | kurang | senang |
Takut = | galak | susah | ajak | diskusi |

C. *TF-IDF*

Metode TF-IDF merupakan metode pembobotan *term* yang banyak digunakan sebagai metode pembanding terhadap metode pembobotan baru. Pada metode ini, perhitungan bobot *term* t dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inverse Document Frequency*.

Pada *Term Frequency* (tf), terdapat beberapa jenis formula yang dapat digunakan yaitu [14] *Tf biner* (*binery tf*), hanya memperhatikan apakah suatu kata ada atau tidak dalam dokumen, jika ada diberi nilai satu, jika tidak ada akan diberi nilai nol. Tf murni (*raw tf*), nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya, jika muncul Lima kali maka kata tersebut akan bernilai lima. Pada Tf logaritmik, untuk menghindari dominansi dokumen yang mengandung sedikit kata dalam query, namun mempunyai frekuensi yang tinggi.

$$Tf = 1 + \log (tf) \quad (2.1)$$

Sedangkan untuk Tf normalisasi, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen.

$$Tf = 0.5 + 0.5 \times \left[\frac{tf}{\max tf}\right] \quad (2.2)$$

Inverse Document Frequency (idf) dihitung dengan menggunakan formula

$$Idf_j = \log (D / df_j) \quad (2.3)$$

Dimana, D adalah jumlah semua dokumen dalam koleksi, DF_j adalah jumlah dokumen yang mengandung term t_j.

Menurut Defeng [15] Jenis formula yang akan digunakan untuk perhitungan *term frequency* (TF) yaitu TF murni (*raw* tf). Dengan demikian rumus umum untuk TF-IDF adalah penggabungan dari formula perhitungan *raw* tf dengan formula IDF (rumus 2.3) dengan cara mengalikan nilai *term frequency* (TF) dengan nilai *inverse document frequency* (IDF):

$$W_{ij} = tf_{ij} \times idf_j$$
$$w_{ij} = tf_{ij} \times \log (D / df_j) \quad (2.4)$$

Keterangan :

w_{ij} adalah bobot *term* tj terhadap dokumen d_i
tf_{ij} adalah jumlah kemunculan term t_j dalam dokumen di
D adalah jumlah semua dokumen yang ada dalam database
df_j adalah jumlah dokumen yang mengandung term tj (minimal ada satu kata yaitu term tj)

Berdasarkan rumus 2.4, berapapun besarnya nilai tf_{ij}, apabila D = df_j, maka akan didapatkan hasil 0 (nol) untuk perhitungan IDF. Untuk itu dapat ditambahkan nilai 1 pada sisi IDF, sehingga perhitungan bobotnya dirumuskan menjadi pada rumus 2.5 .

$$w_{ij} = tf_{ij} \; x \; (\log\left(\frac{D}{tf_j}\right) + 1) \quad (2.5)$$

D. *Vector Space Model*

Vector Space Model adalah suatu model yang digunakan untuk kemiripan antara suatu dokumen dan suatu query dengan mewakili setiap dokumen dalam sebuah koleksi sebagai sebuah titik dalam ruang (*vector* dalam ruang *vector*) [16]. Poin yang berdekatan ruang di ruang ini memiliki kesamaan *semantic* yang dekat dan titik yang terpisah jauh memiliki kesamaan *semantic* yang semakin jauh. Kesamaan antara *vector* dokumen dengan *vector query* tersebut dinyatakan dengan *cosinus* dari sudut antar keduanya [17].

Dalam metode *Vector Space Model* bobot dari setiap *term* yang didapat dalam semua dokumen dan *query* dari *user* harus dihitung lebih dulu. *Term* adalah suatu kata atau suatu kumpulan kata yang merupakan ekspresi verbal dari suatu pengertian. Perhitungan bobot tersebut dilakukan melalui persamaan nomor (2.6).

$$\textit{Term weight}\ w_I = tf_i * \log \frac{D}{df_i} \quad (2.6)$$

Dimana Tfi adalah frekuensi *term* atau banyak *term* I yang ada pada sebuah dokumen n (*Term Frekuency*), Dfi adalah frekuensi dokumen atau banyak dokumen yang mengandung *term* I *(inverse Dokumen Frequency*) dan D adalah jumlah semua dokumen.

