

USULAN TUGAS AKHIR

1. IDENTITAS PENGUSUL

NAMA : Hari Setiawan
NRP : 05111440000156
DOSEN WALI : Darlis Herumurti, S.Kom., M.Sc
DOSEN PEMBIMBING : 1. Arya Yudhi Wijaya, S.Kom., M.Kom
2. Rully Soelaiman, Ir., M.Kom

2. JUDUL TUGAS AKHIR

“Pengenalan Emosi pada Teks Menggunakan Metode Ensemble”

3. LATAR BELAKANG

Kognisi dan emosi manusia adalah bawaan sejak lahir dan sangat berarti dalam aspek alami manusia. Penelitian dalam bidang Kecerdasan Buatan mencoba untuk mengeksplorasi dan mendapatkan pemahaman lebih baik dari mekanisme di balik kebiasaan manusia yang bertujuan untuk memberikan sistem dan aplikasi komputer kemampuan untuk mengenali aspek – aspek alami dari manusia, seperti emosi. Emosi merupakan faktor utama kecerdasan manusia yang memberikan karakteristik indikatif perilaku manusia, mewarnai cara komunikasi manusia dan dapat memainkan peran penting dalam interaksi komputer dan manusia. Peran dari emosi pertama kali diteliti oleh Picard, yang memperkenalkan konsep komputasi afektif [1], menunjukkan pentingnya emosi dalam interaksi komputer manusia dan menggambarkan arah untuk penelitian interdisipliner dari berbagai bidang, seperti ilmu komputer, ilmu kognitif, dan psikologi. Tujuan komputasi afektif adalah untuk memungkinkan komputer mengenali status emosi dan perilaku manusia dan menjembatani kesenjangan antara emosi manusia dan komputer

dengan mengembangkan sistem dan aplikasi yang bisa menganalisa, mengenali, dan beradaptasi sesuai dengan keadaan emosi pengguna [2]

Emosi manusia dapat diekspresikan melalui berbagai media, seperti ucapan, ekspresi wajah, gerak tubuh, dan data tekstual. Cara yang paling umum bagi orang - orang untuk berkomunikasi dengan orang lain dan dengan sistem komputer adalah melalui teks, yang merupakan bentuk komunikasi utama baik itu dalam web ataupun sosial media. Selama beberapa tahun terakhir kemunculan web dan maraknya sosial media telah benar – benar mengubah cara komunikasi manusia karena mereka menyediakan sarana baru yang menghubungkan semua orang di seluruh dunia dengan informasi, acara, dan berita secara *real time*. Selain itu, mereka telah benar – benar mengubah peran pengguna; mereka telah mengubah peran pengguna dari pencari informasi pasif sederhana dan konsumen untuk produsen aktif [3]. Setiap hari, sejumlah besar artikel dan pesan teks ditulis di berbagai situs, blog, portal berita, toko online, jaringan sosial dan forum. Jumlah konten tekstual pada web memerlukan metode otomatis untuk menganalisa dan mengekstrak pengetahuan [4].

Menganalisa konten web dan pesan tekstual masyarakat dengan tujuan untuk menentukan status emosi mereka adalah topik yang sangat menarik dan menantang di area *microblogging* [5]. Aliran data tekstual yang besar dan terus menerus di web dapat mencerminkan perasaan, pendapat, dan pemikiran para penulis di berbagai fenomena mulai dari peristiwa politik di seluruh dunia. Hal itu bisa menyampaikan status emosi seseorang dan informasi substansial tentang keyakinan dan sikap mereka (Qiu et al, 2012). Analisis data tekstual diperlukan untuk lebih dapat lebih dalam memahami status dan perilaku emosional seseorang, dan dalam garis ini dapat memberikan faktor yang sangat indikatif terhadap sikap masyarakat menuju berbagai topik dan juga dapat menggambarkan status emosional dari sebuah komunitas, kota, atau bahkan negara. Dari sudut pandang manusia sendiri, menganalisis pesan teks dari orang yang spesifik dapat memberikan faktor yang sangat indikatif mengenai situasi emosional orang tersebut, perilaku orang tersebut, dan juga memberikan petunjuk lebih dalam untuk menentukan kepribadiannya. Selanjutnya, sehubungan dengan berita, artikel dan komentar dari orang – orang, dengan topik yang berkaitan, analisis dari komentar – komentar orang pada sebuah topik spesifik dapat memberikan informasi yang sangat bermanfaat tentang sikap publik, perasaan dan perilakunya mengenai berbagai topik dan acara. Pada hal ini, model emosi dapat digunakan untuk mengetahui apa yang orang rasakan mengenai sebuah entitas yang diberikan seperti film, ataupun acara tertentu.

