

Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier

Ni Made Ayu Juli Astari¹, Dewa Gede Hendra Divayana², Gede Indrawan³

Program Studi Ilmu Komputer Program Pascasarjana Universitas Pendidikan Ganesha
e-mail: ¹ayujuliastari@yahoo.com, ²hendra.divayana@undiksha.ac.id, ³gindrawan@undiksha.ac.id

Diajukan: 3 Agustus 2020; Direvisi: 25 September 2020; Diterima: 7 Oktober 2020

Abstrak

Virus Corona menjadi permasalahan internasional pada tahun 2020. Hal ini sangat berdampak bagi kehidupan masyarakat. Pemerintah Indonesia mengambil peran dalam menekan peningkatan jumlah penderita virus Corona dengan cara membatasi kegiatan masyarakat di luar rumah. Salah satu dampak yang signifikan dari Virus Corona adalah di sektor perekonomian. Oleh karena itu, perlu dilakukan analisis sentimen untuk menentukan kecenderungan opini masyarakat terhadap dampak virus Corona. Twitter merupakan salah satu platform yang digunakan oleh masyarakat untuk mengekspresikan kondisi terkini setelah virus Corona merambah. Tujuan dari penelitian ini adalah memperoleh analisis dokumen text untuk mendapatkan sentimen positif atau negatif masyarakat. Data yang digunakan merupakan dokumen tweet dari Twitter mengenai dampak virus Corona. Data yang terkumpul dibagi untuk digunakan sebagai data latih dan data uji proses klasifikasi. Metode yang digunakan untuk klasifikasi dalam penelitian ini adalah Metode Naive Bayes Classifier. Hasil klasifikasi dievaluasi menggunakan accuracy dan error rate dengan tujuan mengetahui keakuratan dokumen setelah diklasifikasi menjadi sentimen positif atau negatif. Hasil penelitian menunjukkan metode Naive Bayes mampu mengklasifikasi dokumen tweet dengan akurasi 67% dan error rate sebesar 33%. Percobaan dengan menggunakan 3 jumlah data berbeda (100, 200, dan 500) menghasilkan selisih nilai akurasi yang tidak jauh berbeda yaitu 0,02. Hal ini menunjukkan metode Naive Bayes untuk klasifikasi data tweet terkait dampak virus Corona menghasilkan performa yang stabil. Nilai accuracy yang diperoleh cukup baik dan penelitian selanjutnya bisa dikembangkan dengan memperhitungkan unsur semantik pada dokumen tweet.

Kata kunci: Analisis sentimen, Dampak Corona, Klasifikasi, Naive Bayes, Twitter.

Abstract

Corona virus became an international problem in 2020. This condition has a significant impact on people's lives. The Indonesian government limits outside activities with the aim of anticipating and reducing the number of sufferers of the Corona virus in Indonesia. Corona virus is very impacting on the economy in society. Therefore, sentiment analysis needs to be done to determine the tendency of public opinion towards the impact of the Corona virus. Twitter is one of the platforms used by the public to express the latest conditions after the Corona virus. The purpose of this study is to analyze text documents to get positive or negative sentiment. The data used are tweet documents from Twitter about the impact of the Corona virus. The collected data is divided into training data and test data. The method used for classification in this study is the Naive Bayes Classifier Method. The results of the classification are evaluated using accuracy and error rate with the aim of knowing the accuracy of the document after being classified as positive or negative sentiment. The results show the Naive Bayes method is able to classify tweet documents with 67% accuracy and 33% error rate. Experiments using 3 different amounts of data (100, 200, and 500) resulted in a difference in accuracy values that were not much different, i.e. 0.02. This shows that the Naive Bayes method for classifying tweet data related to the impact of the Corona virus produces stable performance. The accuracy value obtained is good and further research can be developed by considering the semantic term features in tweet documents.

Keywords: Sentiment analysis, Corona impact, Classification, Naive Bayes, Twitter.

