

ANALISIS SENTIMEN OPINI PUBLIK MENGENAI COVID-19 PADA TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN KNN

Muhammad Syarifuddin

Pendidikan Teknologi Informasi

Universitas Brawijaya

www.ub.ac.id

syariff@student.ub.ac.id/syarifuddinm79@gmail.com

Abstract— Public aspirations, opinions and criticisms are very difficult to convey directly, due to several constraints such as limited time, space, and even human resources. However, social networking does not rule out the possibility of this, one of which is Twitter. Twitter is a social network that is often used by people to express opinions, criticisms and aspirations. But the opinions on Twitter are many different meanings from each “netizen”, one of which is about the opinion of COVID-19. Therefore we need an analysis of public opinion sentiments to harmonize and provide a new perspective on an issue of COVID-19, the method used is the Naïve Bayes algorithm and KNN, this study focuses on comparing the results of the classification of the Naïve Bayes and KNN methods, as well as identifying opinion trends community on twitter. Subjects were taken using Twitter API as many as 1098 opinions with the keyword “COVID-19”, starting from the classification of positive or negative opinions, data cleansing, preprocessing, until the final results were obtained. From the test results it is known that the Naïve Bayes method has a higher value, one of which is an accuracy level of 63.21%, while the KNN method is 58.10%, and there is also a tendency for public opinion on Twitter to lean positively, this can be seen from the number of positive opinions of 610 while negative 488, supported by the results of precision testing in the Naïve Bayes method with a higher positive value than negative, namely 66.40%: 58.94%.

Keywords: Public Opinion, Covid-19, Twitter, Naïve Bayes, KNN

Abstrak— Aspirasi, opini dan kritik masyarakat sangat sulit disampaikan secara langsung, karena beberapa kendala seperti keterbatasan waktu, ruang, bahkan SDM. Namun, jejaring sosial tidak menutup kemungkinan akan hal tersebut, salah satunya adalah *twitter*. *Twitter* merupakan salah satu jejaring sosial yang sering digunakan oleh masyarakat untuk mengungkapkan opini, kritik hingga aspirasi. Namun opini pada *twitter* banyak sekali makna yang berbeda dari setiap *netizen*, salah satunya adalah tentang opini COVID-19. Oleh karena itu perlu suatu analisis sentimen opini masyarakat guna menyelaraskan dan memberi pandangan baru mengenai suatu isu tentang COVID-19, metode yang digunakan adalah algoritma *Naïve Bayes* dan KNN, penelitian ini berfokus pada perbandingan hasil klasifikasi metode *Naïve Bayes* dan KNN, serta mengetahui kecenderungan opini masyarakat di *twitter*. Subjek diambil menggunakan API *twitter* sebanyak 1098 opini dengan kata kunci “COVID-19”, Dimulai dari klasifikasi opini positif atau negatif, data *cleansing*, *preprocessing*, hingga didapatkan hasil akhir. Dari hasil pengujian diketahui metode *Naïve Bayes* memiliki nilai lebih tinggi, salah satunya adalah tingkat *accuracy* sebesar 63.21%, sedangkan metode KNN sebesar 58.10%, dan didapatkan pula kecenderungan opini masyarakat di *twitter* condong positif, hal tersebut dapat dilihat dari jumlah opini positif sebesar 610 sedangkan negatif 488, ditunjang dengan hasil pengujian *precision* di metode *Naïve Bayes* dengan nilai positif lebih tinggi dari pada negatif yaitu 66.40%: 58.94%.

Kata Kunci: Opini Publik, Covid-19, Twitter, Naïve Bayes, KNN

PENDAHULUAN

Akhir-akhir ini masyarakat dihebohkan dengan virus baru dari Wuhan, China yaitu virus corona atau disebut COVID-19 (*Corona Virus Desaease* 2019), virus ini masuk di Indonesia sejak awal maret lalu hingga saat ini dan diprediksi akan