Namun, pengembangan dari sistem dan aplikasi untuk menganalisis bahasa alami secara otomatis dengan tujuan untuk mengetahui isi sentimentalnya adalah proses yang sangat sulit. Beberapa penelitian telah menunjukkan bahwa menganalisis dan mengetahui emosi dari dokumen teks sangatlah kompleks dan merupakan permasalahan pada bidang *NLP (natural language processing)* secara keseluruhan.

Dalam paper ini, kami menyajikan sebuah sistem *ensemble classifier* untuk sentimen analisis dari data tekstual. Skema *ensemble* tersebut bertujuan untuk secara efektif menggabungkan beberapa macam algoritma klasifikasi dan pembelajaran untuk mengatasi kekurangan dari setiap algoritma. Sistem ini terdiri atas tiga algoritma pembelajaran utama, dua diantaranya adalah algoritma statistik dan sisanya menggunakan *knowledge based classifier tool*. Ketiga algoritma tersebut digabungkan

menggunakan sistem voting. Algoritma statistik yang digunakan adalah *naive Bayes* dan *maximum entropy* dimana dilatih pada data *ISEAR (Internasional Survei on Emotion Antecedents and Reaction)* [6] *dataset*. Sedangkan *Knowledge based tool* menganalisis struktur kalimat menggunakan peralatan seperti *Stanford parser* untuk mengetahui kata yang dibutuhkan dan menggunakan *WordNet Affect*, sumber leksikal untuk mengetahui kata yang mewakili emosi – emosi yang dicari. Kemudian dari kata – kata tersebut dapat diketahui kekuatan emosional setiap kata dan menentukan status emosional kalimat tersebut berdasarkan grafik ketergantungan kalimat itu dalam pendekatan dimana keseluruhan keadaan emosi suatu kalimat didapatkan dari kedekatan bagian – bagian emosional dari kalimat tersebut. Sistem klasifikasi ini melakukan analisis sentimen pada tingkat kalimat, oleh karena itu kalimat yang baru pada awalnya akan dipisah menjadi beberapa kalimat dan kemudian baru dimasukkan ke dalam sistem klasifikasi ini, dimana kita akan mengekstraksi fitur – fiturnya, yang direpresentasikan dalam bentuk *bag-of-words*, dan kemudian dikirim kepada algoritma klasifikasi statistik. Sistem klasifikasi *ensemble* ini menentukan apakah sebuah teks mengandung makna emosional atau netral, dan dalam kasus jika dia emosional maka akan ditentukan emosi apa yang terdapat didalamnya.

Hasil tugas akhir ini diharapkan dapat memberikan metode yang lebih baik untuk deteksi emosi pada teks sehingga dapat menyelesaikan permasalahan di atas dengan optimal dan diharapkan dapat memberikan kontribusi pada perkembangan ilmu pengetahuan dan teknologi informasi.

4. RUMUSAN MASALAH

Rumusan masalah pada tugas akhir ini adalah sebagai berikut :

1. Bagaimana melakukan deteksi emosi menggunakan metode *ensemble*?
2. Bagaimana melakukan deteksi emosi menggunakan metode *ensemble* untuk mengurangi jumlah emosi yang salah terdeteksi?
3. Bagaimana menghitung akurasi deteksi emosi menggunakan metode klasifikasi *ensemble*?

5. BATASAN MASALAH

Batasan masalah pada tugas akhir ini adalah sebagai berikut :

1. Metode untuk deteksi emosi yang digunakan dengan menggunakan metode *ensemble* yang menggabungkan algoritma *naive bayes*, *max entropy*, dan *knowledge based tool*.
2. Percobaan akan dilakukan menggunakan Python.
3. Dataset yang digunakan untuk percobaan adalah dataset yang tersedia untuk umum (publik) yaitu dataset yang diambil dari International Survei on Emotion Antecedents and Reaction dataset.

6. TUJUAN PEMBUATAN TUGAS AKHIR

Tujuan dari pembuatan tugas akhir ini yaitu untuk mengetahui akurasi pada metode *ensemble* yang menggabungkan algoritma *naive bayes*, *max entropy*, dan *knowledge based tool*.

