

## DETEKSI EMOSI PADA TEKS MENGGUNAKAN ALGORITMA NAÏVE BAYES

Akhmad Fadjeri<sup>1</sup>, Kardilah Hidayat<sup>2</sup>, Dwi Riska Handayani<sup>3</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Ma'arif Nahdlatul Ulama Kebumen

<sup>2,3</sup>Program Studi S2 Teknik Informatika, Universitas Amikom Yogyakarta

<sup>1</sup>akhmadfadjeri@umnu.ac.id, <sup>2</sup> karilah.hidayat@students.amikom.ac.id, <sup>3</sup> riska.1116@students.amikom.ac.id

### Abstract

*Proper use of emotions at the moment and the right atmosphere can affect the outcome of human activities. Emotions are usually not expressed blatantly by humans (implicitly) and also caused by certain events or situations. The text depicts patterns of events or situations that cause emotions to be revealed with words clearly (explicitly). Text is also the main media in communication using computers (Computer-mediated communications) such as email, blogs and social media. According to the nature of emotions can be classified into two namely positive emotions and negative emotions. Emotion detection is a new and moderately researched field of research, particularly in the linguistic field. In this research the emotion detection will be applied to the status of Facebook using the Naïve Bayes method and the weighted word Count Vector. Before the Count Vector method has been applied, the text will be the test data and the training data in applying the Prapremprosesan which includes folding case, Stopword removal, stemming. The results of this study are the accuracy value of 87,7%...*

*Keywords: Emotion Text, Count Vector, Naïve Bayes*

### Abstrak

Penggunaan emosi yang tepat pada waktu dan suasana yang tepat dapat mempengaruhi hasil kegiatan yang dilakukan oleh manusia. Emosi itu biasanya dinyatakan secara implisit oleh manusia dan juga disebabkan oleh peristiwa atau situasi tertentu. Teks menggambarkan pola peristiwa atau situasi yang menyebabkan emosi dapat terungkap dengan kata-kata secara jelas (eksplisit). Teks juga menjadi media utama dalam komunikasi menggunakan computer (Computer-Mediated Communication) seperti email, blog dan media sosial. Menurut sifatnya emosi dapat diklasifikasikan menjadi dua yaitu emosi positif dan emosi negative. Deteksi emosi merupakan bidang penelitian baru dan sedang banyak diteliti khususnya pada bidang linguistik. Dalam penelitian ini deteksi emosi akan diterapkan pada status facebook dengan menggunakan metode Naïve Bayes dan pembobotan Count Vector. Sebelum diterapkannya metode pembobotan Count vector, teks yang akan menjadi data uji dan data latih di lakukan penerapan prapremprosesan yang meliputi case folding, stopword removal, stemming. Hasil dari penelitian ini adalah nilai akurasi sebesar 87,7%.

Kata kunci: Emosi Teks, Count Vector, Naïve Bayes

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



## 1. Pendahuluan

Penggunaan emosi yang tepat pada waktu dan suasana yang tepat dapat mempengaruhi hasil kegiatan yang dilakukan oleh manusia. Emosi itu biasanya dinyatakan secara implisit oleh manusia dan juga disebabkan oleh peristiwa atau situasi tertentu. Teks menggambarkan pola peristiwa atau situasi yang menyebabkan emosi dapat terungkap dengan kata-kata secara jelas (eksplisit). Teks juga menjadi media utama dalam komunikasi menggunakan computer (Computer-Mediated Communication) seperti email, blog dan media sosial. Menurut sifatnya emosi dapat

diklasifikasikan menjadi dua yaitu emosi positif dan emosi negatif. Yang termasuk emosi negative diantaranya marah, sedih, cemburu, malu, cemas, iri, jijik, takut, kasihan dan rasa bersalah. Sedangkan yang termasuk dari emosi positif adalah harapan, bangga, lega, gembira dan cinta.

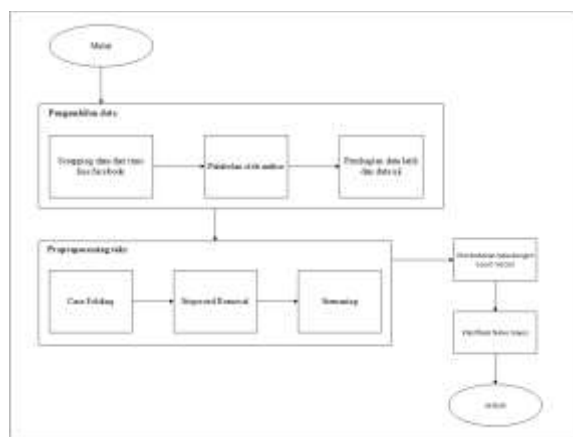
Deteksi emosi pada teks dari media sosial adalah bidang penelitian yang memperoleh minat tinggi terutama untuk kepentingan analisis emosi. Menurut hasil survey yang telah dilakukan, dari tahun 2014-2017 terdapat 6 dari 10 paper yang diterbitkan meneliti tentang deteksi emosi pada teks (Al-Saaqa et al., 2018). Untuk metode klasifikasi yang digunakan juga bermacam-macam, diantaranya

*Logistic Regression, Naïve Bayes, Support Vector Machine, Random Forest, K-Means, dan Lexical Based* (Dandannavar et al., 2018). Diantara beberapa metode tersebut Naïve Bayes termasuk kategori tertinggi nomor 3 untuk tingkat akurasi. Hasil survey penelitian dari tahun 2014-2018 juga mengatakan bahwa dataset yang digunakan sebagian besar menggunakan twitter. Dalam penelitian ini dataset yang kami gunakan adalah media social facebook dimana dalam facebook mempunyai fitur update status yang memungkinkan pengguna membagikan gambar, video dan terutama teks yang akan tampil pada timeline pengguna terkait. Dataset yang digunakan peneliti adalah dataset yang telah dikumpulkan oleh peneliti sebelumnya (Rohman et al., 2019).

Berdasarkan latar belakang peneliti sebelumnya peneliti mengusulkan metode untuk deteksi emosi pada teks yang dataset berasal dari media social facebook menggunakan metode Naïve Bayes dan Pembobotan Count Vector.

## 2. Metode Penelitian

Tahapan yang dilakukan untuk mendapatkan hasil penelitian ini dapat dilihat pada gambar 1 dapat dilihat pada Gambar 1.



Gambar 1 Metode Penelitian

Pada penelitian ini data yang digunakan adalah data update status dari media social facebook. Data ini adalah data yang dikumpulkan oleh peneliti sebelumnya (Rohman et al., 2019) dengan melakukan pelabelan langsung oleh *author*. Setelah dataset didapatkan maka proses selanjutnya adalah preprocessing teks. Preprocessing yang dilakukan berupa proses case folding, stopwords removal, dan stemming. Selanjutnya setelah selesai dilakukan preprocessing teks maka dilanjutkan proses pembobotan dengan

metode count vector. Setelah mendapatkan hasil dari perhitungan vector maka selanjutnya adalah penerapan algoritma naïve bayes. Dan yang terakhir adalah perhitungan akurasi dari klasifikasi tersebut.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Dataset

Dataset yang digunakan oleh peneliti ini berasal dari media social facebook dengan jumlah 100 update status. Selanjutnya dataset tersebut dibagi menjadi data latih dan data uji dengan perbandingan 80:20.

Tabel 1 Contoh data update status sebelum *preprocessing*

No	Update status	Keyword	Emosi
1	selamat atas pernikahannya resya ☐☐	Selamat, pernikahan	Bahagia
2	berusaha untuk tidak mengeluh, pasti terlewat, bismillah, semangat ❤️	Berusaha, Semangat	Bahagia
3	Saat lelah, bukan kau yang menjadi tempat beristirahat. Sedangkan kau? Kau lelah, ia yang kau cari.	Lelah	Sedih
4	Mohon maaf lahir dan batin ☐☐	Maaf	Sedih
5	senang sekali bisa kenal dengan guru guru disini, masya allah sangat menerima kami dengan sangat baik ☐☐☐☐ memah ami apa yang kami lewati dan kerjakan sekarang, dan memperlancar tugas kami, semoga buhan pian sehat sehat dan lancar rejekinya.	Senang, baik	Bahagia

### 3.2 Preprocessing Teks

#### 3.2.1 Case Folding

Pada tahap ini, semua huruf akan diubah menjadi lowercase atau huruf kecil.

Tabel 2 contoh penerapan case folding

No	Update status	Case Folding
1	selamat atas pernikahannya resya □□	selamat atas pernikahannya resya □□
2	berusaha untuk tidak mengeluh, pasti terlewati, bismillah , semangat ♥	berusaha untuk tidak mengeluh, pasti terlewati, bismillah , semangat ♥
3	Saat lelah, bukan kau yang menjadi tempat beristirahat. Sedangkan kau? Kau lelah, ia yang kau cari.	saat lelah, bukan kau yang menjadi tempat beristirahat. sedangkan kau? kau lelah, ia yang kau cari.
4	Mohon maaf lahir dan batin □□	mohon maaf lahir dan batin □□
5	senang sekali bisa kenal dengan guru guru disini, masya allah sangat menerima kami dengan sangat baik □□□□ memaha mi apa yang kami lewati dan kerjakan sekarang, dan memperlancar tugas kami, semoga buhan pian sehat sehat dan lancar rejekinya	senang sekali bisa kenal dengan guru guru disini, masya allah sangat menerima kami dengan sangat baik □□□□ memaha mi apa yang kami lewati dan kerjakan sekarang, dan memperlancar tugas kami, semoga buhan pian sehat sehat dan lancar rejekinya

### 3.2.1 Stopword Removal

Pada tahap ini, dihilangkan karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan.

Tabel 3 contoh penerapan stopword removal

No	Update status	Stopword removal
1	selamat atas pernikahannya resya □□	selamat pernikahannya resya
2	berusaha untuk tidak mengeluh, pasti terlewati, bismillah , semangat ♥	berusaha mengeluh terlewati bismillah semangat

3	Saat lelah, bukan kau yang menjadi tempat beristirahat. Sedangkan kau? Kau lelah, ia yang kau cari.	lelah bukan menjadi tempat beristirahat sedangkan lelah cari
4	Mohon maaf lahir dan batin □□	maaf lahir batin
5	senang sekali bisa kenal dengan guru guru disini, masya allah sangat menerima kami dengan sangat baik □□□□ memahami apa yang kami lewati dan kerjakan sekarang, dan memperlancar tugas kami, semoga buhan pian sehat sehat dan lancar rejekinya	senang kenal guru disini masya allah menerima baik memahami lewati kerjakan sekarang memperlancar tugas semoga buhan pian sehat lancar rejekinya

### 3.2.2 Stemming

Pada tahap ini mengubah token yang berimbuhan menjadi kata dasar, dengan menghilangkan semua imbuhan yang ada pada token tersebut.

Tabel 4 contoh penerapan stemming

No	Update status	Stemming
1	selamat atas pernikahannya resya □□	selamat nikah resya
2	berusaha untuk tidak mengeluh, pasti terlewati, bismillah , semangat ♥	usaha ngeluh lewat bismillah semangat
3	Saat lelah, bukan kau yang menjadi tempat beristirahat. Sedangkan kau? Kau lelah, ia yang kau cari.	lelah bukan jadi tempat istirahat sedang lelah cari
4	Mohon maaf lahir dan batin □□	maaf lahir batin

