

ANALISIS SENTIMEN TERHADAP PENERAPAN SISTEM PLAT NOMOR GANJIL/GENAP PADA *TWITTER* DENGAN METODE KLASIFIKASI *NAIVE BAYES*

Nanang Ruh yana

STMIK NUSA MANDIRI

Jl. Kramat Raya No.18 RT.07/RW.08 Kwitang, Senin - Jakarta Pusat

14002086@nusamandiri.ac.id

ABSTRAK

Penerapan sistem Ganjil Genap pada rambu lalu lintas merupakan kebijakan pemerintah terutama untuk mengurangi kemacetan di ibu kota, dengan cara tersebut pemerintah memberikan solusi kepada pengguna jalan terutama angkutan umum karena di ibu kota banyak pengguna kendaraan pribadi terutama mobil yang bisa membuat kemacetan parah, karena makin banyak para pengguna kendaraan pribadi pada saat hari kerja. Dalam penelitian ini mencoba untuk menganalisis hasil sentimen pada *twitter* untuk kebijakan ganjil genap tersebut, sehingga dapat mengelompokkan pengguna *twitter* menjadi opini positif dan negatif, penelitian ini menggunakan metode *data mining* untuk klasifikasi dengan algoritma *Naive Bayes Classifier*, *Naive Bayes Classifier* (NBC) merupakan metode yang berdasarkan atas probabilitas *bayes* untuk melakukan pengelompokan data. Penelitian ini menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap, dalam penelitian ini menghasilkan *accuracy* 86,67%, *precision* 71,43% dan *recall* 80,00%.

Kata kunci : *Analisis Sentimen, Ganjil Genap, Klasifikasi, Naive Bayes.*

1. PENDAHULUAN

Aktivitas ibu kota Jakarta sangat padat dan juga sangat ramai lalu lintasnya pada saat orang-orang melakukan aktivitas di pagi hari. Dengan demikian menyebabkan kemacetan lalu lintas. Dikarenakan juga kurang kesadaran para pengguna kendaraan pribadi terhadap transportasi umum, sehingga kemacetan di Jakarta sangat tidak bisa dihindari.

Berdasarkan data Gabungan Industri Kendaraan Bermotor Indonesia (Gaikindo), mencatat sepanjang 2017 penjualan mobil mencapai 1,079 juta. Angka itu naik 1,6% dari tahun sebelumnya (Sugianto, 2018). Dengan melihat data penjualan kendaraan tersebut tidak bisa dipungkiri semakin ramai kendaraan di jalanan, maka Pemerintah DKI dan Direktorat Lalu Lintas (Ditlantas) Polda Metro Jaya menerapkan sistem plat nomor ganjil/genap pada kendaraan untuk mobil. Kemacetan yang terjadi di Jakarta hampir di setiap ruas jalan di Jakarta dikarenakan sarana dan prasarana transportasi umum kurang memadai. Untuk masalah transportasi

dalam kemacetan karena buruknya layanan angkutan umum inilah yang memicu peningkatan pengguna kendaraan pribadi di Jakarta. Selain itu juga kurangnya kesadaran pribadi untuk menggunakan kendaraan umum.

Kebijakan penerapan sistem ganjil/genap di Jakarta untuk mengurangi kepadatan lalu lintas. Dengan angka penjualan mobil yang ada, otomatis pengendara pribadi semakin meningkat untuk mengantisipasinya maka diberlakukan untuk sistem ganjil/genap pada setiap plat nomor kendaraan khususnya mobil. Dan tingkat kedisiplinan serta kesadaran yang semakin hari semakin tidak beraturan sehingga tidak terkendalinya. Penataan dan pengaturan penggunaan jalan dan fasilitas pendukungnya yang belum maksimal untuk para pengguna jalan.

Dengan adanya penerapan sistem ganjil/genap banyak sekali tanggapan masyarakat ada yang berpendapat positif dan ada yang berpendapat negatif, terutama pada sosial media *twitter* yang orang bebas berpendapat untuk berpendapat apa saja.

Berdasarkan pembahasan tersebut maka penulis mencoba melakukan penelitian tentang pendapat analisa sentimen pada *twitter* dalam penerapan sistem ganjil / genap sehingga dapat mengklasifikasi pendapat dari masyarakat di sosial media *twitter*. Dalam penelitian sebelumnya ada beberapa tentang sentimen klasifikasi sosial media *twitter* terhadap tokoh publik pada *twitter* (Hidayatullah & Sn, 2014).

2. METODELOGI

Penelitian ini bertujuan untuk mengklasifikasi sentimen pada *twitter* terhadap penerapan sistem ganjil / genap sehingga dapat bisa memberikan hasil apakah penerapan ganjil / genap tanggapan masyarakat baik atau buruk di media sosial *twitter*.