7. MANFAAT TUGAS AKHIR

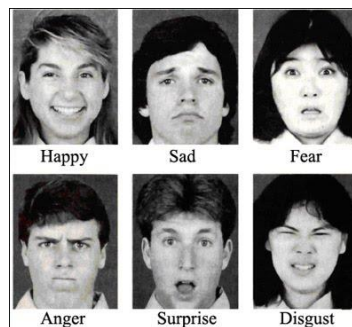
Tugas akhir ini diharapkan dapat membantu memberikan metode yang lebih baik dari beberapa metode deteksi emosi pada teks yang telah ada.

8. TINJAUAN PUSTAKA

a. Model Emosi

Emosi adalah sebuah perasaan kuat yang didapatkan dari sebuah situasi, suasana hati atau hubungan dengan orang lain [7]. Sebuah cara bagaimana emosi dapat direpresentasikan adalah sebuah aspek dasar dari sistem pengenalan emosi [8]. Model yang paling populer untuk merepresentasikan emosi adalah model kategorikal dan dimensional. Model kategorikal mengasumsikan bahwa terdapat beberapa emosi diskrit dan dasar dengan jumlah yang terbatas, dimana setiap emosi tersebut melayani tujuan tertentu. Disisi lain, model dimensional memiliki cara yang berbeda dan merepresentasikan emosi dalam sebuah pendekatan dimensional.

Model kategorikal yang paling populer dan banyak digunakan adalah model emosi Ekman [9], dimana membagi emosi dasar manusia menjadi: marah, menjijikkan, ketakutan, kebahagiaan, kesedihan, dan terkejut. Emosi – emosi ini dikarakteristikan sebagai *universal*, sebagaimana mereka diekspresikan dalam cara yang sama tanpa memandang perbedaan budaya dan era.



Gambar 1 Model Emosi Ekman

b. *Sentiment Analysis*

Sentiment analysis adalah penggunaan *natural language processing*, *text analysis*, *computational linguistics*, dan biometrics untuk secara sistematis mengidentifikasi,

mengeksrak, mengkuantitaskan, dan mempelajari keadaan afektif dan informasi subjektif. *Sentiment analysis* sangat banyak digunakan pada materi – materi yang berhubungan dengan ulasan dan respon survei, sosial media, dan perawatan kesehatan.

Secara umum, *Sentiment analysis* bertujuan untuk menentukan sikap narasumber, penulis, atau subjek lain sesuai dengan topik yang berhubungan atau reaksi emosional terhadap sebuah dokumen, interaksi, atau acara. Sikap tersebut dapat berupa anggapan atau evaluasi, keadaan afektif, atau komunikasi emosional yang diharapkan.

c. Metode *Ensemble*

Metode *ensemble* adalah sebuah cara untuk menggunakan lebih dari satu algoritma pembelajaran untuk memperoleh performa prediktif yang lebih baik dari pada satu algoritma apapun. Tujuan dari *ensemble* adalah untuk mengambil keuntungan dari kelebihan salah satu algoritma dan meminimalkan dampaknya.

Ada banyak alasan untuk mendesain, mengembangkan, dan menggunakan sistem klasifikasi *ensembles* seperti disebutkan oleh *Dietterich* [10]. Dari pandangan statistik, dengan membangun skema *ensemble* dari *classifier* yang telah dilatih, algoritma tersebut dapat mengambil rata – rata dari vote mereka dan mengurangi risiko kesalahan. Meskipun menggunakan *classifier* yang berbeda dan menghasilkan performa yang baik, ketika hanya satu yang dipilih, hasilnya tidak akan memiliki kemampuan generalisasi yang baik untuk data yang baru.

d. Naive Bayes classifier

Dalam *machine learning*, *naive Bayes* classifier adalah keluarga dari sistem klasifikasi probabilistik sederhana berdasarkan penerapan teori *Bayes* dengan asumsi independen yang kuat antar fitur. *Naive Bayes* adalah teknik sederhana untuk membangun sebuah *classifier*, dimana menetapkan label kelas pada tiap data permasalahan, direpresentasikan sebagai vektor berisi nilai fitur, dimana label – label kelasnya diambil dari beberapa kumpulan data terbatas.