Setelah itu untuk mengetahui tingkat kemiripan antar dokumen nilai cosinus dari sudut antar *vector* dokumen dengan *vector query* dihitung melalui persamaan nomor (2.2).

$$\cos\theta_{Di} = \text{Sim}\ (Q\ D_i) \quad (2.7)$$

Dimana

$$sim\ (QD_i) = \frac{\sum j\ wq,j\ w\ i,j}{\sqrt{\sum i w^2_{i,j}} \sqrt{\sum j w^2_{i,j}}} \quad (2.8)$$

Sim (Q, Di) = nilai kesamaan antara sebuah dokumen I dengan query Q
$w_{Q,j}$ = bobot *term* j pada *query* Q

w_{ij} = bobot *term* j pada dokumen i

Hasil *cosinus* tersebut di urutkan dari nilai kesamaan yang terbesar ke nilai yang terkecil. Hasil terbesar memiliki kedekatan yang lebih baik dengan *user query* dibandingkan nilai kesamaan yang lebih kecil [18].

II. METODE PENELITIAN

A. Data

Data yang digunakan adalah data kuisioner yang didapat dari hasil kuesioner mahasiswa dan dikosultasikan dengan phisikolog untuk dibuat sebagai data set emosi dan dijadikan data pelatihan. Data set tersebut dari teks bahasa Indosesia yang telah diproses melalui pos-tagging; Chunking data set yaitu pemenggalan kalimat menjadi klausa; dan Tag set, kumpulan kelas kata yang digunakan dalam bahasa Indonesia.

B. Variabel dan parameter

Dalam pembobotan teks setelah melakukan pemengkalan kalimat, maka variabel - variabel yang digunakan adalah *term frekuensi* (Tf), yaitu banyaknya jumlah kata atau term (t) yang sama yang muncul dalam sebuah dokumen (d) DF - dokumen frekuensi yaitu jumlah dokumen yang memuat term(t) dalam seluruh dokumen (N) IDF: *invers* nilai dari DF.

Variable yang digunakan untuk mengklasifikasi dan pengujian emosi adalah *true positive* (TP) yang mewakili jumlah emosi positif dan hasilnya kelas positif. True Negative (TN) adalah mewakili jumlah emosi negative dan hasilnya kelas negative. *False positive* (FP) yang mewakili emosi negatif dan hasilnya positif sedangkan *False Negatif* (FN) yang mewakili emosi positif dan hasilnya kelas negative. Proses validasi dengan jumlah degmentasi pada proses *K-fold cros validation*.

Adapun kategori untuk parameter emosi yaitu Emosi senang: kenikmatan: bahagia, gembira, ringan, puas, riang, senang, terhibur, bangga, kenikmatan indrawi, takjub, rasa terpesona, rasa terpenuhi, kegirangan luar biasa, dan senang sekali [10]. Kecewa: patah hati, haru biru, kecil hati, putus asa, bersedih hati, merasa tidak berdaya, menyedihkan dan penuh penderitaan [19]. Takut- cemas, takut, gugup, khawatir, waswas, perasaan takut sekali, khawatir, waspada, sedih tidak tenang; ngeri takut sekali, dan kecut [10].

