

Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial *Twitter*

Fajar Sodik Pamungkas^{a,*}, Iqbal Kharisudin^a

^a Universitas Negeri Semarang, Semarang 50229, Indonesia

*Alamat Surel: fajarswodik@students.unnes.ac.id

Abstrak

Pandemi Covid-19 sangat berdampak diberbagai sektor kehidupan masyarakat, keadaan yang memaksa masyarakat untuk melaksanakan *physical distancing* merubah pola hidup masyarakat. Hal tersebut membuat berbagai pendapat atau tanggapan masyarakat terhadap pandemi Covid-19 yang dituangkan dalam media sosial. Untuk mengetahui sentimen tanggapan masyarakat tersebut perlu dilakukan analisis sentimen dengan algoritma machine learning. Pada penelitian ini dilakukan analisis sentimen tanggapan masyarakat Indonesia terhadap pandemi Covid-19 pada media sosial Twitter menggunakan algoritma *Support Vector Machine* (SVM), *Naive Bayes*, dan *K-Nearest Neighbor*, yang kemudian ketiga algoritma tersebut dibandingkan mana yang paling baik untuk mengklasifikasikan data tanggapan. Berdasarkan tingkat rata-rata akurasi dengan menggunakan evaluasi model *10-Fold Cross Validation*, diperoleh kesimpulan bahwa algoritma SVM memiliki akurasi yang lebih tinggi daripada *Naive Bayes* dan KNN dengan rata-rata akurasinya sebesar 90,01% pada SVM dengan kernel *linear*, 79,20% pada *Naive Bayes* dengan jumlah *laplace* adalah 1, dan 62,10% pada KNN dengan jumlah K adalah 20 dan menggunakan kernel *optimal*.

Kata kunci:

Covid-19, Sentimen, SVM, *Naive Bayes*, KNN.

© 2021 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Covid-19 membuat pemerintah menerapkan kebijakan *physical distancing*. *Physical distancing* diartikan sebagai jaga jarak atau mengurangi kontak fisik secara langsung dan selalu membersihkan diri dengan tata cara yang sudah ditentukan (WHO, 2020). Penerapan *physical distancing* ini berdampak pada seluruh sektor pemerintahan, selain itu juga berdampak pada pola hidup masyarakat yang berubah. Dengan fenomena tersebut, masyarakat banyak yang memberikan berbagai macam pendapat baik itu pujian atau keluh kesah yang dipublikasikan di berbagai media sosial, salah satunya adalah Twitter. Tanggapan tersebut perlu digolongkan ke dalam bentuk tanggapan positif atau negatif menggunakan analisis sentimen. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tesktual secara otomatis untuk mendapat informasi (Pang & Lee, 2008). Dalam analisis sentimen diperlukan adanya klasifikasi, klasifikasi digunakan untuk mengelompokkan tanggapan mana yang memiliki sentimen positif dan sentimen negatif.

Menurut Fithriasari *et al.* (2020) Terdapat beberapa algoritma dalam klasifikasi untuk mengklasifikasikan teks antara lain *support vector machine*, *naive bayes*, serta *K-nearest neighbor*. Pada dasarnya *support vector machine* bekerja dengan mencari *hyperplane* terbaik untuk membagi kelas, *naive bayes* bekerja dengan cara mencari nilai probabilitas bersyarat terbesar dari masing-masing kelas, sedangkan *K-nearest neighbor* bekerja dengan mencari sejumlah k pola yang terdekat.

Penelitian terbaru tentang analisis sentimen terhadap pandemi Covid-19 pada media sosial Twitter di Arab Saudi yang dilakukan oleh Alhajji *et al.* (2020) menyimpulkan bahwa *naive bayes* dapat digunakan untuk klasifikasi teks dengan akurasi yang cukup tinggi yaitu sebesar 89%. Namun dalam penelitian

To cite this article:

Pamungkas, F. S., & Kharisudin, I. (2021). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial *Twitter*. *PRISMA, Prosiding Seminar Nasional Matematika* 4, 628-634

Fithriasari, Mayasari, Iriawan, & Winahju, (2020) menyimpulkan bahwa algoritma *support vector machine* memiliki nilai AUC yang lebih baik daripada *naive bayes* pada studi kasus evaluasi pemerintah Surabaya pada media sosial Twitter. Sedangkan jurnal penelitian Huq, Ali, & Rahman, (2017) dalam pengklasifikasian data tweet menyimpulkan bahwa algoritma *K-nearest neighbor* memiliki tingkat akurasi tertinggi dibandingkan dengan algoritma *support vector machine*.