1. Pendahuluan

Virus Corona merupakan virus yang menyerang pernapasan dan penyebarannya sangat cepat hingga membuat khawatir seluruh dunia. Virus Corona berasal dari Kota Wuhan, China. Pada awal tahun

2020 virus ini menyebar ke berbagai negara, termasuk Indonesia [1], [2]. Dewasa ini pemerintah Indonesia terus melakukan berbagai cara untuk meminimalisir orang yang terinfeksi virus Corona. Salah satunya yaitu dengan mengeluarkan kebijakan pembatasan kegiatan di luar rumah, seperti kegiatan sekolah, *work from home*, bahkan ibadah.

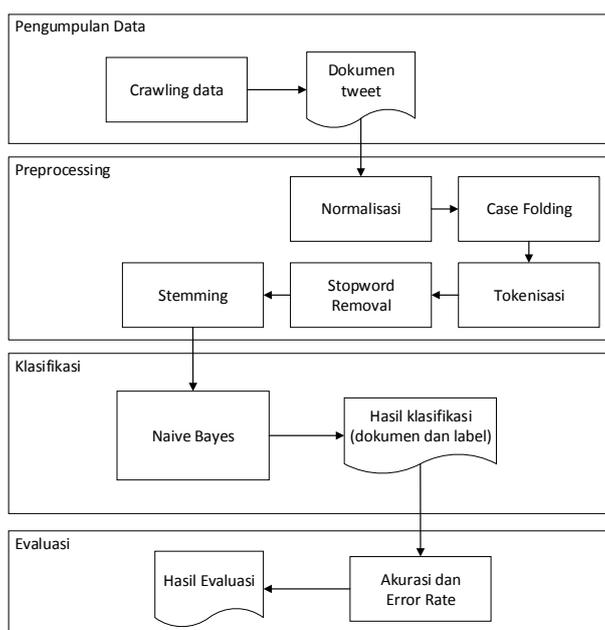
Penggunaan media sosial telah menyebar luas dan sangat pesat dalam berbagai macam lapisan masyarakat. Tidak hanya sebagai media untuk bersosialisasi, dan berkomunikasi namun juga untuk menyampaikan aspirasi dan mempresentasikan tentang apa saja yang terjadi dan dirasakan masyarakat [3]. Banyak masyarakat yang mengutarakan pendapatnya melalui media sosial Twitter dalam menanggapi wabah virus Corona [4]. Hal ini dapat dijadikan sebagai sebuah acuan untuk mengetahui sentimen masyarakat terhadap dampak yang dirasakan oleh masyarakat mengenai virus Corona.

Analisis sentimen mengolah data secara otomatis untuk memperoleh sentimen sebuah kalimat [5]. Hal ini dilakukan untuk melihat pendapat atau kecenderungan opini masyarakat apakah cenderung berpandangan positif atau negatif [4], [6], [7]. Terdapat beberapa penelitian sebelumnya yang membahas analisis sentimen dokumen teks. Penelitian [8] menggunakan analisis sentimen untuk menilai kepuasan seseorang melalui opini berdasarkan dokumen teks. Dokumen dipecah berdasarkan teks yang penyusun, kemudian masing-masing teks dinilai untuk menentukan opini memiliki sentimen positif atau negatif. Penelitian [9] menggunakan sentimen analisis untuk mengetahui tingkat kepuasan masyarakat terhadap produk dan layanan dari SAMSAT Kota Malang. Dokumen yang digunakan adalah *tweet* masyarakat yang membahas produk dan layanan tersebut. Pada penelitian tersebut, Naive Bayes menunjukkan hasil yang relatif tinggi. Penelitian [10] memanfaatkan metode Hybrid Cuckoo Search untuk menganalisis sentimen dengan menggunakan data berupa opini pada sosial media Twitter. Penelitian lainnya menggunakan Naive Bayes dan K-Means untuk menganalisis sentimen pada Twitter dengan topik pemilihan presiden 2019 [11]. Penelitian [12] melakukan sentimen analisis terhadap kasus pemilihan calon legislatif 2014 melalui data Facebook Page dengan memanfaatkan metode Naive Bayes. Metode lain yang dikembangkan untuk sentimen analisis adalah Textblob [4] dan Backpropagation [13].