Dari proses parameter yang digunakan dalam pengujian sistem yaitu:

- *Accuracy* : rasio antara jumlah data yang terklasifikasi dengan benar dengan data seluruhnya

$$Accuracy = \frac{TP+TN}{TP++TN+FP+FN} \quad (8)$$

- *Precission* : rasio banyaknya data yang diklasifikasikan dengan benar dibagi dengan jumlah total data yang berhasil diklasifikasikan dalam satu kelas kata emosi yang sama

$$Precision,\ p = \frac{TP}{TP+FP} \quad (9)$$

- *Recall* : rasio jumlah data yang diklasifikasikan dengan benar dibagi dengan jumlah total data dalam kelas sebenarnya

$$Recall,\ r = \frac{TP}{TP+FN} \quad (10)$$

- F- meansure : nila rata-rata harmonik dari Precission dan recall

$$F\text{-}Measure = 2\frac{precision\,.\,recall}{precision+recall} \quad (11)$$

ROC: penggambaran hasil klasifikasi dalam bentuk dua dimensi maka parameter pengujian sistem tersebut dapat dengan persamaan berikut [20].

C. Deteksi Frasa

Untuk memudahkan proses pemahaman bahasa manusia ke dalam bahasa komputer adalah melakukan *preprosesing* yaitu proses *case folding, tokenizing, filtering dan stemming*. Proses *tokenizing* yang memecah kalimat menjadi kata tunggal membuat komputer menganggap semua kata sama. Maka perlu ada frasa yang dibentuk dari salah satu kata yang tidak penting yang masuk dalam daftar stoplist dan melakukan penggabungan dua kata. Dengan hal itu diperlukan sebuah aturan yang berisi kombinasi kelas kata yang berpeluang besar dalam membentuk frasa. Proses tokenisasi berbasis deteksi frasa menggunakan Hidden Markov Model (HMM) POS-Tagger untuk memudahkan klasifikasi menggunakan metode VSM. Label yang digunakan berdasarkan tag set data untuk memudahkan komputer mengenali ciri – ciri frase dalam kombinasi label katanya [21]. Hasil pelabelan kata seperti pada tabel I.

TABEL I
HASIL PELABELAN KATA

No.	Kalimat	Hasil
1.	dosennya bergaya santai sehingga materi mudah dicerna	dosennya/RB bergaya/VBI santai/NN sehingga/SC materi/NN mudah/JJ dicerna/VBI
2.	galak pemarah pokoknya menakutkan sekali	galak/NN pemarah/NN pokoknya/RB menakutkan/JJ sekali/RB
3.	kalau terlambat masuk pasti kena hukuman jadi malas kuliah	kalau/SC terlambat/VBI masuk/VBT pasti/JJ kena/VBT hukuman/NN jadi/JJ malas/JJ kuliah/NN
4.	dosennya galak sehingga takut untuk bertanya	dosennya/VBT galak/NN sehingga/SC takut/VBT untuk/IN bertanya/VBT
5.	jam perkuliahan sering kosong rem blong	jam/NN perkuliahan/NN sering/JJ kosong/JJ rem/NN blong/JJ

Hasil dari proses pelabelan tersebut, maka setiap data kata beserta labelnya diubah kedalam bentuk array berpasangan yang akan dicocokkan dengan data pelatihan kata-kata yang mempunyai pola frasa, baik itu frasa verbal (kata kerja), frasa nominal (kata benda), frasa ajektiva (kata sifat)

III. HASIL DAN PEMBAHASAN

A. Pengelompokan sumber Data

Tahapan yang paling awal dilakukan dalam proses deteksi emosi mahasiswa adalah pengelompokan sumber data, yaitu dengan cara mengelompokkan kedalam tabel yang meliputi beberapa kategori emosi senang, takut dan kecewa. Sumber data tersebut didapat dari kuesioner mahasiswa berupa komentar yang telah dikonsultasikan dengan psikolog untuk menentukan kelas dari masing-masing komentar. Contoh pengelompokan data tersebut seperti terlihat pada tabel II.

