

Rancangan Klasifikasi Twitter Pada Perguruan Tinggi

Thomas E Tarigan^{*1}, Robby C Buwono², Sri Redjeki³

^{1,2,3}STMIK AKAKOM; Jl Raya Janti No 143 Yogyakarta, 0274-486664

e-mail: ¹tarigan@akakom.ac.id, ²robbycokro@akakom.ac.id, ³dzeky@akakom.ac.id

Abstrak

Tujuan dari penelitian ini adalah melakukan perancangan atau pemodelan klasifikasi twitter pada peningkatan pengelolaan perguruan tinggi. Hasil rancangan ini akan membantu perguruan tinggi untuk mengetahui opini publik secara langsung dan terus berubah melalui media sosial. Penelitian ini memanfaatkan twitter untuk mendapatkan klasifikasi opini atau pendapat yang dapat memberikan feedback terhadap pengelolaan perguruan tinggi. Metode klasifikasi yang digunakan pada penelitian ini adalah Naïve Bayes Classifier (NBC). Data yang digunakan adalah data twitter dengan keyword yang berhubungan dengan salah satu obyek penelitian yaitu perguruan tinggi swasta bidang IT. Proses crawling data twitter sangat dipengaruhi oleh keyword yang kita gunakan pada saat crawling. Hasil crawling akan dilakukan proses preprosesing sehingga data text akan diklasifikasi menjadi 3 kelas yaitu positif, netral dan negatif. Penelitian ini menggunakan data sebanyak 1000 tweet yang digunakan untuk keperluan data training. Kemudian 20 data tweet digunakan untuk testing. Hasil penelitian memberikan akurasi NBC sebesar 75%

Kata kunci— Klasifikasi, NBC, Perguruan Tinggi, Twitter

Abstract

The purpose of this study is to design or model the Twitter classification of improving the management of tertiary institutions. The results of this design will help universities to find out public opinion directly and keep changing through social media. This research provides twitter to get an opinion rating or opinion that can provide feedback on the management of tertiary institutions. The classification method used in this study is Naïve Bayes Classifier (NBC). The data used is twitter data with keywords related to one of the research objects, namely private IT colleges. The process of crawling Twitter data is very much needed by the keywords we use at the time of crawling. The results of crawling will be carried out the preprocessing process so that the text data will be classified into 3 classes, namely positive, neutral and negative. This study uses 1000 data tweets used for data training purposes. Then 120 data tweets are used for testing. The research results give an accuracy of NBC above 75%

Keywords—Classification, NBC, Higher Education, Twitter

1. PENDAHULUAN

Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Popularitas media jejaring sosial terus mengalami peningkatan dalam beberapa tahun terakhir. Media jejaring sosial seperti *Twitter*, *Facebook*, dan *Youtube* merupakan beberapa media perangkat komunikasi

terpopuler yang ada di kalangan pengguna internet saat ini [1]. Oleh karena itu media jejaring sosial banyak dimanfaatkan oleh Perguruan Tinggi untuk memberikan informasi seputar Perguruan Tinggi. Pengguna twitter di Indonesia pada tahun 2018 mengalami kenaikan 11% dibandingkan tahun sebelumnya. Hal ini semakin banyaknya pengguna media social di Indonesia menggunakan twitter untuk menuliskan opini mereka secara langsung di media social. Twitter terus mengalami peningkatan jumlah pengguna sejak kemunculannya pada tahun 2006. Data statistika menunjukkan bahwa twitter memiliki lebih dari 1 Milyar pengguna dimana sebanyak 313 juta diantaranya adalah pengguna aktif. Banyak peneliti yang telah memberikan perhatian lebih pada media sosial. Analisis data menggunakan media sosial tidaklah mudah karena faktor ketidaklengkapan dan sifatnya yg dinamis [2] [3].

Twitter adalah jejaring sosial yang tumbuh paling cepat sejak 2006. Twitter adalah salah satu layanan microblogging real-time dengan menyediakan hanya 140 karakter pendek tetapi dapat memberikan wawasan atau makna yang cukup [4] [5]. Keberadaan twitter yang telah digunakan secara luas oleh berbagai lapisan masyarakat dapat dilihat sebagai sebuah refleksi yang baik dimana keberadaan twitter dapat mempresentasikan apa yang sedang menjadi tren pembicaraan dan hal apa yang sedang menarik untuk dibahas.

Penelitian yang membahas penggunaan media social untuk ekstraksi opini sebuah perguruan tinggi antara lain oleh Imam Fahrur dkk. Pada penelitian ini dikembangkan sistem *opinion mining* untuk menganalisis opini publik pada perguruan tinggi. Pada subproses *document subjectivity* dan *target detection* digunakan *Part-of-Speech (POS) Tagging* menggunakan *Hidden Markov Model (HMM)*. Pada hasil proses *POS Tagging* kemudian diterapkan *rule* untuk mengetahui apakah suatu dokumen termasuk opini atau bukan, serta untuk mengetahui bagian kalimat mana yang merupakan objek yang menjadi target opini. Dokumen yang dikenali sebagai opini selanjutnya diklasifikasikan ke dalam opini negatif dan positif (subproses *opinion orientation*) menggunakan *Naïve Bayes Classifier (NBC)*. Penelitian tentang penggunaan *NBC* sebagai metode pengklasifikasi teks telah dilakukan oleh SM Kamaruzzaman dan Chowdury Mofizur Rahman [6] [7] pada tahun 2004. Dari proses pengujian secara kualitatif disebutkan bahwa teks bisa diklasifikasikan dengan akurasi yang tinggi.

Beberapa tulisan mengenai media social untuk melakukan estimasi atau prediksi berbasis teks saat ini banyak dilakukan oleh peneliti bidang social maupun eksak. Big data salah satu bidang ilmu yang berkembang saat ini telah memberikan jalan untuk terus melakukan eksplorasi penelitian di media social. Salah satu media social yang banyak di eksplorasi adalah twitter. Beberapa penelitian yang dijadikan rujukan peneliti adalah oleh Imam Fahrur dkk tahun yang mengembangkan sistem untuk ekstraksi opini melalui twitter untuk melihat opini public terhadap sebuah perguruan tinggi. Ahmad Fathan Hidayatullah dan Azhari SN pada tahun 2014 dengan judul Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter [8]. Rizky Maulana dan Sri Redjeki 2016 membahas tentang analisis sentiment untuk tokoh di Indonesia menggunakan SVM berbasis Cloud Computing [9]. Tulisan lain dari Mutiah 2017 melakukan analisis sentimen yang digunakan untuk mereview restoran menggunakan teks Bahasa Indonesia.

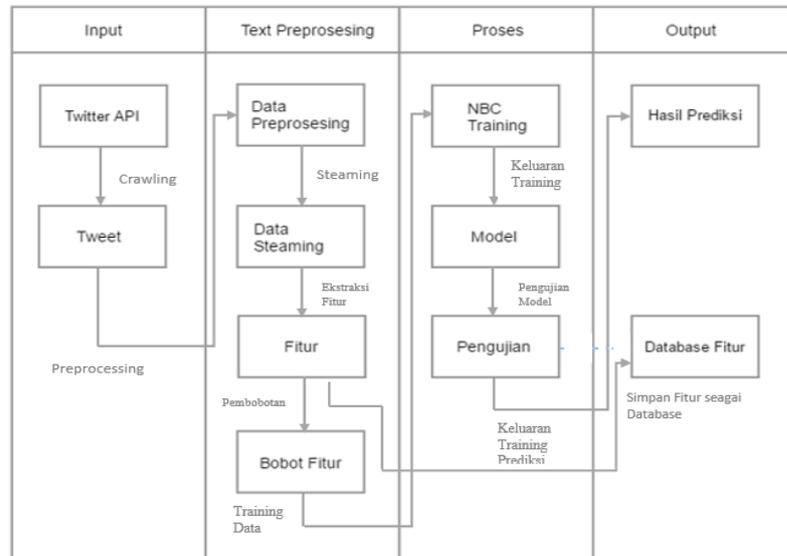
Penelitian ini melakukan analisis dan juga klasifikasi sentimen terhadap salah satu media social perguruan tinggi swasta bidang IT di Yogyakarta. Adapun metode yang digunakan dalam klasifikasi kategori sentiment adalah *Naïve Bayes Classifier*. Obyek yang ingin diklasifikasi bukan berada pada level dokumen melainkan kata pada kalimat twitter, serta mengklasifikasikan apakah *tweet* tersebut termasuk twitt positif, netral atau negatif.

2. METODE PENELITIAN

Penelitian ini menggunakan data twitter dengan keyword nama dan hal-hal yang terkait dengan perguruan tinggi yang akan di lakukan ekstrasi opini publik. Perguruan tinggi yang dijadikan obyek penelitian ini adalah perguruan tinggi bidang IT yang ada di Yogyakarta. Jumlah data twitter yang dikumpulkan sebanyak 1000 data twitter yang sudah bersih dan diberikan

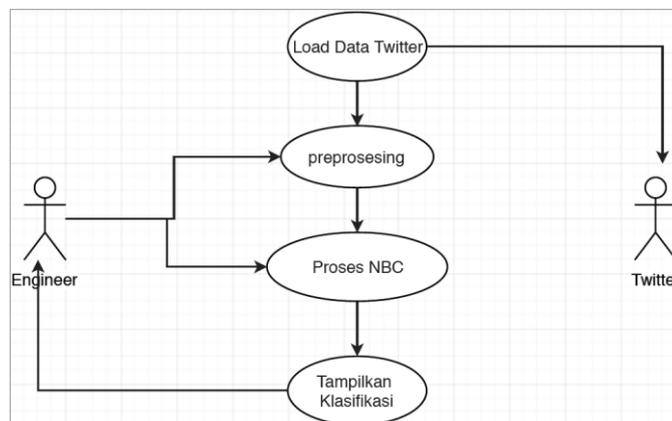
labelling class secara manual. Pemberian labelling class dilakukan dengan menggunakan kamus data kata dasar yang mempunyai makna sentimen positif, netral dan negatif.

Alur mekanisme penelitian ditunjukkan pada gambar 1 berupa blok diagram sistem. Pada gambar 1 terdapat 4 blok yang menunjukkan tahapan yaitu input yang terdiri dari 2 sub tahapan yaitu menggunakan twitter API untuk melakukan crawling data dari tweet yang ada. Keyword menggunakan kata-kata yang terkait dengan perguruan tinggi yang akan dijadikan obyek penelitian. Pada tahap berikutnya adalah preprocessing yang didalamnya terdapat 4 sub tahapan yaitu data preprosesing, proses steaming, ekstraksi fitur dan pemberian bobot masing-masing fitur dan pemberian labelling class. Data yang dihasilkan pada tahapan ini akan digunakan untuk tahapan proses. Pada tahapan ketiga atau proses ini mulai dilakukan pelatihan dan pengujian menggunakan metode klasifikasi yaitu Naïve Bayes Classifier.

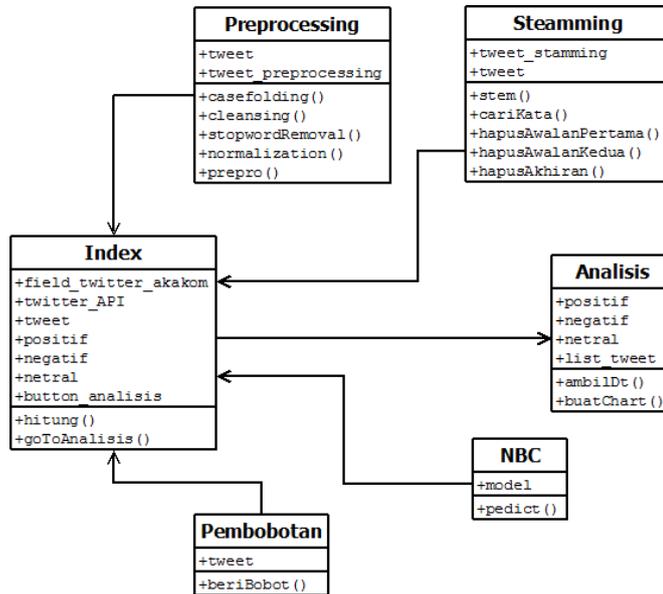


Gambar 1. Blok Diagram Sistem

Terdapat 3 klasifikasi pada sistem ini yaitu sentimen atau opini positif, netral dan negatif. Tahapan terakhir yaitu tahapan menghasilkan keluaran yang berupa prediksi, dimana user dapat memasukkan sembarang tweet atau opini dan akan diperoleh hasil prediksi klasifikasinya. Deskripsi hubungan antara sistem dan aktor yang dikembangkan pada penelitian ini ditunjukkan pada gambar 2. Terdapat 2 aktor yaitu engineer dan twitter API sebagai faktor eksternal sedangkan faktor internal sistem terdapat 4 fungsionalitas pokok yaitu load data twitter, preprocessing, proses menggunakan algoritma Naïve Bayes Classifier dan hasil penelitian berupa klasifikasi hasil analisis sentiment menjadi 3 kelas yaitu positif, netral dan negatif.