Berdasarkan latar belakang tersebut, pada penelitian ini akan diajukan analisis sentimen mengenai dampak virus Corona melalui dokumen Twitter. Metode yang digunakan dalam penelitian ini adalah Naive Bayes dengan pertimbangan tingginya akurasi hasil klasifikasi pada penelitian sebelumnya. Metode ini diharapkan mampu menghasilkan nilai akurasi yang tinggi untuk klasifikasi dokumen sosial media pada kasus dampak virus Corona.

2. Metode Penelitian

Penelitian ini dilakukan dengan menggunakan beberapa modul besar, yaitu pengumpulan data, *preprocessing*, klasifikasi, dan evaluasi sistem. Rancangan secara detail bisa dilihat pada Gambar 1.



Gambar 1. Desain sistem analisis sentimen *tweet* dampak virus Corona.

2.1. Pengumpulan Data

Pengumpulan data merupakan proses untuk memperoleh dokumen yang akan digunakan sebagai *input* ke sistem. Pada penelitian ini menggunakan dokumen *tweet* yang membahas mengenai dampak virus Corona. Data yang digunakan adalah dokumen tweet masyarakat pada Twitter yang membahas tentang dampak virus Corona. Hal ini diketahui dari penggunaan tagar yang berkaitan dengan virus Corona seperti #dampakCorona, #newNormal, #dampakVirusCorona, #workFromHome, #covid19, dan #diRumahSaja. Data yang terkumpul kemudian diberikan label kelas yaitu positif atau negatif oleh para ahli. Untuk memberikan label sentimen suatu dokumen, digunakan beberapa *term* untuk menandakan indikasi sentimen dokumen tersebut. Misal *tweet* dengan sentimen positif akan mengandung kata-kata seperti semangat, semoga, tenang, hebat, dll. Sedangkan *tweet* dengan sentimen negatif akan mengandung kata-kata seperti bohong, hancur, mati, dll.

2.2. Preprocessing

Preprocessing dokumen dilakukan untuk mempersiapkan dokumen dalam teks mentah hingga siap untuk dilakukan analisis teks sesuai dengan metode yang digunakan. Tujuan dari proses ini adalah untuk meningkatkan akurasi klasifikasi ulasan perangkat lunak [14]. Beberapa teknik yang digunakan untuk *preprocessing* adalah normalisasi, *case folding*, tokenisasi, *stopwords removal*, dan *stemming* [15], [16]. Penjelasan *preprocessing* data Twitter yang digunakan pada penelitian ini ditampilkan pada Tabel 1.

Tabel 1. Penjelasan *preprocessing* data.

Proses	Penjelasan
Normalisasi	Membersihkan fitur-fitur yang tidak diperlukan dalam pengambilan data, seperti URL, <i>username</i> , tanggal, dan lain-lain.
<i>Case folding</i>	Mengonversi huruf kapital pada teks ulasan ke dalam huruf kecil. Tahap ini bertujuan agar dokumen teks ulasan memiliki bentuk standar
Tokenisasi	Teks dipilah ke dalam <i>token</i> , yaitu bagian terkecil teks yang dapat dimengerti dalam bentuk kata.
<i>Stopwords removal</i>	Menghilangkan kata-kata yang sering muncul dalam dokumen, akan tetapi tidak memberikan arti yang signifikan dan tidak berpengaruh secara semantik.
<i>Stemming</i>	Menghilangkan imbuhan-imbuhan yang terdapat pada sebuah kata.