Untuk beberapa tipe dari model probabilitas, *naive Bayes classifiers* dapat dilatih dengan sangat efisien pada pengaturan *supervised learning*. Dalam banyak implementasi, estimasi parameter untuk *naive Bayes* model menggunakan metode *maximum likelihood*; dengan kata lain, kita bisa menggunakan *naive Bayes* tanpa menerima probabilitas Bayesian. Meskipun desain dari sistem klasifikasi ini tampaknya terlalu sederhana, *naive Bayes* classifier telah bekerja cukup baik pada banyak situasi kompleks dunia nyata. Sebuah keuntungan dari *naive Bayes* adalah kita hanya membutuhkan sejumlah kecil data latihan untuk dapat memprediksi dengan baik.

Secara abstrak, *naive Bayes* adalah sebuah model probabilitas kondisional: dimana diberikan sebuah vektor $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_n$ yang merepresentasikan beberapa n fitur (variabel independen), kita tetapkan itu kedalam instansi probabilitas ini

$$p(C_k | \mathbf{x}_1, \dots, \mathbf{x}_n)$$

Permasalahan dengan formulasi diatas adalah jika jumlah dari fitur n besar atau jika sebuah fitur dapat memiliki nilai yang besar. Oleh karena itu, formula probabilitas kondisional diatas dapat didekomposisi menjadi

$$p(C_k | x) = \frac{p(C_k)p(x | C_k)}{p(x)}$$

e. **Maximum Entropy classifier**

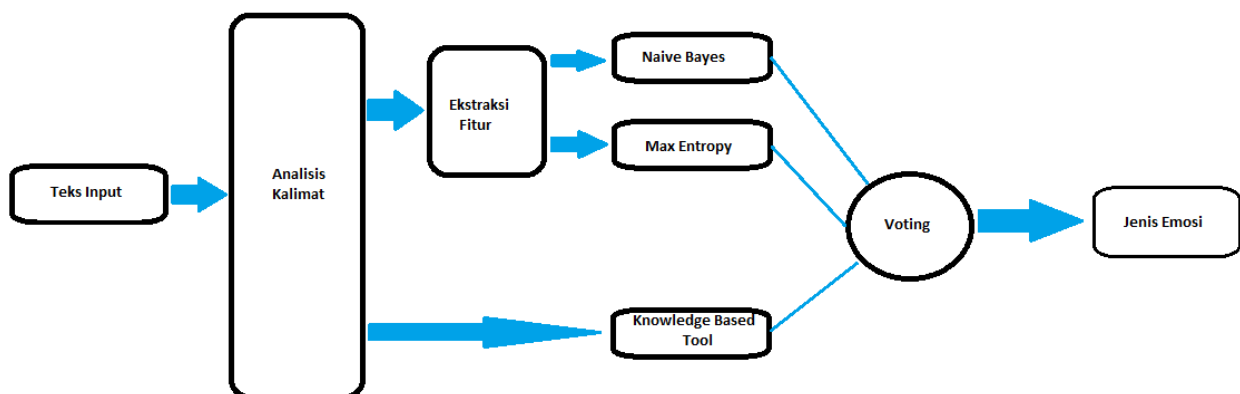
Dalam statistika, *maximum entropy* atau yang disebut juga *multinomial logistic regression* adalah sebuah metode klasifikasi yang menggeneralisasi *logistic regression* pada permasalahan multiclass, dimana memiliki hasil lebih dari dua kelas.

Multinomial logistic regression digunakan jika variabel dependen di dalam permasalahan adalah nilai nominal (kategorikal, atau memiliki nilai – nilai yang telah didefinisikan) dan dimana terdapat lebih dari dua kategori.

9. RINGKASAN ISI TUGAS AKHIR

Tujuan dari tugas akhir ini adalah mengetahui performa dari penggunaan metode *ensemble* untuk menggabungkan metode *Naive Bayes*, *Max entropy*, dan *knowledge based tool* untuk deteksi emosi.

Beberapa tahap dilakukan untuk mendapatkan hasil deteksi eksudat. Alur proses ditunjukkan pada gambar 2. Berikut tahap-tahap pada metode yang diusulkan:



Gambar 2. Proses Deteksi Emosi

a. Representasi fitur

Sebuah cara bagaimana sebuah dokumen dapat dianalisis dan bagaimana dia direpresentasikan adalah hal yang penting dalam performa sebuah algoritma *machine learning*. Pada sistem ini, untuk representasi teks *natural language*, kami menggunakan teknik representasi *bag-of-words* (BOW). BOW lebih sering digunakan karena

kemudahannya dalam proses klasifikasi. Teknik ini banyak digunakan dalam implementasi *text mining* yang dikombinasikan dengan penghapusan *stop-words* dan pemotongan dari kata yang berguna. Dalam pendekatan ini, sebuah dokumen dianggap sebagai sebuah kumpulan kata yang tidak urut, dimana sebuah posisi dari kata – kata dalam dokumen tidak penting. Dalam sistem ini, sebuah kalimat baru awalnya ditokenisasi dan dipecah menjadi beberapa kata dan kemudian setiap kata melalui proses *lemmatization*, dimana *lemmatization* adalah pemotongan kata dalam bahasa tertentu menjadi bentuk dasar pengenalan fungsi setiap kata dalam kalimat. Kemudian pada akhirnya fitur dari kalimat tersebut dimasukkan kedalam *base learner*. Contoh dari kalimat input yang digunakan ada pada Tabel 1.

Kalimat Input	Jenis Emosi
One day my room-mate started imitating me, doing what I had done áa few hours back. I got very angry and started beating him up	Kemarahan
Before knowing my marks I was not sure that I wanted to go to á that clas	Ketakutan
When fell in love with my present girlfriend.	Kebahagiaan
A child, 9 years old, died in a stupid way. He fell from a place á which was 1 metre high	Kesedihan

Tabel 1. Contoh Kalimat Input

b. Pemungutan suara *ensemble classifier*

Ensemble classifier disini menyesuaikan pendekatan pemungutan suara pada umumnya untuk membuat keputusan klasifikasi berdasarkan hasil dari setiap *classifier* dasar. Setiap *classifier* memiliki suara bahwa, untuk setiap kata dalam kalimat termasuk dalam kelas yang ditentukan oleh *classifier*. Sistem pemungutan suara dianggap sebagai metode yang paling sederhana dan intuitif untuk menggabungkan hasil dari beberapa *classifier* [11]. Pada umumnya, pemungutan suara menghitung suara untuk setiap kelas dari tiap input dan memilih kelas yang paling dominan. Gagasan untuk memilih beberapa *classifier* untuk membuat sebuah *ensemble* daripada menggunakan semua *classifier* telah ditangani dengan cara yang berbeda. Secara teoretis, jika sebuah *classifier* dasar dapat membuat kesalahan yang independen, dapat dibuktikan bahwa pemungutan suara lebih cocok dan dapat mengalahkan performa dari sebuah *classifier* saja.

c. *Naïve Bayes classifier*

Naïve Bayes adalah sebuah model sederhana untuk klasifikasi dan dapat mencapai performa yang bagus dalam kategorisasi teks. Hal ini berdasarkan teori *Bayes* dan merupakan sebuah pendekatan klasifikasi berdasarkan probabilitas yang menganggap bahwa kumpulan kata dalam dokumen dihasilkan melalui mekanisme probabilitas. Umumnya, unit leksikal dari sebuah kumpulan tulisan telah diberi label menurut kategori tertentu dan juga telah diproses secara komputasional. Selama proses ini, setiap dokumen diperlakukan sebagai *bag-of-words*, dan begitu juga dokumennya dianggap tidak memiliki struktur internal, dan tidak ada hubungan diantara kata. Sebuah fitur universal dari klasifikasi *naïve Bayes* adalah asumsi kondisional yang independen. *Naïve Bayes* mengasumsikan bahwa kata – kata yang ada saling independen dan oleh karena itu setiap individu kata dianggap sebagai indikasi dari emosi yang telah ditetapkan. Formula *Bayesian* menghitung probabilitas dari kelas yang didefinisikan, berdasarkan fitur dari dokumen tersebut dan dihitung menggunakan:

$$P(cs) = \frac{P(c)P(s|c)}{p(s)}$$

dimana $P(c)$ adalah probabilitas bahwa sebuah kalimat termasuk kedalam kategori c , $P(s)$ adalah probabilitas dari kemunculan dari sebuah kalimat, $P(s|c)$ adalah probabilitas bahwa kalimat tersebut termasuk dalam kategori c dan $P(c|s)$ adalah probabilitas bahwa kalimat yang diberikan termasuk dalam kategori c . Istilah $P(s|c)$ dapat dihitung dengan sebagai berikut:

$$P(sc) = \prod_{1 \leq k \leq n} P(S_k|C)$$

dimana $P(S_k|C)$ merepresentasikan probabilitas bahwa sebuah kata S_k muncul pada kategori c dan n merepresentasikan panjang dari kalimat s .

