

Analisis Sentimen Terhadap Kebijakan Pemerintah Tentang Larangan Mudik Hari Raya Idulfitri di Indonesia Tahun 2021 Menggunakan Metode *Naïve Bayes*

Abdul Aziz¹, Fauziah², Iskandar Fitri³

Teknik Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
Jl. Sawo Manila No.61, Pasar Minggu, Jakarta Selatan 12520, (021) 7806700
abddul.aziz.97@gmail.com, fauziah@civitas.unas.ac.id, iskandarfitri66@gmail.com

Abstract

Social media as a place to access and disseminate information has grown very rapidly, one of which is Twitter. Twitter, as a place for information flow, is a rich source for seeking public opinion and sentiment analysis. Twitter in this study was used as a source to obtain data about the 2021 homecoming in Indonesia. The purpose of this study is to determine public satisfaction with government policies regarding the ban on going home in Indonesia in 2021. The data to be processed is Indonesian-language tweets, the keywords are #mudik and #dilarangmudik, the length of data collection is 1 week, with lots of data generated as many as 1000. Sentiment analysis in this study using the Naïve Bayes Classification method. The steps in this study are first crawling Twitter data which is then stored in csv format, second preprocessing which consists of tokenizer, case folding, cleansing and stop removal, third Naïve Bayes classification which will be carried out after going through the Pre-processing stage, where the results of the classification tweets tend to be positive or negative or neutral. The results of this study obtained an accuracy of 56.52% with each positive sentiment value of 62.28%, negative sentiment as much as 46.72% and neutral sentiment as much as 66.50%.

Keywords: Sentiment Analysis, Nave bayes, Text Mining

Abstrak

Media sosial sebagai tempat untuk mengakses dan menyebarkan informasi telah berkembang sangat pesat, salah satunya adalah Twitter. Twitter sebagai tempat arus informasi, merupakan sumber yang kaya untuk mencari pendapat masyarakat dan analisis sentimen. Twitter pada penelitian ini digunakan sebagai sumber untuk mendapatkan data tentang mudik tahun 2021 di Indonesia. Tujuan dari penelitian ini adalah untuk mengetahui kepuasan masyarakat terhadap kebijakan pemerintah tentang larangan mudik di Indonesia tahun 2021. Data yang akan di proses merupakan tweet berbahasa Indonesia, kata kunci yaitu #mudik dan #dilarangmudik, lama pengumpulan data adalah 1 minggu, dengan banyak data yang dihasilkan sebanyak 1000. Analisis sentimen pada penelitian ini menggunakan metode Naïve Bayes Classifier. Langkah-langkah pada penelitian ini pertama crawling data Twitter yang selanjutnya di simpan dalam format csv, kedua preprocessing yang terdiri dari tokenizer, case folding, cleansing dan stop removal, ketiga klasifikasi Naive bayes yang akan dilakukan setelah melalui tahap Pre-processing, dimana hasil klasifikasi tweet berkecenderungan positif atau negatif atau netral. Hasil dari penelitian ini di dapatkan akurasi 56,52% dengan masing-masing nilai preicision sentimen positif sebanyak 62,28%, sentiment negatif sebanyak 46,72% dan sentimen netral sebanyak 66,50%.

Kata kunci: Analisis Sentimen, Naïve Bayes, Text Mining

1. PENDAHULUAN

Mudik merupakan kegiatan pulang ke kampung halaman yang identik dengan hari raya keagamaan seperti hari raya Idulfitri. Momen mudik digunakan sebagai sarana untuk berkumpul dengan sanak saudara dan orang tua. Berbagai macam Transportasi baik umum maupun pribadi bisa digunakan untuk alat transportasi ketika mudik. Kegiatan mudik banyak ditemukan di negara-negara dengan sebagian besar penduduk beragama Islam seperti Malaysia, Bangladesh dan Indonesia. Pada tahun 2021 seiring dengan meningkatnya jumlah korban terinfeksi virus Corona maka pemerintah membuat kebijakan larangan mudik yang dimuat dalam “Surat Edaran (SE) Satuan Tugas Penanganan Covid-19 No. 13 Tahun 2021”, kebijakan ini diambil sebagai upaya pemerintah dalam menangani dan menekan laju penularan virus covid-19. Kebijakan yang diambil pemerintah mendapat respon positif dan juga negatif dari masyarakat. Banyak kalangan masyarakat menyambut baik kebijakan pemerintah untuk meniadakan mudik lebaran tahun 2021 sebagai upaya pencegahan penularan Covid-19. Namun banyak juga masyarakat yang tidak puas bahkan menentang kebijakan tersebut, khususnya perantau.

