

Avis Tantra Mukti - Analisis Sentimen Warganet Terhadap Keberadaan Juru Parkir Liar Menggunakan Metode Naïve Bayes Classifier

by Layanan Perpustakaan UHAMKA

Submission date: 09-Nov-2023 10:54AM (UTC+0700)

Submission ID: 2222428855

File name: AVIS_TANTRA_TI_TURNITIN_KE-2_-_Avis_Tantra_Mukti.docx (544.79K)

Word count: 3850

Character count: 23191



Analisis Sentimen Warganet Terhadap Keberadaan Juru Parkir Liar Menggunakan Metode Naïve Bayes Classifier

Avis Tantra Mukti, Firman Noor Hasan*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia

Email: ¹avistantra01@gmail.com, ²firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Abstrak– Masih tingginya angka pengangguran di Indonesia menyebabkan banyak sekali dampak, salah satunya adalah munculnya juru parkir liar. Kondisi tersebut terus didukung dengan terciptanya area parkir para pelaku usaha untuk para pengunjungnya. Area parkir yang disediakan sering kali gratis, bahkan memiliki tulisan yang menyatakannya. Akan tetapi, ada saja oknum yang memanfaatkan lahan parkir gratis tersebut untuk mendapatkan penghasilan. Banyak sekali sentimen dari warganet terhadap fenomena juru parkir liar di media sosial. Oleh sebab itu dalam penelitian ini digunakan sebuah analisis tanggapan berupa sentimen warganet terhadap juru parkir liar di media sosial X dengan menggunakan *Naïve Bayes*. Tujuan utama penelitian ini adalah memahami perasaan masyarakat terhadap keberadaan juru parkir liar yang beroperasi di area parkir. Dataset yang digunakan dalam analisis ini sebanyak 905 yang diambil dari media sosial X dan dianalisis menggunakan algoritma *Naïve Bayes* dengan melakukan klasifikasi sentimen ke dalam kategori positif, netral, dan negatif terhadap keberadaan juru parkir liar yang beroperasi. Hasil dari penelitian ini warganet merasa sangat kesal, marah, dan terganggu dengan keberadaan juru parkir liar yang beroperasi. Hal ini dibuktikan dengan hasil sentimen negatif yang mendominasi sebesar 93% dari keseluruhan data atau sebanyak 841 sentimen negatif terhadap fenomena ini.

Kata Kunci: Analisis Sentimen; Juru Parkir Liar; Media Sosial X; Tweet; *Naïve Bayes Classifier*

Abstract– The still high rate of poverty in Indonesia causes many impacts, one of which is the emergence of illegal parking attendants. This condition continues to be supported by the creation of a parking area for business actors for their visitors. The parking areas provided are often free, and even have signs saying so. However, there are individuals who use the free parking space to earn income. There are many netizens' sentiments regarding the phenomenon of lying parking attendants on social media. Therefore, in this research, an analysis was used in the form of netizen sentiment towards illegal parking attendants on social media X using *Naïve Bayes*. The main objective of this research is to understand the public's feelings towards the existence of illegal parking attendants operating in the parking area. The dataset used in this analysis was 905 taken from social media. The results of this research netizens felt very annoyed, angry and disturbed by the presence of illegal parking attendants operating. This is proven by the results of negative sentiment which dominates 93% of the total data or as many as 841 negative sentiments regarding this phenomenon.

Keywords: Sentiment Analysis; Illegal Parking Officer; Social Media X; Tweets; *Naïve Bayes Classifier*

1. PENDAHULUAN

Di era digital ini, peran media sosial sangat digemari karena dapat digunakan untuk berbagi pengalaman dan pendapat mengenai beragam topik secara terbuka tanpa adanya batasan termasuk keresahan yang dialami oleh penggunanya [1]. Seiring berkembangnya teknologi, kini aplikasi didukung dengan adanya versi *mobile* sehingga berbagi pengalaman penggunaan lebih mudah dan fleksibel untuk dilakukan di mana pun kapan pun [2] asal terhubung dengan internet [3]. Sementara itu, manusia sebagai makhluk ekonomi yang perlu melakukan transaksi tidak lepas dari yang namanya praktik jual beli. Pengguna media sosial khususnya X sering kali mengunggah pengalaman keseharian dan opininya [4]. Tak terkecuali pengalaman jual beli, khususnya perihal parkir. Perkara parkir pada kegiatan jual beli sering kali dibagikan melalui media sosial X terlepas pengalaman menyenangkan maupun pengalaman buruk.