d. *Alat klasifikasi Knowledge-based*

Dalam pendekatan *knowledge-based* dan dengan peralatan yang telah dikembangkan, berlawanan dengan pendekatan statistik, yang mencoba menganalisis dan mengekstrak *knowledge* dari setiap kalimat dengan tujuan untuk mengetahui status sentimental. Arsitektur dari peralatan yang digunakan tercantum pada Gambar 3. Alat ini melakukan sentimen analisis pada tingkat kalimat. Sistem ini menggunakan *Tree Tagger* [12], sebuah *part-of-speech tagger*, untuk menentukan peran gramatikal tiap kata dalam kalimat dan bentuk dasarnya, dan juga *Stanford parser* [13] untuk menganalisis struktur kalimat dan membuat pohon dependensi sesuai pada hubungan katanya. Alat *Named Entity Recognizer(NER)* digunakan untuk mendeteksi nama yang cocok dan entitas bernama yang muncul dalam sebuah kalimat dengan tujuan untuk membantu analisis kalimat dan spesifikasi bagaimana bagian emosional dihubungkan dengan entitas sebuah kalimat,

seperti orang. Kata – kata yang diketahui menyampaikan sebuah emosi dapat terdeteksi menggunakan sumber leksikal dari *knowledge base (KB)* yang digunakan dan setiap kata emotional yang terdeteksi kemudian dianalisis dengan menggunakan peralatan – peralatan tersebut dan hubungannya dan cara mereka berinteraksi dengan kata pada kalimat tersebut akan ditentukan. Berdasarkan hubungan kata, mengidentifikasi hubungan dari suatu jenis spesifik dari sebuah kata emosional dengan kata kuantifikasi, dengan tujuan untuk menentukan kekuatan emosionalnya. Pada akhirnya, *emotion extractor unit* menentukan status emosional keseluruhan pada suatu kalimat berdasarkan bagian emotional dari kalimat tersebut.

Dengan cara ini, peralatan untuk kalimat *natural language* ini seperti dibawah :

1. Menggunakan *tree tagger* untuk menentukan lemma dan peran grammatikal dari sebuah kata.
2. Menggunakan *Stanford parser* untuk menganalisis struktur kalimat dan mendapatkan dependensi dan pohon dependensinya.
3. Menggunakan *NER* untuk mengenali nama entitas dan nama orang.
4. Untuk setiap kata digunakan *knowledge base* untuk menentukan apakah itu emosional atau tidak, jika iya,
 - 1.1 Analisis hubungannya
 - 1.2 Periksa jika terjadi modifikasi hubungan dengan kata kuantifikasi, analisis itu dan tentukan kekuatan emosinya.
 - 1.3 Analisis phon dependensinya, kenali pola/struktur kalimat dan berdasarkan hasilnya, tentukan isi emosional dari kalimat tersebut.

Alat *knowledge base* digunakan untuk menyimpan informasi yang berhubungan dengan kata – kata emosional yang diketahui menyampaikan emosi. Hal ini bergantung pada sumber leksikal *WordNet Affect*, sebuah ekstensi dari *Wordnet* yang sudah banyak digunakan yang juga diperpanjang dengan menambahkan beberapa kata kata emosional dan peran grammatikal mereka. Sebagai tambahan, *knowledge base* ini menyimpan kata – kata kuantifikasi, yang merupakan jenis spesial dari kata – kata yang dapat mengukur dan dapat menilai konten dari kata – kata emosional.

Node – node dari pohon dependensi adalah kata – kata dari kalimat itu dan *edge*-nya menentukan hubungan yang telah ada diantara kata. Untuk setiap kata, peran grammatikalnya dalam kalimat dan cara dia berinteraksi dengan kata lain akan ditentukan. Hubungan antara dua kata ditandai dengan adanya sebuah *edge* dan tipe interaksi yang tepat ditandai dengan nama *edge*-nya. Contohnya nsubj (she, kissed) adalah sebuah subjek hubungan nominal antara dua kata, mendefinisikan bahwa kata ‘she’ adalah subjek dari kata ‘kissed’.

1. Perumusan dari satuan emosional

Setelah kalimat *natural language* telah dianalisis, kemudian sistem mengenali kata – kata emosional dan merumuskan satuan emosional. Untuk melakukannya, sistem ini menggunakan sumber leksikal untuk mengenali kata –

kata yang diketahui menyampaikan konteks emosional. *Knowledge base*, seperti yang telah disebutkan diatas, menyimpan informasi tentang (a) kata – kata emosional dan (b) kata – kata kuantifikasi.

