
Lexicon-based Emotion Detection for Academic Questionnaire Results

Deteksi Emosi Berbasis Leksikon untuk Hasil Kuesener Akademik

Amir Hamzah¹

¹ Informatika, Institut Sains & Teknologi AKPRIND Yogyakarta, Indonesia

¹amir@akprind.ac.id

*: *Penulis korenspondensi (corresponding author)*

Keywords: emotion; lexicon-based; detection

Abstract

Purpose : This study aims to apply the emotion detection method to the student commentary text of the questionnaire based on the emotion lexicon. Emotion labels from comments will increase the accuracy of the results of the questionnaire analysis. Another objective of this research is to see the effectiveness of Emolex's emotional lexicon for emotion detection in academic questionnaire texts.

Methodology: The research data consisted of comments and suggestions from students on the learning evaluation questionnaire at IST AKPRIND 2014-2017 as many as 3,975 comments. The types of emotions detected were 8 types of emotions, namely anger, anticipation, disgust, fear, happiness, sadness, surprise and confidence. The emotion lexicon used is NRC Emolex. The first step of detection after the pre-processing stage is to extract the emotional features using the emotion lexicon list. The second step is emotion detection by calculating the largest weight of the extracted emotional features as emotion labels for the comments.

Result: The results showed that from all the data detected, 46.7% could be identified with the emotional label. Of the emotion detected, the 3 highest percentages of emotional labels are Sadness (18.5%), Joy (17.1%) and Fear (14.0%). From the research revealed that the performance of emolex for emotion detection is still not satisfactory. This is very likely because only 37% of Emolex libraries have emotion labels.

State of the art: We found weaknesses in the NRC Emolex lexicon as a result of translating from English, among others, many lexicons have not been labeled with emotion and the effect of translating into two or three words in Indonesian so that it cannot be used.

Kata kunci: emosi; basis leksikon; deteksi

Abstrak

Tujuan: Penelitian ini bertujuan menerapkan metode deteksi emosi pada teks komentar mahasiswa hasil kuesener berbasis pada leksikon emosi. Label emosi dari komentar akan meningkatkan akurasi dari hasil analisis kuesener. Tujuan lain dari penelitian ini adalah melihat sejauh mana efektivitas leksikon emosi Emolex untuk deteksi emosi teks kuesener akademis.

Metode: Data penelitian berasal dari komentar dan saran mahasiswa pada kuesener evaluasi pembelajaran pada IST AKPRIND tahun 2014-2017 sebanyak 3.975 komentar. Jenis emosi yang dideteksi adalah 8 jenis emosi, yaitu marah, antisipasi, jijik, takut, bahagia, sedih, terkejut dan yakin. Leksikon emosi yang digunakan adalah NRC Emolex. Langkah pertama deteksi setelah tahap pre-processing adalah ekstrak fitur emosi menggunakan daftar leksikon emosi. Langkah kedua adalah deteksi emosi dengan cara menghitung bobot terbesar dari fitur emosi yang terekstrak sebagai label emosi bagi komentar tersebut.

Hasil: Hasil penelitian menunjukkan bahwa dari seluruh data yang dideteksi, 46,7% dapat diketahui label emosinya. Dari yang diketahui label emosinya 3 prosentase tertinggi ada pada label Sadness (18,5%), Joy(17,1%) dan Fear (14,0%). Dari penelitian terungkap bahwa kinerja emolex untuk deteksi emosi masih belum memuaskan. Hal ini sangat mungkin disebabkan karena baru 37% pustaka Emolex memiliki label emosi.

State of the art: Ditemukan kelemahan leksikon NRC Emolex sebagai hasil translate dari bahasa inggris antara lain, banyak leksikon belum diberikan label emosi dan efek translate menjadi dua atau tiga kata dalam bahasa Indonesia sehingga tidak dapat digunakan.

1. Pendahuluan

Emosi memegang peran penting dalam tingkah laku seseorang dalam berkomunikasi. Pada kenyataannya komunikasi melalui media apapun akan senantiasa melibatkan emosi dari pihak-pihak yang berkomunikasi [1]. Emosi mempengaruhi perilaku manusia [2]. Stabilitas emosi bahkan sangat berpengaruh pada kualitas komunikasi antar manusia [3] dan pada komunikasi manusia-komputer [4]. Saat ini komunikasi menggunakan teks menjadi bagian penting dalam berbagai bidang. Melimpahnya teks menjadi tantangan tersendiri untuk menguak emosi yang terkandung dalam teks [4]. Komunikasi teks ini juga terjadi dalam evaluasi pembelajaran di Perguruan Tinggi. Dalam setiap akhir pembelajaran biasanya pihak administrasi pendidikan

