

DETEKSI KOMENTAR NEGATIF DI INSTAGRAM MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Maria Bellanar Ismiati

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Katolik Misi Charitas
Jl. Bangau No.60, Palembang-Sumatera Selatan.

*Email: bella@ukmc.ac.id

Abstrak

Komentar yang diberikan oleh pengguna (dalam dunia maya disebut netizen) dapat mengandung makna positif atau negatif. Akan tetapi sering sekali netizen tidak menyadari apakah komentar yang diberikan negatif atau positif dikarenakan belum banyak sistem untuk mendeteksi isi komentar tersebut masuk di kategori positif atau negatif. Selain itu, dari sisi netizen, tidak ada yang salah saat ingin berkomentar atau mencurahkan isi hati. Penelitian ini dibuat untuk membantu para pengguna saat ingin berkomentar agar tidak selalu terjerumus ke kategori komentar negatif. Di Indonesia, apabila diketahui komentar netizen tersebut mengarah ke negatif, seringkali harus berurusan dengan hukum dan kepolisian. Metode yang digunakan dalam mendukung penelitian ini adalah Naive Bayes Classifier untuk menentukan apakah komentar yang dimasukkan termasuk kategori positif atau negatif. Hasil akurasi yang didapatkan adalah sebanyak 76,7% dimana ada 23 pengguna dari 30 yang hasil kategori komentarnya sesuai dengan harapan.

Kata kunci : komentar, positif, negatif, naive bayes classifier, akurasi

1. PENDAHULUAN

1.1. Latar Belakang

Teknologi pada jaman sekarang semakin berkembang dan canggih di berbagai bidang kehidupan, salah satunya di media sosial. Media sosial merupakan beberapa media perangkat komunikasi terpopuler di masyarakat saat ini (Aliandu, 2012). Salah satu contoh media sosial yang semakin diminati oleh masyarakat adalah Instagram. Instagram merupakan aplikasi untuk mengirimkan informasi dengan cepat, yakni dalam bentuk foto yang berupa mengelola foto, mengedit foto, dan berbagi (share) ke jejaring sosial yang lain (Ghazali, 2016). Berdasarkan tujuan penggunaan instagram tersebut, aplikasi ini dapat digunakan dengan menggunakan teknologi internet yang menjadi dasar aktivitas dari media sosial ini.

Berdasarkan survey dari #StatusOfMinde yang dipublikasikan oleh United Kingdom's Royal Society for Public Health, Instagram dianggap sebagai media sosial yang paling buruk bagi kesehatan mental dan jiwa. Survey tersebut melibatkan masukan dari 1.479 orang muda (usia 14-25) dari seluruh Inggris Raya yang dilakukan pada Februari - Mei 2017. Media sosial Instagram banyak disukai karena dapat menjadi platform untuk menampilkan ekspresi diri melalui foto, video, bahkan status, akan tetapi Instagram juga berkaitan dengan tingkat kecemasan yang tinggi, depresi, bullying, dan FOMO (fobia ketinggalan berita di jejaring sosial). (Anna, 2017).

Permasalahan di atas dikaitkan dengan foto-foto yang di-upload di Instagram. Selain fungsi Instagram untuk mengirim foto, Instagram juga dapat digunakan untuk membuat dan mengirim status seseorang ke dunia maya agar dapat dilihat oleh orang lain yang menjadi temannya di Instagram serta dapat juga memberikan komentar terhadap foto/status yang ditulis oleh teman-temannya. Komentar yang dibuat dan dikirimkan oleh seseorang tersebut biasanya merupakan ungkapan isi hati orang tersebut terhadap kehidupan pribadinya ataupun suatu hal yang sedang menjadi topik pembicaraan masyarakat. Isi komentar tersebut terkadang tidak disadari oleh orang yang membuat, apakah mengandung konten negatif atau tidak. Hal tersebut juga didukung oleh banyaknya media sosial yang belum menggunakan suatu aplikasi untuk mendeteksi apakah isi komentar dari para netizen (sebutan untuk orang-orang yang saling bersosialisasi di dunia maya) tersebut mengandung konten negatif atau tidak. Apabila komentar negatif tersebut sudah terlanjur di-upload di media sosial dan dibaca oleh masyarakat, maka seringkali apabila sudah melebihi batas maka orang yang membuat komentar tersebut dapat dilaporkan ke polisi dan parahnya dapat langsung dipenjara.

Penelitian ini akan membantu masyarakat yang ingin mengirim komentar yang mereka buat agar tidak mengandung konten negatif sehingga dapat terhindar dari kasus kejahatan di dunia maya. Berdasarkan permasalahan yang sudah dijelaskan di atas, penelitian ini berjudul Deteksi Komentar Negatif Di Instagram Menggunakan Naive Bayes Classifier. Tidak semua komentar di Instagram digunakan dalam penelitian ini, hanya ± 50 komentar yang dijadikan objek penelitian. Tahapannya adalah pengumpulan data, *case folding*, *tokenizing*, *filtering*, *stemming*, klasifikasi dengan Naive Bayes Classifier, dan hasil klasifikasi. Dengan menggunakan algoritma Naive Bayes tersebut, maka penelitian ini diharapkan dapat membantu masyarakat untuk menentukan apakah komentar yang dibuat mengandung konten negatif atau tidak.

1.2. Landasan Teori

1.2.1. Information Retrieval (IR)

Information Retrieval (IR) merupakan bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. *Information Retrieval* merupakan suatu pencarian informasi (biasanya berupa dokumen) yang didasarkan pada suatu *query* (*inputan user*) yang diharapkan dapat memenuhi keinginan *user* dari kumpulan dokumen yang ada. Dalam bentuk yang paling sederhana, sebuah *query* merupakan suatu *keywords* (kata kunci) dan dokumen yang mengandung *keywords* merupakan dokumen yang dicari dalam sistem IR (Manning, Raghavan, & Schütze, 2009).

Ada empat tahapan proses pokok dalam *Information Retrieval* (IR), yaitu :

- Text Preprocessing*, tahap ini melakukan analisis semantik (kebenaran arti) dan sintaktik (kebenaran susunan) terhadap teks.
- Text Transformation* / Transformasi teks atau pembentukan atribut mengacu pada proses untuk mendapatkan representasi dokumen yang diharapkan.
- Feature Selection* / Pemilihan fitur (kata) merupakan tahap lanjut dari pengurangan dimensi pada proses transformasi teks dimana akan dipilih hanya kata-kata yang relevan dan merepresentasikan isi dari suatu dokumen.
- Pattern Discovery* merupakan tahap penting untuk menemukan pola atau pengetahuan (*knowledge*) dari keseluruhan teks. (Eko, 2011)

1.2.2. Ekstraksi Dokumen

Cara yang digunakan dalam mempelajari suatu data teks, adalah dengan terlebih dahulu menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen. Sebelum menentukan fitur-fitur yang mewakili, diperlukan tahap *preprocessing* yang dilakukan secara umum dalam teks mining pada dokumen, yaitu dimulai dari proses *case folding*, *tokenizing*, *filtering*, dan *stemming*.

1.2.3. Naive Bayes Classifier (NBC)

Menurut Hamzah, metode NBC mempunyai 2 tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sedapat mungkin menjadi representasi dokumen. Selanjutnya penentuan probabilitas *prior* bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi, ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi (Hamzah, 2012).

Jika diasumsikan secara lebih konkrit, terdapat koleksi dokumen sebagai berikut $D = \{d_i \mid i=1,2, \dots, |D|\} = \{d_1, d_2, \dots, d_{|D|}\}$ dan koleksi kategori $V = \{v_j \mid j=1,2, \dots, |V|\} = \{V_1, V_2, \dots, V_{|V|}\}$. Klasifikasi NBC dilakukan dengan cara mencari probabilitas $P(V=v_j \mid D=d_i)$, yaitu probabilitas kategori v_j jika diketahui dokumen d_i . Dokumen d_i dipandang sebagai tuple dari kata-kata dalam dokumen, yaitu $\langle a_1, a_2, \dots, a_n \rangle$, yang frekuensinya diasumsikan sebagai variabel random dengan distribusi probabilitas *Bernoulli* (McCallum dan Nigam, 1998). Selanjutnya klasifikasi dokumen adalah mencari nilai maximum dari:

$$V_{map} = \operatorname{argmax}_j P(V_j | a_1, a_2, \dots, a_n) \quad (1)$$

Dengan menerapkan Teorema Bayes persamaan 1 dapat ditulis seperti berikut:

$$V_{map} = \operatorname{argmax} \frac{P(a_1, a_2, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, \dots, a_n)} \tag{2}$$

Karena nilai $P(a_1, a_2, \dots, a_n)$ untuk semua v_j besarnya sama maka nilainya dapat diabaikan, sehingga persamaan (3) menjadi:

$$V_{map} = \operatorname{argmax} P(a_1, a_2, \dots, a_n | V_j) P(V_j) \tag{3}$$

Sehingga persamaan 4 dapat ditulis seperti berikut:

$$V_{map} = \operatorname{argmax} P(V_j) \prod_i P(a_i | V_j) \tag{4}$$

Nilai $P(v_j)$ ditentukan pada saat pelatihan, yang nilainya didekati dengan:

$$P(V_j) = \frac{|doc_j|}{|contoh|} \tag{5}$$

Dimana $|doc_j|$ adalah banyak dokumen yang memiliki kategori j dalam pelatihan sedangkan $|contoh|$ banyaknya dokumen dalam contoh yang digunakan untuk pelatihan.

Untuk nilai $P(a_i | V_j)$, yaitu probabilitas kata a_i dalam kategori j ditentukan dengan:

$$P(a_i | V_j) = \frac{n_k + 1}{n + |vocabulary|} \tag{6}$$

Dimana n_k adalah frekuensi munculnya kata a_i dalam dokumen berkategori v_j , dan $|vocabulary|$ adalah banyaknya kata dalam contoh pelatihan.

1.2.4. Konten / Komentar Negatif

Konten negatif adalah gambar porno, pelecehan, pencemaran nama baik dan status yang di-publish oleh seseorang, mengandung arti SARA dan tidak mendidik (Kominform, 2006).

1.3. Penelitian Terdahulu

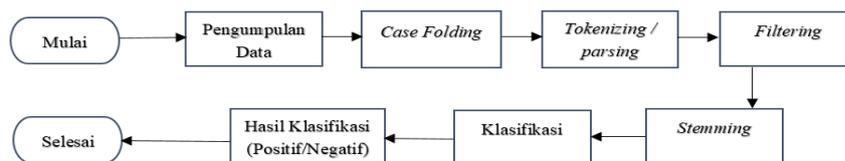
Berdasarkan penelitian Hu dkk yang meneliti tentang *Recognition of Pornographic Web Pages by Classifying Texts and Images* dengan menggunakan algoritma C4.5, SVM, NBC. Hasil yang didapatkan adalah akurasi klasifikasi sebesar 99.3% untuk klasifikasi *continuous pages* (300 teks porno, 300 teks tentang seks, dan 1000 teks normal). Hasil selanjutnya sebesar 91.6% untuk klasifikasi *discrete text* (1000 teks porno dan 2000 teks normal) (Hu dkk., 2007).

Selanjutnya Abidin dkk meneliti tentang *System of Negative Indonesian Website Detection Using TF-IDF and Vector Space Model* dengan menggunakan algoritma TF-IDF dan Vector Space Model. Hasil Akurasi yang dihasilkan pada penelitian tersebut adalah 82.80% dengan data latih 100 dan data uji 183 (Abidin dkk., 2014).

Penelitian berikutnya dilakukan oleh Dan dkk dengan judul *Research of Text Categorization on Weka* dengan membandingkan beberapa algoritma klasifikasi. Hasil yang didapatkan adalah SVM dan Naïve Bayes memiliki performa yang lebih baik dibandingkan metode lain. Performa klasifikasi juga dipengaruhi oleh *corpus category/class* (kondisi data) (Dan dkk., 2013).

2. METODOLOGI

Flowchart penelitian yang mendukung penelitian ini adalah sebagai berikut :



Gambar 1. Flowchart Penelitian

Berdasarkan Gambar 1 di atas, berikut dijelaskan secara lebih detail mengenai langkah-langkah pembuatan penelitian ini.

a) Pengumpulan Data

Pada tahap ini dilakukan pencarian data yang digunakan dalam penelitian ini, yaitu ±50 komentar instagram secara acak dari instagram penulis. komentar instagram ini tidak hanya diambil saat orang-orang menulis status sambil *upload* gambar/foto, tetapi juga komentar dari status yang dibuat melalui *insta story*. Komentar positif berjumlah 25 dan komentar negatif berjumlah 25.

b) *Case folding*

Case folding adalah proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf “a” sampai dengan “z” yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter (melebihi batas). Contoh status yang digunakan seperti dalam tabel 1.