Knowledge base juga menyimpan informasi tentang kata – kata kuantifikasi yang menghitung dan memodifikasi kekuatan dari kata – kata emosional dengan berinteraksi dengan mereka. Oleh karena itu, kami mengembangkan sebuah daftar kata kuantifikasi, seperti: {very, some, all, hardly, less, dan lain – lain}. Sebuah kategori spesial dari kata – kata kuantifikasi adalah kata – kata yang menunjukkan negasi. Kata negasi, ketika muncul pada suatu kalimat dapat membalik sifat dari kata yang berinteraksi dengannya. Contoh dari kata negasi adalah: {none, No., not, never, nobody}.

Dampak Modifikasi	Kata
Tinggi	Very, great, huge, extensive
Sedang	Hardly, quite
Rendah	Little, less
Kebalikan	No, not

Tabel 2. Contoh kata – kata kuantifikasi

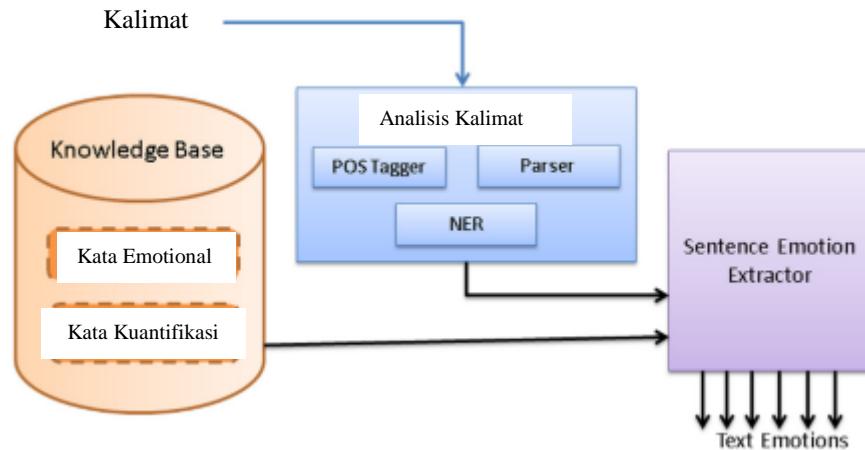
Pada akhirnya anggap kasus negasi seperti ‘is not very furious’, dimana kata emosional ‘furious’ terdeteksi menunjukkan ‘kemarahan’. Kata ‘very’ memiliki kuantifikasi yang tinggi pada emosi dan kata negasi terdeteksi membalik kuantifikasi menjadi rendah dan oleh karena itu bagian emosional pada kalimat ini ditentukan menjadi rendah (20) ‘kemarahan’.

2. Menentukan konten emosional kalimat

Setelah kata emosional dari sebuah kalimat dikenali dan kekuatan mereka telah ditentukan, sistem menentukan keseluruhan konten dari kalimat. Untuk melakukannya, dia menganalisis struktur kalimatnya. Khususnya, dia mengenali dan menganalisis pola dari kalimat tersebut yang terdiri atas kata *verb* utama, objek dan subjek dari kalimat tersebut. Maka pola “Subjek-Verb-Objek” diekstrak dari kalimat berdasarkan dependensinya. Pola ini menjadi tulang punggung dari struktur kalimat dan mempertahankan arti inti dari kalimat itu. Bahkan, menganalisisnya dapat membantu sistem dalam memahami interaksi dari bagian kalimat yang menghubungkan bagian – bagian emosionalnya. Maka sistem mengolah struktur kalimat melalui langkah – langkah sebagai berikut:

- 1.) Analisis dependensi dari kalimat dan ekstrak pola subjek-verb-objek

- 2.) Untuk setiap peran gramatikal dari pola
 - 2.1 Tentukan apakah itu adalah bagian emosional,
 - 2.2 Analisis hubungannya dengan bagian emosional (jika ada),
 - 2.3 Tentukan konten emosionalnya.



Gambar 3 gambaran arsitektur dari alat klasifikasi knowledge-based.

- e. Tentukan arah emosional

Menurut Russels model dua dimensi dari afeksi, emosi dapat dipresentasikan dalam ruang dua dimensi, dimana satu dimensi merepresentasikan polaritas emosi dan dimensi lain aktivasi emosi. Polaritas dimensi mengkaraktersistikkan emosi sebagai positif atau negatif, sebagaimana aktivasi mengkaraktersistikkan emosi sebagai teraktivasi atau tidak. Pemetaan ruang dimensi emosi tercantum pada Gambar 4.