5	senang sekali bisa kenal dengan guru guru disini, masya allah sangat menerima kami dengan sangat baik □□□□ memahami apa yang kami lewat dan kerjakan sekarang, dan memperlancar tugas kami, semoga buhan pian sehat sehat dan lancar rejekinya	senang kenal guru disini masya allah nerima baik paham lewat kerja sekarang lancar tugas semoga buhan pian sehat lancar rejeki
---	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------

### 3.3. Pengujian

Dalam penelitian ini pengujian dilakukan dengan menggunakan data sebanyak 550 data dengan pembagian data latih sebanyak 306 dan data uji sebanyak 244. Data tersebut berasal dari media sosial facebook yang telah dikumpulkan oleh peneliti sebelumnya. Dari data yang telah dikumpulkan tersebut kemudian dilakukan pelabelan manual oleh *author* yang dikelompokkan menjadi 7 emosi, yaitu bahagia, sedih, marah, jijik, terkejut atau kaget, takut, senang. Metode pembobotan kata Count Vector dan metode klasifikasi Naïve Bayes mampu menghasilkan akurasi dengan nilai 87,7%. Nilai presisi, *recall* dan *F1-Measure* dapat dilihat pada table berikut.

Tabel 5. Nilai presisi, *recall* dan *F1-Measure* dari masing-masing emosi

Emosi	Presisi	Recall	F1-Measure
Senang	0.95	0.93	0.94
Takut	1.00	0.67	0.80
Kaget	1.00	0.33	0.50
Jijik	1.00	0.64	0.78
bahagia	0.80	0.95	0.87
Sedih	0.88	0.97	0.92
Marah	1.00	1.00	1.00

Sedangkan waktu yang dibutuhkan untuk melakukan ekstraksi pada data latih dan data uji adalah 0.009237 s dan 0.002449 s.

## 4. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode Naïve Bayes dan pembobotan count vector dapat menghasilkan kinerja yang baik dalam melakukan klasifikasi emosi pada teks. Hal ini dapat dilihat dari hasil nilai presisi, *recall* dan *f1-measure* dimana dari 306 data latih dan 244 data uji semua emosi dapat terdeteksi

Pada proses pengujian tingkat accuracy yang dihasilkan oleh metode naïve bayes dan

metode fitur ekstraksi count vector adalah 87,7%. Untuk nilai *recall* tertinggi adalah emosi marah dengan nilai 1.00 atau 100%, nilai presisi tertinggi adalah emosi marah, takut, kaget dan jijik yaitu 1.00 atau 100% dan nilai *F1-Measure* tertinggi pada emosi marah adalah 1.00 atau 100%.

## DAFTAR PUSTAKA

- Aji Andika, L., & Amalia Nur Azizah, P. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1), 34–41.
- Al-Saaqa, S., Abdel-Nabi, H., & Awajan, A. (2018). A Survey of Textual Emotion Detection. *2018 8th International Conference on Computer Science and Information Technology, CSIT 2018*, 136–142. <https://doi.org/10.1109/CSIT.2018.8486405>
- Dandannavar, P. S., Mangalwede, S. R., & Kulkarni, P. M. (2018). Social Media Text - A Source for Personality Prediction. *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018*, 62–65. <https://doi.org/10.1109/CTEMS.2018.8769304>
- Rohman, A. N., Utami, E., & Raharjo, S. (2019). Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing. *Eksplora Informatika*, 9(1), 70–76. <https://doi.org/10.30864/eksplora.v9i1.277>
- Rustiana, D., & Rahayu, N. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 8(1), 113–120. <https://doi.org/10.24176/simet.v8i1.841>
- Saptono, R., & Maret, U. S. (2017). CLASSIFICATION OF CUSTOMERS EMOTION USING NAIVE BAYES CLASSIFIER ( Case Study : Natasha Skin Care ). *Itsmart*, 6(2), 92–97. <https://doi.org/10.20961/itsmart.v6i2.17328>
- Simanjuntak, Steven Roy. (2017). Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB). *Jurnal Telematika*, 12(01). <http://journal.ithb.ac.id/telematika/article/view/154>