Algoritma klasifikasi memiliki keunggulan dan kelemahannya masing-masing dalam mengklasifikasikan data berbentuk teks, diantara *support vector machine*, *naive bayes*, serta *K-nearest neighbor* algoritma mana yang paling tepat mengklasifikasikan data tanggapan masyarakat Indonesia pada media sosial twitter. Selain itu bagaimana visualisasi hasil klasifikasi data tersebut.

Penelitian ini akan memvisualisasikan dan membandingkan analisis sentimen tanggapan masyarakat terhadap pandemi Covid-19 pada media sosial Twitter dengan algoritma klasifikasi *support vector machine*, *naive bayes*, serta *K-nearest neighbor*

2. Metode

Data yang digunakan dalam penelitian ini adalah data tanggapan masyarakat Indonesia terhadap pandemi Covid-19 melalui media sosial *Twitter* dengan rentang waktu 1 Maret 2020-31 Juli 2020 dimana setiap bulannya diambil sebanyak 2000 data tanggapan. Jadi total data yang digunakan dalam penelitian ini adalah 10000 data tanggapan. Metode pengumpulan data yang digunakan adalah dengan teknik *crawling text*. *Crawling text* adalah teknik pengumpulan informasi berupa *text* yang ada dalam web (Saputra, 2017). *Crawling text* dilakukan langsung pada media sosial *Twitter* dengan menggunakan *package* pada *python* yaitu *Twint* terhadap pencarian “Covid-19” dengan bahasa Indonesia. Data yang telah diperoleh kemudian akan di analisis sentimen menggunakan beberapa algoritma klasifikasi, adapun proses tahap analisis yang dilakukan adalah sebagai berikut.

2.1. Text Mining

Darmanik, (2014) berpendapat *text mining* adalah proses menambang teks untuk menemukan suatu informasi yang berguna dalam koleksi dokumen teks sehingga diperoleh pola, tren, atau keterhubungan antar teks. Hasil *text mining* yang berupa teks, yang pada umumnya terdapat *noise* yang tinggi maka perlu dilakukan *text preprocessing*.

Hal pertama yang dilakukan dalam *text preprocessing* adalah *case folding* yaitu penyamaan *case* dalam dokumen menjadi *lowercase*, kemudian dilakukan *spell normalization* yaitu perbaikan kata-kata dalam penyingkatan, lalu dilakukan *tokenizing* yaitu pemecahan kalimat menjadi kata per kata, setelah itu dilakukan *filtering* yaitu mengeluarkan kata yang tidak penting, dan yang terakhir dilakukan adalah *Stemming* yaitu mencari kata dasar pada setiap kata.

2.2. Analisis Sentimen

Analisis sentimen adalah bagian dari ilmu *data mining* yang memiliki tujuan untuk menganalisis dan mengekstrak data tekstual yang berupa pendapat, evaluasi, sikap, emosi, penilaian, dan sentimen seseorang terhadap suatu barang, orang, organisasi, dan masalah. Pada penelitian ini proses pelabelan sentimen pada tanggapan dilakukan dengan menghitung jumlah kata positif dan negatif pada setiap tanggapan, apabila jumlah kata positif lebih banyak maka termasuk tanggapan positif, namun apabila jumlah kata negatif lebih banyak termasuk tanggapan negatif.

2.3. Klasifikasi

Ada beberapa algoritma klasifikasi yang digunakan dalam penelitian ini, yang pertama adalah algoritma *support vector machine*. *Support vector machine* (SVM) merupakan salah satu algoritma klasifikasi yang sering digunakan untuk analisis sentimen. Tujuan dari metode *Support vector machine* (SVM) adalah menemukan *hyperplane* yang paling optimum (Suyanto, 2018). *Hyperplane* yang optimum adalah sebuah *hyperplane* yang jika tepat berada di tengah-tengah kedua kelas sehingga memiliki jarak paling jauh ke data-data terluar di kedua kelas.