Platform media sosial khususnya *Twitter* menjelang Idulfitri banyak diramaikan dengan topik pembahasan “mudik” masyarakat dari semua kalangan memberikan opini serta pandangan mereka tentang kebijakan pemerintah meniadakan mudik hari raya Idulfitri tahun 2021 di Indonesia melalui tweet yang mereka unggah. *Twitter* sebagai tempat arus informasi merupakan sumber yang kaya untuk belajar tentang pendapat orang dan analisis sentiment. Analisis sentimen kini telah menjadi pendekatan yang dominan digunakan untuk mengekstrak sentimen dan penilaian tentang topik tertentu dari sumber online, yang terbagi menjadi tiga kategori yaitu positif, negatif dan netral

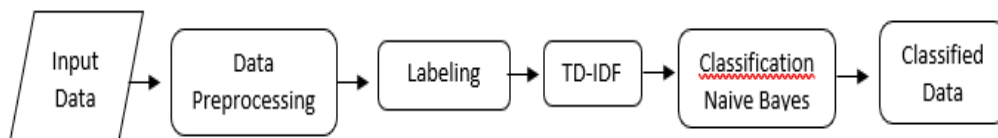
Text mining adalah proses mengambil informasi dari data yang berbentuk teks untuk mengetahui permasalahan pada sebuah topik tertentu [1]. Pada analisis sentiment Text mining mampu menganalisa dan mengidentifikasi jenis emosional pada sebuah pernyataan [2].

Naïve Bayes Classifier merupakan salah satu teknik pembelajaran text mining. Metoda klasifikasi *Naïve Bayes* bersumber pada teorema Bayes, yaitu klasifikasi yang memanfaatkan metode statistik serta probabilitas yg diciptakan oleh Thomas Bayes [3]. *Naïve Bayes* dapat memperkirakan peluang di masa yang akan datang dengan mempelajari pengalaman di masa lalu [4]. Opini yg sangat kuat (naif) tentang independensi dari masing-masing peristiwa adalah ciri khas *Naive Bayes Classifier* [5].

Penelitian ini berfokus pada analisis klasifikasi sentimen data dengan menerapkan *Naive Bayes Classifier* untuk menghasilkan opini yang mengandung sentimen positif, negatif atau netral yang bertujuan untuk memprediksi kepuasan masyarakat terhadap kebijakan pemerintah tentang larangan mudik di Indonesia tahun 2021.

2. METODOLOGI PENELITIAN

Gambar 1 merupakan rangkaian proses pada penelitian. Data yang dikumpulkan (*crawling data*) merupakan opini yang diambil dengan berdasarkan tanggapan masyarakat mengenai kebijakan pemerintah tentang larangan mudik hari raya Idulfitri tahun 2021 melalui media sosial *Twitter*. setelah data set terkumpul selanjutnya dilakukan proses *Pre-processing* yang terdiri dari tokenizing, cleansing, case folding dan stopword removal. Kalimat opini yang telah melalui proses *Pre-processing* berupa kalimat dalam tulisan bahasa Indonesia yang mudah dimengerti. Setelah tahap *Pre-processing* selanjutnya kalimat akan diberikan label (*labeling*) dan akan melewati tahap pembobotan kata (TF-IDF), pemberian label dilakukan secara manual pada dataset. Proses terakhir adalah klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.



Gambar 1. Proses Metodologi Penelitian.

2.1. Input Data

Data *Twitter* tentang opini masyarakat mengenai kebijakan pemerintah tentang larang mudik di Indoneisa tahun 2021 akan diambil dengan menggunakan metode pengumpulan tweets yang disediakan oleh API *Twitter* (*crawling data*) lalu disimpan dalam format csv. Data yang diambil adalah data tweets dengan hashtag #mudik dan #dilarangmudik, pengumpulan data dilakukan selama 1 minggu dari tanggal 8 Juli 2021 sampai dengan 15 Juli 2021 dengan jumlah *tweet* yang terkumpul adalah 1000 buah.

