

Identifikasi Komentar Negatif Berbahasa Indonesia Pada Instagram Dengan Metode K-Means

Yoannes Romando Sipayung¹, Reny Sulistyowati², Iwan Setiawan Wibisono³

^{1,2} Jurusan S1 Teknik Informatika, Universitas Ngudi Waluyo

Email : mando19april@gmail.com¹, rensuyatin@gmail.com², loyal.wb99@gmail.com³

Abstract - Comments given by users or people on social media such as Instagram are varied, some are positive or negative. In terms of netizens, there is nothing wrong when you want to comment or pour out your heart. In Indonesia, denying netizens comments related to negativity, agreeing to agree with the law and the police. This research will help people who want to send comments that they make so as not to contain negative content so they can avoid cyber crime. The method used in supporting this research is the K-Mean method, to determine whether the comments entered are positive or negative. In this study, the data used were 40 comments that were used as research objects.

Keywords: Negative Comments, Instagram, K-Means

I. PENDAHULUAN

Instagram merupakan salah satu aplikasi media sosial berbasis web dan mobile yang khusus digunakan untuk mengunggah gambar / foto. Instagram merupakan situs media sosial yang semakin banyak digunakan terutama oleh para artis/aktor Indonesia. Para pengguna Instagram akan mengunggah foto-foto kegiatan mereka dan kemudian untuk masing-masing foto / gambar tersebut akan dapat diberi caption, tagging akun Instagram lainnya, lokasi tempat kejadian foto tersebut diunggah, edit foto sesaat sebelum diunggah langsung dari aplikasi smartphone, dan hashtag tertentu agar foto tersebut makin banyak dilihat orang lain. Aplikasi Instagram dikembangkan untuk smartphone, baik platform iOS, Android, ataupun Windows Phone dan bersifat gratis. Instagram berbasis web dapat diakses di <http://www.instagram.com>.

Salah satu hal yang menyebabkan Instagram banyak digunakan adalah kemudahannya untuk mengunggah foto langsung dari smartphone, mengingat pengguna media sosial kebanyakan adalah orang muda dan sangat menyukai selfie. Namun disamping kelebihan tersebut tentu terdapat kekurangan yang cukup mengganggu yaitu banyaknya komentar yang dapat dikategorikan sebagai komentar spam terhadap suatu post foto yang diunggah pada Instagram. Komentar spam akan semakin banyak terhadap Instagram artis/orang terkenal karena follower-nya juga semakin banyak.

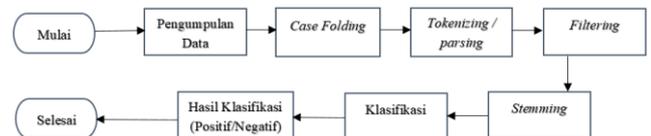
Beberapa peneliti sudah melakukan penelitian tentang Instagram, yaitu: Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes” oleh Antonius Rachmat, dan Yuan Lukito, pada tahun 2017. Metode yang digunakan adalah Naive Bayes. Penelitian lainnya adalah “Deteksi Komentar Negatif Di Instagram

Menggunakan Algoritma Naive Bayes Classifier”, oleh Maria Bellanier Ismiati, pada tahun 2018, yang menggunakan metode Naive Bayes Classifier.

II. METODE PENELITIAN

A. Diagram Alur Metode Penelitian

Terdapat beberapa tahapan dalam proses penelitian ini, yang dapat dilihat dalam diagram alur dibawah ini:



Gambar 1 Diagram Alur Metode Penelitian

B. Pengumpulan Data

Tahap pengumpulan data digunakan untuk mengumpulkan data yang berhubungan dengan penelitian yang akan dilakukan. Tahap ini akan menggunakan Instagram untuk pengambilan data dari Instagram yang berasal dari 5 (lima) artis Indonesia yang memiliki follower lebih dari 1 juta. Pada tahap ini juga termasuk pre-processing data. Data uji akan disimpan dalam basis data yang berisi data username IG artis, posting IG, tanggal posting, dan semua komentar dari 10 status terbaru, siapa yang berkomentar, dan tanggal komentar.

C. 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).

D. Tokenizing

Tokenizing / parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.

E. Filtering

Filtering adalah tahap mengambil kata-kata penting dari hasil token. 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.

F. Stemming

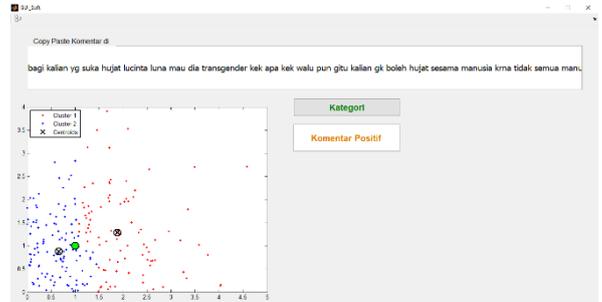
Stemming merupakan sebuah proses ekstraksi dari sebuah kata yang digunakan untuk mendapatkan kata dasar dari suatu kata, proses stemming dilakukan setelah proses preprocessing pada data mentah yang diambil dari instagram.

G. Klasifikasi

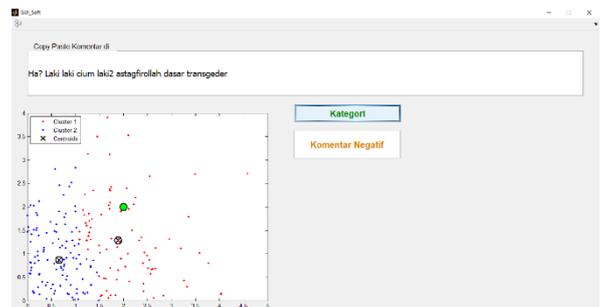
Proses klasifikasi dilakukan dengan menggunakan metode K-Means. Hasil klasifikasi metode tersebut untuk mendapatkan kesimpulan akhir bahwa komentar input mengandung unsur negatif atau positif.

III. HASIL DAN LUARAN YANG DICAPAI

Pada penelitian ini dilakukan proses percobaan terhadap metode K-Means. Hasil klasifikasi komentar di Instagram sesuai dengan yang diharapkan oleh peneliti. Pernyataan tersebut dibuktikan pada gambar-gambar di bawah ini, yaitu setelah dimasukkan beberapa komentar-komentar dari Instagram dan dilakukan cek, maka akan ditampilkan kategori komentar yang dimasukkan, apakah komentar positif atau negatif.



Gambar 2 Hasil Klasifikasi Komentar Positif



Gambar 3 Hasil Klasifikasi Komentar Negatif

IV. KESIMPULAN

Dari percobaan yang telah dilakukan dapat disimpulkan bahwa metode yang digunakan pada penelitian ini mampu mendeteksi komentar negatif dari instagram berbahasa Indonesia. Metode ini bergantung pada data training dan list badwords yang dapat ditambah dan diperbaiki untuk mendapatkan akurasi yang lebih tinggi dan membuat program mengenali kata hinaan yang lebih beragam.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada LPPM Universitas Ngudi Waluyo yang telah mendanai penelitian ini.

REFERENSI

- [1] 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.
- [2] Aliandu, P., (2012), Analisis Sentimen Tweet Berbahasa Indonesia di Twitter, Tesis, Universitas Gadjah Mada, Yogyakarta.
- [3] Anna, Lusia 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.
- [4] 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,
- [5] Dan, Li, dkk, (2013), Research of Text Categorization on WEKA, Third International Conference on Intelligent System Design and Engineering Applications.
- [6] Eko, Nugroho., (2011), Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp, Skripsi, Universitas Brawijaya. Malang.
- [7] Ghazali, M., (2016), Buat Duit Dengan Facebook dan Instagram : Panduan Menjana Pendapatan dengan Facebook dan Instagram, Publishing House, Malaysia, Hal. 8.
- [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.
- [9] 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.
- [10] Lusia 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.
- [11] Manning, C.D., Raghavan, P., Schutze, H., (2009), Introduction to Information Retrieval, Cambridge, England: Cambridge University Press.
- [12] McCallum, A. dan Nigam, Kamal., (1998), A Comparison of Event Models for Naïve Bayes Text Classification, Thesis, University Pittsburgh, Pennsylvania.