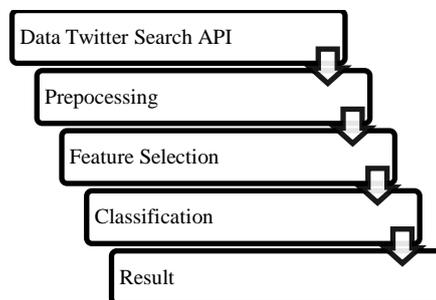
2.1. Rancangan Sistem

Rancangan sistem yang akan dilakukan pada penelitian ini dibagi menjadi beberapa bagian, *preprocessing*, *feature selection*, dan *classification methods*. Pada gambar 1 memperlihatkan yang di aplikasikan secara lebih rinci. Bagian sistem berinteraksi dengan pengguna aplikasi secara langsung akan memberikan pilihan kepada *user* dengan pencarian *hashtag* #ganjilgenap sebagai *query*, pengambilan data dengan proses token API *twitter* pada aplikasi *Rapid Miner*.

Data *tweet* yang dikumpulkan antara lain *id*, *create-at*, *from-user*, *from-user-id*, *to-user*, *to-user-id*, dan *text*. Dalam prose pengambilan data *twitter* dengan API *Rapid Miner* secara otomatis akan mengambil data *tweet* yang mengandung *hashtag* #ganjilgenap. Data *tweet* yang dihasilkan dari *Rapid Miner* dan di *export* ke format *excel* dan akan melewati proses yang telah disebutkan diatas. yaitu proses pelabelan. Tahap pelabelan *tweet* dilakukan secara manual sesuai dengan sentimen yang sudah ditentukan sebelumnya yaitu sentimen positif untuk *hashtag* #ganjilgenap berkomentar positif, dan sentimen negatif untuk *hashtag* #ganjilgenap berkomentar negatif.

Setelah seluruh data memiliki kelas masing-masing kemudian dilakukan pemisahan data menjadi dua bagian yaitu *data training* dan *data testing*. Data latih digunakan untuk mengklasifikasi *tweet* pada

kelas sentimennya dan jumlah datanya lebih banyak.



Gambar 1. Rancangan *System*

2.2. *Preprocessing*

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, Varma, & A.Govardhan, 2012). Tahapan teks *preprocessing* pada penelitian ini diantaranya:

1. Menghapus URL
URL ([http:// www.situs.com](http://www.situs.com)) dan email (nama@situs.com) dihapus di tahapan ini.
2. Mengganti *Emoticon*
Proses *convert emoticon* mengganti *emoticon* yang terdapat pada *tweet* dengan kata yang mencerminkan *emoticon*. Daftar emosi *emoticon* dapat dilihat pada Tabel 1.

Tabel 1. Konversi *Emoticon*

<i>Emoticon</i>	Konversi
:) :-) :) :-)) =)) =))	Senyum
:D :-D =D	Tawa
:- (:(Sedih
;-) ;)	Kedip
:-P :P	Ejek
:-/ :/	Ragu
: :-	Haru

3. **Menghapus Karakter Khusus *Twitter***
Proses ini dilakukan dengan menghapus karakter khusus *twitter* seperti *hashtag* (*#hashtag*), *username* (*@username*), dan karakter khusus (misal : RT, yang menunjukkan bahwa *user* melakukan *retweet*).
4. **Tokenisasi**
Tokenisasi yaitu mengumpulkan semua kata yang muncul dan menghilangkan baca maupun simbol apa pun yang bukan huruf (Muthia, 2017)
5. **Cek Nama *Hashtag***
Proses pengecekan nama *hashtag* untuk mengecek tidak terjadi kesalahan penulisan.
6. ***Case Folding***
Proses *case folding* menyeragamkan bentuk huruf menjadi huruf besar atau huruf kecil.
7. **Menggantikan Kata Tidak Baku (*Slangword*)**
Proses ini mengganti kata-kata dalam *tweet* yang tidak baku menjadi kata baku yang telah dikenal dalam bahasa Indonesia.
8. ***Stemming***
Stemming adalah tahap mencari akar kata dengan menghilangkan imbuhan pada sebuah kata (Feldman, 2013)
9. ***Stopword*** merupakan kata-kata tidak berpengaruh terhadap proses klasifikasi.
10. **Menggabungkan kata Negasi**
Proses ini dilakukan untuk mendeteksi negasi yang terdapat dalam *tweet*.

2.3. **Feature Selection**

Proses pemilihan dan ekstraksi fitur yang digunakan dalam penelitian ini antara lain :

1. ***Unigram***
Unigram feature extractor merupakan cara paling sederhana dalam mendapatkan fitur dari *tweet* (Go, Bhayani, & Huang, 2009). Proses ekstraksi *unigram* dilakukan dengan mengekstrak kata per kata dalam dokumen.