2.3. Klasifikasi Naive Bayes

Sebuah dokumen dapat diklasifikasi ke dalam kategori tertentu berdasarkan kata-kata yang terkandung dalam dokumen tersebut. Kumpulan kata pada dokumen dapat digunakan untuk menentukan kategorinya karena kata-kata tersebut dinilai memiliki makna tertentu [17]–[19]. Salah satu algoritma yang digunakan untuk klasifikasi teks adalah Naive Bayes. Algoritma ini menggunakan perhitungan probabilitas dan statistik dalam pengoperasiannya [11], [20]. Metode klasifikasi Naive Bayes digunakan untuk mengambil keputusan dengan melakukan prediksi suatu kasus berdasarkan hasil dari klasifikasi yang telah diperoleh. Pada penelitian ini, metode Naive Bayes digunakan untuk menentukan sentimen dari dokumen *tweet*. Proses ini berdasarkan metode probabilitas dan statistik [5], [11], [12]. Notasi klasifikasi Naive Bayes ditampilkan pada Persamaan 1.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Di mana:

A: Hipotesis data merupakan suatu *class* spesifik.

B: Data dengan kelas yang masih belum diketahui.

P(A|B): Probabilitas hipotesis berdasar kondisi.

P(A): Probabilitas hipotesis.

P(B|A): Probabilitas berdasarkan kondisi pada hipotesis.

P(B): Probabilitas B.

2.4. Evaluasi

Evaluasi terhadap metode dilakukan untuk mengetahui performa pendekatan yang diusulkan. Pada penelitian ini, pendekatan yang diusulkan dievaluasi menggunakan perhitungan *accuracy* dan *error rate*. Pengukuran dilakukan dengan membandingkan hasil klasifikasi dari pendekatan yang diusulkan dengan data *ground truth*.

Tabel 2. Kondisi *confusion matrix*.

Kondisi		Nilai Sebenarnya	
		Relevan	Tidak Relevan
Nilai Prediksi	Tampil	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
	Tidak Tampil	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Perhitungan *accuracy* dan *error rate* memanfaatkan metode *confusion matrix* [21]. Kondisi *confusion matrix* ditampilkan pada Tabel 2. Metode ini terdiri dari *True Positive (TP)* yaitu jumlah data kelas positif yang diklasifikasikan sebagai kelas positif dan *True Negative (TN)* merupakan jumlah data kelas negatif yang diklasifikasikan sebagai kelas negatif. Sedangkan *False Positive (FP)* merupakan jumlah data kelas negatif yang diklasifikasikan sebagai kelas positif, dan *False Negative (FN)* adalah jumlah data kelas positif yang diklasifikasikan sebagai kelas negatif. Perhitungan *accuracy* menggunakan Persamaan 2 dan *error rate* dihitung menggunakan Persamaan 3.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \tag{2}$$

$$error\ rate = \frac{FP+FN}{TP+FP+TN+FN} \times 100\% \tag{3}$$

2.5. Skenario Pengujian

Pengujian dilakukan dengan menggunakan jumlah data yang bervariasi. Penelitian ini menggunakan tiga variasi jumlah data, yaitu 100, 200, dan 500 data. Hal ini bertujuan untuk mengetahui akurasi dari metode yang diterapkan pada jumlah data yang berbeda-beda. Pengujian akan dilakukan sebanyak 10 kali pada masing-masing varian jumlah data untuk memperoleh rata-rata hasil pengujian. Pada masing-masing variasi jumlah data, data akan dibagi menjadi data latih dan data uji. Rasio yang digunakan adalah 75% data latih dan 25% digunakan sebagai data uji.

3. Hasil dan Pembahasan

Hasil penelitian ditampilkan sesuai dengan urutan proses penelitian. Urutan proses pada penelitian ini adalah pengumpulan data, *preprocessing* data, klasifikasi, dan evaluasi.

3.1. Data

Data yang terkumpul sebanyak 796 data. Hasil penentuan label menghasilkan 290 *tweet* dengan sentimen positif dan 506 *tweet* dengan sentimen negatif. Data yang sudah diberi label sentimen akan digunakan sebagai *ground truth* pada penelitian ini. Contoh dokumen *tweet* yang dikumpulkan ditampilkan pada Tabel 3.