Algoritma selanjutnya adalah *naive bayes*. *Naive bayes* adalah metode *machine learning* yang menggunakan teorema kuno warisan abad ke-18 yang ditemukan oleh Thomas Bayes (Suyanto, 2018). Dalam teori tersebut suatu probabilitas bersyarat dinyatakan dengan persamaan sebagai berikut.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dimana X adalah bukti, H adalah hipotesis, $P(H|X)$ adalah probabilitas bahwa hipotesis H benar untuk bukti X atau dengan kata lain $P(H|X)$ merupakan probabilitas *posterior* H dengan syarat X . Namun dalam *machine learning* X adalah sebuah *tuple* atau objek data, H adalah hipotesis dan atau dugaan bahwa *tuple* X adalah kelas C (Suyanto, 2018).

Algoritma klasifikasi terakhir yang digunakan dalam penelitian ini adalah *K-nearest neighbor*. *K-nearest neighbor* atau yang biasa disebut dengan KNN merupakan algoritma klasifikasi yang mengelompokkan data baru berdasarkan jarak data baru ke beberapa data atau tetangga terdekat (Santosa, 2007). Algoritma ini bekerja dengan cara mencari sejumlah k pola (diantara semua pola latih yang ada di semua kelas) yang terdekat dengan pola masukan, kemudian menentukan kelas keputusan berdasarkan jumlah pola terbanyak diantara k pola tersebut (Suyanto, 2018).

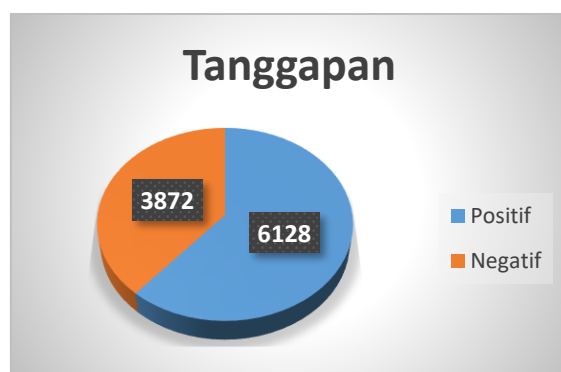
Ketiga algoritma klasifikasi tersebut akan dibandingkan dengan melihat performa masing-masing algoritma dalam mengklasifikasikan data tanggapan menggunakan teknik evaluasi *K-fold cross validation*. *K-fold cross validation* merupakan metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak (Pipit Pitria, 2014).

2.4. Visualisasi

Visualisasi data yang dilakukan menggunakan *wordcloud* dan asosiasi kata. *Wordcloud* merupakan sebuah sistem yang memunculkan visualisasi kata-kata dengan memberikan penekanan pada frekuensi kemunculan kata terkait dalam wacana tertulis (Qeis, 2015). Asosiasi kata adalah teknik yang digunakan untuk menemukan asosiasi (hubungan) dari beberapa kombinasi item dalam suatu data. Perhitungan asosiasi dilakukan pada kata yang paling sering muncul pada dokumen (Pratiwi & Widodo, 2017).

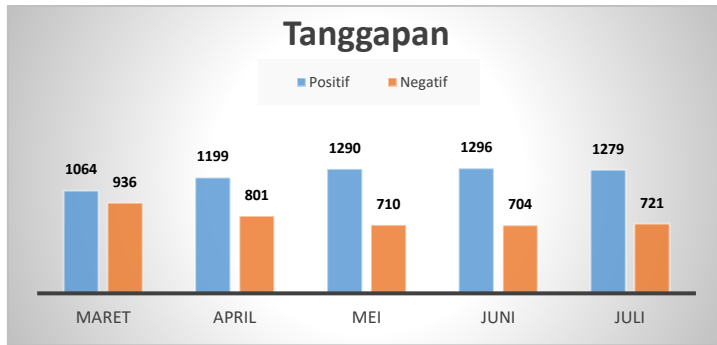
3. Hasil dan Pembahasan

Data tanggapan yang telah diperoleh kemudian dilakukan *text preprocessing* agar *noise* pada data text berkurang. Kemudian data tanggapan dikategorikan kedalam dua kategori yaitu tanggapan positif dan tanggapan negatif. pengkategorian didasarkan pada jumlah kata positif dan negatif pada tiap tanggapan, apabila kata positif lebih banyak maka tanggapan akan dikategorikan sebagai tanggapan positif, begitu juga pada tanggapan negatif. Secara keseluruhan, pengklasifikasian data tanggapan yang diperoleh dapat dilihat pada gambar berikut.