2. ***Negation***

Negasi merupakan sesuatu yang dikenal dalam semua bahasa dan biasanya negasi digunakan untuk mengubah polaritas dari suatu pernyataan (Blanco & Moldovan, 2011).

2.4. **Classification Methods**

Metode yang penulis gunakan adalah metode klasifikasi *Naive Bayes* yang merupakan teknik *machine learning* yang populer untuk klasifikasi teks serta memilih *performance* yang baik.

3. LANDASAN TEORI

3.1. ***Twitter***

Twitter merupakan salah satu media sosial yang berbasis *microblogging*, dimana penggunaannya dapat mengirim sebuah pesan yang disebut dengan *tweet*.

3.2. **Klasifikasi**

Klasifikasi adalah proses menentukan suatu objek ke dalam suatu kelas atau kategori yang telah ditentukan. Penentuan objek dapat menggunakan suatu model tertentu beberapa model yang bisa digunakan antara lain : *classification (IF-THEN) rules*, *decision trees*, formula matematika atau *neural networks* (Han, Kamber, & Pei, 2012).

3.3. **Analisis Sentimen**

Analisis sentimen adalah tugas menemukan opini dari penulis tentang suatu entitas tertentu.

Untuk melakukan analisis sentimen ada beberapa algoritma yang dapat digunakan salah satunya adalah algoritma *Naive Bayes*.

3.4. ***Naive Bayes***

Naive Bayes Classifier (NBC) merupakan metode yang berdasarkan atas probabilitas bayes untuk melakukan pengelompokan data (Domingos, P., dan Pazzani, M. 1997).

Tahapan Algoritma *Naive Bayes* :

1. Hitung probabilitas bersyarat/*likelihood* :

$$P(x | C) = P(x_1, x_2, \dots, x_n | C)$$

$C = class$

$x =$ vektor dari nilai atribut n

$P(x_i|C)$ = proporsi dokumen dari *class* C yang mengandung nilai atribut x_i .

- Hitung probabilitas *prior* untuk tiap *class*:

$$P(C) = \frac{N_j}{N}$$

N_j = jumlah dokumen pada suatu *class*

N = jumlah total dokumen

- Hitung probabilitas *posterior* dengan rumus:

$$P(C|x) = \frac{p(x|C)p(c)}{p(x)}$$

Menurut pendapat Santoso (Santosa, 2007) Dengan kata-kata yang lebih umum, rumus *Bayes* bisa diberikan sebagai berikut :

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

3.5. Rapid Miner

Rapid miner merupakan *software tools open source* untuk *data mining*. *Rapid Miner* dioperasikan pada sebuah lingkungan untuk *machine learning*, *data mining*, *text mining*, *predictive analytics* (Setyawan & Nugroho, 2014).

4. HASIL DAN PEMBAHASAN

Hasil dan Pembahasan dari penelitian ini akan dijelaskan secara detail dengan menggunakan klasifikasi teks menggunakan Algoritma *Naive Bayes*.

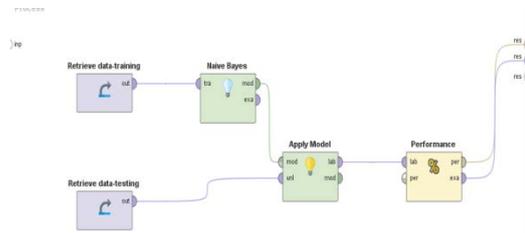
Dalam pengumpulan data untuk penelitian ini penulis mengambil data set dari *twitter* melalui API dengan Aplikasi *Rapid Miner* dengan *query hashtag #ganjilgenap* setelah mendapatkan *dataset* dari *twitter* setelah itu *export* ke format excel diberi nama "twitter ganjil genap.xls" data yang telah diberi label positif negatif dari data training dan data testing Untuk dilakukan proses klasifikasi.

Gambar 2. ExampleSet Data Training

Gambar 2. ExampleSet Data Testing

Setelah *dataset* yaitu *data training* dan *data testing* sudah didapatkan selanjutnya dilakukan proses *preprocessing* sebagaimana dijelaskan pada metodologi penelitian antar lain menghapus *url*, mengganti *emoticon*, tokenisasi, cek nama *hashtag #ganjilgenap*, *case folding*, mengganti *slangword*, *steaming* dan *stopword*.

Dengan proses yang telah dilakukan maka selanjutnya ekstraksi fitur dan menerapkan metode klasifikasi dengan *naive bayes*, sehingga dapat dilihat pada gambar untuk proses di *tools rapid miner*.