Tabel 3. Contoh dokumen *tweet*.

Label	Dokumen
Positif	arif
	@andika_arf
	May 16
	Semoga New Normal berdampak baik bagi kehidupan kita kali ini. #Covid19 #RakyatDukungNewNormal
Negatif	Ferry Gunawan
	@ferrygunawannn_
	May 16
	Covid19 membuat perekonomian kita hancur!!! Fiks masa depan jadi suram! #Covid19

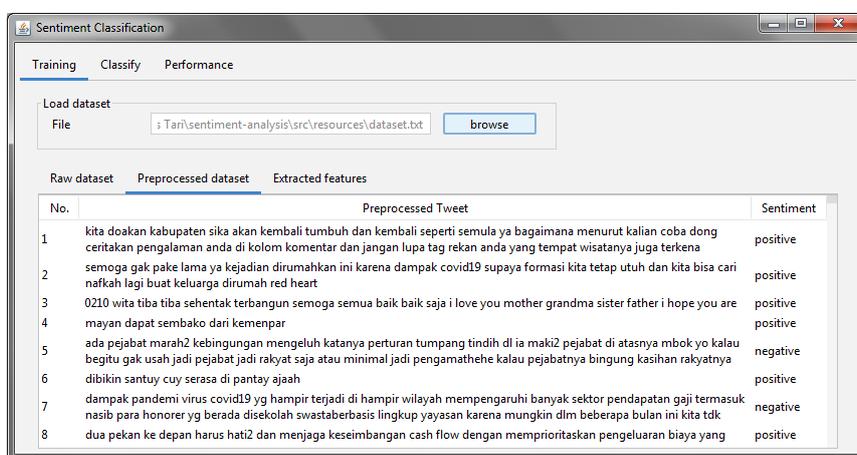
3.2. Preprocessing Data

Penelitian ini menggunakan beberapa teknik untuk *preprocessing* data, yaitu normalisasi, *case folding*, tokenisasi, *stopwords removal*, dan *stemming*. Contoh *preprocessing* data Twitter yang digunakan pada penelitian ini ditampilkan pada Tabel 4.

Tabel 4. Contoh *preprocessing* data.

Proses	Contoh
Data mentah	arif @andika_arf May 16 Semoga New Normal berdampak baik bagi kehidupan kita kali ini. #Covid19 #RakyatDukungNewNormal
Normalisasi	Semoga New Normal berdampak baik bagi kehidupan kita kali ini
Case folding	semoga new normal berdampak baik bagi kehidupan kita kali ini
Tokenisasi	“semoga”, “new”, “normal”, “berdampak”, “baik”, “bagi”, “kehidupan”, “kita”, “kali”, “ini”
Stopwords removal	“semoga”, “new”, “normal”, “berdampak”, “baik”, “bagi”, “kehidupan”, “kali”
Stemming	“semoga”, “new”, “normal”, “dampak”, “baik”, “bagi”, “hidup”, “kali”

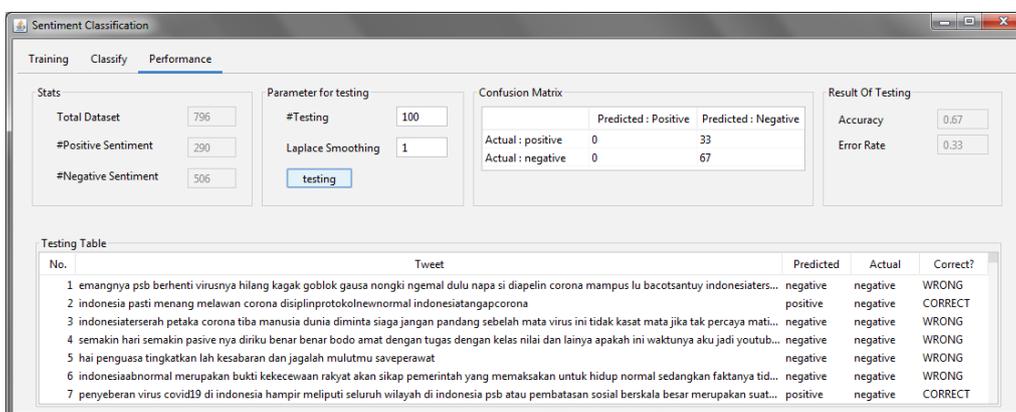
Semua teknik *preprocessing* dilakukan setelah data dikumpulkan. Setelah *preprocessing* dilakukan, maka semua data siap digunakan untuk analisis. Langkah selanjutnya adalah membagi data keseluruhan menjadi data latih dan data uji. Jumlah data uji ditentukan sebelum proses klasifikasi dilakukan. Tahap *preprocessing* pada program dapat dilihat pada Gambar 2.