Pemetaan ini memungkinkan sistem untuk menentukan polaritas dari sebuah kalimat berdasarkan konten emosional didalamnya. Sistem ini mengadaptasi dari pemetaan emosional ini dan bergantung padanya untuk menentukan polaritas emosional dari sebuah kalimat berdasarkan konten emosional dari sebuah kalimat.



Gambar 4. Pemetaan emosi pada ruang Russel

10.METODOLOGI

a. Penyusunan proposal tugas akhir

Penyusunan proposal Tugas Akhir ini adalah tahap awal untuk memulai pengerjaan Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan untuk Pengenalan Emosi dalam Teks menggunakan Metode Ensemble.

b. Studi literatur

Pada tahap ini akan dilakukan pendalaman studi literatur yang bersumber utama dari *paper* yang berasal dari jurnal internasional bereputasi yaitu *sciencedirect* dengan judul “*Recognizing emotions in text using ensemble of classifiers*”. Selain itu akan digunakan sejumlah referensi yang diperlukan dalam pembuatan aplikasi, yaitu mengenai penggabungan beberapa algoritma klasifikasi yaitu *Naive Bayes* dan *Max Entropy*.

c. Analisis dan desain perangkat lunak

Tahap analisis dan desain meliputi deteksi calon kandidat dokumen. Untuk membangun sistem deteksi emosi pada teks ini harus melalui beberapa tahap yaitu mendapatkan fitur *bag-of-words* dari setiap dokumen terlebih dahulu. Setelah itu melatih sistem klasifikasi *naive Bayes* dan *Max Entropy* dengan fitur – fitur yang telah didapatkan. Kemudian menggunakan *knowledge-based tool* untuk melengkapi klasifikasi pada *ensemble*. Terakhir adalah melakukan perhitungan kemiripan dengan menggunakan *presicion*, *recall*, dan *accuracy* untuk mendapatkan hasil yang akurat.

d. Implementasi perangkat lunak

Implementasi untuk melakukan percobaan ini dibangun dengan menggunakan bahasa pemrograman dan kakas bantu Python dengan fungsi yang sudah tersedia di dalamnya.

e. Pengujian dan evaluasi

Metode Analisis:

- Ekstraksi fitur *bag-of-words* dari tiap dokumen.
- Latih tiap model klasifikasi pada *ensemble* menggunakan fitur yang telah ada.

Implementasi:

- *Tools* yang digunakan adalah Python.

Pengujian:

- Masukkan beberapa matriks data *test* atau masukkan satu data dokumen.
- Ekstraksi fitur pada data *test* dan data *training* menggunakan metode *ensemble*, kemudian proses klasifikasi data *test* dengan data *training*.
- Sistem akan mengeluarkan hasil akurasi yang dihasilkan dengan menggunakan metode *ensemble*.

f. Penyusunan Buku Tugas Akhir

Pada tahap ini dilakukan penyusunan laporan yang menjelaskan dasar teori dan metode yang digunakan dalam tugas akhir ini serta hasil dari implementasi aplikasi perangkat lunak yang telah dibuat. Sistematika penulisan buku tugas akhir secara garis besar antara lain:

1. Pendahuluan
 - a. Latar Belakang
 - b. Rumusan Masalah
 - c. Batasan Tugas Akhir
 - d. Tujuan
 - e. Metodologi
 - f. Sistematika Penulisan
2. Tinjauan Pustaka
3. Desain dan Implementasi
4. Pengujian dan Evaluasi
5. Kesimpulan dan Saran
6. Daftar Pustaka

11. JADWAL KEGIATAN

Tahapan	2017				2018											
	Desember				Januari				Februari				Maret			
Penyusunan Proposal																
Studi Literatur																
Perancangan Sistem																
Implementasi																
Pengujian dan Evaluasi																

- [16] “Digilib Petra,” [Online]. Available:
http://digilib.petra.ac.id/viewer.php?page=10&submit.x=20&submit.y=17&qual=high&submitval=next&fname=%2Fjiunkpe%2Fs1%2Finfo%2F2006%2Fjiunkpe-ns-s1-2006-26402056-6118-aplikasi_image-chapter2.pdf. [Diakses 12 June 2017].