

Klasifikasi Emosi Pada Sosial Media Menggunakan *Support Vector Machine* dan N-Gram

Riska Dwi Handayani¹, Ahmad Fauzi Anggi A K², Yusuf Wahyu Setiya Purtra³

¹Program Studi Manajemen Informatika, STMIK Bina Patria

²Program Studi Teknik Informatika, STMIK Bina Patria

³Program Studi Sistem Informasi, STMIK Bina Patria

¹riska@stmikbinapatria.ac.id, ²fauziariesta@stmikbinapatria.ac.id, ³yusuf@stmikbinapatria.ac.id

1. Abstract

The large number of Facebook social media users in Indonesia provides an opportunity for researchers to study various things through text, one of which is the translation of emotions through status update text which is the main feature of the social media. The status update feature is often used by the public to share information, express opinions on something or express opinions on experiences that are a trend of public attention. In addition, the status update feature is also often used by the public to share the emotions they are currently feeling. To analyze data from Facebook status updates which are irregular in nature, it is necessary to do text analysis using the text mining method.

The author proposes to conduct research on Facebook status updates using the text mining method. Where in this study using N-Gram level feature extraction and the classification method used is the Support Vector Machine (SVM) method. From the tests that have been carried out using this method it produces an accuracy value of 53%. The highest precision values are generated by several emotions, namely fear, shock, disgust, and anger with a value of 100%. Furthermore, the highest recall value is generated by the Happy emotion with a value of 100%. And finally, the highest f1-score is generated by the emotion of shock with a value of 67%.

Keywords: Text Mining, Classification, Emotions, Support Vector Machine (SVM), N-Gram, Social Media.

2. Abstrak

Banyaknya pengguna media sosial facebook di Indonesia memberikan peluang kepada para peneliti untuk mengkaji berbagai hal melalui teks, salah satunya adalah penerjemahan emosi melalui teks update status yang merupakan fitur utama dari media sosial tersebut. Fitur update status sering digunakan oleh masyarakat untuk membagikan informasi, menyampaikan pendapat terhadap sesuatu ataupun mengekspresikan opini pengalaman yang menjadi trend perhatian masyarakat. Selain itu fitur update status juga sering digunakan oleh masyarakat untuk membagikan emosi yang saat ini dirasakan. Untuk menganalisis data dari update status facebook yang sifatnya tidak teratur maka perlu dilakukan analisis teks dengan menggunakan metode text mining.

Penulis mengusulkan untuk melakukan penelitian terhadap update status facebook dengan menggunakan metode text mining. Dimana dalam penelitian ini menggunakan ekstraksi fitur N-Gram level serta metode klasifikasi yang digunakan adalah metode *Support Vector Machine (SVM)*. Dari pengujian yang telah dilakukan dengan menggunakan metode tersebut menghasilkan nilai akurasi sebesar 53%. Nilai presisi tertinggi dihasilkan oleh beberapa emosi yakni takut, kaget, jijik, dan marah dengan nilai 100%. Selanjutnya nilai recall tertinggi dihasilkan oleh emosi Bahagia dengan nilai 100%. Dan yang terakhir nilai f1-score tertinggi dihasilkan oleh emosi kaget dengan nilai 67%.

Kata kunci: : Text Mining, Klasifikasi, Emosi, Support Vector Machine (SVM), N-Gram, Sosial Media

This is an open access article under the [CC BY-SA](#) license.



3. Pendahuluan

Pesatnya perkembangan teknologi saat ini mengakibatkan banyak faktor berubah, salah satunya adalah cara berkomunikasi antar satu orang dengan orang yang lain. Salah satu teknologi yang populer dikalangan masyarakat adalah media sosial. Media

sosial seperti Facebook, Twitter, Instagram, Youtube dan Google+ sering sekali dimanfaatkan untuk sarana bertukar informasi, menyampaikan pendapat atau pandangan terhadap sesuatu dan mengekspresikan opini, pengalaman serta hal lain yang menjadi trend perhatian pada masyarakat [1].

Facebook merupakan media social yang sangat populer di kalangan masyarakat Indonesia. Berdasarkan data yang didapat dari Statista, terdapat 195,2 juta pengguna facebook di Indonesia dimana 31,8 pengguna berusia 18 dan 24 tahun [2]. Facebook mempunyai fitur update status yang bisa digunakan oleh pengguna yang akan membagikan peristiwa atau informasi yang mereka punya. Akan tetapi fitur itu tidak hanya digunakan untuk membagikan informasi saja tapi banyak juga pengguna yang menggunakan fitur tersebut untuk menyampaikan emosi yang pengguna rasakan.

Emosi merupakan keadaan kompleks dari pikiran yang dipengaruhi oleh peristiwa eksternal, perubahan fisiologis, atau hubungan dengan orang lain. Dengan tidak adanya kontak tatap muka untuk mengklasifikasi ekspresi wajah dan intonasi dalam suara, opsi alternatifnya adalah menguraikan emosi dari teks di layanan sosial media [3].

Untuk mengklasifikasi emosi dari teks pada layanan sosial media twitter dengan data yang tidak terstruktur maka perlu dilakukan analisis teks salah satunya dengan menggunakan Text Mining [3]. Text Mining merupakan proses untuk memperoleh informasi berkualitas tinggi dari suatu teks [4].

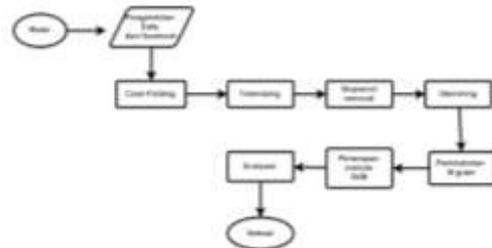
Terdapat beberapa peneliti yang telah menggunakan metode text mining ini dalam menemukan sebuah pola unik diantara yang pertama adalah penelitian yang dilakukan oleh Fajar Romadoni, Yuyun Umaidah dan Betha Nurina Sapi tentang text mining analisi sentiment menggunakan metode Support Vector Machine menghasilkan nilai akurasi sebesar 98,7% [5]. Kemudian penelitian lain tentang text mining juga dilakukan oleh Eva Rahma Indriyani, Paradise dan Merlinda Wibowo dengan membandingkan dua metode yaitu Naïve Bayes dan Support Vector Machine untuk menganalisis sentiment. Hasil dari penelitian tersebut adalah metode support vector machine memiliki nilai akurasi lebih besar yaitu 87,27% [6]. Peneliti lain yaitu Dwi Ardiada, Made Sudarma dan Dwi Giriantari juga melakukan penelitian text mining pada social media dan menghasilkan metode support vector machine memiliki nilai akurasi lebih tinggi dari pada metode KNN yaitu dengan nilai 81% [3].

Berdasarkan latar belakang dan penelitian sebelumnya maka peneliti mengusulkan untuk melakukan penelitian tentang text mining pada social media dengan menggunakan ekstraksi fitur N-Gram level dan metode support vector machine.

Selanjutnya klasifikasi emosi ini dapat bermanfaat diberbagai bidang terutama untuk perusahaan atau organisasi yang akan melakukan perekrutan karyawan. Perusahaan atau Organisasi tersebut dapat melakukan tracking terhadap pelampiasan emosi yang dilakukan oleh calon karyawan tersebut.

4. Metode Penelitian

Metode yang digunakan untuk menemukan hasil dari penelitian ini adalah sesuai dengan yang ada pada gambar dibawah ini .



Gambar 1. Flowchart Proses Penelitian

Peneliti mendapatkan dataset dari peneliti sebelumnya [7]. Dan dataset dari peneliti tersebut sudah berupa data yang berlabel. Selanjutnya setelah mendapat data tersebut, proses selanjutnya adalah tahap *preprocessing*. Peneliti menggunakan kombinasi teks preprocessing yaitu *tokenizing*, *stopword removal*, *stemming* yang sudah dibuktikan oleh peneliti sebelumnya mempunyai waktu tercepat dalam menghasilkan ringkasan teks Bahasa Indonesia [8]. Dan pada penelitian ini ditambahkan tahap *case folding* pada tahap awal preprocessing karena dataset yang didapat sangat bervariasi. Setelah selesai tahap preprocessing masuk ke tahap pembobotan dengan menggunakan TF-IDF Ngram level lalu setelah pembobotan selesai dilanjutkan penerapan metode *support vector machine* dan yang terakhir dilakukan evaluasi terhadap hasil penelitian.

5. Hasil dan Pembahasan

5.1. Dataset

Pada penelitian ini, dataset yang diperoleh berjumlah 1000 dataset. Selanjutnya dataset tersebut dibagi menjadi data latih dan data uji dengan perbandingan 80:20 sehingga menghasilkan 800 data latih dan 200 data uji. Berikut beberapa contoh dataset yang telah diperoleh.



Gambar 2. Contoh Dataset

Untuk penulisan jenis emosi dalam dataset tersebut dipisahakan dengan menggunakan tanda pemisah default dari computer masing-masing seperti titik, koma atau titik koma.

5.2. Tahap Preprocessing

5.2.1. Case Folding

Case Folding adalah tahap penghapusan spasi kosong dan perubahan huruf capital ke huruf kecil. Berikut contoh hasil dari penghapusan spasi kosong.

```
In [13]: runfile('C:/Users/user/untitled0.py', wdir='C:/Users/user')
/Masih ada beberapa Provinsi di Negara Indonesia yang menerapkan PSBB untuk menangani virus Corona.
```

Gambar 3. Menghapus Spasi

Selanjutnya dibawa ini adalah contoh hasil perubahan huruf capital ke huruf kecil atau (*lower case*).

```
In [9]: runfile('C:/Users/user/untitled0.py', wdir='C:/Users/user')
masih ada beberapa provinsi di negara indonesia yang menerapkan psbb untuk menangani virus corona.
```

Gambar 4. Lower Case

5.2.2. Tokenizing

Setelah selesai tahap *case folding*, selanjutnya adalah tahap tokenizing. Tokenizing adalah tahap pemecahan kalimat menjadi kata. Strategi umum yang dilakukan pada tahap tokenizing adalah memotong kata pada white space atau spasi dan membuang karakter tanda baca. Tahap tokenizing membagi urutan karakter menjadi kalimat dan kalimat menjadi token [9]. Berikut adalah contoh hasil tokenizing.

```
In [15]: runfile('C:/Users/user/untitled0.py', wdir='C:/Users/user')
['Masih', 'ada', 'beberapa', 'Provinsi', 'di', 'Negara', 'Indonesia', 'yang', 'menerapkan', 'PSBB', 'untuk', 'menangani', 'virus', 'Corona']
```

Gambar 5. Hasil Tokenizing

5.2.3. Stopword Removal

Setelah data selesai di terapkan tokenizing, maka selanjutnya adalah proses penghapusan kata. Stopword Removal ini juga sering disebut tahap filtering yaitu penghapusan kata-kata yang sangat umum [10]. Kata yang termasuk dalam stopwords contohnya adalah yang, dan, di, itu, dengan, untuk, tidak, dari, dalam, akan, pada, ini, juga, saya, serta, adalah, bahwa, lain, kamu, dan lain lain. Berikut adalah contoh hasil *stopword removal*.

```
In [19]: runfile('C:/Users/user/untitled0.py', wdir='C:/Users/user')
Masih beberapa Provinsi Negara Indonesia menerapkan PSBB menangani virus Corona
```

Gambar 6. Hasil Stopword Removal

5.2.4. Stemming

Kemudian tahap yang terakhir adalah Stemming. Stemming merupakan proses perubahan seluruh kata menjadi kata dasarnya saja atau kata lain penghapusan imbuhan kata. Contoh hasil dari Stemming adalah sebagai berikut.

```
In [21]: runfile('C:/Users/user/untitled0.py', wdir='C:/Users/user')
masih ada beberapa provinsi di negara indonesia yang terap psbb untuk tangan virus corona
```

Gambar 7. Hasil Stemming

5.3. Tahap Ekstraksi Fitur

Selanjutnya masuk ke tahap ekstraksi fitur. Dalam penelitian ini ekstraksi fitur yang digunakan adalah TF IDF N-Gram level. Dari N-gram tersebut yang digunakan adalah tipe bigram dan trigram. N-Gram mendefinisikan kombinasi kata-kata tetangga atau panjang huruf n dalam teks tertentu. Sebuah N-Gram mewakili kelompok kata atau karakter n (ditunjukkan sebagai gram menunjukkan tata bahasa) yang mengikuti satu sama lain. N-gram banyak digunakan dalam penambangan data dan tugas pengolah kata. Misalnya Jika "Saya bekerja di Kantor" berarti $n = 4$ (jumlah kata dari kalimat). Kata tersebut digunakan untuk membuat indeks tentang seberapa sering kata-kata mengikuti satu sama lain. N-gram dapat direpresentasikan dalam formula berikut: [2672-6117-1-PB].

$$Ngrams_k = X - (N - 1) \quad (1)$$

5.4. Hasil Pengujian

Pengujian dalam penelitian ini dilakukan dengan menguji 1000 data update status facebook Bahasa Indonesia yang selanjutnya dibagi menjadi data latih sebanyak 800 data serta data uji sebanyak 200 data dan kategori emosi juga ditentukan yaitu senang, sedih, takut, marah, jijik, bahagia, kaget. Penelitian diawali dengan melakukan tahap preprocessing teks, lalu dilanjutkan penerapan ekstraksi fitur n-gram level dan yang terakhir adalah penerapan metode *support vector machine*.