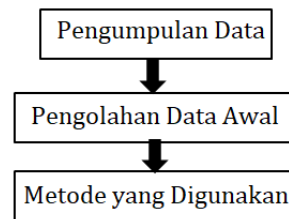
terus berlanjut hingga waktu yang belum diketahui (Azizah, 2020). Beberapa cara telah dilakukan oleh pemerintah, institusi, hingga kalangan masyarakat untuk memutus rantai penyebaran virus corona. Sampai saat ini, COVID-19 telah memakan korban jiwa di Indoensia hingga mencapai 872 orang (Merdeka, 2020). Hal tersebut membuat

kebanyakan orang merasa was-was dan resah bahkan ketakutan akan virus ini, Disisi lain kinerja pemerintah dinilai kurang tepat dalam menangani hal tersebut, hal tersebut membuat masyarakat umum ingin mengungkapkan segala pendapat, aspirasi dan kritikan, namun keterbatasan waktu dan ruang membuat aspirasi masyarakat tidak tersampaikan, Disisi lain jejaring sosial merupakan salah satu tempat yang cocok untuk mengungkapkan segala hal yang ada, salah satunya adalah *twitter*, *twitter* adalah jejaring sosial yang memungkinkan pengguna mengirim karakter hingga 140 karakter yang sering disebut *twit* atau kicauan (Suryono et al., 2018). Namun setiap *twit* dari *netizen* tidak mengandung makna yang selaras, hal tersebutlah yang perlu dilakukan analisis terhadap opini *netizen* di *twitter* mengenai COVID-19 dengan klasifikasi positif dan negatif, analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial yang didalamnya terdapat mengandung pelayanan umum, serta isu terkini. Disisi lain, analisis sentimen dapat digunakan untuk mengevaluasi suatu kinerja, pelayanan dan sebagainya. Adapun menurut Fink pada tahun 2011, jenis analisis sentimen yang sering digunakan dalam kalangan peneliti yaitu, analisis sentimen dalam bentuk dokumen atau kalimat (Fink et al., 2011).

Disisi lain terdapat teks *mining* yang dapat bekerja dalam komputer dengan tujuan mengolah informasi lama secara eksplisit sehingga menghasilkan temuan informasi baru. Hal ini selaras dengan pengertian data *mining* menurut Pertiwi data *mining* merupakan pengambilan pola tertentu yang menarik didalam data, bedanya dengan teks *mining* yang cenderung dengan data berbentuk teks, sedangkan data *mining* dataanya berbentuk angka atau numerik (Pertiwi, 2019). Penelitian ini membahas klasifikasi *twitts* di *twitter* yang mengandung nilai positif dan negatif mengenai COVID-19 menggunakan metode *Naïve Bayes* dan KNN (*K-Nearest Neighbor*). Hasil dari penelitian ini akan memberi gambaran kepada masyarakat umum apakah COVID-19 cenderung ke opini positif atau negatif, serta membandingkan hasil tingkat akurasi metode *Naïve Bayes* dengan KNN.

BAHAN DAN METODE

Tahap yang dilakukan dalam penelitian ini terdiri dari beberapa tahap, yang dapat dilihat pada Gambar 1 dibawah ini:



Gambar 1. Bagan Tahap Penelitian

1. Pengumpulan Data

Pengumpulan data dilakukan dengan jejaring sosial *twitter* dengan mencari kata kunci "COVID-19" yang dapat diambil di *website* resmi *twitter* yaitu www.twitter.com, untuk sampel yang diambil sebanyak 1098 *tweets* dari semua populasi yang tersebar tentang COVID-19 di Indonesia, dan diambil secara *random* atau acak. Adapun salah satu contoh *tweet* opini positif dan negatif dari www.twitter.com bisa dilihat pada Gambar 2 dan Gambar 3 berikut:



Sumber: (Syarifuddin, 2020a)

Gambar 2. Opini Positif Tentang COVID-19



Sumber: (Syarifuddin, 2020a)