Sesuai peraturan, lahan parkir adalah suatu keharusan bagi pelaku usaha sesuai undang-undang yang berlaku dalam peraturan presiden nomor 112 tahun 2007 [5]. Hal ini berguna untuk mengantisipasi terjadinya kemacetan karena pembeli atau pengunjung yang datang. Setiap pelaku usaha memiliki hak masing-masing untuk membuat kebijakan apakah area parkir yang disediakan gratis atau berbayar. Akan tetapi tidak jarang area parkir gratis yang disediakan oleh pelaku usaha tidak sesuai kenyataannya sehingga membuat pembeli atau pengunjung merasa kurang nyaman. Hal tersebut dikarenakan maraknya muncul oknum juru parkir liar yang memanfaatkan situasi ini [6] ditambah keberadaannya diperkirakan dilindungi oleh pihak lain yang sama-sama mengambil keuntungan [7]. Juru parkir liar kerap melakukan pungutan [8] kepada pengguna parkir ketika mereka akan pergi meninggalkan area parkir. Pungutan liar sendiri telah dilarang dalam undang-undangan dan dapat dikenai pidana [9]. Maraknya oknum yang bermunculan juga dikarenakan masih tingginya jumlah pengangguran di Indonesia yang mencapai ratusan juta jiwa [10]. Hal tersebut terjadi karena pertumbuhan tenaga kerja tidak sejalan dengan tumbuhnya lapangan pekerjaan [11].

Dengan banyaknya pengguna media sosial X yang telah memiliki pengalaman tentang keberadaan juru parkir liar, maka banyak sekali tweet di media sosial X. Dalam konteks ini, analisis sentimen sangat berguna sebagai alat untuk memahami pemikiran netizen terhadap munculnya para juru parkir liar. Hal ini juga bisa sekaligus mengetahui respon netizen terhadap kinerja aparat penegak hukum untuk mengatasi kegiatan ilegal ini. Dengan menganalisis emosi, pendapat, dan sikap yang diungkapkan oleh pengguna media sosial, maka analisis sentimen



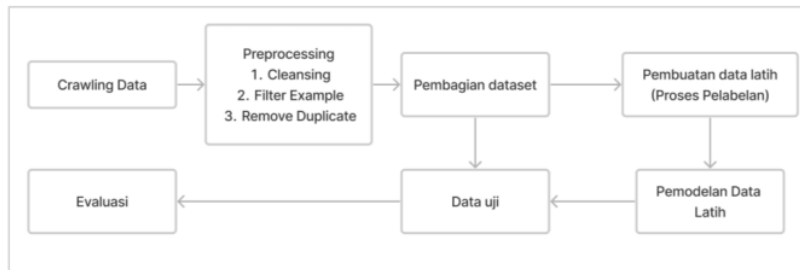
memungkinkan bisnis dan organisasi untuk mengumpulkan informasi dan mengambil keputusan untuk strategi bisnis selanjutnya [12].

Penelitian terdahulu terkait analisis sentimen telah dilakukan sebelumnya. Diantaranya adalah penelitian untuk menganalisis sentimen masyarakat terhadap dampak pandemic akibat virus corona dengan menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Penelitian ini menggunakan data sebanyak 1104 yang diambil dari media sosial Twitter. Hasil penelitian tersebut menunjukkan tingkat akurasi sebesar 81,07% menggunakan *Naïve Bayes Classification* dan 79,96% menggunakan *Support Vector Machine* [13]. Penelitian lainnya membahas sentimen terhadap program kampus merdeka menggunakan algoritma *Naïve Bayes*. Pada penelitian ini menggunakan data sebanyak 501 opini publik yang dari media sosial Twitter. Hasil penelitian tersebut menunjukkan tingkat akurasi sebesar 57% dengan opini positif sebanyak 272 dan opini negatif sebanyak 229 [14]. Penelitian analisis sentimen lainnya membahas tentang ulasan pelanggan pada Aplikasi Fore Coffe yang tersaji pada Google Play Store menggunakan metode *Naïve Bayes*, juga mendapatkan hasil akurasi yang tinggi sebesar 74,28% dari 1801 data yang diproses [15].

Berdasarkan penelitian terdahulu yang telah dilakukan dan fenomena yang terjadi, belum pernah ada analisis sentimen yang berkaitan tentang keberadaan juru parkir liar di masyarakat. Oleh sebab itu, penelitian ini bertujuan untuk menganalisis opini masyarakat Indonesia mengenai keberadaan juru parkir liar dengan menggunakan metode *Naïve Bayes* sebagai algoritma untuk klasifikasi data. Hal ini bertujuan untuk mengetahui opini masyarakat dengan adanya juru parkir liar sekaligus menjadi bahan evaluasi bagi penegak hukum untuk mengatasi kegiatan ilegal tersebut. Sumber dataset yang digunakan berasal dari tweet media sosial X dengan menggunakan kata kunci "tukang parkir liar" dalam rentang waktu 5 Oktober 2023. Total tweet yang berhasil dikumpulkan sebanyak 928 tweet. Melalui proses analisis, penelitian ini mengklasifikasikan data menjadi tiga jenis sentimen, yaitu sentimen positif, netral, dan negatif.

2. METODOLOGI PENELITIAN

Dalam pengembangan penelitian ini, digunakan metode *Naïve Bayes* untuk melakukan analisis sentimen. Terdapat diagram alir yang menggambarkan langkah-langkah dalam penelitian ini agar terstruktur dan dapat terukur. Diagram tersebut sesuai yang tertera pada gambar 1.