Gambar 2. Diagram Usecase Sistem Klasifikasi



Gambar 3. Class Diagram Sistem

Pada gambar 3 menunjukkan class diagram sistem yang dikembangkan pada penelitian ini. Terdapat 6 class yang digunakan pada sistem yaitu class preprosesing, class steaming, class indexing, class analisis, class bobot dan class NBC. Sistem pada sisi *client* digunakan untuk mempresentasikan data kedalam beberapa bagian, yaitu presentase sentimen dan data tweet, setelah mendapatkan data presentase sentimen dan data tweet maka aplikasi *client* akan mempresentasikannya kedalam bentuk *pie chart*. Pada bagian *server* terdapat beberapa jenis kelas, yaitu kelas *index* yang merupakan kelas utama perangkat lunak, serta terdapat kelas *preprocessing*, *steaming*, dan pembobotan.

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dibahas mengenai hasil training dan testing dari model yang dipakai yaitu Naïve Bayes Classifier (NBC). Jumlah data training yang digunakan sebanyak 1000 data tweet yang telah dilakukan preprosesing.

Tabel 1. Tweet dan Klasifikasi

No.	Kata	Kategori
1.	Fasilitas kampus lengkap	Positif
2.	Dosen ramah semua	Positif
3.	Mahasiswa jogja	Netral
4.	Kuliah padat	Netral
5.	Dosen sering terlambat	Negatif
6.	Kuliah susah sekali	Negatif

Tabel 1 adalah contoh tweet yang sudah bersih dan diberi labelling secara manual. Masing-masing data training akan dihitung nilai probabilitasnya terhadap kemungkinan ketiga

class yang ada menggunakan Bayesian. Data twitter yang telah diperoleh ditampilkan dalam dashboard gambar 4.

No.	Data Twit	Sentimen	Aksi
1	terbukti rating game ini memiliki 4 bintang di play store. Game ini dibuat khusus untuk dimainkan dengan VR	positif	Perbarui Hapus
2	game yang cukup menguji adrenalin ini cukup digemari para gamers dengan nyali tinggi, wow banget kan	netral	Perbarui Hapus
3	Roller Coaster VR merupakan salah satu games favorite yang dimainkan dengan VR	netral	Perbarui Hapus
4	Akakom membuka Pendaftaran mahasiswa baru gelombang 1 pada tanggal 1 mei sampai 30 juni 2017.	netral	Perbarui Hapus
5	stmk akakom me-retweet Kementerian Kominfo/Verified Masing-masing agar BACKUP DATA ke Storage TERPISAH	positif	Perbarui

Gambar 4. Dashboard Data

Untuk kamus data juga ditampilkan dalam dashboard gambar 5. User dapat melakukan editing data apabila ada perubahan kamus data. Menu ini untuk memudahkan manajemen kamus data yang mempunyai kecenderungan bertambah.

No.	Dictionary Word	Dictionary Stopword	Dictionary Sentiment	Aksi
1	a	Ya	netral	Perbarui Detail Hapus
2	ab	Ya	netral	Perbarui Detail Hapus
3	aba	Ya	netral	Perbarui Detail Hapus
4	aba-aba	Ya	netral	Perbarui Detail Hapus
5	abad	Ya	netral	Perbarui Detail Hapus
6	abadi	Bukan	positif	Perbarui Detail Hapus
7	abadiah	Ya	netral	Perbarui Detail Hapus
8	abah	Ya	netral	Perbarui Detail Hapus
9	abahai	Bukan	netral	Perbarui Detail Hapus

