

ANALISIS SENTIMEN TANGGAPAN PUBLIC MENGENAI E-TILANG MELALUI MEDIA SOSIAL YOUTUBE MENGUNAKAN ALGORITMA NAIVE BAYES

Zalsa Nanda Aulia¹, Gilang Kuncoro Jati², Imam Santoso³

¹Program Studi Sarjana Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika

²Program Studi Sarjana Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika

³Universitas Teknologi Muhammadiyah Jakarta

zalsa@stikomcki.ac.id gilang@stikomcki.ac.id imam.santoso@utmj.ac.id

ABSTRAK

E-tilang merupakan sistem tilang elektronik yang digunakan oleh pemerintah dalam penegakan hukum lalu lintas. Dalam penelitian ini melakukan analisis sentimen tentang sistem e-Tilang atau opinion mining untuk mengelompokan ragam komentar masyarakat yang memberikan kesan positif, negatif atau netral. Media sosial youtube menjadi salah satu objek untuk menyampaikan opini karena user friendly, topik ter-update, dan terbuka mengakses komentar. Opini pada Youtube dikumpulkan, lalu dilakukan tahapan preprocessing, selanjutnya dengan seleksi membantu mengurangi noise yang disebabkan oleh label-label yang tidak relevan. Untuk Data set yang dikelola sebanyak 500 data. Algoritma naive bayes pengklasifikasian tidak membutuhkan adanya pemodelan maupun uji statistik. Tahap selanjutnya adalah klasifikasi sentimen dengan algoritma Naïve Bayes Penelitian ini menghasilkan accuracy 79,44%.

Kata kunci: Analisis sentimen, E-Tilang, Naive Bayes

ABSTRACT

E-ticketing is an electronic ticket system used by the government in enforcing traffic laws. In this study, sentiment analysis was carried out about the e-Tilang system or opinion mining to classify various public comments that gave a positive, negative or neutral impression. YouTube social media is one of the objects for conveying opinions because it is user friendly, topics are updated, and comments are open to access. Opinions on Youtube are collected, then the preprocessing stage is carried out, then selection helps reduce noise caused by irrelevant labels. For data sets that are managed as many as 500 data. The naive Bayes classification algorithm does not require any modeling or statistical tests. The next stage is sentiment classification using the Naïve Bayes algorithm. This study produces an accuracy of 79.44%.

Keywords: Sentiment analysis, E-ticketing, Naive Bayes

1. PENDAHULUAN

E-tilang merupakan sistem tilang elektronik yang digunakan oleh pemerintah dalam penegakan hukum lalu lintas. Algoritma Naive Bayes adalah metode klasifikasi yang berdasarkan teorema Bayes dengan asumsi bahwa setiap fitur adalah independen satu sama lain. Metode ini telah banyak digunakan dalam analisis sentimen karena kecepatan dan keandalannya dalam mengklasifikasikan teks. Penerapan e-Tilang merupakan langkah baru yang diambil pemerintah dalam mewujudkan Good Governance agar pelayanan publik yang diberikan dapat lebih efektif, efisien, transparan dan akuntabel. E-Tilang merupakan digitalisasi proses tilang, dengan memanfaatkan teknologi diharapkan seluruh proses tilang akan lebih efisien

Sistem e-Tilang menjadi solusi mendisiplinkan para pengendara kendaraan bermotor dari banyaknya melakukan pelanggaran berlalu-lintas. Keberadaan e-Tilang juga menjadi solusi mencegah kenakalan penegak hukum dari pungutan liar, istilah damai ditempat, hingga akuntabilitas uang denda. Efektif dan efisien sistem e-Tilang menimbulkan ragam komentar pada masyarakat. Perkembangan website 4.0 dan tingginya penggunaan media sosial dapat dimanfaatkan untuk menjadi sumber informasi dan pengambilan keputusan, dengan melakukan text mining yang bersumber dari komentar pada media sosial ataupun website dapat melakukan analisis sentimen. Hal yang dilakukan dalam penelitian ini adalah melakukan text mining pada postingan mengenai sistem e-Tilang pada media sosial Youtube menggunakan algoritma klasifikasi Naive Bayes.

