



## Analisis Sentimen untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19

Fajar Sodik Pamungkas<sup>✉</sup>, Iqbal Kharisudin

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia  
Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang, 50229

### Info Artikel

Sejarah Artikel:  
Diterima  
Disetujui  
Dipublikasikan

#### Keywords:

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

### 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*, *Naive Bayes*, dan *K-Nearest Neighbor*, tahap analisis data yaitu *text preprocessing*, pelabelan data, pengklasifikasian menggunakan ketiga algoritma kemudian dibandingkan mana yang paling baik untuk mengklasifikasikan data tanggapa. 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% 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*

### Abstract

The Covid-19 pandemic has had a profound impact on various sectors of people's lives, a situation that forces people to carry out *physical distancing* to change people's lifestyle. This has resulted in various opinions or responses from the public regarding the Covid-19 pandemic which have been poured out on social media. To find out the sentiment of the public response, a sentiment analysis is needed with a machine learning algorithm. In this study, a sentiment analysis of the responses of the Indonesian people to the Covid-19 pandemic was carried out on Twitter social media using the *Support Vector Machine*, *Naive Bayes*, and *K-Nearest Neighbor* algorithms, the data analysis stages were *text preprocessing*, data labeling, classification using all three algorithms then compared where which is best for classifying response data. Based on the average level of accuracy using the *10-Fold Cross Validation* model evaluation, it is concluded that the SVM algorithm has higher accuracy than *Naive Bayes* and KNN with an average accuracy of 90.01% with a *linear* kernel, 79.20% in *Naive Bayes* with the number of *laplace* is 1, and 62.10% on the KNN with the number of K is 20 and uses the *optimal* kernel

#### How to cite:

Pamungkas, F. S., Kharisudin I. 2020. Analisis Sentimen untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19. *UNNES Journal of Mathematics*. 7(2):1-11.

## PENDAHULUAN

Di awal tahun 2020 dunia digemparkan dengan merebaknya virus baru yaitu *coronavirus* jenis baru (SARS-CoV-2) dan penyakitnya disebut *Coronavirus Disease 2019* (Covid-19). Oleh karena virus tersebut setiap negara memiliki strategi yang berbeda dari negara lain. Salah satu strategi yang digunakan oleh beberapa negara adalah penerapan *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. Twitter merupakan salah satu media sosial yang memiliki jumlah pengguna terbanyak di Indonesia dengan 6,43 juta pengguna aktif bulanan (Websindo, 2019).

Tanggapan yang terdapat dalam Twitter merupakan tanggapan yang berupa teks. Maka digunakan *text mining* dan analisis sentimen dalam melakukan analisis. *Text mining* mempunyai definisi sebagai menambang data yang berupa teks, sedangkan 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.

Terdapat beberapa algoritma dalam klasifikasi, antara lain *Support Vector Machine*, *Naive Bayes*, dan *K-Nearest Neighbor*. Penelitian terbaru tentang analisis sentimen terhadap pandemi Covid-19 pada media sosial Twitter di Arab Saudi yang dilakukan oleh Alhajji, Khalifah, Aljubran, & Alkhalifah, (2020) menyimpulkan bahwa *Naive Bayes* dapat digunakan untuk klasifikasi teks dengan akurasi yang cukup tinggi yaitu sebesar 89%. Namun dalam penelitian 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 kelemannya 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*

## 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 dilakukan *text preprocessing*. Hal pertama yang dilakukan dalam *text preprocessing* adalah *case folding* yaitu proses penyamaan *case* pada dokumen menjadi *lowercase*, setelah itu dilakukan *spell normalization* yaitu proses perbaikan singkatan, kemudian dilakukan *tokenizing* yaitu proses pemecahan dokumen menjadi kata perkata, lalu dilakukan proses *filtering* yaitu proses penghapusan kata yang tidak diperlukan pada dokumen, dan yang terakhir adalah *stemming* yaitu mencari kata dasar pada tiap kata dalam dokumen.

Setelah dilakukan *text preprocessing* dilakukan pelabelan data. Pelabelan data dilakukan dengan menghitung jumlah kata positif dan negatif pada tiap tanggapan. Apabila jumlah kata positif lebih besar daripada jumlah kata negatif maka tanggapan akan dikategorikan menjadi tanggapan positif, begitu juga sebaliknya.