Gambar 3. Opini Negatif Tentang COVID-19

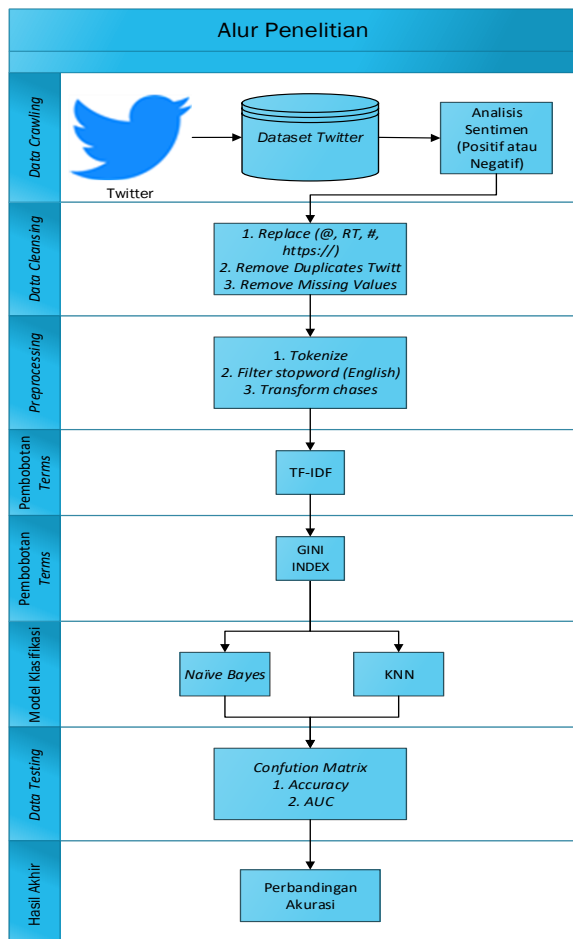
2. Pengolahan Data Awal

Pada tahap ini data akan di klasifikasikan apakah data tersebut dalam kategori positif

ataupun negatif, kemudian dilakukan *cleansing* data. Pada tahap ini akan merubah beberapa karakter pada *twitter*, seperti “#” menjadi “*hashtag*”, “@” menjadi “*At*”, “RT” menjadi “*retweet*” dan lain sebagainya. Selain itu, terdapat beberapa atribut yang tidak relevan seperti “☺”, akan dibuang keberadaannya karena akan mempengaruhi tingkat akurasi dari pengolahan data (Negara et al., 2016).

3. Metode yang Digunakan

Dalam penelitian ini metode yang digunakan yaitu metode klasifikasi dengan algoritma *Naïve Bayes* dan *KNN*, sebenarnya banyak metode klasifikasi seperti *Decision Tree*, *Neural Network* dan *Support Vector Machine*, dan *Machine Learning* (Wibowo & Winarko, 2014). Namun, peneliti hanya akan fokuskan pada 2 (dua) metode itu saja. Adapun alur penelitian pada Gambar 4 berikut:



Sumber: (Syarifuddin, 2020b)

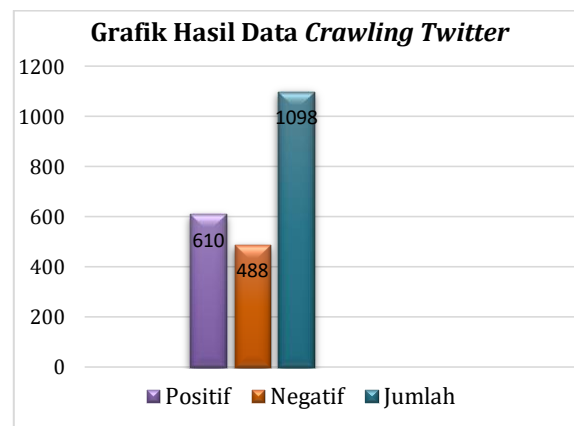
Gambar 4. Alur Penelitian Secara Detail

Dari *flowchart* Gambar 4 diatas dapat diketahui bahwa penelitian ini dimulai dari pencarian data di *twitter* menggunakan API *Twitter*, kemudian didapatkan kumpulan data

