

Faktor Exacta 12 (4): 315-322, 2019
p-ISSN: 1979-276X
e- ISSN: 2502-339X
DOI: 10.30998/faktorexacta.v12i4.5205

Nooraeni, dkk-Analisis Sentimen Publik Terhadap.....

**ANALISIS SENTIMEN PUBLIK TERHADAP SISTEM ZONASI
SEKOLAH MENGGUNAKAN DATA *TWITTER* DENGAN METODE
*NAÏVE BAYES CLASSIFICATION***

**RANI NOORAENI
AMIRAH BALQIS SAFIRUDDIN
AULIA FATIN AFIFAH
KRISNA DWI AGUNG WIJAYA
NADA NABILA ROSYAD**

Statistika Sosial dan Kependudukan, Politeknik Statistika STIS
Jl. Otto Iskandardinata No. 64C Jakarta 13330

Email: raninoor@stis.ac.id, mrhbalqis@gmail.com, auspiciousaulia@gmail.com,
krisnawijaya011@gmail.com, nabilarosyad15@gmail.com

Abstract. *School zonation (Zonasi Sekolah) is an education system that regulates the acceptance process of new students according to the student's residence area so there will be no schools' labelling as favorite or non-favorite schools. School zonation system is only applied in public schools, namely schools that are managed by the government. In addition, this system has been implemented in various countries such as United State, Canada, Japan, and Australia. School zonation system has both positive and negative impacts for the students. In the middle of this year, the Twitter timeline was enlivened with this issue. Many Twitter users wrote their opinion towards the issue in the form of support, suggestion, and criticism. Public criticism on the school zonation system is something that must be considered as a suggestion elaborate this new system. The writer is interested in researching the public opinions toward the implementation of school zonation system by using Twitter data in the period of June – July 2019. From the public's tweets, the public sentiment will be analyzed towards such implementation.*

Keywords: Zonation System, Sentiment, Twitter Data.

Abstrak. Zonasi adalah sistem pengaturan proses penerimaan siswa baru sesuai dengan wilayah tempat tinggal agar tidak ada sekolah yang dianggap favorit dan non-favorit. Sistem zonasi hanya diterapkan di sekolah negeri yaitu sekolah yang dikelola pemerintah. Selain itu, sistem ini telah diterapkan di berbagai negara seperti Amerika Serikat, Kanada, Jepang, dan Australia. Sistem zonasi memiliki dampak positif dan negatif bagi peserta didik. Pada pertengahan tahun ini, Twitter diramaikan dengan isu tentang penerapan sistem zonasi. Pengguna Twitter banyak yang menuliskan pendapatnya mengenai isu tersebut berupa dukungan, masukan, dan juga kritikan. Kritikan masyarakat terhadap sistem zonasi merupakan hal yang harus diperhatikan untuk menyempurnakan sistem zonasi yang masih terbilang baru. Peneliti tertarik untuk meneliti opini masyarakat mengenai sistem zonasi menggunakan data Twitter periode bulan Juni – Juli 2019. Dari cuitan-cuitan masyarakat tersebut, akan dilakukan analisis sentimen public terhadap penerapan sistem zonasi sekolah.

Kata Kunci: Sistem Zonasi, Sentimen, Data Twitter

PENDAHULUAN

Salah satu tujuan dari berdirinya bangsa Indonesia tertuang dalam Pancasila sila ke-5 yaitu "Keadilan sosial bagi seluruh rakyat Indonesia". Keadilan sosial meliputi keadilan dalam ekonomi, kesehatan, pendidikan, dan lain-lain. Selama 74 tahun Indonesia merdeka, tujuan tersebut belum sepenuhnya tercapai, terutama dalam ranah pendidikan. Salah satu masalah pendidikan di Indonesia adalah kurangnya pemerataan antar daerah. Hal tersebut disebabkan oleh berbagai faktor seperti kondisi geografis, infrastruktur, ekonomi, kurangnya tenaga pendidikan,

dan masih banyak faktor lainnya (Aristo, 2019). Faktor-faktor tersebut menyebabkan masyarakat melabeli sekolah dengan sekolah favorit dan non-favorit.

Seperti yang telah diketahui secara umum, pendidikan merupakan hal penting dalam pemerintahan negara, negara belum bisa dikatakan maju apabila pendidikannya masih terbelakang dan belum merata. Pemerataan pendidikan sangat penting karena secara tidak langsung terjadi upaya pemerataan kualitas tenaga kerja dalam jangka panjang yang diharapkan pemerataan pembangunan dapat tercapai.