Berdasarkan hasil pengujian yang telah dilakukan maka dihasilkan nilai akurasi sebesar 53%. Selain nilai akurasi nilai presisi, recall dan f1-score juga dihasilkan dalam penelitian ini. Selanjutnya detail hasil dari pengujian menggunakan ekstraksi fitur n-gram dan metode support vector machine dapat dilihat pada gambar 8 dibawah ini.

```
-----Support Vector Machine TF IDF NGram-----
train time: 0.136s
test time: 0.007s
accuracy: 0.533
precision recall f1-score support
takut 1.00 0.26 0.42 19
kaget 1.00 0.50 0.67 6
jijik 1.00 0.33 0.50 6
bahagia 0.49 1.00 0.66 90
sedih 0.75 0.05 0.10 55
marah 1.00 0.13 0.23 23
accuracy 0.53 199
macro avg 0.57 0.38 0.43 199
weighted avg 0.70 0.53 0.43 199
```

Gambar 8. Hasil Pengujian

6. Kesimpulan

Dari hasil pengujian diatas dapat disimpulkan bahwa, penelitian ini menghasilkan nilai akurasi cukup rendah yakni 53%. Nilai presisi tertinggi dihasilkan oleh beberapa emosi yakni takut, kaget, jijik, dan marah dengan nilai 100%. Selanjutnya nilai recall tertinggi dihasilkan oleh emosi Bahagia dengan nilai 100%. Dan yang terakhir nilai f1-score tertinggi dihasilkan oleh emosi kaget dengan nilai 67%.

Saran untuk penelitian selanjutnya adalah dapat dilakukan penelitian tentang pengaruh kombinasi teks preproccesing dalam melakukan klasifikasi emosi pada sosial media.

7. Daftar Pustaka

- [1] M. Cindo and D. P. Rini, *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Literatur Review: Metode Klasifikasi Pada Sentimen Analisis*. 2019. [Online]. Available: <https://seminar-id.com/semnas-sainteks2019.html>
- [2] Hanadian Nurhayati-Wolff, "Share of Facebook users in Indonesia as of August 2022, by age group."Diakses 28 Desember 2022
- [3] I. M. D. Ardiada, M. Sudarma, and D. Giriantari, "Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour," *Majalah Ilmiah Teknologi Elektro*, vol. 18, no. 1, p. 55, May 2019, doi: 10.24843/mite.2019.v18i01.p08.
- [4] A. Deolika and E. Taufiq Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *Jurnal Teknologi Informasi*, vol. 3, no. 2, 2019.
- [5] F. Romadoni, Y. Umaidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 2, pp. 247–253, Jul. 2020, doi: 10.32736/sisfokom.v9i2.903.
- [6] E. R. Indriyani, P. Paradise, and M. Wibowo, "Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter," *Jurnal Media Informatika Budidarma*, Vol. 6, No. 3, P. 1545, Jul. 2022, Doi: 10.30865/mib.v6i3.4220.
- [7] A. N. Rohman, E. Utami, and S. Raharjo, "Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing," *Eksplora Informatika*, vol. 9, no. 1, pp. 70–76, 2019, doi: 10.30864/eksplora.v9i1.277.
- [8] H. Najjichah, A. Syukur, and H. Subagyo, "Pengaruh Text Preprocessing Dan Kombinasinya Pada Peringkat Dokumen Otomatis Teks Berbahasa Indonesia," 2019. [Online]. Available: <http://research.fithri Selva Jumeilah>
- [9] Fithri Selva Jumeilah, "Support Vector Machine (SVM) untuk Pengkategorian Penelitian," 2017. [Online]. Available: <http://jurnal.iaii.or.id>
- [10] H. Benbrahim, M. Bramer, and Y. Llp, "Text and Hypertext Categorization," 2009.