Gambar 1. Diagram Alir Tahapan Penelitian

Untuk memulai penelitian, tahap pertama adalah *crawling data* dari internet yang berfokus pada media sosial X. *Crawling data* adalah pengambilan data pada internet. Pengambilan data melalui media sosial Twitter karena X adalah salah satu media sosial yang mengizinkan pengguna untuk berbagi foto, video, hingga pendapat dengan beropini secara bebas [16]. Pengambilan data di media sosial X melalui Google Colab menggunakan bahasa Python dengan *output* dalam bentuk file CSV. Setelah itu, data yang telah berhasil diambil akan dilakukan pemrosesan berupa *preprocessing*. Pada tahap ini data akan dirapikan dengan cara menghapus bagian data yang berupa *URL*, *mention*, *hashtag*, karakter karakter yang tidak diperlukan ([!~?.,: " # * @ % \$ & () -]), data kosong, data duplikat, dan *whitespace*. Tahap ini bertujuan untuk memudahkan peneliti dalam melakukan pelabelan sentimen.

Setelah dataset telah rapi dari karakter karakter yang tidak diperlukan, maka siap dilanjutkan ke tahap *labeling*. Pada tahap ini data akan diberikan label sentimen apakah data mengandung sentimen positif, negatif, ataupun netral. Pelabelan ini menggunakan bantuan Microsoft Excel secara manual. Pelabelan ini dilakukan hanya sebagian saja dari keseluruhan data untuk disebut sebagai data latih. Sebanyak 543 data akan dilabeli dan 362 dibiarkan saja dengan pembagian data menggunakan perbandingan 60:40 dari keseluruhan data. Nantinya data latih inilah yang akan dijadikan contoh menggunakan metode *Naïve Bayes* untuk menentukan data sisanya yang belum dilabeli atau yang disebut data uji. Setelah proses *Naïve Bayes* selesai dalam memproses data uji, maka data latih yang telah dilabeli di awal akan digabung dengan hasil data uji yang telah didapatkan. Setelah data latih dan data uji digabung, maka dilakukan evaluasi untuk mengukur akurasi dari hasil proses yang telah dilakukan. Metode yang digunakan untuk menghitung tingkat akurasi adalah *confusion matrix* dengan 3 parameter, yakni *accuracy*, *precision*, dan *recall*.



3. HASIL DAN PEMBAHASAN

10

3.1 Data Penelitian

Penelitian ini menggunakan dataset yang berasal dari media sosial X. Dataset diperoleh menggunakan bahasa Python melalui Google Colab dengan proses *crawling*. Total data yang terkumpul sebanyak 928 tweet pada tanggal 5 Oktober 2023. Proses *crawling* sesuai pada gambar 2 berikut ini.



Gambar 2. Proses *Crawling* Data

Peneliti menggunakan kata kunci "tukang parkir liar" untuk menyaring tweet yang akan diambil dari media sosial X sesuai yang terdapat pada gambar 3. Setelah data berhasil diambil, kemudian data disimpan sementara pada file yang bernama *jukir_liar* dalam format CSV. Setelah proses *crawling* selesai, kemudian file CSV akan diunduh dan akan dilanjutkan kepada proses *preprocessing*.

```
# Crawl Data

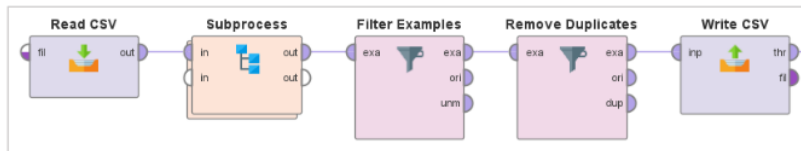
filename = 'jukir_liar.csv'
search_keyword = 'tukang parkir liar'
limit = 999

!npx --yes tweet-harvest@latest -o "{filename}" -s "{search_keyword}" -l {limit} --token ""
```

Gambar 3. *Crawling* Data X Menggunakan kata kunci "Tukang Parkir Liar"

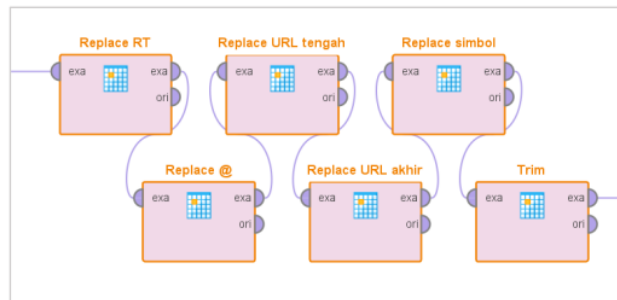
3.2 *Preprocessing*

Tahap *Preprocessing* berfungsi menyeleksi data agar menjadi data yang beraturan. Data yang telah diambil kemudian akan dibersihkan sesuai kebutuhan [17]. Tahap ini menggunakan bantuan aplikasi RapidMiner. Beberapa operator digunakan pada tahap ini seperti yang disajikan pada gambar 4.



