Sentimen Analisis Terhadap Kaum Gay Pada Twitter Menggunakan Metode Naive Bayes

by Tri Listyorini

Submission date: 10-Aug-2021 09:03AM (UTC+0700)
Submission ID: 1629760359
File name: erhadap_Kaum_Gay_Pada_Twitter_Menggunakan_Metode_Naive_Bayes.pdf (337.85K)
Word count: 1633
Character count: 10304

Sentimen Analisis Terhadap Kaum *Gay* pada *Twitter* Menggunakan Metode *Naïve Bayes*

Agnes Pramitha[1], Endang Supriyati[2], Tri Listyorini[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muria Kudus
[1]apramitha58@gmail.com

***Abstrak*— Saat ini *gay* menjadi fenomena isu yang banyak diperbincangkan. Isu-isu *gay* tersebut beredar di media sosial yang menimbulkan berbagai macam perspektif bagi masyarakat. Apalagi akhir-akhir ini banyak kasus *gay* di Indonesia yang menimbulkan pro dan kontra. Untuk mengetahui perspektif masyarakat terhadap *gay* berdasarkan sentimen masyarakat melalui *Twitter*, maka penulis mencoba untuk membuat analisa sentimen berupa pandangan positif, negatif, maupun netral dengan menggunakan metode *Naïve Bayes*. Analisa dilakukan menggunakan *tools* berupa *Microsoft Excel* dan *Pycharm* dengan bahasa pemrograman *Python*. *Hasil* analisa yang dilakukan menggunakan klasifikasi *Naïve Bayes* menunjukkan polaritas sentimen negatif 150,5%, netral 12,6%, dan positif 0,4%. Dari 1635 data yang diambil, *tweet* yang bernilai negatif sesuai hasil polaritas sebanyak 1505, dengan tingkat akurasi sebesar 93%.**

***Kata Kunci*—** *Sentiment Analisis, Gay, Twitter, Python, Pycharm Naïve Bayes.*

I. PENDAHULUAN

Masyarakat Indonesia sedang mengalami era perubahan budaya. Perubahan budya terjadi dalam kehidupan sosial masyarakat, berupa dampak positif dan negatif. Dampak negatif terasa dalam hidup masyarakat dan budaya adalah lunturnya nilai dan norma, etika dan keyakinan agama di masyarakat.

Perkembangan kaum *gay* di Indonesia mencerminkan pergeseran budaya yang telah terjadi dan telah menarik perhatian masyarakat. Kebebasan berbicara kaum *gay* di media sosial *Twitter* memiliki berbagai pengaruh terhadap masyarakat. Melalui media sosial, *gay* dapar dengan mudah berinteraksi dengan sesama *gay* secara bebas.

Gay tidak disebabkan oleh penyakit genetik, ketidakseimbangan hormon, penyakit mental atau kejahatan. Namun karena pengaruh lingkungan dan pengalaman pribadi, *gay* ini muncul. Dengan cara ini, orang sering berinteraksi dengan lelaki *gay* bisa menjadi *gay*. [1]

Istilah *gay* baru-baru ini banyak diperbincangkan di Indonesia melalui media sosial. Berawal dari pemberitaan yang bersumber dari Magdalen.co yang menyebutkan bahwa kelompok *LGBT* di Indonesia mengikuti Paris Parade (Festival *LGBT)* pada tanggal 29 Juni 2019. [2]

Tidak hanya kasus tersebut, bahkan terdapat kasus lagi yang baru-baru ini disiarkan di Liputan6.com yaitu berita *seks gay* saat penggrebekan di Jakarta pada tanggal 3 September 2020. Banyak pasangan-pasangan *gay* yang terciduk di wilayah Jakarta tersebut. [3]

Berdasarkan kasus-kasus yang telah terjadi, maka penelitian ini akan menganalisis sentimen masyarakat terhadap kaum *gay* pada *Twitter* menggunakan metode *Naïve Bayes.* Analisa dilakukan menggunakan *tools* berupa *Microsoft Excel* dan *Pycharm* dengan bahasa pemrograman *Python.* Hasil klasifikasi nantinya berupa sentimen positif, negatif, dan netral.

II. METODOLOGI PENELITIAN

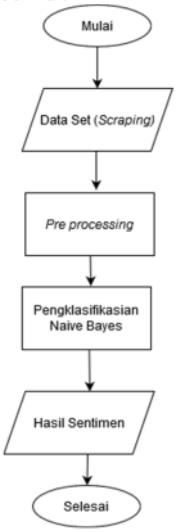
Penelitian ini mengembangkan sentimen analisis untuk mengelompokkan *tweet* yang dibuat oleh pengguna *Twitter* bernilai positif, negatif, atau netral. Metode yang digunakan untuk menganalisis sentimen *tweet* kaum *gay* adalah metode *Naïve Bayes.*

A. Metode Pengumpulan Data

Data yang diperoleh untuk melakukan penelitian ini adalah dengan cara *scraping* yaitu data yang diambil langsung dari *tweet* menggunakan *Twitter API*. Pengumpulan data dilakukan dengan menggunakan *keyword gay*. Jumlah data yang diambil kurang lebih berjumlah 1000 data.

B. Perancangan Program Analisis

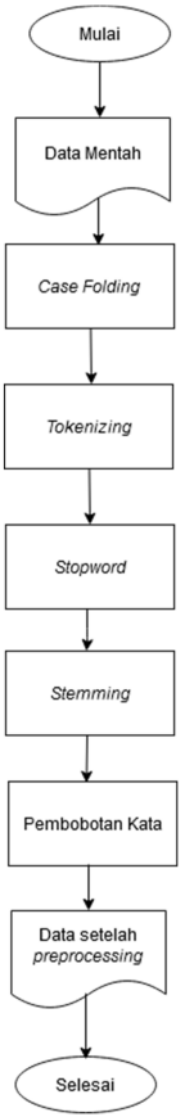
Flowchart atau alur secara umum pada program dapat dilihat pada gambar diagram alir berikut:



Gbr 1. Diagram Alir Program

Pada Gbr 1 tersebut menjelaskan bahwa alur program dimulai dengan mengambil data (*scraping*) dari *Twitter* melalui *Python* dan *Pycharm*. Kemudian data diolah dengan melakukan *preprocessing*. Pada tahap *preprocessing* data diolah melalui Microsoft Excel. Setelah itu masuk ke tahap klasifikasi *Naïve Bayes* yang mana akan didapatkan hasil sentimen berupa positif, negatif, maupun netral.

1) Preprocessing: Data dari *scraping* selanjutnya akan di *Pre-processing* sebelum dapat diklasifikasikan. Alurnya adalah sebagai berikut :



Gbr 2. Diagram Alir Tahap *Preprocessing*

Gbr 2 menjelaskan bahwa alur tahap *preprocessing* dari data yang masih mentah dilakukan tahap *case folding* yaitu proses menghilangkan tnada baca. Setelah itu adalah tahap *tokenizing* yaitu proses pemisahan teks menjadi potongan-potongan (token) untuk dianalisa. Selanjutnya adalah tahap *stop word* yaitu proses mengambil kata-kata penting dari hasil token. Kemudian tahap *stemming* yaitu proses mengubah kata menjadi bentuk sederhana. Yang terakhir yaitu proses pembobotan kata, dilakukan dengan cara menghitung bobot (*term*). Setelah itu akan diperoleh data yang telah dilakukan *preprocessing*.

2) Algoritma *Naïve Bayes* umumnya digunakan untuk melakukan prediksi sentimen yang muncul pada data yang masih belum memiliki sentimen. Metode *Naïve Bayes* dapat mengelola data dalam jumlah besar dengan hasil akurasi yang tinggi.

Untuk menghitung probabilitas, persamaan ini menerangkan hubungan antara probabilitas hipotesis A dengan probabilitas B [4]. Persamaan umum yang dimiliki pada algoritma klasifikasi *Naïve Bayes* dapat dilihat pada persamaan (1) berikut:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

A : Hipotesis data suatu *class* spesifik
B : Data dengan *class* yang masih belum diketahui.
P(A|B) : Probabilitas hipotesis berdasar kondisi.
P(A) : Probabilitas A.
P(B|A) : Probabilitas berdasar kondisi pada hipotesis.
P(B) : Probabilitas B.

Persamaan (1) menjelaskan tentang algoritma yang akan digunakan dalam penelitian ini. Algoritma ini dipilih karena tingkat akurasi yang cukup tinggi [3].

III. HASIL DAN PEMBAHASAN

1) Scraping adalah teknik yang digunakan untuk mengambil data dari *Twitter* yang menyediakan *API (Application Program Interface)*. *Scraping* dilakukan dengan menggunakan *tools* yaitu *Python* dan *Pycharm* [5]. *Keyword* untuk analisa data adalah *Gay*. Dengan *scraping* didapatkan data sejumlah 1635 data. Hasil data *scraping* yaitu sebagai berikut:

TABEL I
DATA *SCRAPING*

Keyword	**Tanggal**	**Username**	***Tweet***
Gay	04/01/2021 15:12:05	sseapurple	Aneh bgt. Suka korea dibilang plastik, suka thailand dibilang gay, jadi atheis dibangga banggain. Kenapa ga hidup di jalan

			masing-masing aja, sibuk bgt kakak itu ngurusin hidup orang.

2) Preprocessing dilakukan untuk membersihkan data, menghilangkan *tweet* yang berulang, dan membakukan kalimat yang tidak baku sehingga mendapatkan data bersih. Tahap *preprocessing* adalah sebagai berikut:

a. *Case Folding* adalah proses mengubah semua huruf yang sebelumnya berupa huruf kapital diubah menjadi huruf kecil semua. [6]
b. *Tokenizing* atau proses tokenisasi adalah proses membagi kalimat menjadi beberapa bagian yang disebut token. Token dapat dibentuk dalam kata, frase, atau yang lainnya.[7]
c. *Stop word* atau proses menghilangkan kata-kata yang tidak memiliki pengaruh terhadap makna dan sentimen dari sebuah kalimat. Proses ini mempermudah untuk mendeteksi *keyword* atau kunci di setiap kalimat.[1]
d. *Stemming* adalah proses mengubah kata menjadi bentuk sederhana. *Stemming* dapat mengurangi semua variasi dari satu kata menjadi bentuk sederhana yang sama. Tujuan dari *stemming* adalah untuk mengubah kata yang mirip menjadi bentuk yang sama. Misalnya kata "pemberani" diubah menjadi kata "berani".[8]
e. Pembobotan kata dilakukan dengan cara menghitung bobot (*term*) kata dari dalam dokumen. Dengan menyatakan frekuensi munculnya suatu *term* dan banyaknya jumlah dokumen dimana sebuah *term* itu muncul.[9]

Hasil dari *preprocessing* adalah sebagai berikut:

TABEL 2
DATA *PREPROCESSING*

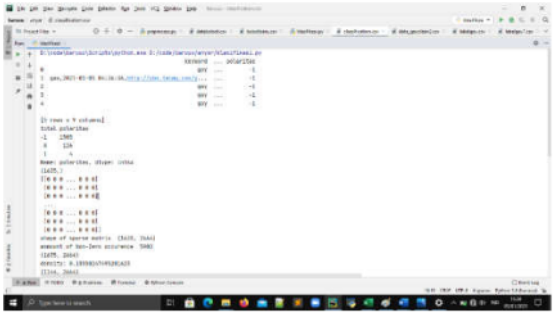
***Dataset* Awal**	Aneh bgt. Suka korea dibilang plastik, suka thailand dibilang gay, jadi atheis dibangga-banggain. Kenapa ga hidup di jalan masing-masing aja, sibuk bgt kakak itu ngurusin hidup orang.
Hasil *Text Pre-processing*	aneh suka korea dibilang plastik suka thailand dibilang gay atheis bangga banggain hidup jalan sibuk kakak ngurusin hidup orang

3) Klasifikasi algoritma *Naïve Bayes* yang dilakukan dalam analisa ini menggunakan perbandingan *70*:30 antara data *training* dengan data *testing*. Dengan melakukan *training* pada *dataset* yang ada akan menghasilkan sebuah model pada data *testing*. Data yang telah selesai dilakukan klasifikasi akan menghasilkan polaritas dan sentimen analisis sebagai berikut:

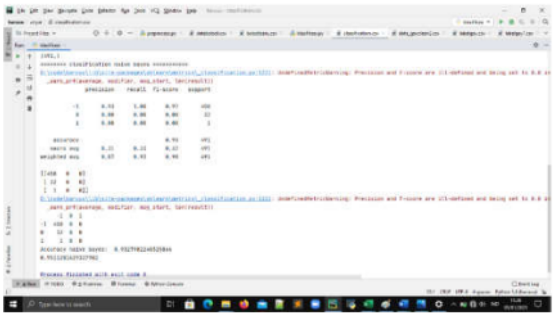
TABEL 3
DATA HASIL KLASIFIKASI

Sentimen	Polaritas
Negatif	1505
Netral	126
Positif	4

Hasil klasifikasi dapat dilihat seperti gambar di bawah ini:



Gbr 3. Hasil Polaritas



Gbr 4. Hasil Klasifikasi dan *Confussion Matrix*

Hasil klasifikasi yang terdapat pada tabel 3 merupakan polaritas dari setiap *class* yang ada yaitu negatif, netral, dan positif. Ditunjukkan bahwa polaritas pada *class* negatif terdapat 1505 *dataset*, netral 126 *dataset*, dan positif 4 *dataset.* Pada gambar 3 menunjukkan polaritas yang ada di dalam program. Sedangkan gambar 4 berisi hasil klasifikasi yang memiliki akurasi 93% dan *confussion matrix* yang menunjukkan data matriks dari setiap *class* setelah dilakukan *testing* menggunakan model.

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari sentimen analisis menggunakan klasifikasi algoritma *Naïve Bayes* dengan data yang diambil dari *Twitter* dan *gay* sebagai objek penelitian adalah sebagai berikut:

- Dari 1635 data yang diambil, *tweet* yang bernilai negatif berdasarkan nilai polaritas sebanyak 1505.
- Analisis sentimen dari pengguna *Twitter* tentang *gay* memiliki tingkat akurasi sebesar 93%.
- Terdapat polaritas sentimen negatif sebanyak 150,5%, netral 12,6%, dan positif 0,4%.

REFERENSI

[1] E. N. Novianta, *Digital Digital Repository Repository Universitas Universitas Jember*, vol. 1, no. 3. 2018.

[2] S. Amalia, "Baju Adat, Kuda, Lumping dan Bendera Pelangi:LGBT Indonesia Rayakan Paris Pride," *Magdalene*, 2019. https://magdalen.co/story/kelompok-lgbt-indonesia-ikut-serta-dalam-paris-pride.

[3] D. Prastiwi, "5 Fakta Terbaru Kasus Dugaan Pesta Seks Gay Usai Penggrebekan," *Liputan6*, 2020. https://m.liputan6.com/news/read/4347132/5-fakta-terbaru-kasus-dugaan-pesta-seks-gay-usai-penggrebekan.

[4] H. Simorangkir and K. M. Lhaksmana, "Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier," *e-proceeding of Englineering*, vol. 5, no. 3, pp. 8131–8140, 2018, [Online]. Available: https://openlibrary.telkomuniversity.ac.id/pustaka/files/144621/jurnal_eproc/analisis-sentimen-pada-twitter-untuk-games-online-mobile-legends-dan-arena-of-valor-dengan-metode-na-ve-bayes-classifier.pdf.

[5] M. Y. Aldean, M. D. Hilmawan, R. Indriyati, and J. Lasama, "Analisa Relevansi Tweet terhadap Hashtag dengan Metode Logistic Regression," *Centive*, vol. 2, no. 1, pp. 32–38, 2019.

[6] A. Sentimen, M. Pada, and M. Ridwan, "- 0 3 7'," 2019.

[7] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.

[8] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.

[9] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.

Sentimen Analisis Terhadap Kaum Gay Pada Twitter Menggunakan Metode Naive Bayes

ORIGINALITY REPORT

10%	6%	3%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	text-id.123dok.com Internet Source	2%
2	Muhardi Muhardi. "Sistem E-Inventory Menggunakan Bahasa Pemograman Visual Basic.NET", INTECOMS: Journal of Information Technology and Computer Science, 2019 Publication	1%
3	jurnal.umk.ac.id Internet Source	1%
4	es.scribd.com Internet Source	1%
5	matematika.fst.unair.ac.id Internet Source	1%
6	Fajar Ratnawati. "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter", INOVTEK Polbeng - Seri Informatika, 2018 Publication	1%
7	Submitted to Universitas Dian Nuswantoro	

Student Paper

1%

8 Muhammad Rizki Fahdia, Dwiza Riana, Fachri Amsury, Irwansyah Saputra, Nanang Ruhyana. "Komparasi Algoritma Klasifikasi untuk Orientasi Minat Mahasiswa dalam Penuntasan Studi", JIRA: Jurnal Inovasi dan Riset Akademik, 2021

Publication

1%

9 Submitted to Universitas Negeri Surabaya The State University of Surabaya

Student Paper

1%

10 123dok.com

Internet Source

1%

Exclude quotes	Off	Exclude matches	< 5 words
Exclude bibliography	Off		