Gambar 2. Tahap *preprocessing* data pada program.

3.3. Klasifikasi

Tahap klasifikasi dilakukan setelah tahap *preprocessing* selesai. Proses ini diawali dengan menentukan jumlah data yang akan digunakan. Selanjutnya dipilih sejumlah *tweet* secara acak sesuai dengan jumlah yang ditentukan. Setelah proses klasifikasi selesai dijalankan, maka akan tampil sejumlah *tweet* yang ditentukan sebagai data uji disertai dengan hasil prediksi sentimen. Selain itu, *tweet* akan disertai label aktual, yaitu label yang diberikan secara manual oleh tenaga ahli. Selanjutnya akan ditentukan apakah hasil prediksi dan label aktual sesuai atau tidak. Tampilan program klasifikasi ditampilkan pada Gambar 3.



Gambar 3. Tahap klasifikasi pada program.

Skenario pengujian menggunakan tiga varian jumlah data berbeda. Hal ini bertujuan untuk mengetahui performa metode pada masing-masing jumlah data. Pada masing-masing varian, dilakukan sepuluh kali pengujian untuk mengetahui rata-rata hasil pengujian. Hasil pengujian berdasarkan skenario pengujian ditampilkan pada Tabel 5.

Tabel 5. Hasil pengujian klasifikasi.

Jumlah Data	Percobaan ke-	Accuracy	Error Rate
100	1	0.72	0.28
	2	0.73	0.27
	3	0.65	0.35
	4	0.63	0.37
	5	0.69	0.31
	6	0.64	0.36
	7	0.67	0.33
	8	0.65	0.35
	9	0.66	0.34
	10	0.70	0.3
		Rata-rata	0.67
200	1	0.63	0.37
	2	0.69	0.32
	3	0.61	0.40
	4	0.64	0.37
	5	0.68	0.33
	6	0.64	0.36
	7	0.63	0.37
	8	0.67	0.34
	9	0.66	0.34
	10	0.69	0.32
		Rata-rata	0.65
500	1	0.62	0.38
	2	0.65	0.35
	3	0.65	0.35
	4	0.65	0.35
	5	0.64	0.36
	6	0.67	0.33
	7	0.65	0.35
	8	0.67	0.33
	9	0.66	0.34
	10	0.67	0.33
		Rata-rata	0.65

Berdasarkan hasil pengujian pada Tabel 5, dapat dilihat hasil rata-rata *accuracy* pada masing-masing varian data tidak jauh berbeda yaitu 0,67 pada jumlah data 100 dan 0,65 pada jumlah data 200 dan 500. Peningkatan jumlah data tidak berbanding lurus dengan jumlah *error rate* karena nilai *error* terendah justru diperoleh pada percobaan 100 data. Hal ini menunjukkan jumlah data tidak berpengaruh signifikan terhadap hasil klasifikasi Naive Bayes untuk data *tweet* mengenai dampak virus Corona.

