

#AkuGalau: Korpus Bahasa Indonesia untuk Deteksi Emosi dari Teks

Julius Bata
Program Studi Sistem informasi, Fakultas Teknik
Universitas Katolik Indonesia Atma Jaya, Jakarta
e-mail: julius.victor@atmajaya.ac.id

ABSTRAK

Deteksi emosi dari teks merupakan persoalan klasifikasi teks berdasarkan jenis emosi. Ketersediaan data berupa korpus emosi memegang peranan penting dalam deteksi emosi. Namun demikian, sebagian besar korpus untuk deteksi emosi tersedia dalam bahasa Inggris. Kondisi tersebut menjadi persoalan utama ketika melakukan deteksi emosi untuk teks Bahasa Indonesia. Korpus teks emosi untuk bahasa Indonesia masih sangat terbatas. Oleh karena itu, penelitian ini berfokus pada pengembangan korpus teks berbahasa Indonesia. Pengembangan korpus emosi berbahasa Indonesia ini sebagai langkah awal dalam penelitian deteksi emosi dari teks bahasa Indonesia. Sumber data yang digunakan untuk mengembangkan korpus adalah tweet. Proses anotasi dilakukan secara otomatis berdasarkan hashtag (#) emosi yang terdapat dalam tweet dengan lima jenis emosi: senang, sedih, marah, takut dan cinta. Penelitian ini menghasilkan korpus teks emosi bahasa Indonesia yang terdiri dari 500 tweet lengkap dengan label emosi pada tingkat *superordinate* dan tingkat *basic*. Percobaan deteksi emosi dilakukan untuk menguji korpus dengan menggunakan metode Naive Bayes. Tingkat akurasi dari percobaan mencapai 82 %, hasil ini menunjukkan bahwa korpus dapat digunakan dalam deteksi emosi.

Kata kunci: Deteksi emosi, Deteksi emosi dari teks, Korpus emosi Indonesia, Hashtag emosi.

ABSTRACT

Detection of emotions from text is a problem of text classification based on the type of emotion. The availability of an emotional corpus plays an essential role in the detection of emotions. However, most corpus for emotional detection is available in English. This condition is a major problem when developing a system for detecting emotions from Indonesian texts. The emotional text corpus for Indonesian is very limited. Therefore, this research focuses on the development of Indonesian text emotional corpus. The development of such a corpus is the first step in the study of detecting emotions from the Indonesian text. The data source used to develop the corpus is a tweet. The annotation process is done automatically based on the hashtag (#) of emotions contained in a tweet with five types of emotions: happy, sad, angry, afraid, and love. This research produced an Indonesian emotional text corpus consisting of 500 complete tweets with emotional labels at the superordinate and basic levels. Emotion detection experiments were conducted to test the corpus using the Naive Bayes method. The accuracy of the experiments reached 82%, these results indicate that the corpus can be used in text emotion detection.

Keywords: *Emotion detection, Text emotion detection, Indonesian emotion corpus, emotion hashtag.*

PENDAHULUAN

Bidang penelitian Analisis Sentimen memiliki dua sub bidang penelitian utama yaitu 1) *opinion mining* dan 2) *emotion mining* [1]. Beberapa tahun terakhir, bidang *emotion mining* menjadi penelitian yang paling banyak dilakukan. Secara khusus, deteksi emosi menjadi fokus penelitian dalam bidang *emotion mining*. Hal ini karena deteksi emosi dapat diterapkan pada berbagai bidang seperti *e-learning*, analisis perilaku konsumen dan interaksi manusia-komputer [2].

Berbagai metode telah digunakan untuk melakukan deteksi emosi dari teks. Metode tersebut dapat dikelompokkan menjadi *lexicon-based* dan *machine learning-based* [1][3]. Metode *lexicon-based* menggunakan leksikon emosi, sedangkan pendekatan berbasis *machine learning-based* menggunakan metode pembelajaran mesin untuk menghasilkan model deteksi emosi. Kedua metode tersebut, leksikon maupun *machine learning*, bergantung pada ketersediaan data atau korpus [4].

Korpus adalah kumpulan data teks yang sudah diberi label satu atau lebih jenis emosi. Korpus digunakan sebagai data untuk melatih model deteksi pada metode berbasis *machine learning*. Sedangkan pada metode berbasis leksikon, deteksi dilakukan berdasarkan kamus kata atau leksikon. Umumnya leksikon emosi hanya bekerja dengan baik untuk suatu domain tertentu. Oleh karena itu, korpus emosi dapat digunakan untuk mengembangkan leksikon untuk domain tertentu.

Ketersediaan korpus menjadi sangat penting dalam deteksi emosi dari teks. Mayoritas penelitian deteksi emosi dilakukan untuk bahasa Inggris. Hal ini menyebabkan sebagian besar data maupun *tool* deteksi emosi tersedia untuk bahasa Inggris. Korpus emosi yang tersedia juga

mayoritas dalam bahasa Inggris seperti *Twitter Emotion Corpus* (TEC) [6]. Pada sisi yang lain, penelitian deteksi emosi mulai banyak dilakukan untuk bahasa selain bahasa Inggris. Namun demikian, persoalan utama yang dihadapi ketika mengembangkan sistem deteksi emosi adalah ketersediaan data atau korpus.

Bahasa Indonesia merupakan bahasa yang termasuk dalam golongan “*under-resource language*” dimana sangat sedikit tersedia korpus teks bahasa Indonesia untuk deteksi emosi. Kondisi ini menjadi kendala utama dalam pengembangan sistem deteksi emosi. Oleh karena itu, makalah ini berfokus pada pengembangan korpus teks untuk deteksi emosi. Secara khusus kontribusi makalah ini sebagai berikut :

- Mengembangkan korpus teks untuk deteksi emosi bahasa Indonesia menggunakan data tweet
- Mengukur kinerja korpus dengan melakukan uji coba deteksi emosi

KAJIAN PUSTAKA

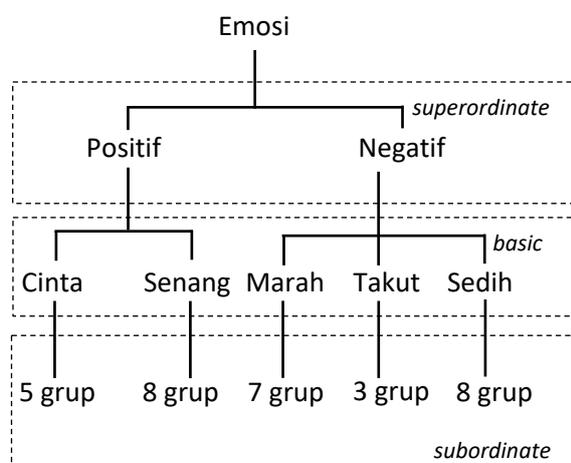
Permasalahan dalam makalah ini terkait dengan pengembangan korpus untuk deteksi emosi. Pada bagian ini akan dijelaskan model emosi yang digunakan serta penelitian – penelitian terkait deteksi emosi dan pengembangan korpus.

A. Model Emosi

Model emosi merupakan dasar dari setiap sistem deteksi emosi. Model emosi menyediakan kategori jenis emosi yang akan digunakan sebagai label dalam deteksi emosi. Secara umum terdapat dua model emosi yaitu kategorial dan dimensional [1][2]. Pada model kategorial, setiap jenis emosi merupakan satu kategori terpisah yang tidak saling terkait. Sedangkan pada model dimensional, emosi dipetakan dalam ruang berdimensi dua

atau lebih. Pada model dimensional, satu jenis emosi memiliki keterkaitan dengan emosi yang lain. Penelitian ini menggunakan model emosi yang diadaptasi dari struktur emosi leksikon bahasa Indonesia [7].

Model struktur emosi leksikon bahasa Indonesia memiliki tiga tingkat emosi [7], yaitu 1) tingkat *superordinate*, 2) tingkat *basic*, dan 3) tingkat *subordinate*. Tiga tingkat model struktur leksikon emosi teks bahasa Indonesia ditunjukkan Gambar 1. Pada tingkat *superordinate* memiliki dua kategori emosi yaitu positif dan negatif. Terdapat lima jenis emosi pada tingkat *basic*, dimana kategori emosi positif memiliki dua jenis emosi: cinta dan senang sedangkan kategori emosi negatif memiliki tiga jenis emosi: marah, takut dan sedih. Pada tingkat *subordinate*, kategori positif terdiri dari 13 grup dan kategori negatif memiliki 18 grup. Setiap grup pada tingkat *subordinate* memiliki sejumlah kata yang menunjukkan jenis emosi tertentu. Sebagai contoh: kategori basic: “cinta” memiliki lima grup, salah satunya adalah cinta yang memiliki beberapa kata kunci seperti “setia”, “kangen”, “kasmaran”, dan “mesra”.



Gambar 1. Struktur emosi teks bahasa Indonesia[7]

B. Deteksi Emosi dan Korpus

Deteksi emosi merupakan proses menentukan jenis emosi dari suatu teks. Diberikan sebuah teks, sistem deteksi emosi akan memilih dan menentukan emosi yang termuat pada teks tersebut. Berbagai penelitian deteksi emosi telah dilakukan, salah satu topik penting dalam penelitian deteksi emosi adalah pengembangan korpus emosi.

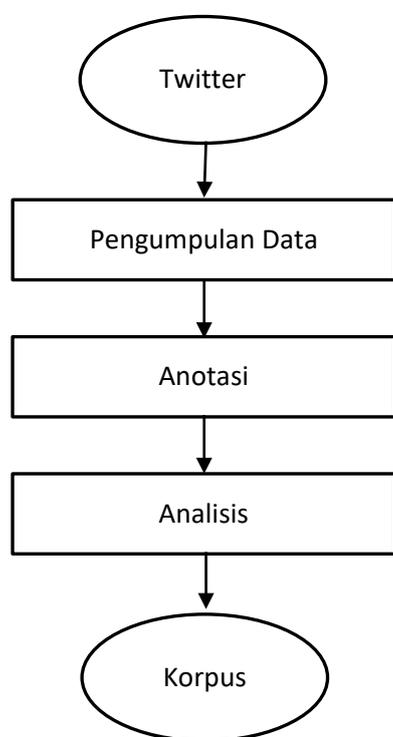
Korpus emosi bahasa Inggris telah dikembangkan menggunakan berbagai data. Data blog digunakan sebagai korpus dalam penelitian [8]. Sejak twitter mulai banyak digunakan, maka korpus juga mulai dikembangkan dengan menggunakan data tweet. Pengumpulan tweet berdasarkan hashtag telah dilakukan oleh [6]. Sebanyak 21.000 tweet dikumpulkan berdasarkan enam jenis emosi Ekman. Pencarian dilakukan menggunakan hashtag: ‘#anger’, ‘#disgust’, ‘#fear’, ‘#happy’, ‘#sadness’, dan ‘#surprise’. Korpus ini disebut *Twitter Emotion Corpus* (TEC).

Sebagian besar penelitian analisis sentimen dan deteksi emosi dilakukan menggunakan teks berbahasa Inggris. Namun demikian, sejumlah penelitian deteksi emosi dan analisis sentimen untuk teks berbahasa Indonesia telah dilakukan. Data tweet digunakan dalam pengembangan korpus [9] dan leksikon [10] untuk analisis sentimen berbahasa Indonesia. Deteksi emosi dari teks bahasa Indonesia juga sudah dilakukan. Deteksi emosi pembaca berita online dilakukan oleh [11]. Penelitian menggunakan korpus berita online yang terdiri dari 1575 artikel. Jenis emosi yang digunakan sebanyak 8 yaitu: senang, terhibur, terinspirasi, tidak peduli, terganggu, sedih, takut dan marah. Pada penelitian yang lain, model emosi Ekman digunakan sebagai kategori dalam deteksi emosi [12]. Data yang digunakan adalah komentar pada situs www.youtube.com. Sebanyak 8115

komentar dikumpulkan dan diberi label emosi oleh dua orang anotator. Data tweet juga sudah digunakan dalam deteksi emosi dari teks berbahasa Indonesia [13]. Tweet dikelompokkan ke dalam lima jenis emosi: senang, marah, takut, sedih dan kaget. Jumlah tweet yang digunakan sebanyak 250 tweet.

METODE PENELITIAN DAN PERANCANGAN

Penelitian ini berfokus pada pengembangan korpus emosi bahasa Indonesia. Korpus emosi merupakan kumpulan data teks beserta label emosi untuk setiap data dalam korpus. Pengembangan korpus terdiri dari tiga tahap utama yaitu: pengumpulan data, anotasi, dan analisa. Tiga tahapan tersebut seperti yang ditunjukkan Gambar 2.



Gambar 2. Tahap pengembangan korpus

A. Pengumpulan data

Pada penelitian ini data yang digunakan berupa data dari media sosial. Media sosial menjadi salah satu sumber data untuk mengembangkan korpus pada penelitian deteksi emosi. Media sosial yang digunakan yaitu Twitter (www.twitter.com) dan data yang dikumpulkan berupa data tweet. Twitter menyediakan sejumlah API (*Application Programming Interface*) yang dapat digunakan oleh pengembang untuk mengakses data tweet. Salah satu API yang dapat digunakan untuk mengakses tweet yaitu *Streaming API*. *Streaming API* memungkinkan untuk mengakses tweet secara *real-time*. Akses *Streaming API* menggunakan *library* Tweepy yang berbasis bahasa pemrograman Python. Parameter pencarian yang digunakan dalam penelitian ini adalah sejumlah kata kunci yang terkait dengan suatu jenis emosi (contoh: #senang, #marah).

Terdapat satu tahapan lagi sebelum masuk pada tahap anotasi. Sebagian besar data tweet mengandung informasi yang tidak berkaitan dengan proses deteksi emosi. Selain itu, terkadang dalam tweet terdapat karakter dalam bentuk unicode. Oleh karena perlu melakukan pra-proses sebelum masuk dalam proses anotasi.

Tahap pra-proses dimulai dengan menghapus nama pengguna yang terdapat didalam tweet. Nama pengguna ditandai dengan karakter '@' pada awal kata. Selanjutnya dilakukan proses penghapusan URL dan kode HTML yang terdapat pada tweet. Setelah itu dilakukan pengecekan simbol *unicode*. Tidak semua karakter *unicode* dihapus, sebagai contoh: simbol "\uD83D\uDE03" yang menunjukkan emoticon "☺" tidak dihapus. Setelah proses konversi karakter *unicode*, dilanjutkan dengan proses tokenisasi. Tokenisasi adalah proses memecah tweet menjadi setiap kata yang menyusunnya.

Proses ini dilakukan untuk memudahkan proses penghapusan *stop-word*. *Stop-word* merupakan kumpulan kata – kata yang memiliki pengaruh kecil terhadap proses deteksi emosi. Pada tahap ini setiap kata penyusun tweet akan dihapus jika kata tersebut terdapat dalam daftar *stop-word*.

B. Anotasi

Tahap berikut setelah pengumpulan data adalah anotasi. Anotasi merupakan proses pemberian label kepada setiap data tweet dalam corpus. Label yang digunakan merupakan jenis emosi seperti “senang”, “sedih”, dan “marah”. Pada umumnya label yang digunakan dalam deteksi emosi dari teks berjumlah enam. Label emosi tersebut sesuai dengan enam jenis emosi dasar yang diusulkan oleh Ekman [14]: marah, jijik, takut, senang, sedih dan kaget. Berbeda dengan penelitian – penelitian sebelumnya, jenis emosi yang digunakan dalam penelitian ini adalah berdasarkan model struktur leksikon emosi bahasa Indonesia [7].

Pada penelitian ini proses anotasi dilakukan untuk dua tingkat emosi yaitu tingkat *superordinate* dan tingkat *basic*. Setiap tweet diberi dua label, label pertama menunjukkan kategori *superordinate*, sedangkan label kedua menunjukkan jenis emosi pada kategori *basic*.

Label emosi setiap tweet ditentukan berdasarkan hashtag (#) yang terdapat dalam tweet. Pada umumnya hashtag (#) digunakan untuk mempertegas atau menunjukkan pendapat serta perasaan. Tweet hanya diberi label jika terdapat hashtag (#) emosi di akhir tweet.

C. Analisis

Tahap terakhir yaitu analisis korpus. Korpus emosi merupakan sumber data yang digunakan untuk melatih model deteksi ketika menggunakan metode

berbasis *machine learning*. Oleh karena itu, salah satu cara untuk mengukur kualitas dari korpus yang dikembangkan adalah dengan menggunakan korpus tersebut dalam sistem deteksi emosi.

Pada penelitian ini analisis dilakukan dengan cara menggunakan korpus dalam proses deteksi emosi secara otomatis. Untuk melakukan hal ini maka dilakukan sejumlah percobaan deteksi emosi. Deteksi emosi menggunakan metode Naive Bayes [15]. Naive Bayes merupakan metode pembelajaran mesin yang berdasar pada teorema Bayes dengan asumsi bahwa setiap fitur tidak saling bergantung satu dengan lainnya. Meskipun pada kasus nyata sangat jarang ditemui data dengan kondisi tersebut, Naive Bayes memberikan unjuk kerja yang baik ketika digunakan dalam klasifikasi teks.

Secara garis besar metode Naive Bayes memiliki dua tahap utama yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses perhitungan bobot probabilitas setiap kata dalam korpus terhadap suatu jenis emosi. Hasil dari tahap pelatihan adalah daftar kata dan bobotnya untuk setiap jenis emosi. Selanjutnya daftar kata dan bobot tersebut akan digunakan dalam proses klasifikasi. Proses klasifikasi bertujuan untuk memberi label jenis emosi kepada teks. Klasifikasi dilakukan dengan menjumlahkan bobot setiap kata dalam teks. Selanjutnya total bobot untuk setiap jenis emosi akan dibandingkan. Jenis emosi untuk data yang baru ditentukan berdasarkan total bobot terbesar. Metode deteksi Naive Bayes selanjutnya diukur unjuk kerjanya dengan menggunakan indikator akurasi. Akurasi dihitung dengan persamaan berikut :

$$Akurasi = \frac{JumlahDeteksiBenar}{JumlahData} \times 100$$

JumlahDeteksiBenar menunjukkan jumlah data yang sesuai antara jenis emosi hasil deteksi dengan jenis emosi data tersebut. JumlahData menunjukkan jumlah keseluruhan data yang digunakan dalam proses uji coba.

HASIL DAN PEMBAHASAN

Makalah ini membahas tentang pengembangan korpus teks bahasa Indonesia untuk deteksi emosi. Teks yang digunakan untuk mengembangkan korpus adalah tweet. Tweet dikumpulkan dari twitter melalui proses *crawling*. Hasil proses *crawling* berupa tweet mentah yang masih memuat karakter – karakter *emoticon* dan juga berbeda – beda format penulisan seperti terlihat pada Gambar 3.

1	1436251842.94::Repvblik-Duri cinta\ud83c\udfb5 @TarraBudir
2	1436251843.16::Aku sedang memilah dan memilih kata, untuk
3	1436251843.6::RT @KamusCewek: Cinta bukan sesuatu yg bisa

Gambar 3. Data hasil *crawling*

Selanjutnya data hasil *crawling* tersebut diproses awal untuk menghasilkan tweet bersih. Proses awal bertujuan untuk membersihkan data tweet dan merubah bentuk kedalam format yang sama untuk setiap data. Hasil dari proses awal adalah data tweet bersih. Selanjutnya data bersih diberi label emosi atau proses anotasi.

Proses anotasi dilakukan secara otomatis yaitu dengan melihat hashtag emosi yang terdapat pada tweet. Sebagai contoh diberikan tweet berikut :

Yeyeee. Akhirnya kita ada juga foto di dalam air. #senang

Tweet tersebut diberi label “positif” dan “senang”. Label positif untuk tingkat *superordinate* dan label senang untuk tingkat *basic*. Pada tahap anotasi juga dilakukan proses manual anotasi. Proses manual perlu dilakukan karena pada beberapa tweet terdapat lebih dari satu

kata kunci hashtag emosi. Selain itu, anotasi manual juga dilakukan untuk menghilangkan tweet yang posisi hashtag emosi tidak berada pada akhir tweet.

Korpus teks berbahasa Indonesia yang dihasilkan seperti yang ditunjukkan Gambar 4.

senang-19 pos senang bangga dong dapat motret biota langka
senang-20 pos senang ini bahagiaku ::kissing_heart bahagia
senang-21 pos senang berusaha saja lebih paham bersabar unt
senang-22 pos senang makasih big bos #oleholeh #bali

Gambar 4. Korpus Bahasa Indonesia

Kolom pertama “senang-19” menunjukkan id, kolom kedua “pos” adalah jenis emosi tingkat *superordinate*, kolom ketiga “senang” menunjukkan jenis emosi tingkat *basic* dan kolom terakhir merupakan isi dari tweet.

Selanjutnya dilakukan proses analisis korpus. Analisis dilakukan dengan melakukan percobaan deteksi emosi. Percobaan deteksi emosi menggunakan korpus dengan lima jenis emosi yaitu marah, cinta, sedih, senang dan takut. Setiap jenis emosi terdiri dari 100 data. Korpus kemudian dibagi menjadi dua data yaitu data latih dan data uji. Data latih digunakan untuk membentuk model deteksi sedangkan data uji digunakan untuk menguji model deteksi. Pada setiap percobaan, data latih berjumlah 90 % dan data uji 10 % dari korpus. Percobaan deteksi dilakukan sebanyak 10 kali dan nilai akurasi akhir merupakan nilai rata – rata dari 10 kali percobaan. Tingkat akurasi yang diperoleh dari percobaan deteksi pada tingkat *superordinate* ditunjukkan Tabel 1.

Tabel 1. Hasil deteksi tingkat *superordinate*

Tingkat	Akurasi
<i>Superordinate</i> (positif, negatif)	82,02 %

Tingkat akurasi yang cukup tinggi yaitu mencapai 82,02 % menunjukkan metode Naive Bayes dapat bekerja dengan baik untuk deteksi emosi pada tingkat *superordinate*. Pada tingkat *superordinate* hanya memiliki dua jenis emosi. Jumlah kelas yang sedikit dapat meningkatkan kinerja dari metode klasifikasi. Selain itu, fitur kata dalam masing – masing kelas sudah dapat menjadi pembeda sehingga hasil klasifikasi memiliki nilai akurasi yang cukup tinggi.

Percobaan selanjutnya adalah melakukan deteksi untuk tingkat emosi *basic*. Pada tingkat *basic*, emosi positif memiliki dua jenis emosi: Senang dan Cinta, sedangkan untuk emosi negatif memiliki tiga jenis emosi: Marah, Sedih dan Takut. Hasil deteksi untuk emosi positif ditunjukkan oleh Tabel 2.

Tabel 2. Hasil deteksi tingkat *basic*

Tingkat	Akurasi
Positif (Cinta, Senang)	83,61 %
Negatif (Marah, Takut, Sedih)	67,92 %

Seperti terlihat pada Tabel 2, nilai akurasi untuk jenis emosi positif juga mencapai nilai yang cukup tinggi yaitu 83,61 %. Seperti halnya pada tingkat *superordinate*, jenis emosi pada tingkat *basic* positif hanya memiliki dua jenis emosi. Berbeda dengan jenis emosi positif, pada jenis emosi negatif kinerja model deteksi mengalami penurunan. Salah satu penyebab hal ini terjadi adalah karena jenis emosi negatif memiliki jenis emosi lebih banyak dibandingkan jenis emosi positif. Jenis emosi negatif memiliki tiga jenis emosi. Selain itu, hasil analisis terhadap fitur kata dalam jenis emosi negatif menunjukkan banyak fitur kata yang memiliki bobot hampir sama untuk ketiga jenis emosi pada jenis emosi negatif. Hal ini juga menunjukkan bahwa kata – kata yang digunakan untuk

mengungkapkan emosi untuk jenis emosi marah, takut, dan sedih memiliki kesamaan. Sebagai contoh, kata “marah” tidak hanya ditemui pada data jenis emosi marah, tetapi juga banyak terdapat dalam data jenis emosi takut.

Tujuan utama pengembangan korpus adalah untuk digunakan dalam proses deteksi emosi. Berdasarkan hasil percobaan maka korpus yang dikembangkan sudah berhasil digunakan dalam proses deteksi. Korpus dapat bekerja dengan baik pada proses deteksi tingkat *superordinate* dan tingkat *basic* untuk jenis emosi positif. Sedangkan untuk emosi negatif, korpus belum menunjukkan kinerja yang cukup baik.

KESIMPULAN

Makalah ini memaparkan hasil penelitian dalam mengembangkan korpus teks berbahasa Indonesia. Korpus ini dapat digunakan dalam penelitian deteksi emosi dari teks bahasa Indonesia. Korpus dikembangkan menggunakan tweet. Proses anotasi dilakukan berdasarkan hastag (tagar #) yang terdapat pada masing – masing tweet.

Sejumlah percobaan deteksi emosi dilakukan dengan menggunakan data korpus. Korpus memiliki 500 data tweet yang telah diberi label emosi baik pada tingkat *superordinate* maupun pada tingkat *basic*. Hasil percobaan menunjukkan proses deteksi pada jenis emosi positif lebih baik jika dibandingkan dengan jenis emosi negatif.

Berdasarkan hasil tersebut maka penelitian selanjutnya akan difokuskan pada menambah jumlah data dalam korpus. Selain itu, penelitian selanjutnya juga akan dilakukan untuk membuat model deteksi emosi. Penelitian berikut juga dapat difokuskan pada jenis emosi negatif yaitu takut, marah, dan sedih karena hasil

deteksi untuk jenis emosi negatif masih rendah.

DAFTAR PUSTAKA

- [1] A. Yadollahi, A.G. Shahraki, and O.R. Zaiane, “Current State of Text Sentiment Analysis from Opinion to Emotion Mining,” *ACM Comput. Surv.*, vol. 50, no. 2, hlm. 1–33, 2017.
- [2] K. Sailunaz, M. Dhaliwal, J. Rokne, and R. Alhaji, “Emotion detection from text and speech: a survey,” *Soc. Netw. Anal. Min.*, vol. 8, no. 1, 2018.
- [3] U. Krcadinac and P. Pasquier, “Synesketch: An Open Source Library for Sentence -Based Emotion Recognition,” *IEEE Trans. Affect. Comput.*, vol. 4, no. 3, hlm. 312 – 325, 2013.
- [4] C. Bosco, V. patti and A. Bolioli, “Developing Corpora for Sentiment Analysis: The Case of Irony and Senty-TUT,” *IEEE Intell. Syst.*, vol. 28, April, hlm. 55 – 63, 2013.
- [5] S. M. Mohammad, “#Emotional Tweets,” dalam First Joint Conference on Lexical and Computational Semantics (*sem), hlm. 246 – 255, 2012.
- [6] P. R. Shaver, U. Murdaya, and R. C. Fraley, “Structure of the Indonesian emotion lexicon,” *Asian Journal of Social Psychology*, vol. 4., hlm. 201 – 224, 2001.
- [7] F. Keshtkar and D. Inkpen, “A Hierarchical approach to mood classification in blogs,” *Nat. Lang. Eng.*, vol. 18, no.1, hlm. 61 – 81, 2012.
- [8] A. F. Wicaksono, C. Vania, B. D. T, and M. Adriani, “Automatically Building a Corpus for Sentiment Analysis on Indonesian Tweets,” dalam Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation, PACLIC 2014, pp. 185–194, 2014.
- [9] C. Vania, M. Ibrahim, and M. Adriani, “Sentiment Lexicon Generation for an Under-Resourced Language,” *Int. J. Comput. Linguist. Appl.*, vol. 5, no. 1, pp. 59–72, 2014.
- [10] A. Sulistya, F. Thung, and D. Lo, “Inferring Spread of Readers’ Emotion Affected by Online News,” Ciampaglia G., Mashhadi A., Yasser T. Soc. Informatics. SocInfo 2017. Lect. Notes Comput. Sci., vol. 10539, pp. 426–439, 2017.
- [11] J. Savigny and A. Purwarianti, “Emotion classification on youtube comments using word embedding,” dalam Proceedings - 2017 International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2017, hlm. 1–5, 2017.
- [12] R. M. Cahyaningtyas and R. Kusumaningrum, “Emotion Detection of Tweets in Indonesian Language Using LDA and Expression Symbol Conversion,” dalam Proc. of 2017 1st Int. Conf. on Informatics and Computational Sciences (ICICoS), pp. 253–258, 2017.
- [13] P. Ekman, “An argument for basic emotions,” *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [14] D. M. Farid, L. Zhang, C. M. Rahman, M. A. Hossain, and R. Strachan, “Hybrid decision tree and naive bayes classifiers for multi-class classification tasks,” *Expert Syst. Appl.*, vol. 41, no. 4 (2), pp. 1937–1946, 2014.