Pemerintah telah banyak melakukan upaya-upaya untuk meningkatkan pemerataan pendidikan di Indonesia. Menurut Pribadi (2017), pemerintah secara nasional telah melakukan upaya-upaya pemerataan pendidikan berupa pembebasan biaya bagi Sekolah Dasar (SD), program Dana BOS, mengalokasikan anggaran pendidikan sebesar 20 persen dari APBN, serta bantuan bagi siswa-siswi SMP dan SMA yang kurang mampu. Selain itu, pada tahun ajaran 2019/2020 ini, pemerintah sedang menerapkan program sistem zonasi sekolah.

Zonasi adalah sistem pengaturan proses penerimaan siswa baru sesuai dengan wilayah tempat tinggal agar tidak ada sekolah yang dianggap favorit dan non-favorit. Dengan begitu, sistem tersebut diharapkan dapat menyetarakan kualitas pendidikan di Indonesia karena pada dasarnya setiap anak memiliki hak yang sama sehingga tidak boleh ada diskriminasi, hak eksklusif, dan kompetisi yang berlebihan untuk mendapatkan layanan pemerintah, dalam hal ini yaitu sekolah. Sistem zonasi hanya diterapkan di sekolah negeri yaitu sekolah yang dikelola pemerintah. Selain itu, sistem ini telah diterapkan di berbagai negara seperti Amerika Serikat, Kanada, Jepang, dan Australia. Sistem zonasi diatur dalam Permendikbud No. 14 Tahun 2018. Sistem zonasi memiliki dampak positif dan negatif bagi peserta didik. Dampak positif di antaranya yaitu adanya pemerataan sekolah sehingga tidak ada sekolah unggulan dan sekolah regular, peserta didik dapat lebih dekat dengan sekolah yang dituju, dan sistem zonasi mendorong kualitas setiap sekolah untuk menjadi lebih baik. Selain itu, Kabid Lalu Lintas Dinas Perhubungan Depok mengatakan bahwa sistem zonasi dapat mengurangi arus pergerakan kendaraan bermotor pribadi yang mengantar jemput peserta didik karena banyak sekolah favorit yang selama ini menyumbang kemacetan saat jam masuk dan pulang sekolah (Media Indonesia, 2019). Sedangkan dampak negatif dari sistem zonasi di antaranya yaitu melanggar hak anak untuk memperoleh pendidikan yang bermutu sesuai dengan UU No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional (UU Sisdiknas) pasal 5 ayat 1, pemalsuan domisili agar lebih dekat dengan sekolah yang diinginkan, dan membuat peserta didik tidak semangat belajar karena menganggap bahwa untuk masuk ke sekolah favorit tidak dibutuhkan kerja keras karena bergantung pada domisili. Dampak-dampak tersebut memunculkan pro dan kontra di masyarakat terkait pelaksanaan sistem zonasi. Selain itu, sistem zonasi masih memiliki beberapa kekurangan seperti belum meratanya mutu sekolah, dan beberapa murid malah diterima di sekolah yang jaraknya lebih jauh dari tempat tinggalnya.

Twitter merupakan salah satu media sosial yang banyak digemari masyarakat. Menurut Antara News, pengguna Twitter Indonesia meningkat 11 persen pada tahun 2018. Twitter merupakan salah satu media sosial yang memungkinkan penggunanya untuk mengirim dan membaca pesan yang berbasis teks. Pengguna Twitter banyak membicarakan isu-isu hangat dari berbagai dimensi, mulai dari *entertainment*, politik, sosial, hingga pemerintahan. Pada pertengahan tahun ini, Twitter diramaikan dengan isu tentang penerapan sistem zonasi. Pengguna Twitter banyak yang menuliskan pendapatnya mengenai isu tersebut berupa dukungan, masukan, dan juga kritikan. Kritikan masyarakat terhadap sistem zonasi merupakan hal yang harus diperhatikan untuk menyempurnakan sistem zonasi yang masih terbilang baru. Kritikan-kritikan ini dapat ditemukan baik berupa lisan maupun tulisan di media sosial seperti Twitter, Facebook dan media sosial lainnya. Data Twitter dapat dijadikan salah satu sumber data penelitian. Periode data yang diambil dari Twitter adalah bulan Juni – Juli 2019 karena pada Penerimaan Peserta Didik Baru (PPDB) dilaksanakan pada periode tersebut.

Penelitian ini bertujuan untuk mengetahui sentimen publik terhadap penerapan sistem zonasi sekolah. Nantinya penelitian ini diharapkan dapat menjadi bahan masukan bagi pemerintah dalam penyempurnaan pelaksanaan sistem zonasi sekolah.

