Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Twitter Menggunakan Naïve Bayes

Tika Adilah M1, Yuris Alkhalifi2, Nissa Almira Mayangky3, Windu Gata4

1,2,3,4 Ilmu Komputer, STMIK Nusa Mandiri Jakarta

Jl. Kramat Raya No.18 Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10420

14002364@nusamandiri.ac.id1, 14002360@nusamandiri.ac.id2, nissa.nky@nusamandiri.ac.id3, windugata@gmail.com3

***Abstrak* – Aturan larangan mudik yang dikeluarkan oleh pemerintah di tengah pandemic virus corona (covid-19) saat ini menjadi topik yang sering diperbincangkan oleh masyarakat, salah satunya di twitter. Twitter merupakan salah satu media sosial yang seringkali digunakan untuk penyampaian pesan berupa pendapat atau opini masyarakat. Analisis sentimen adalah studi komputasional dari opini-opini orang. Analisis sentimen akan mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut. Penelitian ini mencoba untuk menganalisa opini masyarakat mengenai aturan larangan mudik dengan mengklasifikasikan opini kedalam 2 sentimen yaitu positif atau negatif. Klasifikasi dilakukan dengan menggunakan metode *Naïve Bayes Clasifiers.* Metode *Naive Bayes Classifiers* merupakan salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan dokumen latih dan dokumen uji. Hasil pengujian pada tools RapidMiner memperlihatkan bahwa akurasi dengan fitur TF-IDF memberikan hasil akurasi yang baik. Metode Naïve Bayes Classifier dapat menghasilkan akurasi yang tingga yaitu 83,38%. secara keseluruhan penggunaan metode Naive Bayes memiliki performansi yang cukup baik untuk melakukan klasifikasi tweet.**

*Kata Kunci – Analisis Sentimen, Larangan Mudik, Twitter, Naïve Bayes Classifier.*

PENDAHULUAN

Pemerintah Menugaskan Kementerian Perhubungan (Kemenhub) untuk memutuskan melarang mudik lebaran 2020 di tengah pandemic virus corona (covid-19). Aturan larangan mudik diberlakukan dengan mengacu pada Peraturan Menteri Perhubungan Nomor 25 Tahun 2020 tentang Pengendalian Transportasi Selama Musim Mudik Idul Fitri 1441 H dalam rangka pencegahan penyebaran Covid-19. Masyarakat diimbau jangan mudik karena berpotensi menularkan virus corona ke kampung halaman [1]. ini Kebijakan Pemerintah mengenai larangan mudik pada saat ini tidak luput dari komentar publik dalam media sosial twitter.

Twitter sendiri merupakan situs web yang dimiliki dan dioperasikan oleh Twitter, Inc., yang menawarkan jejaring sosial microblog. disebut microblog karena situs ini memungkinkan pengguna untuk mengirim dan membaca pesan blog seperti pada umumnya terbatas hanya 140 karakter yang ditampilkan pada halaman profil pengguna. Twitter memiliki karakteristik penulisan yang unik dan format dengan simbol atau aturan khusus. Pesan di Twitter dikenal sebagai tweet [2]. Jumlah pengguna harian Twitter secara global diklaim meningkat. Pada laporan finansial Twitter kuartal ke-3 tahun 2019, pengguna aktif harian di platform Twitter dicatat meningkat 17 persen, ke angka 145 juta pengguna. Indonesia diklaim menjadi salah satu negara yang pertumbuhan pengguna aktif harian Twitter-nya paling besar Menurut Country Industry Head Twitter Indonesia [3]. Media sosial Twitter seringkali digunakan untuk penyampaian pesan berupa pendapat atau opini masyarakat tentang bahasan tertentu serta bahasan yang menjadi *trending topic* terkini.

Kebijakan larangan mudik saat ini menjadi salah satu hal yang banyak dibicarakan di twitter. Kebijakan ini menuai berbagai macam komentar, mulai dari pujian, kritik, saran, sindiran bahkan ujaran kebencian. Dengan banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Namun dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Salah satu metode yang dapat digunakan untuk menganalisis opini-opini tweet adalah analisa sentimen.