Accuracy tertinggi yaitu 0,73 yang diperoleh pada percobaan kedua dengan menggunakan 100 data. Sedangkan nilai *accuracy* terendah yaitu 0,61 yang diperoleh pada percobaan ketiga menggunakan 200 data. Walaupun masing-masing varian jumlah data menghasilkan performa berbeda, namun perbedaan skor yang dihasilkan tidak jauh berbeda. Perbedaan nilai tertinggi dan terendah tidak terlalu jauh berbeda yaitu 0,12. Sedangkan selisih *accuracy* rata-rata untuk masing-masing varian data adalah 0,02. Hal ini membuktikan sistem yang dibangun yaitu analisis sentimen menggunakan metode Naive Bayes untuk dokumen *Twitter* menghasilkan performa yang stabil dengan menggunakan jumlah data dengan rentang 100 sampai dengan 500 data.

Hasil klasifikasi pada penelitian ini masih kurang maksimal. Setelah dilakukan pengamatan, hal ini dikarenakan kalimat pada dokumen *tweet* yang tidak terstruktur dan banyak mengandung kata yang tidak formal. Hal ini mendukung permasalahan yang dibahas pada penelitian [14]. Selain itu, terlalu luasnya kosa kata yang digunakan oleh pengguna Twitter juga menjadi penyebab kurang maksimalnya hasil *accuracy*. Hal ini berdampak pada sulitnya mengindikasikan sentimen *tweet* tersebut. Untuk penelitian lebih lanjut, kekurangan ini mungkin bisa diatasi dengan memperhatikan unsur semantik dari dokumen yang dianalisis, seperti sinonim, homonim, dan polisemi.

4. Kesimpulan

Penelitian ini mengembangkan analisis sentimen pada dokumen Twitter mengenai dampak virus Corona. Sentimen yang digunakan adalah sentimen positif dan negatif. Proses analisis memanfaatkan metode Naive Bayes untuk klasifikasi dokumen. Data pengujian pada penelitian ini sejumlah 796 dokumen *tweet* dengan 290 sentimen positif dan 506 sentimen negatif. Pengujian dilakukan menggunakan tiga varian jumlah data berbeda, yaitu 100, 200, dan 500 data. Hasil pengujian tertinggi diperoleh dengan menggunakan 100 data dengan *accuracy* sebesar 67% dan *error rate* sebesar 33%. Percobaan dengan menggunakan 3 jumlah data berbeda (100, 200, dan 500) menghasilkan selisih nilai *accuracy* yang tidak jauh berbeda yaitu 0,02. Hal ini menunjukkan metode Naive Bayes untuk klasifikasi data *tweet* terkait dampak virus Corona menghasilkan performa yang stabil.

Beberapa permasalahan yang ditemukan adalah kalimat yang tidak terstruktur dan kata yang tidak formal pada dokumen. Selain itu, kosa kata pada dokumen yang terlalu luas sehingga klasifikasi kurang maksimal. Saran untuk penelitian selanjutnya adalah memperhitungkan unsur semantik pada dokumen.