Gambar 4. Desain Operator pada Tahap *Preprocessing*

- a. Read CSV
Read CSV merupakan operator yang berfungsi untuk membaca file CSV. Pada tahap ini digunakan file CSV yang berisi data tweet telah diunduh dari hasil *crawling* data di internet.
- b. Subprocess



Gambar 5. Desain Beberapa Operator dalam Operator *Subprocess*



Subprocess merupakan operator yang didalamnya dapat berisi kumpulan operator. Pada penelitian ini, Operator *subprocess* diisi dengan operator “Replace” dan “Trim”. Operator *Replace* untuk menghapus dan membersihkan data dari teks dan karakter yang tidak diperlukan. Misalnya, *URL*, *mention*, *hashtag*, dan karakter karakter yang tidak diperlukan ([!~?.,:."#*%\$&()-]). Sedangkan operator *Trim* untuk menghapus spasi berlebih karena adanya proses *cleansing* berupa *replace* sehingga terjadinya perubahan pada data. Desain operator yang digunakan tertera pada gambar 5. Perbandingan data sebelum dan sesudah dilakukan proses *cleansing* ditampilkan pada tabel 1.

Tabel 1. Perbandingan Hasil *Cleansing*

Sebelum	Sesudah
@infoXwor_ Bisa bilang 'gak' sama tukang parkir liar dan pengamen ga niat. Bahkan kadang ga pake bilang. Langsung ngeloyor ato cuekin aja. Ngapain sungkan sama yg tidak tertib. tolong dong ini kalau mau meniadakan belanja online (tiktokshop dll), itu tukang parkir yg gajelas ditiadakan juga.. malesnya belanja offline ya itu krn tukang parkir liar yg gaje itu. mana biaya parkirnya mahal bgtt ðŸŸ²	Bisa bilang gak sama tukang parkir liar dan pengamen ga niat Bahkan kadang ga pake bilang Langsung ngeloyor ato cuekin aja Ngapain sungkan sama yg tidak tertib tolong dong ini kalau mau meniadakan belanja online tiktokshop dll itu tukang parkir yg gajelas ditiadakan juga malesnya belanja offline ya itu krn tukang parkir liar yg gaje itu mana biaya parkirnya mahal bgtt
@tanyarlfs Tp tergantung. Klo kang parkir liar dan mereka cm mintain duit ya kesel, tp kalo tukang parkir beneran ya engga di dunia nyata: tukang parkir liar di internet: affiliator shopee pekerjaan paling dicintai di muka bumi. GUE MELIHAT DIRI GW DI BAYU..... Believe or not, dikantor kan parkir di parkir liar gt. Kalo ngutang boleh krn gw bestie sama yg jaga parkirnya. Dia selalu jagain motor gw yang liputan sampe tengah malem biar gw ga bingung nyarinya. BEST TUKANG PARKIR EVER	Tp tergantung Klo kang parkir liar dan mereka cm mintain duit ya kesel tp kalo tukang parkir beneran ya engga di dunia nyata tukang parkir liar di internet affiliator shopee pekerjaan paling dicintai di muka bumi GUE MELIHAT DIRI GW DI BAYU Believe or not dikantor kan parkir di parkir liar gt Kalo ngutang boleh krn gw bestie sama yg jaga parkirnya Dia selalu jagain motor gw yang liputan sampe tengah malem biar gw ga bingung nyarinya BEST TUKANG PARKIR EVER

c. Filter Example

Kegagalan dan kesalahan sangat mungkin terjadi pada saat *crawling* data, maka dari itu operator ini perlu digunakan untuk akan menyaring dan memilih data yang tidak hilang untuk diolah.

d. Remove Duplicate

Operator ini berfungsi untuk menghapus data yang sama agar menjadi unik (satu) agar tidak ada duplikasi pada data yang akan diolah.

e. Write CSV

Operator ini berfungsi untuk menuliskan kembali hasil dari data yang telah diproses menjadi file dengan format CSV.

3.3 Proses Pelabelan

Dataset yang telah melalui *preprocess* selanjutnya masuk ke tahap pelabelan [18]. Pada tahap ini, semua data yang telah disaring dan dibersihkan dibagi menjadi dua yakni data latih dan data uji. Pada data latih diberikan pelabelan sebagai contoh untuk mengolah data uji nantinya. Dari 928 data yang berhasil dikumpulkan dari internet, hanya sebanyak 905 data yang akan digunakan dalam penelitian ini. Data yang tidak relevan akan dibuang agar penelitian dapat semakin akurat. Peneliti membagi data dengan perbandingan 60:40. Dari 905 data, sebanyak 543 data dijadikan sebagai data latih dan 362 menjadi data uji. Pada data latih, diberikan label positif jika data mengandung opini positif, label negatif jika data mengandung opini negatif, dan label netral jika data mengandung opini netral [19] sesuai yang ditunjukkan pada tabel 2.