Gambar 1. Jumlah tanggapan masyarakat

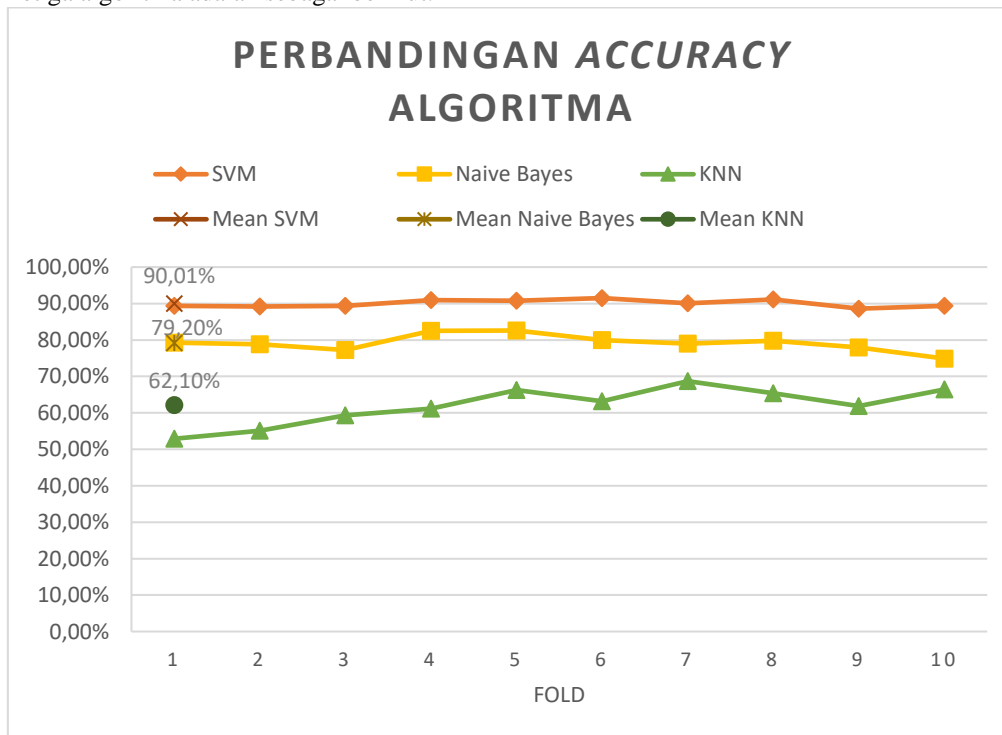
Berdasarkan dari Gambar 1 dari total keseluruhan 10000 data tanggapan 6128 diantaranya merupakan tanggapan masyarakat yang memiliki sentimen positif terhadap pandemi *Covid-19*, sedangkan sisanya 3875 merupakan tanggapan masyarakat yang memiliki sentimen negatif terhadap pandemi *Covid-19*. Untuk perbandingan tanggapan pada masing-masing bulan dapat dilihat pada Gambar berikut.



Gambar 2. Jumlah tanggapan masyarakat berdasarkan bulan

Berdasarkan Gambar 2 dapat diketahui bahwa pada masing-masing bulan jumlah tanggapan yang memiliki sentimen positif lebih besar daripada jumlah tanggapan yang memiliki sentimen negatif, artinya sebagian besar masyarakat pengguna *Twitter* menanggapi pandemi *Covid-19* dengan positif, seperti dengan memberikan informasi pencegahan atau dengan memberikan semangat kepada tenaga medis. Diantara 2000 data pada tiap bulan didapat bahwa bulan Juni merupakan bulan dengan tanggapan positif terbesar dan negatif terkecil, hal ini terjadi dikarenakan seiring meningkatnya jumlah kasus pada bulan Juni.