Data yang sudah dikategorikan, kemudian dilakukan pengklasifikasian menggunakan

beberapa algoritma. Algoritma pertama yang digunakan adalah *support vector machine*. algoritma *support vector machine* atau SVM memiliki tujuan menemukan *hyperplane* yang paling optimum (Suyanto, 2018). *Hyperplane* yang optimum dapat ditemukan dengan mengukur margin *hyperplane* dan mengukur titik maksimalnya. Margin sendiri adalah jarak antara *hyperplane* dengan data terdekat dari masing-masing kelas.

Algoritma kedua yang digunakan adalah *naive bayes*. *Naive bayes* merupakan metode *machine learning* yang menggunakan teorema kuno warisan abad ke-18 yang ditemukan oleh Thomas Bayes (Suyanto, 2018). Algoritma *Naive Bayes* merupakan algoritma yang memanfaatkan metode probabilitas dengan asumsi antar satu kelas dengan kelas yang lain tidak saling tergantung (Manik & Saragih, 2017).

Algoritma terakhir yang digunakan 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). Proses pelatihan KNN menghasilkan  $k$  yang memberikan akurasi tertinggi dalam menggeneralisasi data-data yang akan datang, namun masalahnya sampai saat ini tidak dapat ditentukan secara matematik (Suyanto, 2018).

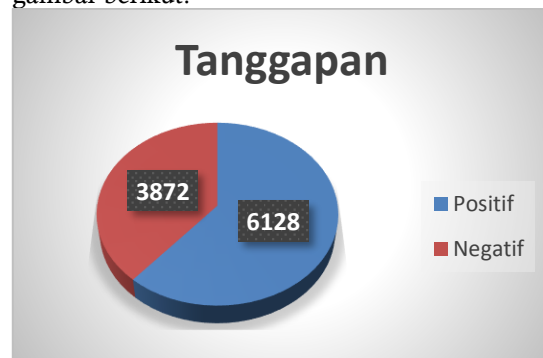
Ketiga algoritma tersebut akan dibandingkan dan dilihat performa kinerjanya dalam mengklasifikasikan data dengan evaluasi model *k-fold cross validation*. Penggunaan *k-fold cross validation* untuk menghilangkan bias pada data. Pelatihan dan pengujian pengujian dilakukan sebanyak  $k$  kali. Pada percobaan pertama, *subset* S1 diperlakukan sebagai data pengujian dan *subset* lainnya diperlakukan sebagai data pelatihan, pada percobaan kedua *subset* S1, S3, ... S $k$  menjadi data pelatihan dan S2 menjadi data pengujian, dan seterusnya (Bramer, 2007).

Visualisasi data tanggapan dilakukan dengan *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). Semakin besar huruf atau kata yang terdapat pada *wordcloud* maka semakin sering pula kata itu muncul. Sedangkan 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).

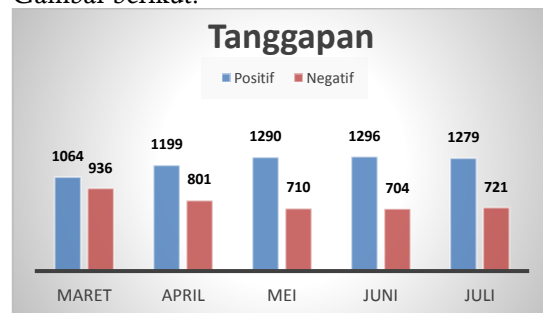
## HASIL DAN PEMBAHASAN

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. 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.