Tabel 2. Pelabelan Sentimen

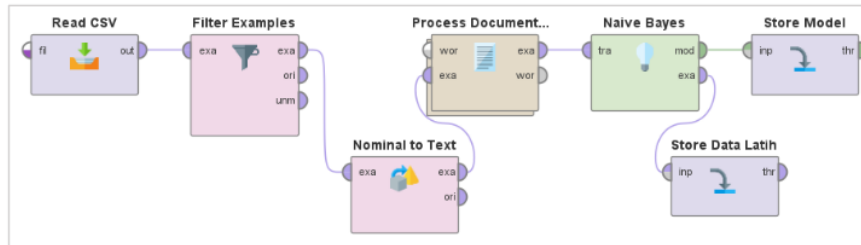
Opini	Sentimen
Bisa bilang gak sama tukang parkir liar dan pengamen ga niat Bahkan kadang ga pake bilang Langsung ngeloyor ato cuekin aja Ngapain sungkan sama yg tidak tertib tolong dong ini kalau mau meniadakan belanja online tiktokshop dll itu tukang parkir yg gajelas ditiadakan juga malesnya belanja offline ya itu krn tukang parkir liar yg gaje itu mana biaya parkirnya mahal bgtt	Negatif
Tp tergantung Klo kang parkir liar dan mereka cm mintain duit ya kesel tp kalo tukang parkir beneran ya engga	Netral



Opini	Sentimen
di dunia nyata tukang parkir liar di internet afiliator shopee pekerjaan paling dicintai di muka bumi	Positif
GUE MELIHAT DIRI GW DI BAYU Believe or not dikantor kan parkir di parkir liar gt Kalo ngutang boleh krn gw bestie sama yg jaga parkirnya Dia selalu jagain motor gw yang liputan sampe tengah malam biar gw ga bingung nyarinya BEST TUKANG PARKIR EVER	Positif

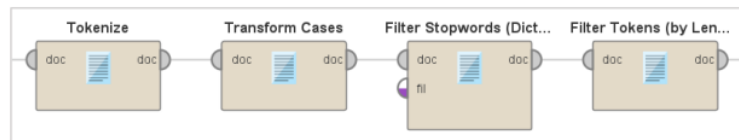
3.4 Pemodelan Data Latih

Metode *Naive Bayes* adalah klasifikasi dengan menghitung probabilitas dengan teorema *Bayes* dengan kombinasi nilai frekuensi dari data [20]. Memiliki hasil yang stabil dan akurat merupakan alasan metode ini banyak digunakan [21]. *Naive Bayes* memiliki keunggulan yakni metodenya yang sederhana dengan hasil yang baik dalam pengklasifikasian teks [22]. Bahkan metode ini lebih unggul jika dibandingkan dengan algoritma *Support Vector Machine* dan *K-Nearest Neighbor* [23]. Setelah melakukan pelabelan pada CSV untuk dijadikan sebagai data latih sebanyak 543 data, maka selanjutnya menjadikan sisa datanya sebanyak 362 sebagai data uji yang akan diproses pada RapidMiner menggunakan metode *Naive Bayes*. Desain dari pembuatan pemodelan data latih menggunakan operator "Store" dapat dilihat pada gambar 6.



Gambar 6. Pembuatan Pemodelan Data Latih

Pertama-tama menggunakan operator "Read CSV" untuk membaca file CSV yang telah disiapkan. File CSV yang digunakan adalah file yang berisi data latih. Kemudian operator dihubungkan dengan operator "Filter Example". Fungsi dari operator ini adalah untuk menyaring data sesuai kebutuhan. Untuk itu, digunakan kondisi menyaring sentimen yang tidak hilang dengan menggunakan opsi "is not missing". Dari operator "Filter Example" selanjutnya akan dihubungkan dengan operator "Process Document from Data", akan tetapi perlu diselipkan operator "Nominal to Text" terlebih dahulu. Operator ini berfungsi untuk mengubah nilai nominal menjadi Text agar operator "Process Document from Data" dapat membaca data dan berfungsi. Di dalam operator "Process Document from Data" terdapat beberapa operator yang digunakan seperti yang disajikan pada gambar 7.



Gambar 7. Operator Yang Digunakan Dalam "Process Document from Data"

- a. **Tokenize**
Operator ini berfungsi untuk memisahkan kata-kata dalam kalimat menjadi individual.

Tabel 3. Perbandingan Hasil Operator *Tokenize*

Tanpa <i>Tokenize</i>	Dengan <i>Tokenize</i>
Bisa bilang gak sama tukang parkir liar dan pengamen ga niat Bahkan kadang ga pake bilang Langsung ngeloyor ato cuekin aja Ngapain sungkan sama yg tidak tertib	'Bisa', 'bilang', 'gak', 'sama', 'tukang', 'parkir', 'liar', 'dan', 'pengamen', 'ga', 'niat', 'Bahkan', 'kadang', 'ga', 'pake', 'bilang', 'Langsung', 'ngeloyor', 'ato', 'cuekin', 'aja', 'Ngapain', 'sungkan', 'sama', 'yg', 'tidak', 'tertib'

- b. **Transform Cases**
Operator ini berfungsi untuk menyeragamkan kapitalisasi pada setiap kata dalam data. Untuk ini digunakan pilihan "lower case" agar seluruh data seragam menjadi huruf kecil.