TABEL II
PENGELOMPOKAN SUMBER DATA

No.	Isi Komentar	Kategori
1.	Dosen ini menyenangkan dan enak diajak berdiskusi	Senang
2.	Materi perkuliahan dirancang sesuai kebutuhan kerja	Senang
3.	Pembahasan materi sangat santai dan diselingi guyon	Senang
4.	Cara mengajarnya sangat mudah dicerna	Senang
5.	Dosennya disiplin dan tepat waktu	Senang
6.	Suara dosennya sangat pelan sekali	Kecewa
7.	Penjelasannya sangat membosankan	Kecewa
8.	Jam perkuliahan sering kosong	Kecewa
9.	Dosennya sering bolos	Kecewa
10.	Setiap pertemuan selalu ada tugas	Kecewa
11.	Dosennya galak sekali, sehingga takut untuk bertanya	Takut
12.	Wajah dosennya saat mengajar sangat menyeramkan	Takut
13.	Galak, pemarah, pokoknya menakutkan sekali	Takut
14.	Kalau terlambat masuk pasti kena hukuman, jadi malas kuliah nih!	Takut
15.	Pas lagi marah, suaranya sangat menggelegar seperti petir menyambar	Takut

B. *Pelatihan pola Frasa*

Proses pelatihan menggunakan 200 kata frase, digunakan untuk membentuk pola pengetahuan dengan memanfaatkan label kata. Hasil dari label kata frase tersebut dijadikan pembanding dengan data array berpasangan, sehingga kata-kata yang membentuk satu entitas seperti pola dari kata frase akan digabung menjadi satu kesatuan. Hasil dari pelatihan pola kata frase seperti pada table III.

TABEL III.
HASIL PELATIHAN KATA FRASE

No.	Input	Hasil
1.	Sepatu kaca	NN/NN
2.	Bunga desa	NN/NN
3.	Rumah besar	NN/JJ
4.	Sedang berbelanja	RB/VBT
5.	Bocah ingusan	NN/VBI
6.	Tas besar	NN/JJ
7.	Sedang menonton	RB/NN
8.	Banting tulang	VBT/NN
9.	Sedang bergandengan	RB/VBI
10.	Tua muda	JJ/JJ

Berdasarkan Tabel III. Hasil dari kata yang sudah terbentuk dalam array dan telah diketahui mempunyai pola kata frase atau tidak, maka proses selanjutnya adalah membuat kata-kata yang memiliki frekuensi lebih dari 1(satu) menjadi kata yang unik berfungsi untuk pengindeks kata. Jika terdapat kata dosen lebih dari 1(satu) maka yang akan diproses dalam perhitungan TF-IDF hanya satu kata saja. Hasil perhitungan TF-IDF seperti pada Gambar 1.

Gambar 1. Hasil perhitungan TF-IDF

Setelah melalui proses tokenizing maka berhasil melakukan pendeteksian pola frasa dengan aturan yang berisi kelas untuk membentuk frasa. Dengan aturan tersebut maka komputer mampu mendeteksi ciri- ciri dari 2 (dua) gabungan kata yang membentuk sebuah frasa seperti Gbr 1, akan tetapi dalam penemuan kata/ term untuk menghitung bobot Tf semakin mengecil atau bernilai 0 (nol). Karena penggabungan 2 kata menjadi kata tunggal.

C. *Pengujian deteksi emosi*

Pengujian pertama yaitu dilakukan dengan memberikan 50 query berupa kalimat. Kalimat tersebut mengambil dari sumber data (Corpus) yang telah tersimpan pada database baik itu kalimat senang, takut ataupun kecewa, dengan tujuan untuk memastikan bahwa metode VSM telah terimplementasi dengan baik pada program. Contoh Hasil pengujian seperti pada tabel IV.

TABEL IV
HASIL PENGUJIAN

No	Kalimat	Tabel	Cos (Senang)	Cos (Takut)	Cos (Kecewa)	Deteksi	Keterangan
1.	Soalnya lebih sulit dari latihannya	Kecewa	0.003870	0	0.014706	Kecewa	Sesuai
2.	jarang masuk tapi tugas banyak	Kecewa	0.002919	0	0.025985	Kecewa	Sesuai
3.	dosennya sangat susah memberi nilai	Kecewa	0	0.001261	0.097708	Kecewa	Sesuai
4.	dosennya pasang wajah cemberut terus	Kecewa	0	0	0.034504	Kecewa	Sesuai