Setelah dikategorikan kemudian dilakukan klasifikasi menggunakan algoritma *support vector machine*, *naive bayes*, dan *K-nearest neighbor* dengan validasi model *10-fold cross validation*. Perbandingan kinerja ketiga algoritma adalah sebagai berikut.



Gambar 3. Diagram perbandingan *accuracy* algoritma

Berdasarkan Gambar 3 Algoritma pertama yang digunakan untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap *Covid-19* pada media sosial *Twitter* adalah *support vector machine*. Setelah dilakukan klasifikasi *support vector machine* menggunakan kernel *linear* dengan evaluasi model *10-fold cross validation* diperoleh rata-rata *accuracy* adalah 0,901 atau 90,1%. Hasil tersebut merupakan hasil yang tinggi, oleh karena itu algoritma *support vector machine* sangat cocok untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap pandemi *Covid-19*. Hal ini dikarenakan algoritma *support vector machine* mempunyai keunggulan dapat mengolah data berdimensi tinggi, tanpa mengalami penurunan performa (Purnamawan, 2015). Keunggulan ini sangat cocok jika algoritma tersebut mengolah data yang kita pakai yaitu data teks yang merupakan data yang berdimensi sangat tinggi.

Algoritma selanjutnya adalah algoritma *naive bayes*. Pada algoritma ini digunakan nilai *laplace correction* adalah 1 dengan evaluasi modelnya adalah *10-fold cross validation*. Kemudian diperoleh hasil rata-rata *accuracy* adalah 0,792 atau 79,2%. Nilai *accuracy* ini terbilang cukup tinggi karena algoritma *naive bayes* memiliki kelebihan algoritma yang sederhana dan kompleksitas perhitungan yang rendah, akan tetapi memiliki kelemahan sifat independensi dari fitur *naive bayes* tidak dapat selalu diterapkan sehingga akan berpengaruh pada tingkat akurasi perhitungan (Suprianto, 2020).

Algoritma yang terakhir untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap pandemi *Covid-19* pada media sosial *Twitter* adalah *K-nearest neighbor*. Nilai parameter *k* yang digunakan adalah 20 dengan kernelnya adalah kernel *optimal*. Setelah dilakukan klasifikasi dengan evaluasi model *10-fold cross validation* diperoleh nilai rata-rata *accuracy* adalah 0,621 atau 62,1%. Nilai ini merupakan nilai yang cukup rendah dikarenakan karena algoritma *K-nearest neighbor* sangat sensitif terhadap fitur-fitur yang kurang relevan, data berderau, data pencilan, dan ukuran ketetanggaan *k*. Selain itu perlu algoritma ini memiliki kompleksitas waktu dan memori yang tinggi setiap melakukan klasifikasi.

Dari ketiga algoritma tersebut maka dapat diketahui bahwa algoritma *support vector machine* merupakan algoritma yang paling cocok untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap pandemi *Covid-19* pada media sosial *Twitter* dibandingkan dengan algoritma *naive bayes* dan *K-nearest neighbor*. Hal ini sesuai dengan penelitian terdahulu yang dilakukan oleh Fithriasari *et al.* (2020) yang menyatakan bahwa algoritma *support vector machine* memiliki *accuracy* lebih tinggi dibandingkan *naive bayes* untuk kasus data evaluasi kinerja pemerintah Surabaya, serta penelitian yang dilakukan oleh Mardiana *et al.* (2019) yang menyatakan bahwa algoritma *support vector machine* memiliki *accuracy* yang lebih tinggi dibanding *K-nearest neighbor* untuk kasus data *Twitter* usaha waralaba. Visualisasi data tanggapan menggunakan *wordcloud* adalah sebagai berikut.