Analisis sentimen dilakukan untuk menentukan apakah opini atau komentar terhadap suatu permasalahan, memiliki kecenderungan positif atau negatif dan dapat dijadikan sebagai acuan dalam meningkatkan suatu pelayanan, ataupun meningkatkan kualitas produk [4]. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di amerika sekitar 20-30 perusahaan yang memfokuskan pada pelayanan analisis sentimen [5].

Salah satu teknik pembelajaran dari *text mining* untuk analisis sentimen adalah *Naive Bayes Classifier.* Metode Naïve Bayes ini salah satu metode terbaik dalam pengklasifikasian [6]. Naïve Bayes menggunakan statistik sederhana berdasarkan teorema bayes yang mengasumsikan keberadaan atau ketiadaan dari suatu fitur tertentu. dari suatu kelas yang tidak berhubungan dengan fitur lainnya [7].

Dalam penelitian ini dipilih metode *Naïve Bayes Classifier* karena dianggap sebagai metode yang berpotensi baik untuk melakukan klasifikasi data dari pada metode klasifikasi lainnya dalam hal akurasi dan komputasi [8]. Penelitian terkait klasifikasi teks sudah banyak dilakukan. Pada penelitian [9] berdasarkan hasil analisis dan pengujian yang telah dilakukan dengan seratus data uji, metode *naive Bayes Classifier* (NBC) dengan *pre*-*processing* yang digunakan menghasilkan rata-rata akurasi 82,30%. Dengan menggunakan NBC pada sistem ini, dapat dihasilkan klasifikasi sentimen positif dan negatif . Sementara pada penelitian yang dilakukan [10] didapatkan bahwa NBC dapat diimplementasikan dengan nilai akurasi 93,00%. Penelitian tersebut membuktikan bahwa metode ini dapat memberikan akurasi cukup besar.

METODE PENELITIAN

Sebuah gambar berisi cuplikan layar

Deskripsi dibuat secara otomatis

Gambar 1.Metode Penelitian

1. Data

*Dataset* yang digunakan dalam penelitian ini adalah data berbahasa Indonesia tmengenai kebijakan larangan mudik di peroleh dari twitter dengan *query* “Larangan Mudik” dengan metode crawling secara otomatis menggunakan RapidMiner.

1. *Preprocessing*

Tahapan ini merupakan tahap awal untuk mengolah data teks menjadi analisa sentimen. Terdapat beberapa metode yang digunakan pada tahap ini diantaranya:

* + - 1. *Transform Cases* digunakan untuk mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf.
      2. *Tokenization* merupakan proses untuk memisahkan data teks menjadi beberapa token.
      3. *Filter Token (by Length)* digunakan untuk pemilihan token dengan ukuran panjang minimal 3 karakter (huruf), walaupun beberapa diantaranya adalah *stopwords.*
      4. *Stopwords* digunakan untuk membuang kata yang terdiri lebih dari 3 huruf dan termasuk dalam *stopwords*.
      5. *Stemming* digunakan untuk mengubah kata berimbuhan menjadi kata dasar.

1. *Weighthing*

Proses pembobotan dilakukan berdasarkan jumlah kemunculan kata-kata dalam suatu dokumen, sehingga dokumen dapat direpresentasikan dalam vektor. Bobot fitur yang digunakan adalah *Term Frequency-Inverse Frekuensi Dokumen* (TF-IDF).

1. Metode Naïve Bayes Classifier

Pada tahap ini data dianalisis, lalu diterapkan model-model yang sesuai dengan jenis data. Model yang diusulkan dalam penelitian ini yaitu *naive bayes.* metode ini merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Data dibagi kedalam data latih (*training data*) dan data uji (*testing data*).

1. Validasi dan Evaluasi

Validasi adalah tindakan membuktikan bahwa suatu proses atau metode memberikan hasil yang konsisten dan mencapai hasil menurut spesifik dan didokumentasikan dengan baik. Proses validasi dalam penelitian ini menggunakan *k-fold cross validasi* dengan k = 2 hingga k = 10

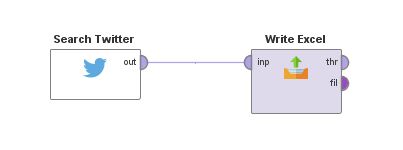
Evaluasi kinerja dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kinerja yang dilakukan oleh sistem. Parameter pengujian yang digunakan untuk mengevaluasi adalah akurasi atau kebenaran tingkat proses klasifikasi yang perhitungannya diperoleh dari tabel matriks

HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai proses yang dijalankan dalam penelitian ini. Ada beberapa langkah yang diterapkan pada penelitian ini:

1. Pengambilan Dataset

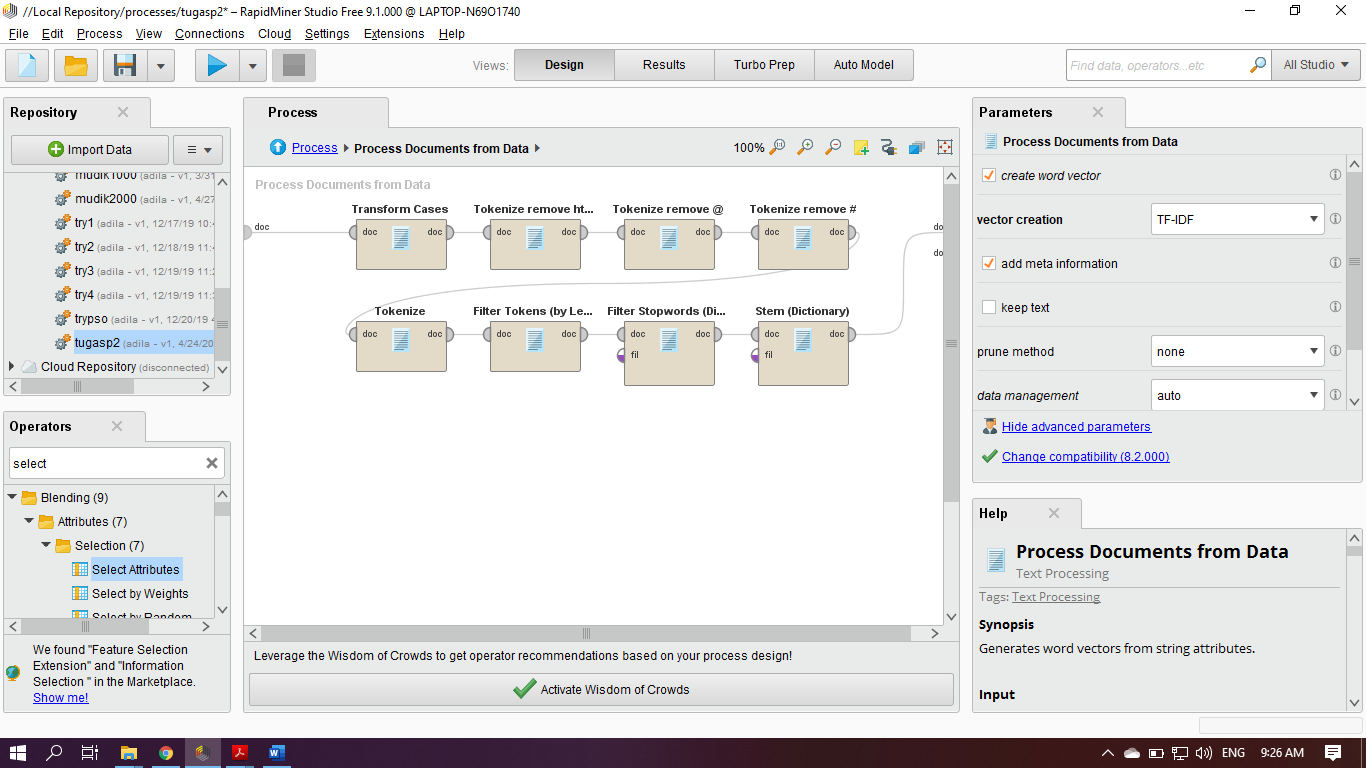
Data yang digunakan dalam penelitian ini merupakan opini berbahasa indonesia mengenai kebijakan larangan Mudik yang dikeluarkan oleh pemerintah. Data diambil dari media soasial twitter. *Crawling* data menggunkan aplikasi rapid Miner dengan *query* ‘Larangan Mudik’ dan diambil sebanyak 1000 data.



Gambar 2.Proses *Crawling* Data

1. *Preprocessing*

Tahapan ini merupakan tahap awal untuk mengolah data teks menjadi analisa sentimen. Berikut tahapan *Preprocessing* yang digunakan:



Gambar 3.*Preprocessing*

1. *Transform Cases*

Pada proses ini, *rapidminer* akan mengubah kapitalisasi karakter (huruf) menjadi kecil untuk semua kata atau huruf. Seteleh melalui tahap ini, seluruh isi komentar akan menjadi non kapital. Proses *transform cases* dapat dilihat pada Tabel 1.

