

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

Tika Adilah M<sup>1</sup>, Yuris Alkhalifi<sup>2</sup>, Nissa Almira Mayangky<sup>3</sup>, Windu Gata<sup>4</sup>

<sup>1,2,3,4</sup>Ilmu Komputer, STMIK Nusa Mandiri Jakarta

Jl. Kramat Raya No.18 Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10420  
14002364@nusamandiri.ac.id<sup>1</sup>, 14002360@nusamandiri.ac.id<sup>2</sup>, nissa.nky@nusamandiri.ac.id<sup>3</sup>,  
windu@nusamandiri.ac.id<sup>3</sup>

**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 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 menggolongkan teks yang ada dalam sebuah kalimat untuk mengetahui pendapat yang tercermin 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 Classifiers*. Metode *Naïve Bayes Classifiers* merupakan 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 tinggi yaitu 83,38%. secara keseluruhan penggunaan metode *Naïve 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 aturan Menteri perhubungan no25 tahun 2020 mengenai pengendalian transportasi selama musim mudik untuk pencegahan penyebaran Covid-19. Masyarakat diimbau untuk tidak pulang kampung karena dikhawatirkan akan berpotensi menularkan virus ke kampung halaman [1]. Kebijakan Pemerintah mengenai larangan mudik pada saat ini tidak luput dari komentar publik dalam media sosial twitter.

Twitter sendiri merupakan situs yang dimiliki dan dikembangkan oleh Twitter, Inc., twitter menawarkan jejaring sosial microblog atau dengan kata lain memungkinkan pengguna mengirim dan membaca pesan terbatas hanya 140 karakter. Twitter mempunyai karakteristik penulisan yang unik dan format dengan simbol atau aturan khusus. Pesan di Twitter dikenal sebagai tweet [2]. Jumlah pengguna harian Twitter di seluruh dunia meningkat. Pada laporan pengguna Twitter kuartal ke-3 pada tahun 2019, pengguna aktif harian Twitter dicatat meningkat 17 persen, jumlah penggunanya mejuju angka 145 juta. Indonesia sendiri merupakan salah satu negara yang pengguna aktif hariannya bertumbuh dengan pesat 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.

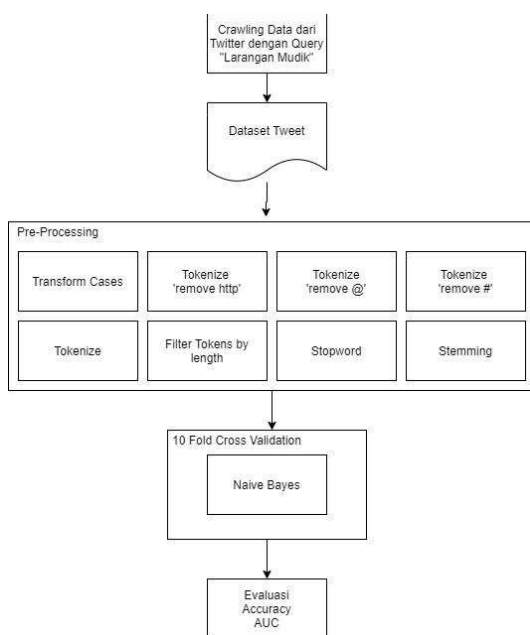
Kebijakan larangan mudik saat ini menjadi salah satu hal yang banyak dibicarakan di twitter. Kebijakan ini menuai berbagai macam komentar, mulai dari pujian, kritik, saran, sindiran bahkan ujaran kebencian. Dengan banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk 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 cara yang dapat digunakan untuk menganalisis opini-opini tweet adalah analisa sentimen.

Analisis sentimen berguna 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]. Besarnya manfaat serta pengaruh dari melakukan analisis sentimen membuat penelitian dan aplikasi mengenai analisis sentimen berkembang pesat. Bahkan di amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pekerjaannya pada pelayanan analisis sentimen [5].