Gambar 4. (a) *wordcloud* tanggapan positif (b) *wordcloud* tanggapan negatif

Pada gambar 4 (a) merupakan *wordcloud* dari keseluruhan tanggapan positif masyarakat Indonesia mengenai pandemi *Covid-19* yang diperoleh pada bulan Maret sampai dengan Juli 2020 melalui media sosial *Twitter*, sedangkan (b) merupakan *wordcloud* dari tanggapan negatif masyarakat Indonesia mengenai pandemi *Covid-19* yang diperoleh pada bulan Maret sampai dengan Juli 2020 melalui media sosial *Twitter*. Berdasarkan *wordcloud* tersebut dapat diketahui bahwa perbincangan positif masyarakat Indonesia mengenai *Covid-19* banyak dikaitkan dengan kata “positif”, “pandemi”, “pasien”, “cegah”, dan “sehat”. Hal itu menunjukkan adanya ajakan untuk mencegah dan hidup sehat disaat pandemi pada tanggapan positif. Sedangkan pada tanggapan negatif perbincangan sering dikaitkan dengan kata “orang”, “sebar”, “pandemi”. “masyarakat”, “rumah”. Hal itu menunjukkan adanya ketakutan masyarakat terhadap penyebaran *Covid-19*.

Asosiasi kata digunakan untuk mengetahui kata apa yang paling sering dikaitkan dengan kata yang muncul pada *wordcloud*. Perhitungan asosiasi menggunakan persamaan korelasi. Semakin besar nilai korelasinya semakin besar pula kemungkinan kata tersebut disebutkan secara bersama-sama. Asosiasi kata dilakukan pada sebagian kata yang sering diperbincangkan yang telah diketahui pada hasil *wordcloud*.

Tabel 1. Asosiasi Kata

	Pasien (positif)	Warga (positif)	Protokol (negatif)	Masker (negatif)			
positif	0,28	Beras	0,23	sehat	0,58	jarak	0,39
tambah	0,24	dampak	0,20	disiplin	0,37	jaga	0,31

demam	0,17	Desa	0,19	adaptasi	0,31	rajin	0,23
isolasi	0,15	keluarga	0,17	tertib	0,25	physical	0,20
rawat	0,15	Bantu	0,17	rantai	0,23	pola	0,20
Gugus (positif)		Penanganan (positif)		Pemerintah (negatif)		Indonesia (negatif)	
tugas	0,81	Cepat	0,38	anjuran	0,26	tumbuh	0,21
cepat	0,33	Tugas	0,30	benarnya	0,25	ekonomi	0,20
bubar	0,31	komite	0,25	laksanakan	0,21	proyeksi	0,20
tim	0,24	presiden	0,24	vital	0,20	penanggulangan	0,17
nasional	0,20	pemulihan	0,24	sasar	0,18	lambat	0,16
Sembuh (positif)				Masyarakat (negatif)			
orang	0,30			dampak	0,27		
positif	0,19			sembako	0,26		
nyata	0,19			giat	0,23		
tinggal	0,19			kondusif	0,21		
total	0,19			bantu	0,18		

Berdasarkan tabel 1 informasi yang dapat diperoleh untuk tanggapan positif adalah pasien positif *Covid-19* yang terus bertambah perlu dilakukan perawatan, kemudian warga yang terdampak antar keluarga harus saling membantu, selain itu pemerintah khususnya gugus tugas harus menangani *Covid-19* dengan cepat dan nyata. Sedangkan untuk tanggapan negatif informasi yang dapat diperoleh dari *wordcloud* dan asosiasi kata adalah upaya pemerintah untuk menanggulangi *Covid-19* ini dinilai lambat terutama di bidang ekonomi, oleh karena itu masyarakat perlu adanya bantuan sembako dari pemerintah.

4. Simpulan

Berdasarkan hasil dalam pembahasan diatas maka dapat disimpulkan Dari 10000 data tanggapan yang diperoleh 6128 tanggapan bersentimen positif dan 3872 tanggapan bersentimen negatif. kemudian jika dilihat berdasarkan perbulannya jumlah tanggapan positif masih lebih besar dibanding dengan tanggapan negatif, dengan tanggapan positif terbesar terjadi pada bulan Juni dan tanggapan negatif terbesar terjadi pada bulan Maret.