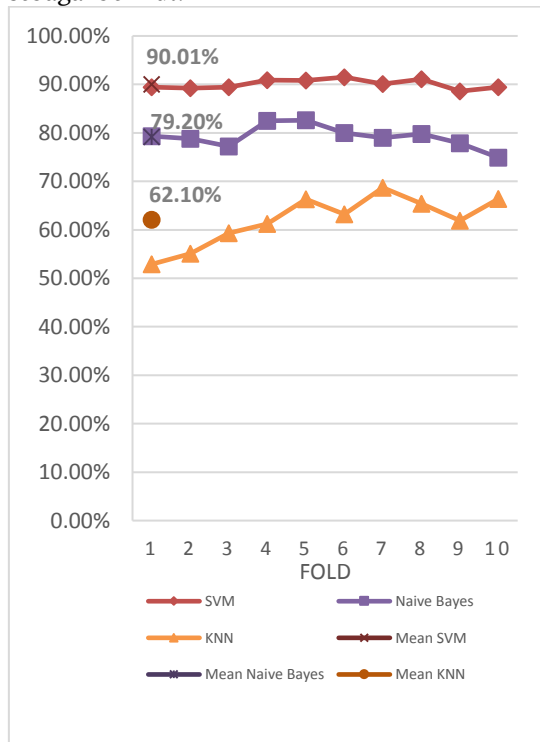


Gambar 2. Jumlah Tanggapan Masyarakat Berdasarkan Bulan

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

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, Syahreva, & Tuslaela, (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. Wordcloud Tanggapan Positif

Berdasarkan wordcloud tersebut dapat diketahui bahwa perbincangan positif masyarakat Indonesia mengenai Covid-19 banyak dikaitkan dengan kata “positif”, “pandemi”, “pasien”, “cegah”, dan “sehat”.



Gambar 5. Wordcloud Tanggapan Negatif

Berdasarkan wordcloud tersebut dapat diketahui bahwa pada tanggapan negatif perbincangan sering dikaitkan dengan kata “orang”, “sebar”, “pandemi”. “masyarakat”, “rumah”.

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.

Hasil dari asosiasi kata pada tanggapan positif adalah sebagai berikut.

Tabel 1. Asosiasi Kata Tanggapan Positif

Pasien (positif)		Warga (positif)	
positif	0,28	beras	0,23
tambah	0,24	dampak	0,20
demam	0,17	desa	0,19
isolasi	0,15	keluarga	0,17

rawat	0,15	bantu	0,17
Gugus (positif)		Penanganan (positif)	
tugas	0,81	cepat	0,38
cepat	0,33	tugas	0,30
bubar	0,31	komite	0,25
tim	0,24	presiden	0,24
nasional	0,20	pemulihan	0,24
Sembuh (positif)			
orang	0,30		
positif	0,19		
nyata	0,19		
tinggal	0,19		
total	0,19		

Berdasarkan hasil asosiasi kata pada tanggapan positif kata “pasien” sering diperbincangkan dengan kata “positif”, “tambah”, “demam”, “isolasi”, dan “rawat” dengan nilai korelasi berkisar antara 0,15 sampai dengan 0,28.

Tabel 2. Asosiasi Kata Tanggapan Negatif

Protokol (negatif)		Masker (negatif)	
sehat	0,58	jarak	0,39
disiplin	0,37	jaga	0,31
adaptasi	0,31	rajin	0,23
tertib	0,25	physical	0,20
rantai	0,23	pola	0,20
Pemerintah (negatif)		Indonesia (negatif)	
anjaran	0,26	tumbuh	0,21
benarnya	0,25	ekonomi	0,20
laksanakan	0,21	proyeksi	0,20
vital	0,20	penanggulangan	0,17
sasar	0,18	lambat	0,16
Masyarakat (negatif)			
dampak	0,27		
sembako	0,26		
giat	0,23		
kondusif	0,21		
bantu	0,18		

Berdasarkan hasil asosiasi kata pada tanggapan negatif kata “protokol” sering diperbincangkan dengan kata “sehat”, “disiplin”, “adaptasi”, “tertib”, dan “rantai” dengan nilai korelasi berkisar antara 0,23 sampai dengan 0,58.



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.

## 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>
- Bramer, M. (2007). *Principles of Data Mining*.  
<https://doi.org/10.1007/978-1-84628-766-4>
- 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.
- Manik, F. Y., & Saragih, K. S. (2017). Klasifikasi Belimbing Menggunakan Naïve Bayes Berdasarkan Fitur Warna RGB. *International Journal of Computer Science and Security*, 11(1).
- 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.
- 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](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.
- Websindo, T. M. (2019). Indonesia Digital 2019 : Media Sosial. Retrieved September 22, 2020, from websindo.com website: <https://websindo.com/indonesia-digital-2019-media-sosial/>
- 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>