Salah satu teknik pembelajaran dari *text mining* untuk analisis sentimen adalah *Naïve 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 mengenai pengklasifikasian teks sudah banyak dilakukan. Pada penelitian [9] hasil analisis dan pengujian yang 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



Gambar 1. Metode Penelitian

#### A. Data

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

#### B. 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 digunakan agar semua kata dalam setiap dokumen dikumpulkan dan dihilangkan tanda baca, dan dihilangkan jika ada simbol atau karakter khusus yang bukan huruf dan memisahkan kalimat

menjadi kata-kata sesuai kebutuhan yang sering disebut token

3. Filter Token (by Length) digunakan untuk pemilihan token dengan ukuran panjang minimal 4 karakter (huruf), walaupun beberapa diantaranya adalah stopwords.
4. Stopwords digunakan untuk membuang kata yang tidak relevan, seperti kata "tetapi", "untuk", "dengan". Kata-kata ini tidak bermakna saat dipisah dengan kata lain.
5. Stemming digunakan untuk mengubah kata berimbuhan menjadi kata dasar.

#### C. 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*).

#### D. 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  $k = 10$

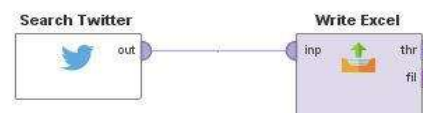
Evaluasi kinerja dilakukan untuk menguji hasil klasifikasi dengan mengukur nilai kinerja yang dilakukan oleh sistem. Parameter pengujian yang digunakan untuk mengevaluasi adalah akurasi atau kebenaran tingkat proses klasifikasi yang perhitungannya diperoleh dari tabel matriks.

### HASIL DAN PEMBAHASAN

Beberapa langkah yang diterapkan pada penelitian ini:

#### A. Pengambilan Dataset

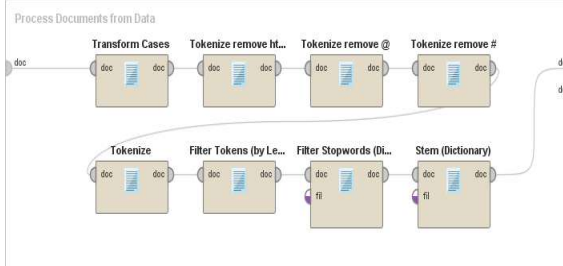
Data untuk penelitian ini merupakan opini berbahasa indonesia mengenai kebijakan larangan Mudik yang dikeluarkan oleh pemerintah. Data diambil dari media sosial twitter. *Crawling* data menggunakan aplikasi rapid Miner dengan *query* ‘Larangan Mudik’ dan diambil sebanyak 1000 data.



Gambar 2. Proses Crawling Data

#### B. 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. Setelah melalui tahap ini, seluruh isi komentar akan menjadi non kapital. Proses *transform cases* ditunjukkan oleh 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 |

2. Tokenization

Proses *tokenization* berfungsi untuk mengilangkan tanda baca, simbol dan karakter yang bukan berupa huruf pada setiap komentar. Proses *tokenization* ditunjukkan oleh tabel 2.

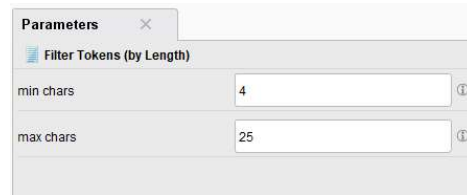
Tabel 2. Tabel Hasil *Tokenization*

| 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 |

3. 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* ditunjukkan oleh Gambar 4.



Gambar 4. Parameter dari Filter tokens (by length)

4. Stopwords

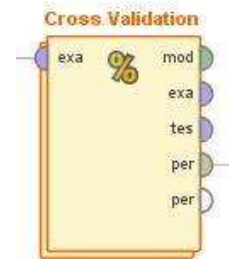
Pada tahap ini akan menyempurnakan tahap *filter token by length*. operator yang digunakan yaitu *Stopword Removal (by Directory)*. Yang dihilangkan yaitu tanda baca, simbol atau karakter khusus yang bukan huruf dan kata-kata yang tidak relevan dengan isi text.