2.2. Data *Pre-processing*

Pre-processing merupakan tahap pertama dalam pengolahan data yang berfungsi untuk membersihkan dataset. Metode yang digunakan pada tahap *Pre-processing* pada peneilitian ini yaitu:

- Tokenizer** merupakan proses membagi data teks menjadi token dan menguraikannya menjadi bentuk kata-kata.
- Cleansing** merupakan proses untuk membersihkan data dengan cara menghilangkan bagian-bagian data yang tidak sesuai atau tidak relevan.
- Case folding** merupakan proses untuk mengubah setiap kata menjadi lower case atau huruf kecil.
- Stopword Removal** merupakan proses menghilangkan konjungsi yang tidak mempunyai arti (stopwords). Penghapusan stopwords ini menggunakan library Sastrawi. Misalnya : “yang”, “di”, “ke”, “atau”, “ada”, “ikut”, “maka”, “dan”, dll.

2.3. Labeling

Labeling data merupakan pemberian label pada masing masing tweet. Pada penelitian ini diberikan 3 label sebagai parameter yaitu, negarif, positif dan netral. Cara menentukan masing-masing label pada tweet adalah sebagai berikut:

- a) Pilih kata/tweet yang mempunyai sentimen positif , sentimen negatif juga sentimen netral.
- b) Apabila kata sentimen positif > kata sentiment negatif, maka label sentimennya adalah positif (1).
- c) Apabila kata sentimen positif < kata sentiment negatif, maka label sentimennya adalah negatif (-1).
- d) Apabila kata sentimen positif = kata sentimen negatif, maka label sentimennya adalah netral (0).

2.4. TF-IDF

Pembobotan kata (TF-IDF) berfungsi untuk mengubah tekstual yang telah melalui tahapan preroessing ke dalam bentuk numeric. Rumus TF-IDF

- a) Hitung *term frequency* ($tf_{t,d}$)
- b) Hitung *weighting term frequency* (W_{tft})
$$W_{tft} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$
- c) Hitung *document frequency* (df)

2.5. Naïve Bayes

Bayes memprediksi data dengan menggunakan probabilitas yang sangat sederhana dan bersumber pada teori Bayes dengan independensi atau ketidak terikatan yang kuat (naif) [6]. Sehingga dapat disimpulkan bahwa Bayes menggunakan model fitur independent dalam melakukan klasifikasi data.

Pada model Naive Bayes, yang dimaksud independensi atau ketidak terikatan terhadap fitur adalah sebuah fitur pada sebuah dataset tidak berhubungan dengan ada atau tidaknya fitur lain dalam dataset yang sama. Prediksi Bayes bersumber pada teori Bayes dengan rumus sebagai berikut:

$$P\left(\frac{H}{E}\right) = \frac{P\left(\frac{E}{H}\right) \times P(H)}{P(E)} \quad (2)$$

Penjelasan :

- 1) $P(H/E)$ Probabilitas akhir bersyarat atau conditional probabilitas, yaitu suatu hipotesis H terjadi jika diberikan bukti (evidence) E terjadi.
- 2) $P(E/H)$ Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H.
- 3) $P(H)$ adalah Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.
- 4) $P(E)$ Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

3. HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan beberapa tahapan pada proses klasifikasi untuk menganalisis sentimen opini terhadap kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia. Dataset berasal dari media sosial *Twitter* yang diambil menggunakan teknik *Crawling* lalu disimpan dalam format *csv*. Tweet tersebut berbahasa Indonesia, yang mengandung sentimen positif, negatif dan netral. Tweet dengan kata kunci atau hashtag #mudik dan #dilarangmudik. Proses klasifikasinya dengan metode *Naïve Bayes*, kemudian diproses oleh pemrograman python pada tools jupyter notebook serta menggunakan aplikasi Rapid Miner.