Hasil klasifikasi data tanggapan masyarakat Indonesia terhadap pandemi *Covid-19* pada media sosial Twitter menggunakan algoritma *support vector machine* dengan kernel *linear* dan evaluasi model *10-fold cross validation* menghasilkan *accuracy* sebesar 90,1%. Jika menggunakan algoritma *naive bayes* dengan *laplace correction* adalah 1 dan evaluasi modelnya *10-fold cross validation* menghasilkan *accuracy* 79,2%. Sedangkan jika menggunakan algoritma *K-nearest neighbor* dengan parameter *k* adalah 20 dan evaluasi modelnya *10-fold cross validation* menghasilkan *accuracy* sebesar 62,1%. Jadi *support vector machine* merupakan algoritma yang paling cocok untuk mengklasifikasikan data tanggapan masyarakat Indonesia terhadap *Covid-19* pada media sosial Twitter dibandingkan dengan *naive bayes* dan *K-nearest neighbor*.

Secara umum tanggapan positif masyarakat Indonesia terhadap pandemi *Covid-19* berkaitan dengan ajakan untuk saling membantu dan saran untuk pemerintah, sedangkan pada tanggapan negatif informasi yang dapat diperoleh dari *wordcloud* dan asosiasi kata adalah upaya pemerintah untuk menanggulangi *Covid-19* ini dinilai lambat terutama di bidang ekonomi, oleh karena itu masyarakat perlu adanya bantuan sembako dari pemerintah.

Daftar Pustaka

- Alhajji, M., Khalifah, A. Al, Aljubran, M. J., & Alkhalifah, M. (2020). *Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19*. (April). <https://doi.org/10.20944/preprints202004.0031.v1>
- Darmanik, R. M. (2014). Pembangunan Aplikasi Pencarian Dokumen Menggunakan Text Mining Berbasis Web. *Jurnal Universitas Atma Jaya Yogyakarta*.

- Fithriasari, K., Mayasari, R. W., Iriawan, N., & Winahju, W. S. (2020). Surabaya Government Performance Evaluation using Tweet Analysis. *MATEMATIKA: MJIAM*, 36(1), 31–42.
- Huq, M. R., Ali, A., & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM. *International Journal of Advanced Computer Science and Applications*, 8(6), 19–25.
- Mardiana, T., Syahreva, H., & Tuslaela, T. (2019). Komparasi Metode Klasifikasi Pada Analisis Sentimen Usaha Waralaba Berdasarkan Data Twitter. *Jurnal Pilar Nusa Mandiri*, 15(2), 267–274. <https://doi.org/10.33480/pilar.v15i2.752>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundation and Trends in Information Retrieval*, 2. <https://doi.org/10.3748/wjg.v22.i45.9898>
- Pratiwi, Y. R., & Widodo, E. D. Y. (2017). *Comparison of Maximum Entropy and Support Vector Machine Methods for Sentiment Analysis of Peralite Product Through Twitter Social Network*. (12), 10–14.
- Pitria, P. (2014). Analisis Sentimen Pengguna *Twitter* pada Akun Resmi *Samsung* Indonesia dengan Menggunakan *Naive Bayes*. Universitas Komputer Indonesia.
- Purnamawan, I. K. (2015). Support Vector Machine pada Information Retrieval. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 12, 173–180.
- Qeis, M. I. (2015). Aplikasi Wordcloud Sebagai Alat Bantu Analisis Wacana. *International Conference on Language, Culture, and Society - ICLCS LIPI*, (November 2015). Retrieved from https://www.researchgate.net/publication/316736417_APLIKASI_WORDCLOUD_SEBAGAI_ALAT_BANTU_ANALISIS_WACANA
- Santosa, B. (2007). *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saputra, P. Y. (2017). Implementasi Teknik Crawling untuk Pengumpulan Data dari Media Sosial Twitter. *Dinamika Dotcom*, 8, 160–168.
- Suprianto, S. (2020). Implementasi Algoritma Naive Bayes Untuk Menentukan Lokasi Strategis Dalam Membuka Usaha Menengah Ke Bawah di Kota Medan (Studi Kasus: Disperindag Kota Medan). *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(2), 125. <https://doi.org/10.30865/json.v1i2.1939>
- Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika Bandung.
- WHO. (2020). Pertanyaan dan jawaban terkait Coronavirus. Retrieved March 20, 2020, from WHO website: <https://www.who.int/indonesia/news/novel-coronavirus/qa-for-public>