5. Stemming

Pada tahap stemming, kata dengan kata hubung diubah menjadi kata dasar dengan menggunakan *indonesian stemming* untuk *review* berbahasa Indonesia.

C. Validasi

pada penelitian ini menggunakan 10 Fold Cross-Validation. 10 Fold cross validation digunakan untuk mendapatkan model terbaik. Menurut [11] 10 Cross-validasi lebih baik daripada validasi Leave-One-Out untuk pemilihan model, dan juga lebih baik daripada K-Fold lainnya. Operator *Cross Validation* K-Fold dtunjukkan pada Gambar 5.



Gambar 5. Operator Cross Validation K-Fold

D. Evaluasi

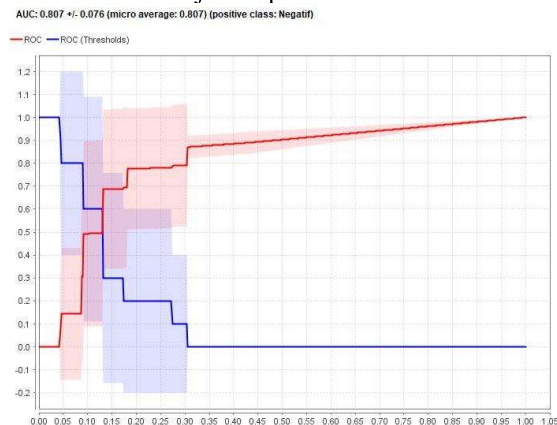
Hasil dari implementasi model *naive bayes* tersebut akan didapat seluruh *rule*. *Rule* yang dihasilkan akan digunakan sebagai dasar prediksi nilai yang akan dilakukan. Sebelumnya *rule* tersebut harus dievaluasi dan divalidasi sehingga diketahui seberapa akurat hasil prediksi yang akan dilakukan. Pada 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 ditunjukkan pada Tabel 3.

Tabel 3. Tabel Hasil Accuracy.

| No |                    | True Positif | True Negatif | Class Precision |
|----|--------------------|--------------|--------------|-----------------|
| 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 twitter memperoleh hasil sebesar 83,83% sedangkan untuk nilai AUC nya sebesar 0.807 ditunjukkan pada Gambar 6.



Gambar 6. AUC Proses NBC

#### KESIMPULAN DAN SARAN

Berdasarkan hasil analisis yang dilakukan terkait dengan sentimen masyarakat mengenai larangan mudik menggunakan metode Naive Bayes, peneliti dapat menarik beberapa kesimpulan bahwa Metode Naive Bayes adalah metode yang cukup baik dalam mengklasifikasikan data mining ataupun text mining. karena algoritma ini dapat menghasilkan nilai akurasi yang cukup tinggi, yang berarti bahwa dari semua tweet mengenai larangan mudik dapat diklasifikasikan oleh metode NBC secara akurat apakah komentar itu negatif atau positif. Untuk kedepannya, diharapkan penelitian ini dapat dikembangkan sehingga hasil akurasi dapat ditingkatkan, penelitian dapat menggunakan lebih banyak dataset dan menggunakan kamus yang selalu diperbarui sesuai dengan pertumbuhan bahasa dan juga metode yang digunakan dapat dikombinasikan dengan metode lainnya seperti metode Support vektor machine (SVM) sehingga bisa mendapatkan hasil yang lebih akurat.

#### 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 Syepri 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.  
 [11] E. Y. Sari, A. D. Wierfi, and A. Setyanto, "Sentiment Analysis of Customer Satisfaction on Transportation Network Company Using Naive Bayes Classifier," *2019 Int. Conf. Comput. Eng. Network, Intell. Multimedia, CENIM 2019 - Proceeding*, vol. 2019-November, 2019.