3.1. Pre-peprocessing Data

Data hasil *Crawling* yang telah disimpan dalam format *csv* selanjutnya diolah dalam proses *Pre-processing*. Metode *Pre-processing* yang di pakai yaitu *tokenizer*, *cleansing*, *case folding* dan *stop removal*. Contoh data hasil pre-processing

Table 1. hasil *Pre-processing*

Sebelum <i>Pre-processing</i>	Sesudah <i>Pre-processing</i>
'Sedih lebaran kali ini tidak ada mudik padahal sehat tpi dilarang mudik. /n Karena lebaran kurang lengkap tanpa mu dik\n#jomblo #lebaran #mudik'	sedih lebaran ini tidak ada mudik padahal sehat tapi larang mudik karena lebaran kurang lengkap tanpa mudik
Jangan mudik demi mencegah penularan covid-19\n#MudikDilarang #mudik #mudik2021 \nTentramkan Negeri https://t.co/xSBTNalp8N	jangan mudik cegah nular covid

Setelah *Pre-processing*, kemudian dilakukan proses pelabelan dengan label 1 pada kata yang mengandung sentimen positif . label 0 pada kata yang mengandung sentimen netral dan label -1 pada kata yang mengandung sentiment negatif. Pada tahap labeling ini didapatkan opini positif sebanyak 360, opini netral sebanyak 340 dan opini negatif sebanyak 300. Proses selanjutnya melakukan pembobotan kata dengan metode TF-IDF, untuk menghitung term pada dokumen.

3.2. Hasil

Metode *Naïve Bayes Classifier* memaksimalkan setiap kemunculan dan frekuensi fitur pada setiap tanggapan/opini untuk mendapatkan tingkat akurasi yang tinggi [7]. Akurasi yang dihasilkan oleh model *Naïve Bayes* dipengaruhi oleh banyaknya data training yang dipakai. Pada penelitian ini data training yang dipakai sebanyak 1000 data dengan 360 data positif, 340 data netral dan 300 data negatif, dalam melakukan pengujian menggunakan teknik sebanyak lima kali atau disebut juga *five-Fold Cross Validation* (5-

Fold), adalah dengan membagi dataset kedalam 5 bagian. Setiap Fold yang akan diuji berjumlah 200 data random yang mempunyai data berbeda. Hasil analisis dari pengujian data menggunakan metode *Naïve Bayes*. Hasil Uji dari masing-masing bagian adalah sebagai berikut:

Accuracy		56.5%	
	true netral	true positif	true negatif
class recall	40.00%	65.22%	57.89%
	class precision		
pred. netral	66.67%		
pred. positif	60.00%		
pred. negatif	47.83%		

Gambar 2. Uji data Fold1

Accuracy		61.1%	
	true netral	true positif	true negatif
class recall	44.44%	93.55%	11.76%
	class precision		
pred. netral	100.00%		
pred. positif	61.70%		
pred. negatif	33.33%		

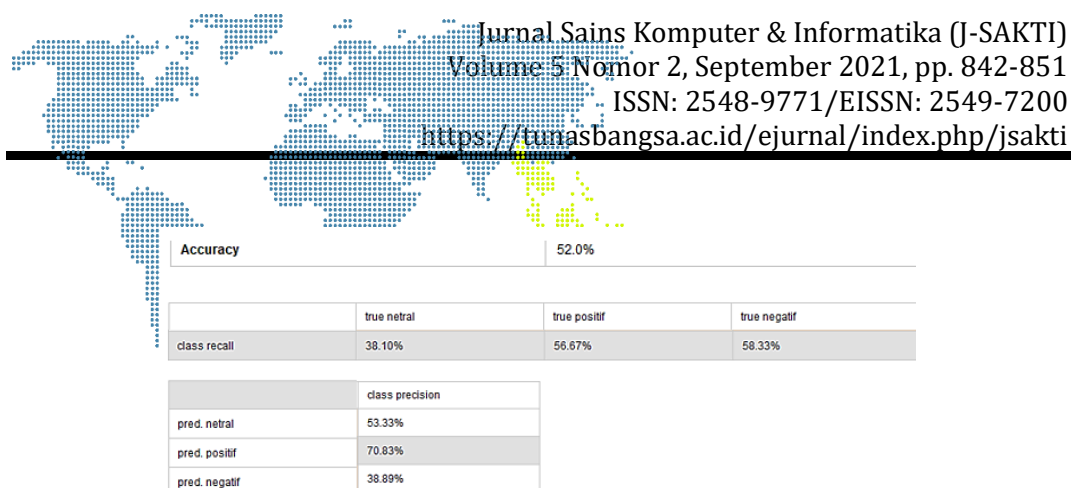
Gambar 3. Uji data Fold2

Accuracy		54.4%	
	true netral	true positif	true negatif
class recall	45.00%	82.61%	21.43%
	class precision		
pred. netral	50.00%		
pred. positif	55.88%		
pred. negatif	60.00%		