Penelitian ini membantu pemerintah dalam melakukan evaluasi program yang dijalankan, memberikan informasi sentiment masyarakat terhadap sistem e-Tilang. Tanpa adanya evaluasi dari masyarakat, pemerintah tidak akan mengetahui keberhasilan sistem yang dijalankan.

2. TINJAU PUSTAKA

A. Analisis Sentimen

Sentiment analysis atau opinion mining merupakan sebuah studi komputasi untuk mencari atau mengenali dan mengekspresikan opini, sentimen, evaluasi, sikap, emosi, subjektifitas, penilaian atau pandangan yang terdapat dalam suatu teks.

Analisis sentimen dapat digunakan untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, institusi, produk atau pada sebuah kondisi tertentu. Nilai dari analisis sentimen bisa dipecah menjadi 3 yakni, sentimen positif, sentimen negatif dan sentimen netral atau diperdalam lagi sehingga dapat menemukan siapa atau kelompok yang menjadi sumber sentimen positif atau sentimen negatif[9].

Langkah-langkah analisis sentimen klasifikasi terhadap data text-mining atau data teks adalah sebagai berikut:

1. Tahap awal: Mengumpulkan data set seperti pendapat masyarakat, penilaian terhadap restoran atau produk dan lain-lain.
2. Pre-processing: Tahapan yang mencakup Tokenization, Stopwords Removal, dan Stemming.
3. Transformation: Pembobotan dari data teks
4. Feature Selection: Tahap mengurangi data yang tidak diperlukan

5. Classification: Tahap klasifikasi teks yang biasanya menggunakan metode seperti: Naive Bayes, K-Nearest Neighbor, Support Vector Machine dan lain-lain.

6. Interpretation/Evaluation: Tahap evaluasi untuk menghitung nilai accuracy dan nilai Area Under the Curve [10]

B. Naive Bayes

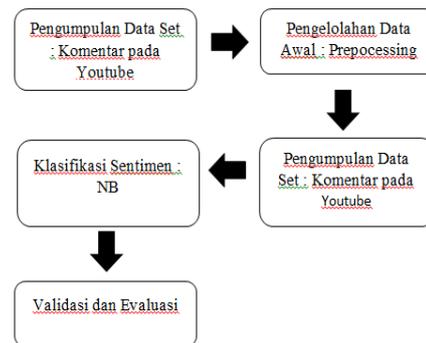
Naive Bayes merupakan metode pengklasifikasian paling populer digunakan dengan tingkat keakuratan yang baik. Banyak penelitian tentang pengklasifikasian yang telah dilakukan dengan menggunakan algoritma ini. Berbeda dengan metode pengklasifikasian dengan logistic regression ordinal maupun nominal, pada algoritma Naive Bayes pengklasifikasian tidak membutuhkan adanya pemodelan maupun uji statistik.

Naive Bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasian probabilitas. Keuntungan algoritma Naive Bayes adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi Naive Bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar.

3. METODE PENELITIAN

Rancangan penelitian yang akan dilakukan dibagi menjadi beberapa bagian, yaitu pengumpulan data, preprocessing, sentimen klasifikasi dengan Naive Bayes Perangkat Lunak yang digunakan untuk melakukan

analisis sentimen yaitu Rapidminer 10.1.



A. Pengumpulan Data Set

Data diambil dengan metode crawling dari media sosial Youtube menggunakan bantuan API key. Pada aplikasi Rapidminer melalui API dengan query e-Tilang. Setelah mendapatkan dataset dari Youtube setelah itu export ke format excel.

B. Pengolahan Data Awal

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada kata, dan data-data yang tidak konsisten. Sebagai contoh terdapat 10 dataset yang akan di kelola. Tahapan preprocessing data diantaranya:

1. Case Folding

Case Folding adalah Proses mengubah semua huruf data teks menjadi huruf kecil.

2. Tokenisasi

Tokenisasi yaitu mengumpulkan semua kata yang muncul dan menghilangkan baca maupun simbol apa pun yang bukan huruf.

3. Stemming

Stemming adalah tahap mencari akar kata dengan menghilangkan imbuhan pada sebuah kata.

4. Stopword

Stopword merupakan kata-kata yang tidak memiliki pengaruh terhadap proses klasifikasi.