Tabel 4. Perbandingan Hasil Operator *Transform Case*

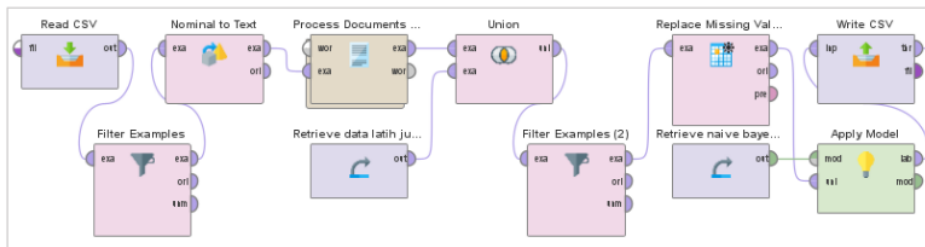
Tanpa <i>Transform Cases</i>	Dengan <i>Transform Cases</i>
Bisa bilang gak sama tukang parkir liar dan pengamen ga niat Bahkan kadang ga pake bilang Langsung ngeloyor ato cuekin aja Ngapain sungkan sama yg tidak tertib	bisa bilang gak sama tukang parkir liar dan pengamen ga niat bahkan kadang ga pake bilang langsung ngeloyor ato cuekin aja ngapain sungkan sama yg tidak tertib

- c. *Filter Stopword (Dictionary)*
Operator ini berfungsi untuk menyaring dan menghapus kata kata pada data yang terkandung dalam file stopwords. File stopwords sendiri bersumber dari kajian yang terdapat pada referensi [24].
- d. *Filter Tokens (by Length)*
Operator ini berfungsi untuk menyaring dan menghapus setiap kata pada data sesuai dari batas maksimal tertentu (1) panjang kata.

Selanjutnya pada operator "Process Document from Data" dihubungkan dengan operator "Naïve Bayes". Setelah itu, Operator "Naïve Bayes" dihubungkan lagi dengan 2 operator "Store". Operator "Store" yang pertama berfungsi untuk menyimpan hasil proses dari operator "Naïve Bayes" yang telah dinamai ulang menjadi "Store Model". Sementara operator "Store" lainnya dinamai dengan "Store Data Latih" untuk menyimpan hasil dari data latih sesuai dengan gambar 6.

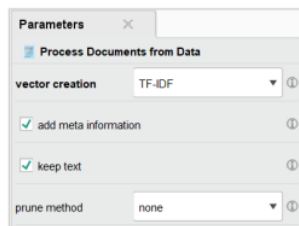
3.5 Pengujian Data Uji

Setelah pembuatan data latih berhasil, selanjutnya adalah Pengujian data uji. Operator RapidMiner yang digunakan kali ini dapat dilihat pada gambar 8 berikut.



Gambar 8. Pengujian Data Uji dengan *Naive Bayes*

Implementasi penggunaan metode *Naive Bayes* diawali dengan menggunakan operator "Read CSV" untuk membaca file CSV yang telah disiapkan, kemudian hubungkan dengan operator "Filter Example". Operator ini diatur pada pilihan "is missing" untuk nantinya sentimen yang masih kosong at(20)elum dilabeli pada data latih yang akan diambil dan diproses. Selanjutnya agar data dapat diproses, diperlukan operator "Nominal to Text" agar operator "Process Document from Data" dapat membaca dan memproses data. Parameter pada operator ini diatur "TF-IDF" pada opsi "vector creation" sesuai dengan pada gambar 9.



Gambar 9. Pengaturan Parameter Pada Operator "Process Document From Data"

Kemudian juga digunakan operator "Union" untuk menggabungkan data yang akan diuji dengan data latih yang telah dibuat. Setelah itu hubungkan kembali dengan operator "Filter Example" dengan kondisi "is missing" karena teks dengan sentimen kosong yang akan diuji dan diproses. Untuk mengisi kekosongan saat pemrosesan data diperlukan operator "Replace Missing Value" dengan kondisi default "zero" agar sistem tidak bingung dengan nilai data yang kosong. Setelah itu tambahkan operator "Apply Model" yang berfungsi untuk menyimpan dan menampilkan hasil sentimen yang dihasilkan dari pemrosesan *Naive Bayes*.