Tabel 1. Proses Case Folding

| Komentar di Instagram | Proses Case Folding |
|--|---|
| Cocoknya BENCONG ini dipenjara saja karena mulutnya itu tak pernah ke sekolahan!!! | cocoknya bencong ini dipenjara saja karena mulutnya itu tak pernah ke sekolahan |
| YANG PENTING DIA CIPTAAN ALLAH, JADI YA HARUS DISYUKURI... | yang penting dia ciptaan allah jadi ya harus disyukuri |

c) *Tokenizing / parsing*

Tokenizing / parsing adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya seperti dalam tabel 2.

Tabel 2. Proses Tokenizing/parsing

| Proses Case Folding | Proses Tokenizing/parsing | Proses Case Folding | Proses Tokenizing/parsing |
|---|--|--|--|
| cocoknya bencong ini dipenjara saja karena mulutnya itu tak pernah ke sekolahan | cocoknya bencong ini dipenjara saja karena mulutnya itu tak pernah ke sekolahan | yang penting dia ciptaan allah jadi ya harus disyukuri | yang penting dia ciptaan allah jadi ya harus disyukuri |

d) *Filtering*

Filtering adalah tahap mengambil kata-kata penting dari hasil token (Tabel 3). Tahap ini bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh stopwords adalah “yang”, “dan”, “di”, “dari”, dan seterusnya.

Tabel 3. Proses Filtering

| Proses Tokenizing/parsing | Proses Filtering : <i>stopword</i> | Proses Tokenizing/parsing | Proses Filtering : <i>stopword</i> |
|--|---|--|------------------------------------|
| cocoknya bencong ini dipenjara saja karena mulutnya itu tak pernah ke sekolahan | cocoknya bencong dipenjara mulutnya sekolahan | yang penting dia ciptaan allah jadi ya harus disyukuri | ciptaan allah disyukuri |

e) *Stemming*

Stemming adalah tahap mencari *root* kata dari tiap kata hasil *filtering*. Pada tahap ini dilakukan proses pengembalian berbagai bentukan kata ke dalam suatu representasi yang sama (Tabel 4). Sebagai contoh, kata bersama, kebersamaan, menyamai, akan di-stem ke *root word*-nya yaitu “sama”.

Tabel 4. Proses Stemming

| Proses <i>Filtering</i> : <i>stopword</i> | Proses <i>Stemming</i> | Proses <i>Filtering</i> : <i>stopword</i> | Proses <i>Stemming</i> |
|---|---|---|--------------------------|
| cocoknya bencong dipenjara mulutnya sekolahan | cocok bencong penjara mulut sekolah | ciptaan allah disyukuri | cipta allah syukur |

f) **Klasifikasi**

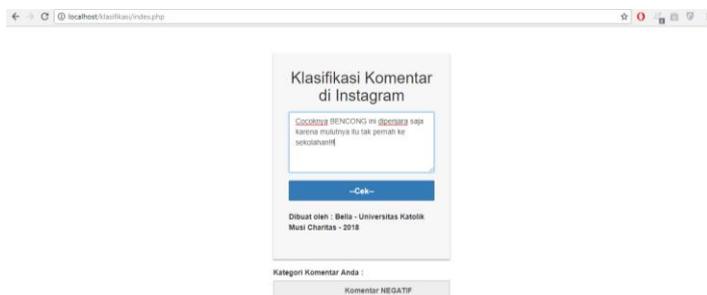
Proses yang dilakukan saat klasifikasi adalah sebagai berikut :

1. Buatlah dataset sesuai dengan kelasnya masing-masing, apakah positif atau negatif;
2. Buatlah data latihnya;
3. Hitunglah probabilitas setiap kata terhadap semua dataset, termasuk komentar yang memiliki kelas positif dan negatif;
4. Hasil klasifikasi ditentukan dari nilai probabilitas yang tertinggi.

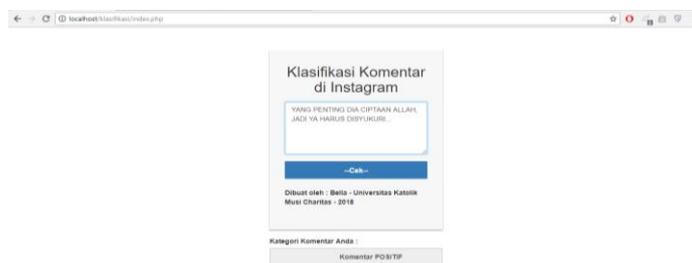
g) Hasil Klasifikasi : (Hasil klasifikasi dijelaskan di bagian Hasil dan Pembahasan)

3. HASIL DAN PEMBAHASAN

Hasil klasifikasi komentar di Instagram sesuai dengan harapan yang diinginkan peneliti. Pernyataan tersebut dibuktikan pada gambar-gambar di bawah ini, yaitu setelah dimasukkan beberapa komentar-komentar dari Instagram dan dilakukan cek, maka akan tampil kategori komentar yang dimasukkan, apakah positif atau negatif seperti dalam Gambar 3 dan 4.