dalam rangka menjaga mutu pembelajaran akan mengadakan evaluasi pembelajaran dengan berbagai cara. Umumnya evaluasi oleh mahasiswa terhadap kinerja dosen dibuat dalam bentuk kuesener berupa pillihan ganda terhadap parameter kinerja seperti persiapan dosen, penyampaian materi, metode pembelajaran dan lain-lain. Analisis data kuesener umumnya menggunakan metode statistik deskriptif. Namun ada data responden yang belum dapat dianalisis dengan metode statistik deskriptif tersebut, yakni komentar/saran yang disediakan dalam kotak saran dan komentar dalam kuesener. Ratusan atau kadang ribuan komentar mahasiswa yang berupa teks bebas tersebut dapat mengungkap opini atau emosi mahasiswa yang sangat berharga bagi evaluasi tambahan. Terlebih dari sisi lain ada kelemahan kuesener dengan jawaban pilihan ganda yang dilakukan berulang, yakni menimbulkan kejenuhan dan berakibat mahasiswa menjawab sekenanya. Sedangkan saran dan komentar umumnya dilakukan oleh mahasiswa yang memang ingin mengungkapkan perasaannya. Penelitian Greaves et.al. (2013) di *English National Health Service website* yang menagkap 6.412 komentar bebas dari pasien yang dirawat. Analisis tentang komentar terkait dengan kebersihan, pelayanan rumah sakit dan berbagai aspek tanggung jawab rumah sakit memberikan hasil kesesuaian antara 81% sampai 89% dibandingkan dengan metode rating kuantitative dengan kuesener [5]. Latar belakang ini memberikan peluang dan tantangan penerapan kajian emosi pada jawaban kuesener yang berupa komentar dan saran. Pendalaman emosi yang terungkap dalam teks jawaban ini akan sangat berguna untuk mengambil kebijakan berikutnya sebagai perbaikan proses pembelajaran.

Deteksi emosi, analisis emosi atau penambangan emosi (*emotion mining*) merupakan kajian yang masuk dalam topik *sentiment analysis* atau *opinion mining* [6]. Analisis sentimen bertujuan mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau kata, yaitu apakah opini yang dikemukakan dalam dokumen, kalimat atau kata tersebut bersifat positif, negatif atau netral. Lebih umum analisis sentimen merujuk pada sikap (*attitude*) seseorang terhadap objek atau topik. Sikap ini dapat berupa evaluasi seperti positif atau negatif, tetapi dapat berupa sikap emosi (*emotional attitude*) seperti marah, kecewa atau senang [2]. Menurut Liu (2012), dalam deteksi atau analisis emosi yang diungkap dalam satu teks dikenal ada 3 pendekatan, yaitu pendekatan *key-word spotting*, *rule based* dan *statistical based* [6],[7]. Pendekatan *key-word spotting* atau biasa disebut *lexical-based* berasumsi bahwa teks yang berisi deretan kata-kata akan dapat ditentukan sentimennya atau kandungan emosinya berdasarkan muatan emosi yang terungkap dalam kata. Untuk keperluan ini diperlukan daftar kata dan emosi yang terkait dengan kata tersebut. Daftar kata dan emosi inilah yang disebut sebagai leksikon emosi [8].

Ekman (1992) menyatakan bahwa pada manusia ada 6 emosi dasar (basic emotion), yaitu anger, disgust, fear, joy, sadness dan trust [9],[10]. Selanjutnya Plutchik (1994) menambahkan 2 status emosi yaitu anticipation dan surprise [1]. Berdasarkan dua kelompok status emosi inilah biasanya peneliti menyusun leksikon emosi [11]. Kebanyakan leksikon emosi masih dalam bahasa inggris. Vania et al. (2014) menyusun sentiment leksikon bahasa Indonesia dengan teknik POS (Part-of-Speech), tetapi masih sebatas setimen positif dan negatif, belum ada label emosinya [12]. Bata et al. (2015) menyusun leksikon emosi dalam bahasa Indonesia yang bersumber dari leksikon dari Shaver et al (2001) [1], [13]-[14]. Leksikon emosi menggunakan 5 status emosi, yaitu senang, cinta, takut, marah, dan sedih. Hasil penyusunan berupa 1165 kata dan status emosinya. Mohammad and Turney (2013) menyusun leksikon emosi dalam bahasa inggris yang diterjemahkan dalam 105 bahasa yang disebut NRC Emotioan Lexicon atau Emolex [15]. Leksion berisi 14.182 dengan 2 status status sentimen, positive dan negative dan

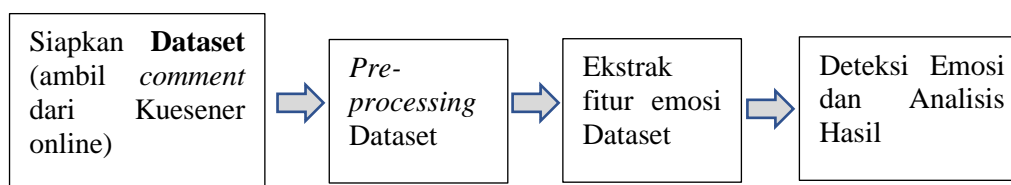
8 status emosi, yaitu anger, anticipation, disgust, fear, joy, sadness, surprise, dan trust. Karakteristik emotion leksikon yang bersifat crowdsourcing, yaitu bertumpu pada khalayak dalam mengembangkan labelnya, memungkinkan Emolex diases oleh banyak peneliti [16].