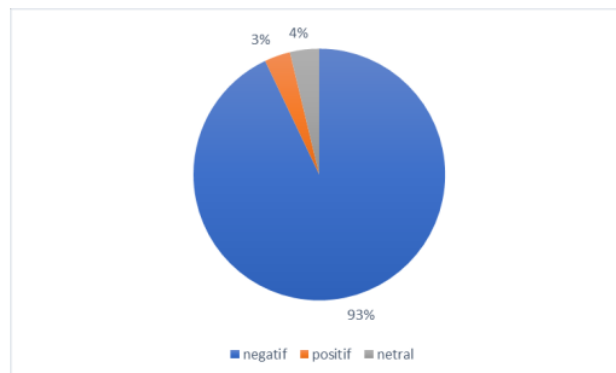


Dataset yang dihasilkan ketika *crawling* data dari media sosial X dengan kata kunci "tukang parkir liar" pada tanggal 5 Oktober 2023 adalah sebanyak 928 data dalam bentuk file CSV. Selanjutnya dilakukan tahap *preprocessing*, dari 928 data hanya diambil dan digunakan sebanyak 905 data yang relevan. Data yang kurang relevan mengandung kata kunci "juru parkir liar" akan tetapi tidak sesuai dengan data yang dibutuhkan dalam penelitian ini. Data yang kurang relevan dihapus dan tidak digunakan agar penelitian semakin akurat. Setelah dataset dimasukkan ke dalam RapidMiner untuk diproses, terdapat frekuensi kata yang paling banyak muncul bentuk *WordCloud*. Pada penelitian ini, peneliti hanya mengambil sebanyak 20 kata dengan frekuensi kemunculan tertinggi. Semakin besar frekuensi munculnya kata, maka semakin besar juga ukuran kata dalam bentuk visualisasi. Seperti yang diperlihatkan pada gambar 10.



Gambar 10. Hasil *WordCloud*

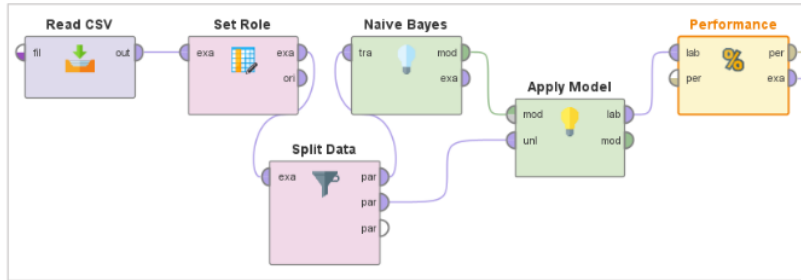
Selanjutnya dataset yang sama diolah juga untuk mendapatkan analisis sentimen. Dari 905 data yang diolah, data dibagi menjadi dua bagian. Sebanyak 543 data dijadikan sebagai data latih, dan 362 dijadikan sebagai data uji. Pembagian jumlah data tersebut berpatokan pada perbandingan 60:40. Dari 905 data yang dianalisis, sebesar 93% atau sebanyak 841 sentimen menyatakan sentimen negatif atas keberadaan juru parkir liar. Sebesar 3% atau sebanyak 30 sentimen yang menyatakan positif dengan keberadaan juru parkir liar. Dan sisanya sebesar 4% atau sebanyak 34 sentimen yang menyatakan netral dengan keberadaan juru parkir liar. Grafik persentase dapat dilihat pada gambar 11.



Gambar 11. Persentase Hasil Sentimen

3.5 Evaluasi

Setelah melakukan penelitian berupa analisis sentimen pada dataset menggunakan metode *Naïve Bayes*, selanjutnya dilakukan pengujian *Confusion Matrix*. Pengujian ini bertujuan menghitung tingkat akurasi berdasarkan evaluasi dan validasi [25]. *Confusion Matrix* menghitung tingkat akurasi dengan tiga parameter, yakni *accuracy*, *precision*, dan *recall* [26]. Perhitungan akurasi ini menggunakan bantuan aplikasi Rapid Miner. Gambar 12 menunjukkan operator operator yang digunakan dalam proses perhitungan *Confusion Matrix* pada RapidMiner.



Gambar 12. Operator-operator yang digunakan untuk menghitung *Confusion Matrix*

Perhitungan *Confusion Matrix* diawali dengan menggunakan operator "Read CSV" yang berisi data latih kemudian dihubungkan dengan operator "Set Role". Operator "Set Role" berfungsi untuk menjadikan kolom hasil sentimen menjadi atribut label. Kemudian dihubungkan dengan operator "Split Data" yang berfungsi membagi data dengan pembagian 60:40 sesuai pada gambar 13.

ratio
0.6
0.4

Gambar 13. Pembagian Data Pada Operator "Split Data"

Pada parameter "Sampling type" menggunakan opsi "automatic" yang berarti data yang digunakan adalah acak. Kemudian dihubungkan kembali dengan operator "Naive Bayes" dan "Apply Model" dan terakhir diakhiri dengan operator "Performance" untuk menyajikan hasil pengukuran akurasi. Hasil *Confusion Matrix* yang terlihat pada gambar 7 Hasil dari *Confusion Matrix* menghasilkan nilai akurasi sebesar 93.37% sesuai pada gambar 14 di bawah ini

accuracy: 92.82%				
	true negatif	true positif	true netral	class precision
pred. negatif	336	12	14	92.82%
pred. positif	0	0	0	0.00%
pred. netral	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	

Gambar 14. Hasil Pengujian Hasil dengan *Confusion Matrix*

Dengan perhitungan sebagai berikut:

a. $Accuracy = \frac{\text{Data Benar}}{\text{Jumlah Data}} = \frac{336}{362} = 0.92817$ atau 92.82% (1)

b. *Precision*

Untuk menghitung *precision*, digunakan tabel 5 dengan rumus sebagai berikut

$Precision = \frac{TP}{TP + FP}$ (2)