Gambar 4. Uji data Fold3

Accuracy		58.6%	
	true netral	true positif	true negatif
class recall	35.71%	68.00%	62.50%
	class precision		
pred. netral	62.50%		
pred. positif	62.96%		
pred. negatif	53.57%		

Gambar 5. Uji data Fold4



Gambar 6. Uji data Fold5

Setelah pengujian selesai dilakukan pada semua fold, berikutnya hitung nilai rata-rata dari nilai precision, recall dan accuracy.

Tabel 2. Data rata-rata hasil Uji

Naïve Bayes Clasifier		
Accuracy	56,52%	
	Precision	Recall
Positive	62,28%	73,21%
Negative	46,72%	42,38%
Neutral	66,50%	45,98%

Pada Tabel 2 dapat di lihat nilai Accuracy data keseluruhan mencapai 56,52%. Nilai precision data opini netral memiliki persentase yang lebih besar dibanding dengan nilai precision opini negatif dan positif yaitu 66,50%. Opini positif mempunyai nilai persentase recall terbesar yaitu 73%, besarnya nilai persentase recall opini positif dapat diartikan sebagai keberhasilan pada sistem dengan hanya mempunyai sedikit data yang salah pada proses klasifikasi. Dari hasil penelitian dapat diketahui sebagai berikut:

- 1) Nilai opini yang paling banyak muncul dalam menanggapi kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia adalah merupakan kategori opini netral dimana opini netral merupakan opini yang memiliki nilai bobot sama besar antara negatif dan positif,
- 2) Nilai opini terbanyak kedua adalah positif, kategori sentimen positif adalah kategori opini masyarakat yang mendukung kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia
- 3) Nilai opini dengan nilai tekecil adalah negatif. Kategori sentiment negatif merupakan kategori opini masyarakat yang menentang atau tidak setuju dengan adanya kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia

3.3. Pembahasan

Penelitian yang dimulai dengan mengumpulkan tweets (crowling) hingga melakukan perhitungan menggunakan metode *Naïve Bayes* masih memiliki banyak permasalahan sehingga membuat penelitian ini tidak bekerja secara maksimal. Permasalahan yang ditemukan diantaranya:

a) Data Tidak Seimbang

Ketidak seimbangan jumlah dataset berakibat pada hasil klasifikasi yang tidak maksimal. Dataset yang digunakan berjumlah 1000 data dengan 360 data opini positif, 300 data opini negatif dan 340 data opini netral. Jumlah data yang tidak seimbang pada masing-masing kategori menyebabkan kesalahan pada sistem saat melakukan proses klasifikasi. misalkan pada kalimat “boleh atau tidak boleh tetap mudik gas”, pada saat tahap labeling kalimat tersebut masuk dalam kategori opini negatif karena menentang kebijakan pemerintah, tapi setelah diproses dan diuji oleh sistem ternyata hasilnya kalimat tersebut masuk dalam kategori opini netral. Jadi dapat di simpulkan bahwa data yang tidak seimbang menyebabkan sistem bekerja kurang baik sehingga penelitian ini bekerja kurang maksimal.

b) Dataset Tidak Tepat

Data set tidak tepat menyebabkan banyak ditemukan kesalahan pada proses klasifikasi kategori pada data uji yang digunakan. Contoh pada kategori positif terdapat kata “mudik” yang berasal dari kata “jangan mudik” kemudian pada kategori negatif juga terdapat kata “mudik” yang berasal dari kata “ayo mudik”. Kata yang sama pada dua kategori yang berbeda ini menyebabkan terjadi kesalahan. contoh lain kalimat “ayo gas mudik jangan takut sama pemerintah” seharusnya kalimat tersebut terklasifikasi kedalam kategori negatif, tetapi pada saat data diolah oleh sistem kalimat tersebut masuk dalam kategori positif, hal ini disebabkan karena dalam kalimat tersebut terdapat kata “mudik” dan kata mudik terdapat dalam kategori positif. Jadi kesimpulannya adalah penggunaan dataset yang kurang tepat menimbulkan permasalahan pada proses klasifikasi dan mengakibatkan kinerja sistem tidak maksimal.