C. Pemodelan

Pada Penelitian ini proses eksperimen menggunakan RapidMiner 10.1. Data Training yang digunakan adalah opini public melalui Media Sosial Youtube yang dikelompokkan menjadi 2 yaitu opini positif dan opini negatif.

D. Klasifikasi Sentimen

Pada Penelitian ini, metode klasifikasi yang diterapkan adalah Naive Bayes Classifier (NBC) merupakan metode yang berdasarkan atas probabilitas bayes untuk melakukan pengelompokan data. Tahapan Algoritma Naive Bayes:

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

C = kelas

x = vektor dari nilai atribut n

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

Hitung probabilitas sebelumnya untuk setiap kelas:

$$P(C) = N_j/N$$

N_j = jumlah dokumen pada suatu

class N = jumlah total dokumen

Hitung probabilitas posterior dengan rumus:

$$P(C | x) = (p(x | C)p(c))/(p(x))$$

E. Validasi dan Evaluasi

Tahap validasi dilakukan dengan menerapkan *10 fold-cross validation*. Proses validasi memiliki dua sub-proses yaitu: *training set* dan *testing set*. Sub-proses *training data* digunakan untuk melatih model algoritma yang sudah ditentukan pada tahap pemodelan dengan data pelatihan yang ada. Setelah model algoritma dilatih pada tahap sub-proses *training*, berikutnya akan dilakukan *testing*. Adapun Evaluasi atau pengujian hasil klasifikasi *Naive Bayes* yang merupakan alat yang dapat digunakan untuk melakukan analisis terhadap seberapa baik klasifikasi yang telah dihasilkan dan dapat mengenali *tuple* dari kelas yang berbeda.

4. HASIL DAN PEMBAHASAN

Dataset diambil dari media sosial Youtube selama 7 hari terakhir dengan jumlah data sebanyak 500 data. Data tersebut kemudian diberikan label positif dan negative sehingga menghasilkan data komentar sebanyak 500 data, data berlabel ini yang dijadikan data set penelitian. Dalam mengelolah data dengan tujuan untuk memperoleh model yang sesuai dengan kebutuhan.

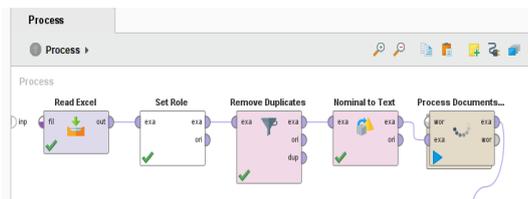
N O	KOMENTAR
1.	Alhamdulillah gak ada lagi polisi nakal buat tilang semuanya.....
2.	Sdh bner kmaren buat ngurangi pungli....tinggal tingkat

	kan pengaturan di jln2 yg rawan kecelakaan, atau yg lawan arus...
3.	Setuju pak..semoga dapat dijalankan oleh petugas jajaran..karna sangat merugikan dari segi waktu dan kenyamanan dijalan
4.	Heran, padahal paling gampang pake kamera. Kasi denda 500rb atau 1jt, kalau ga bayar yah motor ditarik. Pngen liat seberapa kuat pelanggaran lalu lintas mau melanggarnya. Sangat Simple padahal. Bisa nambah pemasukan buat negara.
5.	Mening tilang manual aja gak sih? Tp yg pungli2nya aja ditiadakan. Klau pake ETLE, harus lebih kuat lagi tindakan dan lbih besar lgi denda2nya. Jngn cuma foto doang. Soalnya ETLE tuh bikin pelanggaran2 makin banyak dan makin bnyak jg pelanggar2
6.	@DesaKita2 Ah jancuk kok, aku ming khilaf sitik wae di tilang. Pokokke polisine jeneng e eko
7.	pasti kota lain akan menyusul! dan aku senang makin sedikit yg melanggar lalu lintas karena takut kena e tilang :)
8.	@rosihanazwar @JJRizal Sebenarnya sdh diinisiasi secara nasional ya sejak 2015,

	ujicoba, sampai sekarang resmi sejak Juli 2019. Penindakan e tilang dari bukti rekaman cctv, juga harus segera diproses sesaat setelah kejadian dari rekaman yg "terpantau" . Jd tetap tergantung yg memantau...
9.	Warga Apresiasi Sat Lantas Polres Simalungun Bantu Mengurus Pembayaran Denda E-Tilang
10.	Ngemis duit d jalan d bungkus tilang manual