Saat ini penelitian terkait deteksi emosi dan analisis emosi makin dibutuhkan dalam dunia bisnis [17]. Penelitian telah dilakukan pada teks dengan berbagai bahasa di luar bahasa Inggris pada bahasa dengan abjad non latin seperti deteksi emosi pada teks Punjabi [18] atau teks Arab [9],[16]. Banyak penelitian dilakukan pada deteksi emosi pada media sosial seperti *twitter* [19],[20], komentar *youtube* [21], status *facebook* [15],[22]. Karakteristik penelitian tersebut adalah melakukan deteksi, klasifikasi atau analisis emosi pada teks pendek. Mashal dan Asyani (2017) menemukan bahwa tingkat emosi pada teks pendek dapat membantu mengidentifikasi emosi teks panjang yang memiliki status emosi sama [23]. Jawaban kuesener berupa saran dan komentar memiliki karakteristik sama dengan objek-objek penelitian pada status media sosial. Dengan latar belakang seperti itulah penelitian ini dilakukan.

Tujuan penelitian ini adalah menerapkan deteksi emosi pada jawaban kuesener evaluasi akademik yang berupa saran/komentar. Data yang dijadikan objek adalah saran dan komentar hasil kuesener akademik di IST AKPRIND Yogyakarta tahun 2014 sampai 2017. Jenis emosi yang akan dideteksi adalah *anger, anticipation, disgust, fear, joy, sadness, surprise, dan trust*. Deteksi emosi berbasis pada leksikon yang diambil dari koleksi leksikon *EmoLex*. Tujuan lain adalah mengevaluasi efektifitas *EmoLex* yang untuk penerapan teks bahasa Indonesia.

2. Metode

Metode deteksi opini dari komentar kuesener pada prinsipnya menganggap setiap komentar atau saran mahasiswa sebagai opini yang memiliki sentimen atau muatan emosi. Pengenalan muatan emosi bersandar pada kata-kata yang digunakan mahasiswa yang dikonfirmasi pada pustaka leksikon emosi untuk menentukan jenis emosi yang ada. Adapun langkah penelitian yang ditempuh dapat disajikan seperti pada Gambar 1 berikut ini.



Gambar 1. Diagram Langkah Penelitian

2.1. Pengumpulan Data

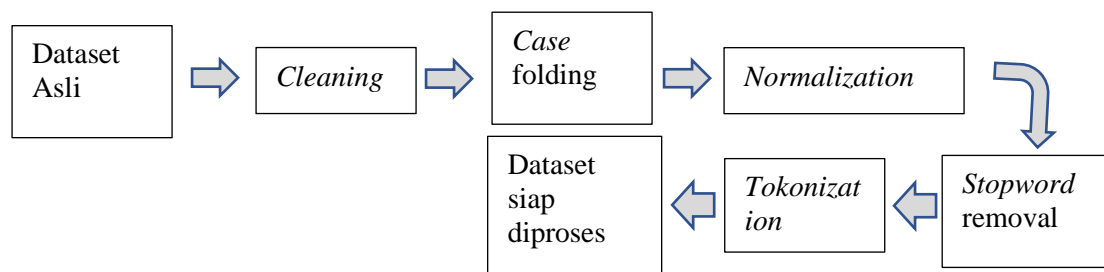
Dataset di dalam penelitian ini diambil dari koleksi opini berupa komentar atau saran mahasiswa dalam kuesener evaluasi pembelajaran IST AKPRIND Yogyakarta tahun 2014 sampai tahun 2017. Setiap komentar atau saran dianggap sebagai dokumen. Data dari sumber kuesener online diperoleh sebanyak 3.975 opini. Opini tersebut masih merupakan data mentah, dapat mengandung karakter-karakter tak berguna, bahasa informal atau salah ketik. Contoh opini yang masih belum diolah adalah seperti Tabel 1 berikut ini.

Tabel 1. Dataset Asal

Opini	Isi
D0001	Pak joko ngajarnya Joss nih!!
D0002	Jika ada kebijakan bebas asap rokok, harusnya difasilitasi dengan tempat merokok (smoking area) di kampus
D0003	Ruang kelas ada coretan yang mengotori, parkir dikasih satpam saja !
D0004	Jika Proyektor Baik Hasil Belajar akan lebih baik !!!
....
D3975	Jika Mengajar baik untuk kedepannya lebih baik lagi!!

2.2. Tahap Pre-processing Data

Koleksi Dataset seperti pada Tabel 1 sebelum dilakukan analisis kandungan emosi diproses awal melalui tahap *pre-processing text*. Tahap proses awal dapat dijelaskan langkahnya seperti pada Gambar 2 berikut ini.