5.	dosennya membuat mahasiswa kecewa,tambah bodoh, pasif	Kecewa	0	0	0.069745	Kecewa	Sesuai
6.	Dosen ini menyenangkan dan sering berdiskusi	Senang	0.042427	0	0	Senang	Sesuai
7.	materi kuliah dirancang secara sistematis	Senang	0.037429	0	0.002278	Senang	Sesuai
8.	pembahasannya ringkas dan santai	Senang	0.036182	0	0	Senang	Sesuai
9.	bahasa yang digunakan sangat mudah dipahami	Senang	0.070205	0	0	Senang	Sesuai
10.	Dosennya kalau menjelaskan gampang dan mudah dicerna	Senang	0.067179	0	0	Senang	Sesuai
11.	wajahnya membikin yang melihat jadi kecil hati	Takut	0	0.028228	0.001726	Takut	Sesuai
12.	Suara pak dosen membikin kita jadi bergidik	Takut	0	0.022438	0	Takut	Sesuai
13.	jangan sering memarahi kami pak	Takut	0.001754	0.020823	0.003345	Takut	Sesuai
14.	banyak tersenyum bu, jangan pasang wajah galak, seperti mau menerkam mahasiswa	Takut	0	0.048424	0.013116	Takut	Sesuai
15.	kalau bapak mengajar seperti itu terus, maka kami sebagai mahasiswa akan merasa mencekam, tertekan tidak dapat berkreasi	Takut	0.001440	0.090052	0.005490	Takut	Sesuai

Kesimpulan dari hasil pengujian 1 (satu) adalah sistem telah mengimplementasikan metode VSM dengan benar, terbukti dengan menguji menggunakan data corpus berhasil mendeteksi emosi sebesar 100%.

Pengujian kedua yang dilakukan adalah pengujian bertingkat dimana data dari database atau corpus akan dibagi menjadi 2 data yaitu data latih dan data uji. Pembagian data tersebut berdasarkan prosentase terhadap jumlah semua keseluruhan data. Jumlah masing-masing data sebanyak 88 data baik itu untuk emosi senang, emosi takut maupun emosi kecewa sehingga total jumlah data adalah 264. Pembagian datanya adalah 90% data latih dan 10% data uji, 75% data latih dan 25% data uji, 50% data latih dan 50% data uji.

TABEL V.
SKENARIO PENGUJIAN DATA

Skenario	Data Training			Jumlah	Data Training			Jumlah
	Senang	Takut	Kecewa		Senang	Takut	Kecewa	
1	79	79	79	237	9	9	9	27
2	66	66	66	198	22	22	22	66
3	44	44	44	132	44	44	44	132

TABEL VI.
HASIL PENGUJIAN DETEKSI EMOSI MENGGUNAKAN METODE VSM

Deteksi Emosi	Jlh Data	∑ Senang	%	∑ Takut	%	∑ Kecewa	%	∑ Tidak terdeteksi	%
Senang	27	9	100	0	0	0	0	0	0
Takut		0	0	9	100	0	0	0	0
Kecewa		1	11	1	11	7	78	0	0
Senang	66	21	95	0	0	1	4,5	0	0
Takut		0	0	21	95	1	4,5	0	0
Kecewa		2	9,1	3	14	17	77	0	0
Senang	132	21	48	7	16	7	16	9	20
Takut		9	20	14	32	15	34	6	14
Kecewa		8	18	5	11	21	48	10	23

TABEL VI.
HASIL PENGUJIAN DETEKSI EMOSI MENGGUNAKAN METODE VSM

Deteksi Emosi	Jlh Data	∑ Senang	%	∑ Takut	%	∑ Kecewa	%	∑ Tidak terdeteksi	%
Senang	27	9	100	0	0	0	0	0	9
Takut		0	0	9	100	0	0	0	0
Kecewa		0	0	1	11	8	89	0	0
Senang	66	22	100	0	0	0	0	0	22
Takut		1	4,5	20	91	1	4,5	0	1
Kecewa		0	0	1	4,5	21	95	0	0
Senang	132	44	100	0	0	0	0	0	44
Takut		1	2,3	43	98	0	0	1	2,3
Kecewa		0	0	1	2,3	43	98	0	0