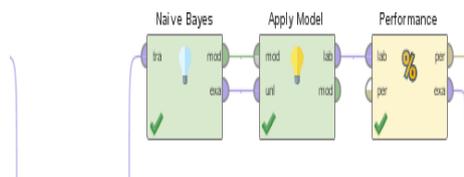
Tabel 1. Sampel Dataset

Data yang telah terkumpul, seperti pada tabel 1 akan melalui tahap preprocessing seperti pada gambar dibawah ini.



Tahap Preprocessing

Tahapan selanjutnya adalah seleksi fitur, lalu tahap selanjutnya melakukan sentimen klasifikasi menggunakan algoritma Naïve Bayes seperti dibawah ini.



Tahap Klasifikasi

Noise yang telah hilang dan data yang telah bersih, selanjutnya melakukan tahapan analisis sentimen, seperti pada gambar dibawah ini.



Tahapan Analisis Sentimen

Dalam penelitian ini algoritma menghasilkan akurasi 79,44% dengan menggunakan metode Naïve Bayes dari pengujian data yang telah dikumpulkan dan dengan diuji dengan data training dan data testing. Dari pengujian hasil data tersebut dengan rapidminer akan mendapatkan class precesion dan class recall, yang dimaksud dengan precesion yaitu perbandingan dokumen yang relevan terhadap total dokumen hasil query sedangkan recall perbandingan jumlah dokumen relevan terambil terhadap total dokumen relevan. Hasil pengujian dapat dilihat pada tabel dibawah ini.

PerformanceVector (Performance) X ExampleSet (Cross Validation) X

Table View Plot View

accuracy: 79.44% +/- 1.63% (micro average: 79.43%)

	true negatif	true positif	class precision
pred. negatif	387	97	79.96%
pred. positif	4	3	42.86%
class recall	98.98%	3.00%	

HASIL PENGUJIAN

5. KESIMPULAN

Penerapan fitur seleksi dalam pengoptimalisasi tingkat akurasi analisis sentiment sistem e-Tilang telah dapat dilakukan dan telah dapat memberikan sebuah peningkatan akurasi terhadap model Naïve Bayes yang merupakan model klasifikasi dan memberikan hasil penelitian dengan hasil accuracy 79,44%.

DAFTAR PUSTAKA

- Feldman, R. (2013).** “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol 56, no 4, pp.82,
- Hemalatha, I., Varma, D. G. P. S., dan A. Govardhan, D. (2012)** “Preprocessing The Informal Data for Efficient Sentiment Analysis,” *International Journal of Emerging Trends & Technology in Computer Science*, Vol 2, no 58.
- Ipmawati, J., Kusrini, dan Taufiq Luthfi, E. (2017).** “Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen”. In *Indonesian Journal on Networking and Security*, Vol. 6.
- Muthia, D. A. (2017).** “Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes, ”*Jurnalilmu Pengetahuan Dan Teknologi Komputer*, vol 2, no 2, p. 39–45.

Pasal 272 Undang – Undang No. 22 Tahun 2009 tentang Lalu Lintas Jalan dan Angkutan Jalan.

Peraturan Pemerintah No 80 Tahun 2012 tentang Tata Cara Pemeriksaan Kendaraan Bermotor di Jalan Dan Penindakan Pelanggaran Lalu Lintas dan Angkutan Jalan.

Oktasari, L., Chrisnanto, Y. H., dan Yuniarti, R, (2016). “Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier,” Prosiding SNST.

Rakhmadani S. (2017). “Analisis Penerapan E-Tilang Dalam Mewujudkan Good Governance di Indonesia,” Prosiding SNaPP2017.

Imam Santoso., Windu Gata., &Atik Budi Paryanti (2019).

Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum Jurnal Rekayasa Sistem dan Teknologi Informasi.penelitian, Pusat Analisis Ekonomi dan Kebijakan Pertanian. Bogor: Kementrian Pertanian.