Gambar 2. Pre-processing Dataset sebelum Analisis

Penjelasan setiap langkahnya adalah sebagai berikut :

1. Tahap *Cleaning*, yaitu proses menghilangkan karakter yang tidak berguna, seperti tanda seru (!), tanda koma (,), titik dua (:), titik-titik (...) atau karakter lain.
2. Tahap *Case Folding*, yaitu merubah semua karakter huruf menjadi huruf kecil
3. Tahap *Normalization*, yaitu proses mengembalikan pada kata standar dari kata-kata yang tidak standar karena gaya penulisan informal, misalnya ‘tdk’, ‘tak’, ‘nggak’, ‘gak’ akan dinormalisasi menjadi ‘tidak’
4. Tahap *Stopword removal*, yaitu menghilangkan kata-kata yang tidak bermakna dan cenderung muncul di semua dokumen sehingga tidak dapat dijadikan fitur pembeda dokumen. Kata-kata seperti itu disebut *stopword*, seperti ‘ini’, ‘itu’, ‘yang’, ‘dengan’, ‘dari’ dan lain-lain harus dihilangkan. *Stop word* diambil dari *transition word* seperti ‘ke’, ‘dari’, ‘sampai’ dan *negation word* seperti ‘enggak’, ‘tidak’ dan lain-lain [12].
5. Tahap *Tokenization*, yaitu memecah dokumen yang telah melewati tahapan-tahapan sebelumnya menjadi token-token. Token inilah yang akan menjadi kandidat fitur dalam deteksi emosi.

Pada tahap *pre-processing* untuk deteksi emosi berbasis leksikon emosi, ada tahap *pre-processing* yang lazim digunakan dalam *sentiment analysis* tetapi tidak dilakukan, yaitu tahap

stemming, yakni merubah kata menjadi kata dasar. Hal ini dilakukan karena dalam leksikon emosi sebuah kata yang dapat berubah karena berimbunan dengan awalan dan atau akhiran sangat mungkin memiliki label emosi yang berbeda. Misalnya dalam EmoLex yang dijadikan acuan, kata ‘cinta’ (translate dari ‘love’) memiliki label emosi yang berbeda dengan kata ‘percintaan’ (translate dari ‘romance’). Kata ‘cinta’ hanya memiliki satu lebel emosi, yaitu *joy*, sedangkan ‘percintaan’ memiliki label emosi, yaitu *anticipate, fear, joy, sadness, surprise* dan *trust*.

Sebagai contoh untuk opini-opini seperti pada Tabel 1 jika dilakukan *preprocessing* maka akan menghasilkan dokumen sebagai kumpulan token seperti pada Tabel 2 berikut ini.

Tabel 2. Dataset setelah *Preprocessing*

Opini	Isi
D0001	‘pak’ ‘joko’ ‘ngajarnya’ ‘joss’
D0002	‘kebijakan’ ‘bebas’ ‘asap’ ‘rokok’ ‘difasilitasi’ ‘tempat’ ‘merokok’ ‘smoking’ ‘area’ ‘kampus’
D0003	‘ruang’ ‘kelas’ ‘coretan’ ‘mengotori’ ‘parkir’ ‘dikasihkan’ ‘satpam’
D0004	‘proyektor’ ‘baik’ ‘hasil’ ‘belajar’ ‘lebih’ ‘baik’
....
D3975	‘mengajar’ ‘baik’ ‘kedepannya’ ‘baik’

2.3. Ekstak Fitur Emosi

Setiap opini yang telah dibersihkan melalui tahap *pre-processing* dapat dipandang sebagai himpunan fitur yang mewakili opini tersebut. Selanjutnya fitur-fitur ini akan dicek apakah terdapat dalam kamus leksikon emosi Emolex (*NRC Emotion Lexicon* diambil dari [15],[24]). Emolex terdiri dari 14.182 leksikon bahasa inggris yang dilengkapi terjemahan dalam 105 bahasa, termasuk bahasa jawa dan bahasa sunda. Jika fitur (kata) yang ada dalam opini terdapat dalam Emolex maka fitur tersebut diekstrak dengan menyimpan label emosinya. Perlu diingat bahwa label emosi tidak selalu unik. Kata ‘marah’ hanya berhubungan dengan label emosi ‘*Anger*’, tetapi kata ‘kebencian’ terhubung dengan label emosi ‘*Anger*’, ‘*Disgust*’, ‘*Fear*’ dan ‘*Sadness*’. Contoh hasil ekstrak fitur kata disajikan seperti pada Tabel 1. Kolom *positive* dan *negative* dihilangkan karena label sentimen bukan bahan kajian. Dalam tabel angka 0 menunjukkan kata tersebut tidak berhubungan dengan label emosi leksikon, sedangkan angka 1 menunjukkan label emosi kata yang bersangkutan. Hasil akhir Dataset yang asalnya berupa kumpulan opini digantikan dengan kumpulan fitur kata dalam opini telah memiliki label emosi.