Dari hasil deteksi emosi Senang, Takut dan kecewa terdapat perbedaan yang sangat jauh. Terlihat pada data uji 132 data dengan menggunakan tanpa frasa hampir rata – rata 98%. Selisih yang sangat jauh antara menggunakan Frasa dengan tanpa Frasa karena pada pendeteksi frasa tersebut melakukan penggabungan kata, sehingga menemukan kata yang di deteksi tidak sesuai pada Frasa yang mana melakukan penggabungan 2 kata dan sekaligus juga menjadi kelemahan Frasa. Kelemahan dari non frasa adalah apabila ada penggabungan kata negative dan positif seperti "tidak baik" maka hasil dari non-frasa akan menghasilkan emosi senang.

D. Nilai Prescision, Recall, F-measure, ROC pada deteksi Emosi dengan Frasa

Pengujian emosi penentuan nilai Prescision, Recall, F-measure, ROC yang sudah dilakukan dimana k=10, hasil pungujian dengan data uji yang berbeda - beda seperti terlihat pada tabel VII.

TABEL VII
NILAI PRESCISION, RECALL, F-MEASURE, ROC PADA DETEKSI EMOSI

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Data	Accuracy
0.889	0.889	0.333	0.889	0.485	0.401	27	70,37%
0.857	0.911	0.305	0.857	0.450	0.452	66	69, 70%
0.123	0.211	0.103	0.137	0.057	0.493	132	48,20%

Dari hasil pengujian tabel VII, terlihat proses pengklasifikasin, terjadi penurunan tingkat accuracy karena semakin kecil data set yang dibuat sebagai data training maka nilai accuracynya semakin rendah.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Adapun kesimpulan dari hasil penelitian ini adalah:

a. Optimasi hasil dari deteksi emosi yang dibuktikan melalui uji coba dengan 2 (dua) tahap yaitu mendeteksi dengan Frasa dan Non-Frasa. Menggunakan metode (TF-IDF) dan *Vector Space Model* (VSM) mampu mendeteksi dokumen emosi teks Bahasa Indonesia dengan data set 90%, hasil deteksi yang diperoleh dengan frasa dan non frasa berhasil dideteksi 100% untuk emosi senang dan Takut. Berbeda dengan emosi kecewa menggunakan frasa 78% sedangkan non frasa 89%, perbandingannya sebesar 11%.

b. Untuk mengukur tingkat keberhasilan dalam penerapan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan VSM (*Vector Space Model*) dalam klasifikasi emosi sangat baik akan tetapi hasil emosi dengan pendeteksian menggunakan Frase masih kurang sempurna. Apabila data set yang digunakan kecil maka hasil pendeteksian emosi sangat buruk, yang hasilnya berbanding terbalik dengan non Frase rata – rata 98% keberhasilannya.

c. Komputer mampu mendeteksi emosi dengan data corpus dengan melakukan pemenggalan kalimat menjadi kata/frase menggunakan Chunk yang dapat mengklasifikasikan emosi senang, kecewa dan Takut. Dari hasil uji coba untuk data *training* 90 % dan data uji 10 % dalam mendeteksi emosi dengan menggunakan Frasa mendapat hasil 92,59%.

B. Saran

Saran yang mungkin perlu dilakukan dalam pengembangan pengembangan penelitian emosi ini adalah ungkapan dalam bahasa Indonesia yang peneliti analisis masih terbatas, penelitian selanjutya diharapkan dapat menganalisis dengan payung ilmu semantik lainnya khususnya semantik *generative* dan semantik kognitif. Penelitian ini bisa dikaji lebih dalam, baik dalam sintaksis dan morfologi bahasa Indonesia.