tweet yang disebut data *set* sebanyak 1098 *tweets*, selanjutnya data *tweet* tentang COVID-19 perlu di klasifikasi apakah bersifat opini positif ataupun negatif, setelah data diklasifikasi tahap selanjutnya yaitu Data *cleansing*, penulis membuang karakter-karakter yang dapat menghambat proses olah data *mining*, dan menghapus *tweet* yang sama (*double*), serta penghapusan data yang memiliki nilai *null*, dilanjut dengan proses *preprocessing*, yaitu merubah beberapa kata asing ke Bahasa Indonesia dan membuang beberapa karakter yang tidak diperlukan, dilanjut dengan pembobotan *terms*, yaitu menggunakan *TIF-IDF* dan *GINI INDEX*, yang merupakan algoritma dalam ilmu komputer yang digunakan dalam menghitung bobot dalam setiap kata yang umum digunakan, yang terdiri model klasifikasi *Naive Bayes* dan *KNN* yaitu, model klasifikasi data dalam data *mining* yang menggunakan aplikasi *rapidminer*. Dilanjut dengan data *testing*, dilakukan dengan aplikasi *rapidminer* sehingga dapat diketahui secara jelas berapa nilai *accuracy*, *presicion*, *recall*, dan *AUC* dari setiap model klasifikasi, dan yang terakhir dilakukan perbandingan dari model klasifikasi *Naive Bayes* dan *KNN* mana yang paling akurat dalam mengklasifikasikan data *tweet* COVID-19.

HASIL DAN PEMBAHASAN

Data *training* yang diujikan diambil di web resmi *www.twitter.com* dengan API *twitter* dengan kata kunci “COVID-19” pada tahun 2020, diambil secara acak dan didapatkan sebanyak 1098 *tweets* populasi, dengan perhitungan opini positif sebanyak 610 sedangkan opini negatif sebanyak 488 opini. Selanjutnya akan diolah dan dilakukan *testing* dan *training* data *set* dan didapatkan nilai *accuracy* dan *AUC*, terdapat juga nilai *recall* dan *precision*. Pada Gambar 5 Grafik dibawah merupakan representasi jumlah populasi data *training* yang penulis gunakan



Sumber: (Syarifuddin, 2020b)

Gambar 5. Hasil Data Crawling Twitter

1. Data Cleansing

Berfungsi membersihkan atau *replace* data atau karakter yang tidak baku yang dapat mengganggu pengolahan data, seperti terlihat pada Tabel 1 dibawah ini:

Tabel 1. Teks Sebelum dan Sesudah di *Cleansing*

Proses <i>Cleansing</i>	
Sebelum cleansing data	@umam_elchoiru: bersatu melawan covid-19. mari sayangi guru ngajimu cc: @na_dirs @adtaufiq @kirekswasta @Nuryoudie @arifsholihan_
Sesudah cleansing data	At umam_elchoiru: bersatu melawan covid-19. mari sayangi guru ngajimu cc: At na_dirs At adtaufiq At kirekswasta At Nuryoudie At arifsholihan_

Sumber: (Pertwi, 2019)

2. Tokenize

Tokenize merupakan proses pemotongan teks menjadi kata, *symbol*, karakter ataupun tanda baca, sehingga menjadi token yang dapat di analisa. berikut pada Tabel 2 merupakan salah satu contoh prosesnya dalam *tweet* COVID-19.

Tabel 2. Teks Sebelum dan Sesudah di *Tokenizing*

Proses <i>Tokenize</i>	
Sebelum di tokenizing	Ternyata maling di mini market ya.. Maling nya bonyok,warganya kerumunan tanpa ada jarak,Akhirnya Covid-19 pun mengintai. 😞 ☐
Sesudah di tokenizing	ternyata maling di mini market ya.. maling nya bonyok,warganya kerumunan tanpa ada jarak, akhirnya covid-19 pun mengintai.

Sumber: (Pertwi, 2019)

3. Filter Stopword (Indonesia)

Tabel 3 dibawah merupakan proses menyaring kata-kata asing atau singkatan menjadi bahasa Indonesia.

Tabel 3. Teks Sebelum dan Sesudah Dilakukan *Filter Stopword*

Proses <i>Filter Stopword</i>	
Teks sebelum dilakukan filter stopword	karena covid-19, anak kos seperti diriku ini jadi rajin masak sendiri. ada yg gini juga ga? wkwk *biar punya kenang2an kalo aku pernah mayan rajin masak ?? sooooo, masak apa hari ini?