Tabel 5. Perhitungan *Precision* Masing Masing Kelas

	negatif	positif	netral
TP	336	0	0
FP	12 + 14	0	0
<i>Precision</i>	$336 / (336 + 26) = 0.92817$	0	0

c. *Recall*

Untuk menghitung *Recall*, digunakan tabel 6 dengan rumus sebagai berikut

$Recall = \frac{TP}{TP + FN}$ (3)

Tabel 6. Perhitungan *Recall* Masing Masing Kelas

	negatif	positif	netral
TP	336	0	0
FN	0 + 0	0	0
<i>Recall</i>	$336 / (336 + 0) = 1$	0	0



4. KESIMPULAN

Setelah dilakukan penelitian menggunakan data dari internet berupa tweet warganet media sosial X dengan metode *Naïve Bayes*, dapat ditarik kesimpulan bahwasannya metode ini mendapatkan hasil akurasi yang tinggi. Hasil penelitian menunjukkan bahwa nilai akurasi untuk memprediksi sentimen mencapai 92.82% dengan tiga hasil kategori, yakni negatif, positif, dan netral. Dari 928 data yang diambil dari internet, hanya digunakan sebanyak 905 data. Sebanyak 24 data tidak dipakai karena kurang relevan dengan sentimen keberadaan juru parkir liar, meskipun data terdapat kata kunci "tukang parkir liar" di dalamnya. Maka dari itu sebanyak 24 data yang kurang relevan dibuang agar proses penelitian semakin akurat.

Adapun hasil dari 905 data yang diproses, sebesar 93% atau sebanyak 841 sentimen menyatakan sentimen negatif atas keberadaan juru parkir liar. Sebesar 3% atau sebanyak 30 sentimen yang menyatakan positif dengan keberadaan juru parkir liar. Dan sisanya sebesar 4% atau sebanyak 34 sentimen yang menyatakan netral dengan keberadaan juru parkir liar. Tingginya sentimen negatif dari warganet dikarenakan juru parkir liar kerap melakukan pungutan liar secara paksa dan sering kali tidak menjaga kendaraan di area parkir sebagaimana mestinya. Selain itu keberadaan juru parkir liar juga dipelopori oleh ormas setempat sehingga keberadaannya sulit sekali untuk ditertibkan oleh pihak yang berwenang.

3

UCAPAN TERIMA KASIH

Terima kasih secara penuh disampaikan kepada pihak-pihak yang telah bersedia membantu dan mendukung terlaksananya penelitian ini.

Avis Tantra Mukti - Analisis Sentimen Warganet Terhadap Keberadaan Juru Parkir Liar Menggunakan Metode Naïve Bayes Classifier

ORIGINALITY REPORT

9%

SIMILARITY INDEX

8%

INTERNET SOURCES

5%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	www.stmik-budidarma.ac.id Internet Source	1%
2	123dok.com Internet Source	1%
3	www.scribd.com Internet Source	1%
4	Firman Noor Hasan. Jurnal Linguistik Komputasional (JLK), 2021 Publication	1%
5	repository.uin-suska.ac.id Internet Source	<1%
6	Submitted to Universitas Sanata Dharma Student Paper	<1%
7	doku.pub Internet Source	<1%
8	ejournal.kemenperin.go.id Internet Source	<1%

9	ojs.stmik-banjarbaru.ac.id Internet Source	<1 %
10	Submitted to Universitas Budi Luhur Student Paper	<1 %
11	app.uff.br Internet Source	<1 %
12	eprints.ummi.ac.id Internet Source	<1 %
13	ocs.unud.ac.id Internet Source	<1 %
14	Mohamad Syafri. Al-Mishbah: Jurnal Ilmu Dakwah dan Komunikasi, 2020 Publication	<1 %
15	djournals.com Internet Source	<1 %
16	docplayer.net Internet Source	<1 %
17	journal.trunojoyo.ac.id Internet Source	<1 %
18	pojoksainskomputer.files.wordpress.com Internet Source	<1 %
19	Muhammad Dwison Alizah, Arifin Nugroho, Ummu Radiyah, Windu Gata. "Sentimen Analisis Terkait Lockdown pada Sosial Media	<1 %

Twitter", Indonesian Journal on Software Engineering (IJSE), 2020

Publication

20

Windy Kusumawati Suhet, Adi Nugroho. "ANALISIS SENTIMEN PENILAIAN MASYARAKAT INDONESIA TERHADAP KONFERENSI G20 DI BALI DENGAN MENGGUNAKAN METODE NAIVE BAYES", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023

Publication

<1 %

21

fr.scribd.com

Internet Source

<1 %

22

jurnal.atmaluhur.ac.id

Internet Source

<1 %

23

jutif.if.unsoed.ac.id

Internet Source

<1 %

24

repository.upnvj.ac.id

Internet Source

<1 %

25

www.slideshare.net

Internet Source

<1 %

26

Ni Made Ayu Juli Astari, Dewa Gede Hendra Divayana, Gede Indrawan. "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes

<1 %

Classifier", Jurnal Sistem dan Informatika (JSI), 2020

Publication

27

publikasi.dinus.ac.id
Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off