Tabel 3. Contoh EmoLex dengan 8 Label Emosi

Word	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Marah	1	0	0	0	0	0	0	0
kebencian	1	0	1	1	0	1	0	0
Senang	0	1	0	0	1	0	0	0
Hina	0	0	1	0	0	0	0	0
Gagal	0	0	0	0	0	1	0	0
prestasi	0	0	0	0	1	0	0	0
mengerikan	0	0	1	1	0	0	0	0
percaya	0	0	0	0	0	0	0	1

Algoritma untuk ekstrak fitur dari koleksi opini dapat dituliskan sebagai berikut

Algoritma 1 :ekstrak fitur emosi

- 1) Start
- 2) $D = \{D_1, D_2, \dots, D_n\}$, $n = \text{number of doc opini}$
 $W = \{w_{i1}, w_{i2}, \dots, w_{im}\}$, $im = \text{number of word in doc } -i$
 $\text{FiturLex} = \{t_1, t_2, \dots, t_L\}$, $L = \text{number of fitur in EmoLex}$
 $\text{VekEmo} = \{v_1, v_2, \dots, v_L\}$, v is vektor contains 8 number (0 or 1 as label of emotion)
- 3) $\text{FiturEkstrak} = []$; $\text{VektorEmosi} = []$;
- 4) For $i=1$ to n do
 - For $j=1$ to im Do
 - If w_{ij} in FiturLex (in order k) do
 - $\text{FiturEksrak}[] \leftarrow t_k$
 - $\text{VektorEmosi}[] \leftarrow v_k$
- 5) Stop

2.4. Deteksi Emosi dari Opini

Algoritma yang digunakan untuk mendeteksi jenis emosi dari opini mengasumsikan setiap fitur kata dalam koleksi opini telah memiliki vektor label dari tahapan ekstrak fitur, yaitu satu baris 8 nilai label emosi seperti pada Tabel 3. Selanjutnya sebuah opini D_i yang akan ditentukan label emosinya dipandang sebagai himpunan fitur yang telah memiliki vektor label emosi, yaitu $D_i = \{f_1, f_2, \dots, f_{ni}\}$ dengan ni adalah cacah fitur terekstraksi dalam opini D_i . Looping dibuat untuk menjumlahkan seluruh nilai skor dari 8 label untuk semua fitur. Jumlah skor untuk ke-8 label tersebut kemudian dicari nilai maksimumnya. Index di mana nilai maksimum skor label emosi berada adalah index dari jenis emosi yang dicari.

Algoritma 2 : Deteksi emosi opini D_i

- 1) Start
- 2) $D_i = \{f_1, f_2, \dots, f_{ni}\}$, $ni = \text{number of fitur emotion of doc } D_i$
 $\text{FiturEkstrak} = \{T_1, T_2, \dots, T_P\}$, $P = \text{number of fitur emotion ekstracted}$

```

    VektorEmosi={v1,v2,..vp}, v is vektor contains 8 number ( 0 or 1 as label
    of emotion)
    3) LabelEmo={Anger, Anticipation, Disgust, Fear,Joy,Sadness, Surprise,
    Trust}
    4) LblEmoSelected=0; TotScoreVek={0,0,0,0,0,0,0,0}
    5) For i=1 to ni do
        For j=1 to 8 Do
            TotScoreVek[j]←TotScoreVek[j]+VektorEmosi[j]
    6) TotScoreVekMax=max(TotScoreVek[1..8])
    7) LblEmoSelected← LabelEmo[j], j is index where TotScoreVekMax is
    located
    8) Stop
    
```

Berikut contoh perhitungan mengikuti algoritma 2 untuk opini D2975 yang berbunyi:

“Jika Mengajarnya baik untuk kedepannya lebih baik lagi!!”

Setelah mengalami tahapan *pre-processing* diperoleh token-token :

‘mengajar’, ‘baik’, ‘kedepannya’, ‘baik’

Setelah melalui tahapam algoritma 1 (seleksi fitur emosi), didapatkan fitur emosi adalah

‘mengajar, ‘baik’, ‘baik’

Maka penjumlahan skor label emosi dari fitur-fitur emosi diperoleh skor maksimum 3 (lihat Tabel 4), yaitu untuk label **Joy**, **Surprise** dan **Trust**. Untuk itu opini tersebut memiliki label 3 jenis emosi.

Tabel 4. Perhitungan Label Emosi untuk opini D3975

	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
mengajar	0	0	0	0	1	0	1	1
Baik	0	1	0	0	1	0	1	1
Baik	0	1	0	0	1	0	1	1
Jumlah skor	0	1	0	0	3	0	3	3

3. Hasil dan Pembahasan

3.1. Hasil deteksi emosi

Hasil deteksi emosi dari koleksi Dataset yang dicobakan menggunakan algoritma 1 dan algoritma 2 adalah seperti pada Tabel 5 berikut ini.

Tabel 5. Hasil Deteksi Emosi Dataset

Jumlah Opini	Hasil Deteksi Emosi	
	Terdeteksi	Tak Terdeteksi
3.975	1.856	2.119
Prosentase	46,7%	53,3%

Dari tabel terlihat bahwa jumlah opini yang dapat diketahui label emosinya hanya 46,7%. Ini merupakan jumlah yang belum terlalu tinggi, masih dibawah 50%. Jika dirunut proses deteksi emosi menggunakan basis leksikon emosi, maka dapat dipahami bahwa keberhasilan emosi hanya tergantung pada dua hal, yang pertama adalah “kualitas” teks opini yang akan dideteksi dan ketersediaan leksikon emosi dalam *Emolex*. Kualitas opini yang dimaksud adalah apakah kalimat opini dapat diuraikan menjadi fitur-fitur yang ada label emosinya dalam *Emolex*. Jika seluruh fitur hasil *tokenization* gagal ditemukan dalam *Emolex*, maka opini tersebut akan gagal dideteksi. Sebagai ilustrasi kita sajikan komentar D0001 berikut :

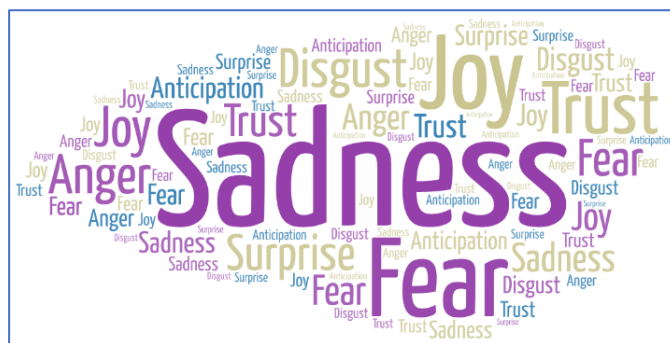
Pak joko ngajarnya Joss nih!!

Setelah melalui proses *pre-processing* dan proses seleksi fitur akan dihasilkan 4 buah token, yaitu : ‘pak’, ‘joko’, ‘ngajarnya’, ‘joss’. Untuk token-token ‘pak’, ‘joko’ dan ‘joss’ tidak dapat difilter oleh langkah *StopWord filtering*. Dengan demikian hanya tinggal fitur ‘ngajarnya’. Pada tahap seleksi fitur menggunakan *Emolex*, ternyata term ‘ngajarnya’ tidak ditemukan dalam *Emolex*. Maka opini tersebut akan gagal terdeteksi label emosinya.

Dari opini yang dapat terdeteksi, yaitu sebanyak 1.856 buah, memiliki sebaran label emosi seperti sersaji pada Tabel 6 dan ilustrasi *wordcloud* pada Gambar 3.

Tabel 6 . Hasil Deteksi Emosi Dataset yang Dikenali

	Label Emosi								Tot Dok
	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	
Jum dok	217	95	236	260	317	343	139	249	217
Prosentase	11,7%	5,1%	12,7%	14,0%	17,1%	18,5%	7,5%	13,4%	100%



Gambar 3. Wordcloud Hasil deteksi Emosi

Dari ilustrasi Gambar 3 terlihat bahwa dari komen dan saran hasil kuesener yang mampu dideteksi menggunakan kamus leksikon *Emolex*, label yang paling menonjol adalah *Sadness*,

disusul *Joy* dan *Trust*. Tentu ini menjadi petunjuk untuk perbaikan model pembelajaran agar emosi positif seperti *Joy* dan *Trust* yang berada pada posisi paling atas.

3.2. Evaluasi pustaka emosi EmoLex

Sebagai sebuah kamus terbuka, *EmoLex* saat ini masih memiliki beberapa kelemahan, terutama untuk bahasa non inggris yang menggunakan melalui mekanisme *translation*. Kelemahan utama yang menyebabkan efektifitas deteksinya kurang tinggi adalah sedikitnya term yang telah diberikan label emosi. Paling tidak ada 4 kelemahan yang dapat disampaikan sebagai evaluasi untuk penggunaan *EmoLex* untuk deteksi emosi berbasis leksikon bahasa Indonesia.

Kelemahan yang **pertama**, dari 14.182 term bahasa inggris, baru 4.463 (36,7%) yang telah diberi label sentimen atau label emosi. Angka tersebut diambil dari bahan NRC *EmoLex* [25]. Pada prinsipnya diperlukan kontributor dari pengguna sesuai dengan bahasa apa yang akan digunakan agar *EmoLex* dapat berdaya guna maksimal. Hal ini sesuai dengan pendekatan yang digunakan yakni *crowdsourcing* yang sangat bertumpu pada peran khalayak untuk memberikan label emosi.

Kelemahan yang **kedua**, terdapat term yang diberi label sentimen namun term tersebut belum diberikan label emosi. Ada bahkan term-term yang sama sekali belum ada label sentimen maupun label emosinya. Tabel 7 berikut contoh term yang belum ada label sama sekali. Beberapa term yang tidak ada label adalah term-term yang sangat akrab dengan istilah dalam komentar mahasiswa dalam pembelajaran. Misalnya term ‘koneksi’, ‘kampus’, ‘merokok’ adalah contoh term yang sangat mungkin muncul dalam komentar mahasiswa dalam evaluasi pembelajaran, misalnya mahasiswa yang megeluh “**koneksi internet kurang bagus**”, akan tidak dapat dideteksi emosi nya karena 4 kata yang muncul tidak ada dalam term atau ada dalam term tetapi tanpa label emosi.

Tabel 7. Translate tanpa Label Emosi

English	Indonesia
Stalls	warung
connection	koneksi
projektor	proyektor
campus	kampus
merokok	merokok

Kelemahan yang **ketiga** yang muncul dan dimungkinkan menurunkan kemampuan deteksi emosi dari kamus leksikon *EmoLex* untuk dokumen bahasa Indonesia adalah masalah efek *translate* bahasa. Kamus *EmoLex* yang aslinya dalam bahasa inggris setiap term adalah **satu kata**, namun ketika *ditranslate* ke bahasa Indonesia tidak semua term *ditranslate* menjadi satu kata bahasa Indonesia, tetapi bisa menjadi dua kata atau bahkan 3 kata. Tabel 8 adalah contoh beberapa term yang *ditranslate* ke bahasa Indonesia menjadi lebih dari satu kata. Dampak dari translasi lebih dari satu kata adalah tahap seleksi fitur emosi bahasa Indonesia akan mengalami kesulitan, karena proses *tokenization* mengharuskan satu term satu nilai label emosi.

Tabel 7. Translate Term Inggris Menjadi Indonesia yang Lebih dari 1 Kata

Term asal	Term Hasil Translate	Jumlah term translate
ablaze	terang benderang	2
aboard	di atas kapal	3
abode	tempat tinggal	2
absent	tidak hadir	2
abstain	menjauhkan diri	2
accidentally	tanpa sengaja	2
adaptable	mudah beradaptasi	2
staid	tenang dan serius	3
backwards	ke belakang	2
untrustworthy	tidak bisa dipercaya	3

Kelemahan yang **keempat** adalah banyaknya term yang tidak ada translasinya pada bahasa Indonesia. Ada sejumlah term, yaitu sebanyak 271 term ($\pm 2\%$) yang hasil translasinya NO TRANSLATION, artinya tidak muncul translasinya dalam bahasa Indonesia. Hal ini dapat berefek pada penurunan kemampuan deteksi emosi dari kamus leksikon *Emolex*.

4. Kesimpulan dan Saran

Penelitian ini telah berhasil menerapkan deteksi emosi berbasis leksikon emosi pada data teks komentar dan saran dari kuesener evaluasi pendidikan yang diambil dari 3.975 komentar dan saran dari evaluasi pendidikan IST AKPRIND tahun 2014-2017. Keberhasilan dalam deteksi emosi menggunakan pustaka leksikon dari *Emolex* hanya mencapai 46,7% dari seluruh opini yang diuji. Dari data opini mahasiswa yang diketahui label emosinya, 3 prosentase tertinggi ada pada label *Sadness* (18,5%), *Joy*(17,1%) dan *Fear* (14,0%). Hal ini memberikan gambaran bahwa hasil ungkapan komen dan saran yang ada respon tertinggi dari respon yang terdeteksi masih bersentimen negatif, yaitu *Sadness* meskipun pada urutan kedua adalah sentimen positif yaitu *Joy*. Dari penelitian penggunaan kamus leksikon emosi *Emolex* terungkap bahwa kinerja *emolex* untuk deteksi emosi masih belum memuaskan. Hal ini sangat mungkin disebabkan karena baru 37% pustaka *Emolex* yang diberikan label emosi. Kekurangan *emolex* sebagai sumber acuan seleksi fitur emosi juga terkendala efek *translastion* inggris-indonesia dari fitur tunggal bahasa Inggris yang tidak semuanya ditranslate menjadi fitur tunggal bahasa Indonesia, melainkan ada yang ditranslate menjadi 2 fitur atau 3 fitur kata bahasa Indonesia yang memunculkan masalah pada tahap konfirmasi fitur pada tahap tokenization di level *pre-processing*.

Daftar Pustaka

- [1] R. Plutchik, *The Psychology and Biology of Emotion*. New York: Harper Collins Publisher, 1994.
- [2] S. M. Mohammad, "Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text," *Emot. Meas.*, pp. 201–237, 2015, doi: 10.1016/B978-0-08-100508-8.00009-6.
- [3] F. Celli and L. Rossi, "The role of Emotional Stability in Twitter Conversations," *Proc. Work. Semant. Anal. Soc. Media, conjunction with EACL 2012*, pp. 10–17, 2012, [Online]. Available: <http://aclweb.org/anthology-new/W/W12/W12-0602.pdf>.
- [4] R. A. Calvo and S. Mac Kim, "Emotions in text : dimensional and categorical models," *Comput. Intell.*, vol. 29, no. 3, pp. 527–543, 2013.
- [5] F. Greaves, C. Excellence, D. Ramirez-cano, and L. J. Donaldson, "Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online," *J. Med. Internet Res.*, vol. 15, no. 11, pp. 145–152, 2013, doi: 10.2196/jmir.2721.
- [6] B. Liu, *Sentiment Analysis and Opinion Mining*, no. May. Morgan & Claypool Publisher, 2012.
- [7] V. Anusha and B. Sandya, "A learning based emotion classifier with semantic text processing," *Adv. Intell. Syst. Comput.*, vol. 320, pp. 371–382, 2015.
- [8] S. Poria, A. Gelbukh, D. Das, and S. Bandyopadhyay, "Fuzzy Clustering for Semi-Supervised Learning — Case study : Construction of an Emotion Lexicon," in *Mexican International Conference on Artificial Intelligence (MICAI 2012) Advances in Artificial Intelligence*, 2012, pp. 23–27.
- [9] M. Abdullah, M. Hadzikadic, S. Shaikh, and N. Carolina, "SEDAT : Sentiment and Emotion Detection in Arabic Text using CNN-LSTM Deep Learning," in *in Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 2018, pp. 835–840.
- [10] P. Ekman, "An Argument for Basic Emotions," *Cogn. Emot.*, vol. 4, no. 3, pp. 169–200, 1992.
- [11] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment Analysis Using Common-Sense and Context Information Sentiment Analysis Using Common-Sense and," *Comput. Intell. Neurosci.*, vol. 2015, no. 6, pp. 715–730, 2015, doi: 10.1155/2015/715730.
- [12] M. Vania, C., Ibrahim, M., & Adriani, "Sentiment Lexicon Generation for an Under-Resourced Language," *Int. J. Comput. Linguist. Appl.*, vol. 5, no. 1, pp. 59–72, 2014.
- [13] J. Bata, Sunyoto, and Pranowo, "Leksikon untuk deteksi emosi dari teks bahasa indonesia," in *Seminar Nasional Informatika 2015 (semnasIF 2015)*, 2015, vol. 2015, no. November, pp. 195–202.
- [14] P. R. Shaver, U. Murdaya, and R. C. Fraley, "Structure of the Indonesian emotion lexicon," *Asian J. Soc. Psychol.*, vol. 4, no. 3, pp. 201–224, 2001.
- [15] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word – Emotion Association

- Lexicon,” *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, 2013.
- [16] M. AL-Al’abeda and M. Al-ayyoub, “A Lexicon-Based Approach for Emotion Analysis of Arabic Social Media A Lexicon-Based Approach for Emotion Analysis of Arabic Social Media Content,” in *Proceedings of the International Computer Sciences and Informatics Conference (ICSIC 2016)*, 2016, pp. 343–351.
- [17] S. Goyal and N. Tiwari, “EMOTION RECOGNITION : A LITERATURE SURVEY,” *Int. J. Technol. Res. Eng.*, vol. 4, no. 9, pp. 1502–1524, 2017.
- [18] S. Grover and A. Verma, “Design for emotion detection of punjabi text using hybrid approach,” in *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016*, 2016, no. 2, pp. 2019–2021.
- [19] A. S. Aribowo, S. Khomsah, U. Pembangunan, N. Veteran, S. Data, and I. Artikel, “Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study : Tweets About Covid-19),” vol. 18, no. 1, pp. 49–60, 2021, doi: 10.31515/telematika.v18i1.4341.
- [20] Muljono, A. S. Winarsih, and C. Supriyanto, “Evaluation of classification methods for Indonesian text emotion detection,” in *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, 2016, pp. 130–133.
- [21] A S Aribowo, H. Basiron, N. S. Herman, and S. Khomsah, “Fanaticism Category Generation Using Tree-Based Machine Learning Method Fanaticism Category Generation Using Tree-Based Machine Learning Method,” *J. Phys. Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1501/1/012021.
- [22] F. Krebs, B. Lubascher, T. Moers, and P. Schaap, “Social Emotion Mining Techniques for Facebook Posts Reaction Prediction,” in *International Conference on Agents and Artificial Intelligence*, 2017, vol. 2, no. 1, pp. 211–220.
- [23] S. X. Mashal and K. Asnani, “Emotion intensity detection for social media data,” in *Proc. Int. Conf. Computing Methodologies and Communication (ICCMC 2017)*, 2017, pp. 155–158.
- [24] A. N. Rohman, E. Utami, and S. Raharjo, “Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing,” *Eksplora Informartika*, 2019. .
- [25] M. Saif, “NRC Emotion Lexicon.” <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.