Gambar 3. Model Klasifikasi *Naive Bayes Rapid Miner*

Dengan menggunakan pemodelan klasifikasi *Naive Bayes* seperti gambar diatas dengan mengklasifikasi 2 jenis yaitu positif dan negatif. Dalam penelitian ini algoritma menghasilkan akurasi 86.67% dengan menggunakan metode *Naive Bayes* dari pengujian data yang telah dikumpulkan dan dengan diuji dengan *data training* dan *data testing*.

Dari pengujian hasil data tersebut dengan *rapid miner* akan mendapatkan *class precision* dan *class recall*, yang dimaksud dengan *precision* yaitu perbandingan dokumen yang relevan terhadap total dokumen hasil *query* sedangkan *recall* perbandingan jumlah dokumen relevan terambil terhadap total dokumen relevan.

Selain *accuracy* ada dua hal tersebut yang dihasilkan dalam klasifikasi teks, maka pengujian data dapat dihasilkan pada gambar 4 .

accuracy: 86.67%

	true positif	true negatif	class precision
pred positif	5	2	71.43%
pred negatif	3	8	100.00%
class recall	100.00%	80.00%	

Gambar 4. Hasil Pengujian *Rapid Miner*

Dapat dilihat *class precision* menghasilkan 71,43% dan *class recall* 80,00%, dan berikut adalah kurva ROC dari hasil pengujian data *tweet* dengan *rapid miner* dengan klasifikasi metode *naïve bayes*.



Gambar 5. Kurva ROC Klasifikasi Metode *Naïve Bayes*

Implikasi penelitian mengarahkan pada beberapa aspek bisa melihat tanggapan dalam kebijakan untuk penerapan ganjil genap sehingga dapat diketahui kecenderungan hasil dari klasifikasi dengan metode *naïve bayes* dengan data-data pada *twitter* dengan penerapan metode klasifikasi tersebut yang membuat kebijakan ganjil genap dapat mengetahui sentimen masyarakat sehingga membantu dalam peraturan ini lebih baik lagi, dan karena keterbatasan waktu dalam penelitian ini, implikasi terhadap penelitian selanjutnya dapat mempergunakan data lebih banyak lagi dari hasil *twitter* tentang ganjil genap dan penerapan metode klasifikasi yang lain dengan melakukan perbandingan sehingga dapat melihat yang lebih baik, dan juga akan bisa memberikan hasil yang dari penelitian yang lebih tepat dalam akurasi.

5. KESIMPULAN

Dari hasil penelitian untuk mencari sentiment dalam hal penerapan sistem ganjil genap untuk tanggapan pada pengguna *twitter* dan dilakukan analisa sentiment dengan label positif dan negatif dengan metode klasifikasi *naïve bayes* pengujian dilakukan dengan *data training* dan *data testing* yang di ambil dari *twitter* menggunakan *software rapid miner* sehingga memudahkan memberikan hasil dari penelitian tersebut dengan hasil *accuracy* 86,67%, presisi 71,43% dan *recall* 80,00%.

Model terbentuk dapat diterapkan untuk melihat sentimen dari para pengguna kendaraan dalam bentuk positif dan negatif. Hal ini dapat membantu dalam penerapan kebijakan pemerintah untuk memberikan solusi terbaik dalam pengaturan lalu lintas di jalan.

DAFTAR PUSTAKA

- Blanco, E., & Moldovan, D. (2011). Some Issues on Detecting Negation from Text. *Twenty-Fourth International FLAIRS Conference*, 228–233. Retrieved from <http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS11/paper/viewFile/2629/3031>
- Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82. <https://doi.org/10.1145/2436256.2436274>
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *Processing*, 150(12), 1–6. <https://doi.org/10.1016/j.sedgeo.2006.07.004>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hemalatha, I., Varma, D. G. P. S., & A.Govardhan, D. (2012). Preprocessing The Informal Data for Efficient Sentiment Analysis. *International Journal of Emerging Trends & Technology in Computer Science*

- (IJETTCS), 1(2), 58. Retrieved from <http://ijettcs.org/Volume1Issue2/IJETTCS-2012-08-14-047.pdf>
- Hidayatullah, A. F., & Sn, A. (2014). Analisis Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. *Seminar Nasional Informatika 2014, 2014*(August 2013), 0–8.
- Muthia, D. A. (2017). Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 39–45.
- Santosa, B. (2007). *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*.
- Setyawan, & Nugroho, Y. S. (2014). Klasifikasi Prestasi Akademik Mahasiswa Fki Ums Menggunakan Metode Decision Tree. *Universitas Muhammadiyah Surakarta*.
- Sugianto, D. (2018). Penjualan Mobil di 2017 hanya Tumbuh 1,6%. Retrieved from <https://finance.detik.com/industri/d-3817970/penjualan-mobil-di-2017-hanya-tumbuh-16>