Tabel 1. Tabel Hasil dari *Transform Cases*

|  |  |
| --- | --- |
| Teks sebelum dilakukan proses | Teks setelah dilakukan proses |
| Mereka yang "mudik pulang kampung" Itu, Yang mereka fikir panjanglah Pak, Wong udah kerjaan ga jalan, mau pakai apa bertahan hidup? Bapak benar, untuk memutuskan rantai Covid-19 ini, langkah bapak larangan mudik. Di Sisi lain, bapak juga harus memperhatikan, Yang gabisa mudik | mereka yang "mudik pulang kampung" itu, yang mereka fikir panjanglah pak, wong udah kerjaan ga jalan, mau pakai apa bertahan hidup? bapak benar, untuk memutuskan rantai covid-19 ini, langkah bapak larangan mudik. di sisi lain, bapak juga harus memperhatikan, yang gabisa mudik |

1. *Tokenization*

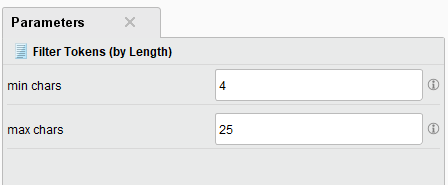
Proses *tokenization* berfungsi untuk mengilangkan tanda baca, simbol dan karakter yang bukan berupa huruf pada setiap komentar. Proses *tokenization* dapat dilihat pada Tabel 2.

Tabel 2.Tabel Hasil *Tokenization*

|  |  |
| --- | --- |
| Teks sebelum dilakukan proses | Teks setelah dilakukan proses |
| Mereka yang "mudik pulang kampung" Itu, Yang mereka fikir panjanglah Pak, Wong udah kerjaan ga jalan, mau pakai apa bertahan hidup? Bapak benar, untuk memutuskan rantai Covid-19 ini, langkah bapak larangan mudik. Di Sisi lain, bapak juga harus memperhatikan, Yang gabisa mudik | Mereka yang mudik pulang kampung Itu Yang mereka fikir panjanglah Pak Wong udah kerjaan ga jalan mau pakai apa bertahan hidup Bapak benar untuk memutuskan rantai Covid-19 ini langkah bapak larangan mudik. Di Sisi lain bapak juga harus memperhatikan Yang gabisa mudik |

1. *Filter Token (by Length)*

Pada tahap ini dilakukan pemilihan token dengan ukuran panjang minimal 4 karakter (huruf), walaupun beberapa diantaranya adalah *stopwords.* Jika kata yang kurang dari 4 huruf namun termasuk dalam *stopwords* akan tetap dibuang, karena tahap selanjutnya adalah pembuangan *stopwords.* Proses *filter token* dapat dilihat pada Gambar 3.



Gambar 4.Parameter dari Filter tokens (by length)

1. *Stopwords*

Pada tahap ini akan menyempurnakan tahap *filter token by length.* Kata-kata yang dibuang yaitu kata-kata yang tidak relevan.

1. *Stemming*

Pada tahap stemming, kata dengan kata hubung diubah menjadi kata dasar untuk *tweet* dalam bahasa Indonesia.

1. Validasi

Menguji *dataset* dengan komposisi data pengujian dan parameter pelatihan yang berbeda dilakukan dengan mendistribusikan data pelatihan dan menguji data dengan komposisi berdasarkan distribusi data dalam validasi silang k-fold dengan nilai k = 2 hingga k = 10. Tes ini bertujuan untuk menentukan nilai k-fold optimal dalam proses klasifikasi opini dalam kasus penelitian ini. Operator *Cross Validation* K-Fold dapat dilihat pada Gambar 5.