Daftar Pustaka

- [1] C. I. Burhanuddin and M. N. Abdi, "Ancaman Krisis Ekonomi Global dari Dampak Penyebaran Virus Corona (COVID-19)," *AkMen J. Ilm.*, vol. 17, pp. 90–98, 2020.
- [2] N. R. Yunus and A. Rezki, "Kebijakan Pemberlakuan Lock Down Sebagai Antisipasi Penyebaran Corona Kebijakan Pemberlakuan Lockdown Sebagai Antisipasi Penyebaran Corona Virus Covid-19 □," *Salam J. Sos. dan Budaya Syar-i*, vol. 7, pp. 227–238, 2020, doi: 10.15408/sjsbs.v7i3.15083.
- [3] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment Analysis in Twitter," *Proc. 11th Int. Work. Semant. Eval.*, 2017.
- [4] C. Kaur and A. Sharma, "Twitter Sentiment Analysis on Coronavirus using Textblob," *EasyChair*, 2020.
- [5] F. Priyono, S. Kanti, I. D. I, I. Amirulloh, E. S. P, and D. Rosiyadi, "Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda Naive Bayes," *J. Electr. Electron. Eng.*, vol. 1, no. 2, pp. 38–45, 2016.
- [6] A. Rasool, R. Tao, K. Marjan, and T. Naveed, "Twitter Sentiment Analysis : A Case Study for Apparel Brands Twitter Sentiment Analysis : A Case Study for Apparel Brands," *J. Phys. Conf. Ser.*, vol. 1176, p. 022015, 2019, doi: 10.1088/1742-6596/1176/2/022015.
- [7] S. Elbagir, "Twitter Sentiment Analysis Based on Ordinal Regression," *IEEE Access* 7, vol. 7, pp. 163677–163685, 2019, doi: 10.1109/ACCESS.2019.2952127.
- [8] S. R. Wardhana, D. Purwitasari, and S. Rochimah, "Analisis Sentimen Pada Review Pengguna Aplikasi Mobile Untuk Evaluasi Faktor Usability," *J. Sist. dan Inform.*, vol. 11, pp. 128–136, 2016.
- [9] I. F. Rozi, E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter (Studi Kasus SAMSAT Kota Malang)," *J. Inform. Polinema*, vol. 4, pp. 149–154, 2018.
- [10] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter Sentiment Analysis using Hybrid Cuckoo Search Method," *Inf. Process. Manag.*, vol. 53, pp. 764–779, 2017, doi: 10.1016/j.ipm.2017.02.004.
- [11] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *J. Eksplora Inform.*, vol. 9, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [12] A. Rachmat, U. Kristen, D. Wacana, Y. Lukito, U. Kristen, and D. Wacana, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *J. Inform. dan Sist. Inf.*, no. November, pp. 26–34, 2017.
- [13] R. Habibi and D. B. Setyohadi, "Analisis Sentimen pada Twitter Mahasiswa Menggunakan Metode Backpropagation," *J. Inform.*, vol. 12, no. 1, pp. 103–109, 2016.
- [14] I. M. M. Parwita and D. Siahaan, "Classification of Mobile Application Reviews using Word Embedding and Convolutional Neural Network," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 10, no. 1, pp. 1–8, 2019.
- [15] J. Prasetyo and O. Siahaan, "Klasifikasi Ulasan Aplikasi pada Toko Aplikasi Bergerak dengan Memanfaatkan Issue Tracker Github," *JUTI J. Ilm. Teknol. Inf.*, vol. 15, pp. 206–213, 2017.
- [16] A. F. Anees, A. Shaikh, A. Shaikh, and S. Shaikh, "Survey Paper on Sentiment Analysis : Techniques and Challenges," *EasyChair*, pp. 2516–2314, 2020.
- [17] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How Can I Improve My App ? Classifying User Reviews for Software Maintenance and Evolution," *2015 IEEE Int. Conf. Softw. Maint. Evol.*, pp. 281–290, 2015, doi: 10.1109/ICSM.2015.7332474.
- [18] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. Gall, "ARdoc : App Reviews Development Oriented Classifier," *Proc. 2016 24th ACM SIGSOFT Int. Symp. Found.*

-
- Softw. Eng.*, pp. 1023–1027, 2016.
- [19] K. Giannakopoulos, “Informative vs . Non-informative Short Message Detection in Social Networks,” *Int. Conf. Big Data Comput. Commun. Inf.*, pp. 165–171, 2017, doi: 10.1109/BIGCOM.2017.55.
- [20] A. R. Chrismanto and Y. Lukito, “Identifikasi Komentar Spam Pada Instagram,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 3, p. 219, 2017, doi: 10.24843/LKJITI.2017.v08.i03.p08.
- [21] S. Dey, S. Wasif, D. S. Tonmoy, and S. Sultana, “A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews,” *2020 Int. Conf. Contemp. Comput. Appl.*, no. May, pp. 217–220, 2020, doi: 10.1109/IC3A48958.2020.233300.