Gambar 3. Hasil Klasifikasi Komentar Negatif



Gambar 4. Hasil Klasifikasi Komentar Positif

Hasil kategori komentar didasarkan pada dataset yang telah dibuat pada sistemnya. Pembahasan mengenai hasil-hasil klasifikasi yang dilakukan sebanyak 30 kali percobaan memberikan hasil akurat sebanyak 76,7% yaitu 23 pengguna yang mendapatkan jawaban sesuai harapannya. Sedangkan 7 pengguna lainnya tidak sesuai dengan harapan untuk kategori

komentarnya. Hal tersebut dikarenakan jumlah dataset yang digunakan hanya 50 komentar dari Instagram sehingga ada beberapa kata ataupun kalimat yang menurut beberapa pengguna negatif, tetapi karena di datasetnya tidak ada maka dimasukkan ke dalam komentar positif.

4. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut :

1. Sistem yang dibuat telah sesuai dengan harapan pengguna karena dapat menampilkan kategori komentar positif ataupun negatif.
2. Hasil akurasi dari 30 kali percobaan sebesar 76,7% yang sesuai dengan harapan pengguna

Saran yang dapat diberikan oleh peneliti adalah sebagai berikut :

1. Semakin banyak dataset, maka hasil yang didapatkan bisa lebih baik dan tingkat akurasi bisa lebih tinggi
2. Dapat dicoba dengan menggunakan metode klasifikasi yang lain seperti NN, KNN, VSM, dan lainnya

DAFTAR PUSTAKA

- Abidin, Zainil, Teguh Bharata Adji, dan Hanung Adi Nugroho, (2014), System of Negative Indonesian Website Detection Using TF-IDF and Vector Space Model, *IEEE International Conference on Electrical Engineering and Computer Science*.
- Aliandu, P., (2012), Analisis Sentimen Tweet Berbahasa Indonesia di Twitter, Tesis, Universitas Gadjah Mada, Yogyakarta.
- Anna, Lusiana Kus. *Instagram, Media Sosial Paling Buruk bagi Kesehatan Mental*. <https://lifestyle.kompas.com/read/2017/07/14/073913720/instagram-media-sosial-paling-buruk-bagi-kesehatan-mental>. Diakses: 15 Mei 2018, jam 20.05.
- Ardiansyah, Ghazali, M.I., (2016), Pengujian Usability User Interface Dan User Experience Aplikasi E-Reader Skripsi Berbasis Hypertext, *Jurnal Ilmiah Teknologi Informasi Terapan*, Vol.2,
- Dan, Li, dkk., (2013), Research of Text Categorization on WEKA, *Third International Conference on Intelligent System Design and Engineering Applications*.
- Eko, Nugroho., (2011), Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp, Skripsi, Universitas Brawijaya. Malang.
- Ghazali, M., (2016), *Buat Duit Dengan Facebook dan Instagram : Panduan Menjana Pendapatan dengan Facebook dan Instagram*, Publishing House, Malaysia, Hal. 8.
- Hamzah, A., (2012), Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis, *SNAST*, Periode III, pp.B-269-B-277.
- Hu, Weiming, dkk., (2007), Recognition of Pornographic Web Pages by Classifying Texts and Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6.
- Lusiana Kus Anna. *Instagram, Media Sosial Paling Buruk bagi Kesehatan Mental*, <https://lifestyle.kompas.com/read/2017/07/14/073913720/instagram-media-sosial-paling-buruk-bagi-kesehatan-mental>. Diakses: 20 Mei 2018, jam 19.30.
- Manning, C.D., Raghavan, P., Schütze, H., (2009), *Introduction to Information Retrieval*, Cambridge, England: Cambridge University Press.
- McCallum, A. dan Nigam, Kamal., (1998), A Comparison of Event Models for Naïve Bayes Text Classification, Thesis, University Pittsburgh, Pennsylvania.
- Santoso, Yohana. *Melindungi Keluarga dari Konten Negatif Dunia Maya*, https://kominform.go.id/content/detail/7606/melindungi-keluarga-dari-konten-negatif-dunia-maya/0/sorotan_media. Diakses: 21 Mei 2018, jam 12.45.