Gambar 5.Operator Cross Validation K-Fold

1. Evaluasi

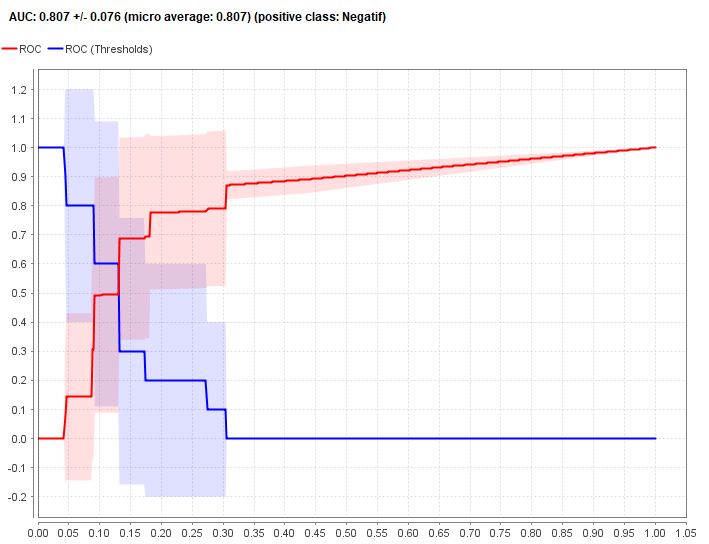
Hasil dari implementasi model *naive bayes* tersebut akan didapat seluruh *rule. Rule* yang dihasilkan akan digunakan sebagai dasar prediksi nilai yang akan dilakukan. Sebelumnya *rule* tersebut harus dievaluasi dan divalidasi sehingga diketahui seberapa akurat hasil prediksi yang akan dilakukan. Pada tahap ini proses evaluasi menggunakan metode *confusion matrix.*

Model *confusion matrix* akan membentuk matriks yang terdiri dari *accuracy, true positif* dan *true negatif, recall* dan *precision.* Berikut merupakan hasil dari *confusion matrix* pada algoritma *naive bayes classifier.* Klasifikasi *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 3.Tabel Hasil Accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No |  | True Positif | True Negatif | Class Precission |
| 1 | Prediction positif | 199 | 34 | 84,41% |
| 2 | Prediction Negatif | 27 | 107 | 79,85% |
| 3 | Class recall | 88.05% | 75.89% |  |

Hasil *accuracy* pengujian *naive bayes* untuk analisis sentimen pada media sosial instagram memperoleh hasil sebesar 83,83% sedangkan untuk nilai AUC nya sebesar 0.807 yang dapat dilihat pada Gambar 6.



Gambar 6. AUC Proses NBC

KESIMPULAN DAN SARAN

Berdasarkan hasil analisa yang dilakukan terkait sentimen terhadap kebijakan larangan mudik dengan menggunakan metode *naive bayes* dapat diambil beberapa kesimpulan bahwa Metode *naive bayes* merupakan metode yang cukup baik dalam pengklasfikasian data mining ataupun *text mining.* hal ini dikarenakan algoritma tersebut dapat menghasilkan nilai akurasi yang cukup tinggi.

REFERENSI

[1] R. R. Ramli, “Berlaku Hari Ini, Simak Fakta-fakta Soal Larangan Mudik Lebaran.”

[2] B. Pratama *et al.*, “Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods,” *J. Phys. Conf. Ser.*, vol. 1201, no. 1, pp. 0–12, 2019.

[3] Kompas.com, “Pengguna Aktif Harian Twitter Indonesia Diklaim Terbanyak,” 2019.

[4] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1750–1757, 2017.

[5] Yislam and I. Budi, “Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter,” no. October, 2016.

[6] R. Y. Hayuningtyas and R. Sari, “Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naïve Bayes Dan Pso,” *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 37–42, 2019.

[7] L. Wilianto, T. H. Pudjiantoro, and F. R. Umbara, “ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DARI KOMENTAR PENGUNJUNG DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER STUDI KASUS JAWA BARAT,” pp. 439–448, 2017.

[8] D. Xhemali, C. J. Hinde, and R. G. Stone, “IJCSI IJCSI Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 4, no. 1, pp. 16–23, 2009.

[9] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepry Maulana Husain, “Sentiment Analysis Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019.

[10] Siswanto, Y. P. Wibawa, W. Gata, G. Gata, and N. Kusumawardhani, “Classification Analysis of MotoGP Comments on Media Social Twitter Using Algorithm Support Vector Machine and Naive Bayes,” *Proc. ICAITI 2018 - 1st Int. Conf. Appl. Inf. Technol. Innov. Towar. A New Paradig. Des. Assist. Technol. Smart Home Care*, pp. 96–101, 2018.