c) Pemahaman Klasifikasi Sentimen

Kesulitan dalam memahami dataset yang diolah juga merupakan penyebab proses klasifikasi tidak berjalan secara maksimal. contoh pada opini yang pada kalimat awal memberikan opini positif atau netral sedangkan di akhir kalimat memberikan opini negatif, ataupun pada kasus sebaliknya. Kesulitan pemahaman klasifikasi ini berakibat pada performa sistem karena sistem akan mendeteksi sentiment pada kalimat yang berada di awal saja.

4. SIMPULAN

- a) Penelitian yang telah dilakukan menggunakan data opini sebanyak 1000 data dan diproses dengan menggunakan metode naïve dengan 5 kali uji data terpisah yang masing-masing data uji sebanyak 200 buah mendapat nilai akurasi keseluruhan sebanyak 56,52%.
- b) Hasil analisis sentimen menunjukkan nilai opini netral menempati peringkat pertama dengan nilai precision 66,50%. Hasil ini menunjukkan bahwa antusias masyarakat tidak terlalu tinggi dalam menanggapi kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia.
- c) Hasil analisis sentimen menunjukkan nilai opini positif menempati posisi kedua dengan nilai precision 62,28% dan masih lebih besar daripada opini negatif yang menempati posisi ketiga dengan nilai precision 46,72%. Hasil ini menunjukkan bahwa kebijakan pemerintah tentang larangan mudik tahun 2021 di Indonesia masih di nilai cukup baik bagi sebagian orang.
- d) Faktor Data yang tidak seimbang, data set yang tidak tepat serta kurangnya pemahaman dalam klasifikasi sentimen mengakibatkan proses klasifikasi tidak berjalan secara maksimal dan hanya mendapatkan nilai akurasi sebanyak 56,52%. Pada penelitian selanjutnya kualitas dataset perlu di tingkatkan agar mendapat hasil yang lebih maksimal.

DAFTAR PUSTAKA

- [1] D. Irma Sukumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. Jurnal Teknik Komputer", Vol 4, No. 1, 157-159, 2018.
- [2] C. W, Dimas Bagus, "Text Mining Pada Media Sosial *Twitter* Studi Kasus: Masa Tenang Pilkada DKI 2017 Putaran 2", Repository Universitas Jember: 26- 28, 2018.
- [3] R. Sari, and Hayuningtyas, R. Y, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website", Indonesian Journal on Software Engineering, Vol 5, No. 2, 52-58, 2019.
- [4] A. Pandhu, and W. Diki, "Analisa sentimen dan Klasifikasi Komentar Positif Pada *Twitter* dengan *Naïve Bayes Classification Sentiment Analysis and Classification of Positive*", JURNAL SWABUMI, Vol 8, No. 2, 115-117, 2020.
- [5] Mulah, Nanlir Salau and Zainon, Wan Mohd Nazme Wan, "Advances in Machine Learning Algorithms for Hate Speech Detection in Social Media", A Review, Vol 9, No. 1, 8836-8837, 2021.
- [6] Pertiwi Astri, "Sentiment Analysis of the Impact of Covid-19 on Indonesia's Economy through Social Media Using the ANN Method. Jurnal Mantik", Vol 4 No. 1, 608-611, 2020

- [7] S. Afrizal, H. N. Irmanda, N. Falih, and I. N. Isnainiyah, "Implementasi Metode *Naïve Bayes* untuk Analisis Sentimen Warga Jakarta Terhadap kehadiran Mass Rapid Transit", *J. Inform.*, Vol. 4221, 155-167, 2019
- [8] Kursimi and Lutfi, Ema Tufik, "Algoritma Data Mining". Yogyakarta: Penerbit ANDI, 2009.
- [9] Mumtaz Fairuzul, "Kupas Tuntas Metodologi Penelitian", Yogyakarta: Pustaka Diantara, 2016.