METODE

Metode Pengumpulan Data

Penelitian ini menggunakan data sekunder yaitu data Twitter. Data yang digunakan pada penelitian ini berasal dari *tweet* dengan kata kunci “sistem”, “zonasi” dan “sekolah” yang dikumpulkan menggunakan *twitterscraper* melalui *anaconda prompt* dari tanggal 1 Juni 2019 hingga 31 Juli 2019. Data yang diperoleh berupa data kotor berjumlah 8387 observasi, 16 variabel dalam format CSV. Selanjutnya, variabel yang akan diolah adalah variabel *tweet*.

Tahap Pre-processing

1. *Cleaning* atau menghilangkan *noise* (html, url, emoji, hashtag, dan enter) pada setiap *tweet*. Tahap ini dilakukan untuk membuang gangguan (*noise*) pada data agar kualitas data menjadi lebih baik sehingga hasil penelitian lebih akurat.
2. Membenahi *slang words*. Tahap ini dilakukan untuk mengonversi kata-kata singkatan yang sering digunakan dalam bahasa *tweet* menjadi kata baku untuk mengurangi bias pada saat dilakukan analisis.
3. *Text Stripping* merupakan tahapan untuk menghilangkan karakter selain huruf serta mengonversi huruf kapital.
4. Menghilangkan duplikasi *tweet* yang mempunyai kesamaan isi agar tidak terjadi redundansi. Dari tahap ini *tweet* yang berjumlah 8387 berkurang menjadi menjadi 7678 *tweet*.
5. *Stemming* merupakan proses mengubah kata-kata dalam *tweet* menjadi bentuk kata dasar untuk mengurangi variasi kosakata agar analisis sentimen menjadi lebih fokus. Algoritma yang digunakan adalah algoritma “Nazief and Andriani”. Algoritma ini memungkinkan penghapusan prefiks, sufiks, dan keduanya. Namun belum bisa menghapus infiks.
6. Tokenisasi merupakan tahap pemisahan *tweet* berdasarkan tiap kata yang menyusunnya.
7. Menyaring kata-kata penting dari hasil tokenisasi dengan algoritma *stopword* yang memungkinkan proses penghapusan kata-kata yang kurang penting atau tidak bermakna seperti “yang”, “kalau”, “adalah”, dsb.

Tahap *pre-processing* diatas mengacu pada Taneyasa, T. & Sitta, Y. (2019). Setelah dilakukan tahap *pre-processing*, selanjutnya dilakukan analisis sentimen dan klasifikasi Naïve Bayes.

Analisis Sentimen

Menurut Kristiyanti (2015), analisis sentimen adalah studi komputasi mengenai pendapat, perilaku, dan emosi seseorang terhadap entitas atau yang menggambarkan individu, kejadian maupun topik. Analisis sentimen adalah proses memahami, mengekstrak, dan mengolah data berupa teks untuk mendapatkan informasi guna mengetahui pendapat terhadap sebuah masalah dan dapat bermanfaat untuk memperoleh gambaran umum persepsi masyarakat terhadap kualitas layanan, pemantauan terhadap sebuah produk, prediksi penjualan, politik dan pengambilan keputusan para investor (Ipmawati, dkk, 2017).

Keuntungan analisis sentimen

- a. *Scalability*, analisis sentimen memungkinkan kita untuk memproses data dengan efisien dan berbiaya rendah.
- b. *Real-time analysis*

- c. *Consistent criteria*, menggunakan kriteria yang objektif saat menganalisis sentimen dibandingkan proses manual karena seseorang dalam menentukan suatu hal apakah positif atau negatif pasti dipengaruhi oleh faktor diri sendiri (pendapat, pengalaman dan lain-lain)

Naïve Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan Decision Tree dan Neural Network. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode Bayes merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (*posterior probability*)

P(H) = Probabilitas hipotesis H (*prior probability*)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Rumus diatas dapat ditulis juga seperti berikut :

$$posterior = \frac{prior \times likelihood}{evidence}$$

Nilai posterior yang dihasilkan nantinya akan dibandingkan dengan nilai posterior kelas lainnya untuk untuk menentukan klasifikasi sampel masuk ke kelas mana.