DAFTAR PUSTAKA

[1] Hirat, R., & Mittal, N. (2015). A Survey On Emotion Detection Techniques using Text in Blogposts. International Bulletin of Mathematical Research Vol 2, Issue 1: pp. 180-187.Andriani, M. (2008). Information Retrieval, Modul Kuliah PemrosesanTeks Fakultas Ilmu Komputer UI semester ganjil 2009.

[2] Sugiono. (2005). Metode Penelitian Administrasi.Bandung. Alfabeta.

[3] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S.M.M., & Williams, H.E. (2007). Stemming Indonesian : A Confix-Stripping Approach. Transaction on Asian Langeage Information Processing. Association for Computing Machinery . New York: Vol. 6, No. 4, Articel 13.

[4] Aman, S. & Szpakowicz, S., 2007. Identifying Expressions of Emotion in Text. In Text, Speech and Dialogue, Lencture Notes in Artificial Intelligence Vol 4629, pp. 196-205.

[5] Ghazi, D., Inkpen, D. & Szpakowicz, S., 2010. Hierarchical approach to emotion recognition and classification in texts. Advances in Artificial Intelligence LNCS Vol 6085, pp. 40-50.

[6] Ghazi, D., Inkpen, D. & Szpakowicz, S., 2014. Prior and contextual emotion of words in sentential context. Computer Speech and Language 28, pp. 76-92.

[7] Mohammad, S.M., 2012a. From once upon a time to happily even after: Tracking emotions in mail and books. Decision Support Systems 53, pp. 730-741.

[8] Bata, J. (2015).Leksikon untuk deteksi emosi dari teks bahasa indonesia Seminar Nasional Informatika 2015 (semnasIF 2015) UPN "Veteran" Yogyakarta, 14 November 2015.ISSN:1979-2328: Pages 195-202.

[9] Hemalatha, I., Varma, P.G., dan Govardhan, A., 2012, Preprocessing the Informal Text for Efficient Sentiment Analysis, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol. 1, July – August 2012, ISSN 2278-6856.

[10] Goleman, D. (2007). Kecerdasan Emosional. Jakarta: PT. Gramedia Pustakan Utama.

[11] Salton, G. (1989). Automatic Text Processing, The Transformation, Analysis, and Retrieval of information by computer. Addison – Wesly Publishing Company, Inc. USA.

[12] Baeza, R.Y. & Neto, R. (1999). Modern Information Retrieval. Addison Wesley-Pearson international edition, Boston. USA.

[13] Tala, F. Z. (2003). A Study of Stemming Efects on Information Retrieval in bahasa Indonesia. Institute for logic, Language ang Computation Universiteit van Amsterdam the Netherlands. https://www.illc.uva.nl/Research/Publications/Reports/MoL-2003-02.text.pdf. (diakses tanggal 20 Oktober 2015).

[14] Mandala, R. (2004). Bahan kuliah sistem temu balik informasi. Institut Teknologi Bandung. Departemen teknik informatika .

[15] Robertson, S. E. (2004). Understanding Inverse Document Frequency: On Theoretical Arguments for IDF. Reprinted from Journal of Documentation 60: 503-520. URL: http://www.staff.city.ac.uk/~sb317/idfpapers/Robertson_idf_JDoc.pdf. (diakses tanggal 20 Oktober 2015).

[16] Turney, P.D. & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. Journal of Artificial Intelligence Research. 37: 141-188.

[17] Ning, L. et al. (2004). Learning Similarity Measures in Non-orthogonal Space. CIKM'04, Washington D.C., U.S.A.

[18] Garcia, E. (2012). The Classic Vector Space Model. Retrieved URL:http://www.miislita.com/term- vector/term-vector-3.html. (diakses tanggal 15 November 2016).

[19] Ekman, P. (1992). An Argument for Basic Emotion.

[20] A. A. Armana, A. B. Putra, A. Purwarianti dan Kuspriyanto, "Syntactic Phrase Chunking for Indonesian Language", Science Direct, pp. 635-640, 2013.

[21] Purwarianti, A. & Wicaksono, A. F. (2010). HMM Based Part-Of-Speech Tagger for Bahasa Indonesia. On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop.