

Studi Pengembangan Model Pengenalan Emosi Pada Teks Media Sosial

Aldy Rialdy Atmadja

Jurnal Algoritma
Sekolah Tinggi Teknologi Garut
Jl. Mayor Syamsu No. 1 Jayaraga Garut 44151 Indonesia
Email : jurnal@sttgarut.ac.id

aldyrialdyatmadja@sttgarut.ac.id

Abstrak – Media sosial mempunyai peranan yang teramat penting dalam dunia teknologi informasi. Aktivitas yang dilakukan pada media sosial dapat menggambarkan seseorang dalam kondisi nyata, tidak jarang pula pengguna media sosial mencurahkan perasaan atau suasana hatinya pada suatu media sosial yang membuat media sosial dapat diukur, salah satunya pengenalan emosi pada teks media sosial. Model pengenalan emosi yang dihasilkan pada penelitian ini digunakan untuk mengenali emosi pada teks dengan tahapan pengumpulan data teks, praproses teks, pemilihan dan ekstraksi fitur, dan klasifikasi emosi. Model ini diharapkan dapat menjadikan baseline dalam riset yang berkenaan dengan klasifikasi teks pada media sosial, khususnya dalam mengenali emosi pada teks media sosial.

Kata Kunci – media sosial, emosi, klasifikasi, pengenalan emosi

I. PENDAHULUAN

Saat ini, media sosial mempunyai peranan yang teramat penting dalam dunia teknologi informasi. Hal tersebut dibuktikan dengan banyaknya pengguna pada beberapa media sosial. Berdasarkan hasil riset yang dikemukakan Asosiasi Pengguna Jasa Internet Indonesia, penggunaan sosial media terbukti menduduki peringkat kedua setelah chatting dalam perilaku pengguna internet di Indonesia[1]. Disamping itu, pada situs internetworldstats.com tercatat pula bahwa Indonesia termasuk peringkat 2 dengan pengguna terbanyak pada media social Facebook di Asia[2].

Hal tersebut merupakan akibat dari kegiatan media sosial yang didukung akses tanpa batas, serta penggunaan *smartphone* untuk media sosial membuat aktivitas dunia maya, seperti *update status*, kegiatan *posting* aktivitas, *tweet*, dan sebagainya mudah dilakukan. Aktivitas yang dilakukan pada media sosial dapat menggambarkan seseorang dalam kondisi nyata, tidak jarang pula pengguna media sosial mencurahkan perasaan atau suasana hatinya pada suatu media sosial. Rutinitas dari penggunaan media sosial tersebut memungkinkan pemantauan terhadap akun media sosial mudah dilakukan. Sehingga beberapa penelitian bermula dari analisis terhadap media sosial seperti sentimen percakapan media sosial [3][4][5], tingkat emosi seseorang [6], sampai pada media sosial digunakan untuk penanggulangan bencana.

Salah satu penelitian dalam melakukan analisis teks pada media sosial yaitu dengan mengukur emosi seseorang melalui media sosial. Pengukuran emosi pada teks media sosial tidaklah mudah, hal ini dikarenakan banyak parameter dalam menentukan emosi tersebut. Menurut Ekman, terdapat beberapa karakteristik yang mempengaruhi emosi seseorang yaitu adanya korelasi dengan kejadian sebelumnya, kejadian yang tidak disangka, ataupun hubungan antar individu [7]. Sehingga bisa dikatakan, analisis terkait emosi merupakan sesuatu yang cukup kompleks.

Beberapa penelitian telah menunjukkan bahwa identifikasi emosi seseorang dapat dilakukan pada teks. Secara general, terdapat dua pendekatan yang dilakukan yang dilakukan dalam klasifikasi emosi, yakni pendekatan dengan menggunakan korpus (*corpus based*) dan pendekatan berbasis pengetahuan

(*knowledge based*). Pendekatan dengan korpus yakni pendekatan yang dilakukan berdasarkan perhitungan statistik dengan melakukan ekstraksi terhadap data korpus menjadi fitur dengan berdasarkan n-grams. Salah satu contoh penelitian dengan pendekatan *knowledge* yakni penelitian Strapparava dan Mihaelcea yang menggunakan *WordNet-Affect* dengan menghitung kemunculan kata pada suatu dokumen (*WordNet-Affect presence*)[8]. Sedangkan pendekatan berbasis pengetahuan yakni identifikasi emosi dilakukan dari kumpulan kata ataupun istilah yang merepresentasikan emosi berupa *synsets*[9].

Dalam mengenali emosi terdapat dua metode yang sering digunakan dalam algoritma pembelajaran (*learning*), yakni menggunakan *supervised learning* dan *unsupervised learning*. *Supervised learning* dapat dilakukan dengan melakukan pelabelan terhadap setiap data ataupun dokumen. Proses ini akan menjadi acuan untuk digunakan dalam proses *learning* dengan algoritma tertentu seperti *Support Vector Machine (SVM)*, *k-Nearest Neighbour (kNN)*, *Decision Tree*, *Bayesian Network*, *Linear Regression*, *Linear Discriminant Analysis*, *Naive Bayes* [10][11][12].

Penelitian ini berfokus pada studi literature dalam pengembangan model untuk pengenalan emosi dalam teks media sosial. Pada paper ini terbagi menjadi lima bagian yaitu pendahuluan, bagian 2 membahas mengenai tinjauan pustaka yang dipergunakan dalam penelitian ini, bagian 3 membahas mengenai metode ataupun usulan yang dilakukan dalam pengembangan model untuk mengenali emosi pada teks media sosial, bagian 4 membahas mengenai hasil dan pembahasan, dan selanjutnya membahas tentang kesimpulan dan saran..

II. TINJAUAN PUSTAKA

Adanya media sosial pada awal mula millennium, memacu terjadinya komunikasi yang terjadi didunia maya. *Internet World Statistics* mencatatkan bahwa pengguna Internet di Indonesia termasuk dalam pengguna terbanyak di benua Asia[2]. Media sosial merupakan salah satu yang sering diakses oleh para pengguna internet di Indonesia, seperti twitter, facebook dan instagram. Dalam media sosial memungkinkan seseorang untuk membuat suatu konten dan membaginya kepada banyak orang. Aktivitas tersebut memberikan kebebasan, sehingga memungkinkan didalamnya terdapat struktur bahasa nonformal, penggunaan singkatan dan bahasa yang tidak standar[4]. Begitu juga dengan pemilihan kata dan istilah yang tidak standar sering digunakan dalam media sosial[6]. Berikut ini merupakan contoh percakapan pada media sosial.

Percakapan 1: Selimut dua biji ga cukup, ini cuaca apa bgt _ _ _
 Percakapan 2: Hmpir mirip quotenya <user> pak :D pkiran yg mlangit
 Percakapan 3: hari ini..sakit fisik, sakit di hati juga..senyum aja deh ya :)

Gambar 1. Contoh Percakapan pada Media Sosial

Pada contoh tersebut percakapan dalam media sosial dapat diekstrak menjadi sebuah informasi dimana didalam percakapan tersebut terdapat emosi. Emosi tersebut dapat diwakili oleh kata-kata yang mempunyai makna emosi tertentu ataupun penggunaan emoticon yang membantu dalam mendeteksi adanya emosi dalam suatu percakapan pada media sosial.

Secara teori, identifikasi emosi dapat dilakukan dengan melihat emosi primer yang muncul, emosi primer ini dapat ditemukan pada manusia dan juga hewan. Emosi primer ini dikenal dengan emosi dasar yang ada pada setiap manusia. Identifikasi emosi yang telah ada saat ini seringkali menggunakan emosi dasar. Menurut Ekman terdapat enam emosi dasar yakni marah (*anger*), jijik (*disgust*), fear (takut), gembira (*joy*), sedih (*sad*), terkejut (*surprise*)[13]. Enam emosi dasar ini sering dipergunakan dalam berbagai penelitian, namun emosi dasar tersebut tidak hanya dikemukakan oleh Ekman, beberapa penelitian lainnya menggunakan kategori emosi yang lain. Pada penelitian tersebut, identifikasi emosi pada penelitian tersebut menggunakan 10 (sepuluh) kategori emosi dasar, yakni

marah (*anger*), jijik (*disgust*), takut (*fear*), bersalah (*guilt*), tertarik (*interest*), gembira (*joy*), sedih (*sadness*), malu (*shame*), terkejut (*surprise*) dan netral[14]. Kedua tipe emosi tersebut ditunjukkan pada Tabel 1 berikut.

Tabel 1. Tipe Emosi yang digunakan dalam Penelitian Klasifikasi Emosi

Ekman (1999)	Neviarouskaya, dkk (2010)
marah (<i>anger</i>), jijik (<i>disgust</i>), takut (<i>fear</i>), gembira (<i>joy</i>), sedih (<i>sad</i>), terkejut (<i>surprise</i>)	marah (<i>anger</i>), jijik (<i>disgust</i>), takut (<i>fear</i>), bersalah (<i>guilt</i>), tertarik (<i>interest</i>), gembira (<i>joy</i>), sedih (<i>sadness</i>), malu (<i>shame</i>), terkejut (<i>surprise</i>) dan netral

Pengklasifikasian emosi dilakukan untuk mengetahui emosi tersebut diekspresikan oleh seseorang melalui teks ataupun percakapan. Emosi pada teks dapat dinilai dengan pandangan subjektif tentang perasaan seseorang berkenaan dengan sesuatu hal. Penentuan emosi dalam suatu dokumen pada teks (artikel, judul berita, teks media sosial) mudah dilakukan ketika hanya terhadap satu emosi saja. Akan tetapi akan menjadi kompleks ketika beberapa emosi yang ditimbulkan dari dokumen tersebut. Berbagai penelitian telah dilakukan untuk dapat melakukan klasifikasi emosi pada teks. Bila dikategorikan ada dua tipe pendekatan yang dilakukan dalam klasifikasi emosi berdasarkan sumber data yang menjadi acuan klasifikasi.

1. Pendekatan dengan memanfaatkan data korpus

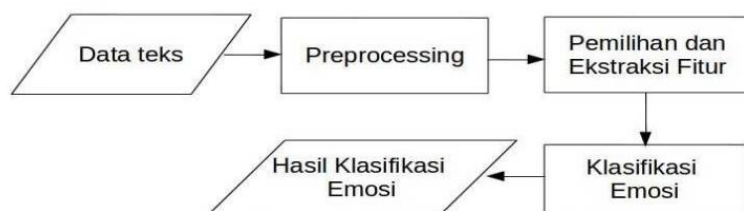
Pendekatan dengan menggunakan data korpus dilakukan dengan melakukan perhitungan statistik terhadap data yakni dengan menggunakan teknik unigram (n-gram) sebagai fitur. Adapun kelebihan pendekatan dengan menggunakan corpus data ini adalah pendekatannya bersifat otomatis yang menyerahkan sepenuhnya terhadap perhitungan data statistik, sehingga tidak membutuhkan para ahli untuk dapat merepresentasikan data tersebut[15]. Beberapa riset tentang klasifikasi emosi yang memanfaatkan pendekatan statistik diantaranya adalah penelitian yang dilakukan oleh Sofiana, dkk yang memanfaatkan fitur unigram, kemudian melakukan perhitungan bobot menggunakan tf-idf. Selanjutnya, algoritma K-Nearest Neighbor (KNN) dipergunakan untuk melakukan klasifikasi emosi[16].

2. Pendekatan dengan memanfaatkan Knowledge-Based

Identifikasi emosi dapat juga dilakukan dengan melakukan pendekatan berdasarkan sinonim atau kumpulan kata untuk menjelaskan emosi tertentu dalam suatu dokumen. WordNetAffect merupakan salah satu contoh dari pendekatan menggunakan knowledge-based. Studi yang dilakukan oleh Strapparava dan Mihalcea mempergunakan WordNetAffect, dimana pada penelitian tersebut diperlukan *baseline* untuk mengetahui adanya emosi dalam suatu teks dengan menghitung frekuensi kemunculan kata emosi berdasarkan WordNet-Affect pada suatu dokumen (*WN-Affect presence*)[9].

III. METODE

Secara umum dalam melakukan pengenalan emosi dalam teks dapat dilakukan seperti tahapan pada Gambar 2 berikut ini.



Gambar 2. Model Pengenalan Emosi dalam Teks

Pada gambar tersebut, model pengenalan emosi dimulai dari sebuah data teks. Data teks tersebut menjadi inputan yang dimasukkan dalam sebuah sistem. Sebelum memulai klasifikasi. Hal yang dilakukan yakni melakukan tahapan praproses. Tahapan tersebut dilakukan untuk menyederhanakan bentuk teks, dan juga menghilangkan teks dari noise dengan tujuan meningkatkan akurasi dari pengenalan emosi. Tahapan selanjutnya ekstraksi fitur dilakukan untuk mendapatkan fitur yang sesuai untuk melakukan klasifikasi teks. Selain itu pemilihan fitur dilakukan dengan tujuan untuk memastikan fitur yang dipilih sesuai dan menghasilkan model yang tepat.

Klasifikasi dilakukan untuk menghasilkan model yang sesuai, model tersebut diuji dengan menggunakan algoritma tertentu, seperti *Support Vector Machine (SVM)*, *k-Nearest Neighbour (kNN)*, *Decision Tree*, *Bayesian Network*, *Linear Regression*, *Linear Discriminant Analysis*, *Naive Bayess*. Setelah model klasifikasi dibuat, maka model tersebut menghasilkan hasil klasifikasi. Untuk memastikan hasil yang maksimal, maka klasifikasi tersebut perlu diuji dengan menggunakan pengukuran akurasi, sehingga model klasifikasi dapat digunakan dan menghasilkan hasil yang optimal.

IV. HASIL DAN PEMBAHASAN

Data merupakan hal yang paling penting dalam melakukan penelitian, beberapa penelitian dalam melakukan klasifikasi emosi menggunakan sumber data media sosial[6]. Media sosial yang dapat digunakan diantaranya facebook, twitter, instagram dan lain sebagainya. Namun yang perlu dipertimbangkan dalam melakukan pengenalan emosi melalui teks ini yakni jumlah dataset, jumlah data set yang banyak akan mempengaruhi hasil klasifikasi yang dihasilkan.

Tahapan selanjutnya yang merupakan tahapan paling penting yakni praproses. Praproses diperlukan untuk memastikan bahwa teks yang akan diproses telah bersih dari noise, yang menyebabkan klasifikasi emosi dapat dilakukan dengan maksimal, sehingga akan meningkatkan akurasi klasifikasi emosi. Tahapan ini diperlukan beberapa proses[5], seperti :

1. Tokenisasi
2. Normalisasi fitur
3. Casefolding
4. Clean number
5. Convert number, convert emoticon, convert word, convert negation
6. Stopword removal.

Langkah selanjutnya adalah melakukan pemilihan dan ekstraksi fitur. Pemilihan fitur dapat dilakukan dengan memilih beberapa fitur seperti dibawah seperti n-gram, posisi n-gram, Kamus Linguistic Inquiry and Word Count (LIWC), Multi Perspective Question Answering (MPQA) Lexicon, WordNetAffect, Part-of-Speech (POS), dan kata sifat (adjectives)[10]. Pembobotan juga dapat digunakan dengan menggunakan TF-IDF, dengan cara melakukan perankingan terhadap fitur yang telah diekstrak[17].

Proses klasifikasi emosi merupakan sebuah proses menemukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui. Klasifikasi dapat dilakukan dengan menggunakan beberapa cara, hal ini berlaku juga untuk klasifikasi emosi . Klasifikasi tersebut dapat dilakukan dengan menggunakan teknik pembelajaran dengan supervised, unsupervised dan juga semi-supervised [18]. Algoritma yang sering kali digunakan pada teknik pembelajaran secara *supervised* diantaranya yaitu *Support Vector Machine (SVM)*, *Decision Tree*, *Bayesian Network*, *Naive Bayess* [11][19]. Selain itu pembelajaran dengan unsupervised dapat dilakukan tanpa menggunakan label yang ada sehingga seringkali disebut dengan klustering[20].

Pada tahapan hasil klasifikasi emosi, maka akan dilakukan pengujian terhadap model yang sudah terbentuk dari proses klasifikasi. Dari model tersebut akan diuji data testing sehingga akan menghasilkan perkiraan kelas dari data test yang diujikan. Hasil penentuan klasifikasi tersebut akan dilakukan pengukuran

terhadap nilai akurasi yang dihasilkan dari model klasifikasi tersebut.

V. KESIMPULAN/RINGKASAN

Model tersebut dipergunakan sebagai *baseline* dalam penelitian untuk mengenali emosi dalam teks media sosial berdasarkan hasil komparasi dengan penelitian klasifikasi emosi terdahulu. Pada dasarnya untuk membuat suatu model pengenalan klasifikasi teks, data merupakan hal yang utama dalam melakukan prediksi terhadap hasil klasifikasi. Praproses merupakan tahapan yang dilakukan untuk memastikan teks bebas dari noise, serta pemilihan fitur perlu dilakukan dalam menghasilkan akurasi yang maksimal. Disamping penggunaan algoritma yang saat ini Support Vector Machine (SVM) masih menjadi acuan terbaik dalam melakukan klasifikasi teks.

Tentu saja model tersebut perlu dikaji ulang sehingga dapat dipergunakan untuk kasus ini saja maupun kasus lain yang erat dengan kaitan analisis terhadap percakapan media sosial. Disisi lain, penggunaan algoritma lain dan metode yang saat ini menjadi riset yang berkembang seperti penggunaan *Deep Learning* untuk klasifikasi dan pengenalan teks menjadi sesuatu yang menarik untuk dikaji secara mendalam.

UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terima kasih kepada kampus Sekolah Tinggi Teknologi Garut yang telah mendukung para dosennya untuk senantiasa mencurahkan tulisannya dalam bentuk karya ilmiah yang dipublikasikan pada khalayak luas.

DAFTAR PUSTAKA

- [1] APJII, "Penetrasi & Perilaku Pengguna Internet Indonesia 2017," *Asos. Penyelenggara Jasa Internet Indones.*, pp. 1–39, 2017.
- [2] "Asia Internet Use, Population Data and 2017, Facebook Statistics - December." [Online]. Available: <https://www.internetworldstats.com/stats3.htm>.
- [3] B. Subaeki, F. Gunawan, and A. R. Atmadja, "Penggunaan Metode Fuzzy Logic untuk Pemantauan Sentimen Brand pada Media Sosial," vol. 1, no. October, pp. 56–62, 2017.
- [4] A. R. Naradhipa and A. Purwarianti, "Sentiment Classification for Indonesian Messages in Social Media," *Int. Conf. Electr. Eng. Informatics*, no. July, pp. 2–5, 2011.
- [5] I. Sunni and D. H. Widyantoro, "Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik," *J. Sarj. Inst. Teknol. Bandung Bid. Tek. Elektro dan Inform.*, vol. 1, no. 2, pp. 200–206, 2012.
- [6] A. R. Atmadja and A. Purwarianti, "Comparison on the rule based method and statistical based method on emotion classification for Indonesian Twitter text," *2015 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2015 - Proc.*, 2016.
- [7] P. Ekman, "An Argument for Basic Emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [8] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," *Proc. 2008 ACM Symp. Appl. Comput. - SAC '08*, p. 1556, 2008.
- [9] A. Valitutti, C. Strapparava, and O. Stock, "Developing affective lexical resources," *PsychNology J.*, vol. 2, no. 1, pp. 61–83, 2004.
- [10] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing Twitter 'Big Data' for Automatic Emotion Identification," *SocialCom/PASSAT*, 2012.
- [11] R. Burget, J. Karásek, and Z. Smékal, "Recognition of emotions in Czech newspaper headlines," *Radioengineering*, vol. 20, no. 1, pp. 39–47, 2011.

- [12] R. Tokuhisa, K. Inui, and Y. Matsumoto, "Emotion classification using massive examples extracted from the web," *Proc. 22nd Int. Conf. Comput. Linguist. - COLING '08*, vol. 1, no. August, pp. 881–888, 2008.
- [13] P. Ekman, "Are There Basic Emotions?," *Psychological Review*, vol. 99, no. 3. pp. 550–553, 1992.
- [14] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "User study on AffectIM, an avatar-based Instant Messaging system employing rule-based affect sensing from text," *Int. J. Hum. Comput. Stud.*, vol. 68, no. 7, pp. 432–450, Jul. 2010.
- [15] S. Aman, "Recognizing emotions in text," *Www-Scf.Usc.Edu*, 2007.
- [16] L. Sofiyana, Z. Abidin, and H. Nurhayati, "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan K-Nearest Neighbor." 2012.
- [17] O. Nurdiana, Jumadi, and D. Nursantika, "Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia," *J. Online Inform.*, vol. 1, no. 1, pp. 59–63, 2016.
- [18] S. Mac Kim, "Recognising Emotions and Sentiments in Text," University of Sydney, 2011.
- [19] W. B. Zulfikar and N. Lukman, "Perbandingan Naive Bayes Classifier dengan Nearest Neighbor untuk Identifikasi Penyakit Mata," *JOIN (Jurnal Online Inform.*, vol. 1, no. 2, pp. 82–86, 2016.
- [20] N. W. S. Saraswati, "Text mining dengan metode naïve bayes classifier dan support vector machines untuk sentiment analysis," 2011.