Akurasi :

Untuk mengukur keakuratan metode maka digunakan penghitungan *accuracy*, dimana semakin besar nilai *accuracy* maka metode semakin bagus. Adapun rumus menghitung akurasi adalah sebagai berikut.

$$accuracy = \frac{jumlah\ tweet\ terklasifikasi\ benar}{jumlah\ tweet\ keseluruhan} \times 100\%$$

Kelebihan:

- Mudah untuk dibuat
- Hasil bagus / akurasi tinggi

HASIL DAN PEMBAHASAN

Karakteristik Data

Penelitian ini menggunakan *twitterscraper* untuk mengambil data *tweets* dengan *keywords* “sistem”, “zonasi”, “sekolah”. *Tweets* yang terambil difilter berdasarkan bahasa Indonesia pada periode waktu 1 Juni sampai 31 Juli 2019. Selanjutnya, data *tweets* yang terambil akan disimpan dalam bentuk file *CSV* yang berisi atribut *username*, *fullname*, *user ID*, *tweet ID*, *tweet url*, *time stamp*, *time stamp epochs*, *replies*, *retweets*, *likes*, *is retweet*, *retweeter username*, *retweeter user ID*, *retweet ID*, *text*, dan *HTML*. Data yang didapatkan sebanyak 8387 *tweets*. Atribut yang digunakan hanyalah atribut *text* yaitu isi *tweets*.

Tabel 1. Contoh Data *Tweets*

<i>Tweets</i>	Klasifikasi menurut Peneliti
Penerimaan siswa baru dgn sistem zonasi tanpa melihat NEM membuat siswa ogah belajar. NEM tdk terpakai utk cari sekolah & hampir pasti lulus	Negatif
Kebijakan Pemerintah menerapkan sistem zonasi berdasarkan jarak rumah ke sekolah dengan dalih untuk pemerataan pendidikan memang benar dapat membuat siswa yang memiliki kemampuan diatas rata rata menjadi menyebar lebih merata.	Positif

Atribut *text* tersebut kemudian dibentuk *wordcloud*. *Wordcloud* merupakan salah satu visualisasi dari data teks, dimana semakin besar ukuran kata yang ditampilkan, maka semakin sering kata tersebut muncul.



Gambar 1. *Wordcloud* sebelum *preprocessing*

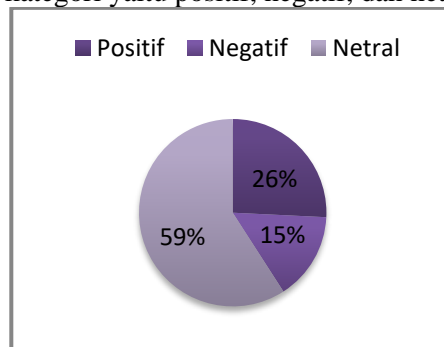
Berdasarkan *wordcloud* di atas, dapat diketahui kata-kata yang paling sering muncul dalam *tweets* yang menulis tentang sistem zonasi tersebut. Kata yang paling sering muncul adalah “sekolah” yaitu sebanyak 11892 kali. Selain itu, terlihat masih terdapat tanda baca, dan kata yang tidak bermakna. Oleh karena itu dibutuhkan adanya *preprocessing*.

Setelah dilakukan *preprocessing*, data *tweets* yang sebelumnya berjumlah 8387 berkurang menjadi 6172 *tweets*. Dari Gambar 2. *Wordcloud* menunjukkan bahwa kata yang paling sering

memberikan tanggapan positif terhadap kebijakan zonasi sekolah lebih besar dibandingkan komentar yang negative.

Naïve Bayes Classification

Metode Naive Bayes adalah *supervised classification* yang memerlukan data *training* dan data *testing*. Data *training* dalam penelitian ini berjumlah 4938 *tweets* dan data *testing* sejumlah 1234 *tweet*. Data training diolah untuk mendapatkan model klasifikasi dengan algoritma Naive Bayes. Setelah itu, model tersebut diterapkan untuk mengklasifikasikan *tweet* pada data *testing* menjadi tiga kategori yaitu positif, negatif, dan netral.



Gambar 4. Diagram Lingkaran Persentase Prediksi Respons terhadap Sistem Zonasi

Berdasarkan Gambar 4 di atas, dapat dijelaskan bahwa proporsi prediksi kategori netral paling banyak yaitu 59 persen. Selanjutnya, kategori positif sebanyak 26 persen dan kategori negatif sebanyak 15 persen. Oleh karena itu, dapat disimpulkan bahwa prediksi respons masyarakat dari data testing terhadap penerapan sistem zonasi sekolah adalah netral. Artinya, masyarakat kebanyakan mengetahui dampak positif dan negatif dari penerapan sistem zonasi sekolah atau dapat dikatakan bahwa banyak masyarakat yang memilih untuk bersikap tidak menentang maupun mendukung.

Akurasi Model

Langkah selanjutnya adalah menampilkan *confusion matrix* untuk mengevaluasi seberapa baik algoritma Naive Bayes mengklasifikasikan *tweet*. Dari tabel *confusion matrix* tersebut kemudian dilakukan penghitungan untuk mendapatkan nilai akurasi yang merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data.