Gambar 5. Dashboard Kamus Data

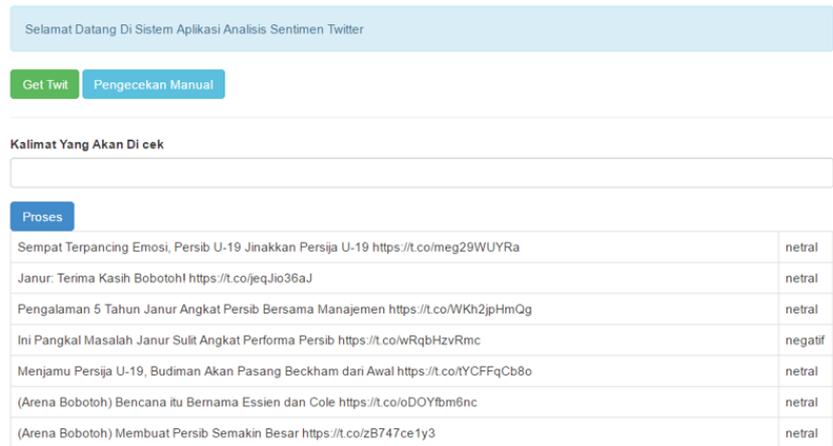


Figure 6. Dashboard Testing Tweet

Gambar 6 digunakan untuk melihat hasil klasifikasi tweet yang inputkan melalui dashboard testing tweet. Hasil klasifikasi akan ditampilkan dalam bentuk daftar tweet dan jenis klasifikasinya.

Dengan menggunakan data training 1000 dan data *testing* sebanyak 20 lalu dilakukan pengujian secara manual dan dibandingkan dengan hasil klasifikasi yang dilakukan oleh *system*. Didapatkan kesalahan sebanyak 5 data *testing* dan benar 15 data *testing*, lalu dari hasil tersebut didapatkan nilai akurasi sebanyak 75%.

4. KESIMPULAN

Beberapa hal yang dijadikan kesimpulan dari hasil penelitian, antara lain :

1. Metode Naïve Bayes Classifier dapat dijadikan sebagai algoritma yang cukup baik untuk melakukan analisis sentiment
2. Hasil Akurasi pengujian klasifikasi dengan metode Naive Bayes Classifier diperoleh sebesar 75%.
3. Kalimat berbahasa Indonesia dengan struktur dan variasi kalimat dan bahasa yang lebih rumit membutuhkan banyak data latih untuk mendapatkan hasil yang baik.
4. Opini publik terhadap sebuah perguruan tinggi dapat dijadikan referensi bagi perguruan tinggi untuk meningkatkan pengelolaan dan kinerja.

5. SARAN

Untuk meningkatkan hasil penelitian lebih baik lagi, hal-hal yang perlu dilakukan untuk pengembangan yaitu :

1. Menambah data pelatihan
2. Membandingkan dengan metode lainnya
3. Melakukan uji kedekatan kata-kata keyword dengan obyek penelitian

DAFTAR PUSTAKA

- [1] D. Chaffey and F. Ellis-Chadwick, *Digital Marketing*, 6th Editio. Pearson, 2016.
- [2] P. Wang, B. W. Xu, Y. R. Wu, and X. Y. Zhou, "Link prediction in social networks: the state-of-the-art," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–38, 2014.
- [3] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein, "Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance," *PLoS Comput. Biol.*, vol. 11, no. 10, pp. 1–15, 2015.
- [4] N. Li and D. D. Wu, "Using text mining and sentiment analysis for online forums

- hotspot detection and forecast,” *Decis. Support Syst.*, vol. 48, no. 2, pp. 354–368, 2010.
- [5] C. Li, J. Bai, L. Zhang, H. Tang, and Y. Luo, “Opinion community detection and opinion leader detection based on text information and network topology in cloud environment,” *Inf. Sci. (Ny)*, vol. 504, pp. 61–83, 2019.
- [6] S. Kamruzzaman *et al.*, “Text Categorization using Association Rule and Na? ve Bayes Classifier,” *Asian J. Inf. Technol.*, vol. 3, no. 9, pp. 657–665, 2004.
- [7] “Manuscript_sriRedjeki.” .
- [8] I. F. Rozi, S. H. Pramono, and E. A. Dahlan, “Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi,” *Electr. Power, Electron. Commun. Control. Informatics Semin.*, vol. 6, no. 1, pp. 37–43, 2012.
- [9] R. Maulana and S. Redjeki, “Analisis Sentimen Pengguna Twitter Menggunakan Metode Support Vector Machine Berbasis Cloud Computing,” *J. TAM (Technol. Accept. Model)*, vol. 6, no. 1, pp. 23–28, 2016.