Tabel 2. *Confusion Matrix*

Prediksi	Aktual		
	Positif	Netral	Negatif
Positif	126	62	5
Netral	48	622	55
Negatif	4	63	249

Berdasarkan Tabel 2. *confusion matrix* , diperoleh nilai akurasi model Naïve Bayes sebesar 0,8079, yang artinya algoritma Naive Bayes cukup baik diterapkan dalam penelitian ini karena dapat mengklasifikasikan 80,79 persen *tweets* dengan benar. Tabel tersebut menunjukkan bahwa dari 178 *tweets* yang sebenarnya positif, terdapat 126 *tweets* yang dikategorikan secara benar positif, sedangkan sisanya termasuk dalam kategori yang tidak sesuai, yaitu 48 *tweets* dikategorikan netral dan 4 *tweets* dikategorikan negatif. Dari 309 *tweets* yang sebenarnya negatif, terdapat 249 *tweets* yang dikategorikan secara benar negatif, sedangkan sisanya

termasuk dalam kategori yang tidak sesuai, yaitu 55 *tweets* dikategorikan netral dan 5 *tweets* dikategorikan positif. Dari 747 *tweets* yang sebenarnya netral, terdapat 622 *tweets* yang dikategorikan secara benar netral, sedangkan sisanya termasuk dalam kategori yang tidak sesuai, yaitu 62 *tweets* dikategorikan positif dan 63 *tweets* dikategorikan negatif.

PENUTUP

Simpulan

Dari hasil wordcloud menunjukkan kata “bagus” dan “setuju” merupakan kata yang paling sering muncul di tweet terkait masalah zonasi sekolah. Sedangkan dari hasil analisis sentimen terhadap system zonasi sekolah yang dilakukan, diperoleh kesimpulan bahwa respons masyarakat terhadap sistem zonasi sekolah secara mayoritas adalah netral yaitu sebanyak 61 %, sedangkan respons positif sebanyak 21% dan respons negatif sebanyak 15%. Respons masyarakat yang mayoritas netral menunjukkan bahwa masyarakat masih banyak memiliki sikap *wait and see* dalam menilai kebijakan sistem zonasi sekolah. Metode Klasifikasi Naïve Bayes cukup baik untuk diterapkan untuk memprediksi sentiment masyarakat terhadap sistem zonasi sekolah karena memiliki akurasi yang cukup tinggi yaitu sebanyak 0,8079.

Saran

Terdapat beberapa hal yang disarankan penulis terutama untuk penelitian selanjutnya, yaitu:

1. Metode pengklasifikasi Naïve Bayes dapat diganti dengan metode lainnya seperti SVM untuk mencari metode pengklasifikasi mana yang lebih baik meskipun dalam penerapannya, SVM lebih kompleks daripada Naïve Bayes.
2. Perlu dilakukan pengamatan lebih lanjut terhadap isi kata-kata dari tweet yang mengandung sentimen positif dan negatif untuk mengetahui lebih lanjut mengenai respons masyarakat terhadap sistem zonasi sekolah.

DAFTAR PUSTAKA

- Aristo, T. J. (2019). *Analisis Permasalahan Pemerataan Pendidikan di Kabupaten Sintang*. Jurnal Akuntabilitas Manajemen Pendidikan, 25-34.
- Ipmawati, Joang., Kusriani, dan Luthfi, E.T. (2017). *Komparasi Teknik Klasifikasi Teks Mining pada Analisis Sentimen*. Indonesian Journal on Networking and Security. 6(1): 28-36.
- KG/J-3. (2019, June 27). Media Indonesia. Retrieved from Zonasi Atasi Kemacetan: <https://mediaindonesia.com/read/detail/243555-zonasi-atasi-kemacetan>.
- Kristiyanti, D.A. (2015). *Analisis Sentimen Review Produk Kosmetik Melalui Komparasi Feature Selection*. Konferensi Nasional Ilmu Pengetahuan dan Teknologi. 1(1): 67-76.
- Pribadi, R. E. (2017). *Implementasi Sustainable Development Goals (SDGs) dalam Meningkatkan Kualitas Pendidikan di Papua*. eJournal Ilmu Hubungan Internasional, 917-932.
- Taneyasa, T., & Sitta, Y. (2019, 01 25). Amazonaws. Diambil kembali dari Text Cleaning Bahasa Indonesia based Twitter Data: http://rstudio-pubs-static.s3.amazonaws.com/461333_4a26820bb8ea4084be104a8ffb67511b.html.