Proses *Filter Stopword*

Teks sesudah dilakukan filter stopword	karena covid-19, anak kos seperti diriku ini jadi rajin masak sendiri. ada yang begini juga ngga? wkwk *biar punya kenang-kenangan kalo aku pernah mayan rajin masak ?? sooooo, masak apa hari ini?
--	--

Sumber: (Pertwi, 2019)

4. Transform Cases

Tabel 4 dibawah merupakan proses merubah huruf kapital menjadi huruf kecil, berikut adalah prosesnya:

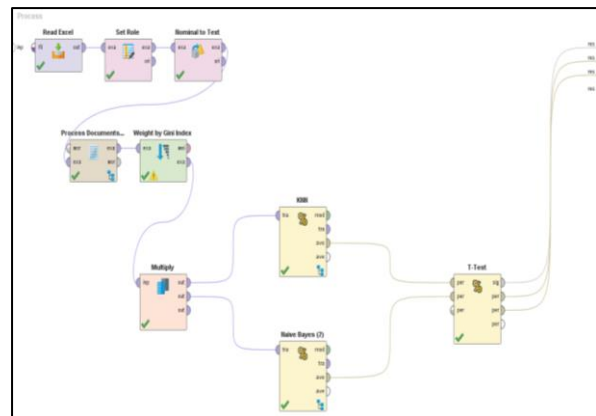
Tabel 4. Teks Sebelum dan Sesudah Dilakukan *Proses Transform Cases*

Proses <i>Transform Cases</i>	
Teks sebelum dilakukan proses <i>Transform Cases</i>	Bahaya merokok atau COVID 19?
Teks sesudah dilakukan proses <i>Transform Cases</i>	bahaya merokok atau covid 19

Sumber: (Pertwi, 2019)

5. Perbandingan Model Klasifikasi *Naïve Bayes* dan KNN

Setelah semua tahap diatas selesai dilakukan, dilanjut dengan mengolah data menggunakan aplikasi, aplikasi yang penulis gunakan adalah *rapidminer Versi 7.1*, data diolah dan didapatkan hasil *design model process* seperti gambar dibawah ini:



Sumber: (Syarifuddin, 2020b)

Gambar 6. Hasil Model Proses dengan Metode *Naïve Bayes* dan KNN

Dari desain model proses pada Gambar 6 diatas jika dijalankan dan berhasil maka akan didapatkan hasil perhitungan sebagai berikut, pada masing-masing model klasifikasi secara detail:

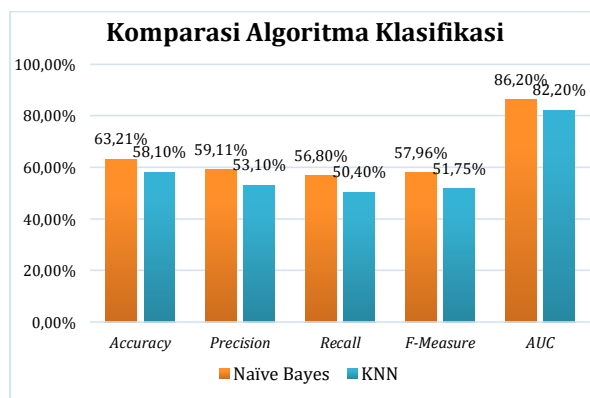
Tabel 5. Klasifikasi Detail

DETAIL PER OPINI						
	Accuracy		Precision		Recall	
	Positif	Negatif	Positif	Negatif	Positif	Negatif
Naive Bayes	66.40%	58.94%	66.40%	58.94%	66.40%	58.94%
KNN	61.83%	53.02%	61.83%	53.02%	61.83%	53.02%

KESELURUHAN					
	Accuracy	Precision	Recall	F-Measure	AUC
Naive Bayes	63.21%	59.11%	56.80%	57.96%	0.862
KNN	58.10%	53.10%	50.40%	51.75%	0.822

Sumber: (Syarifuddin, 2020b)

Lebih sederhana terdapat pada Gambar 7 grafik dibawah ini:



Sumber: (Syarifuddin, 2020b)

Gambar 7. Komparasi Algoritma Klasifikasi

Dari grafik diatas dapat diketahui bahwa hasil klasifikasi model *Naive Bayes* lebih unggul dibandingkan model KNN, ditinjau dari:

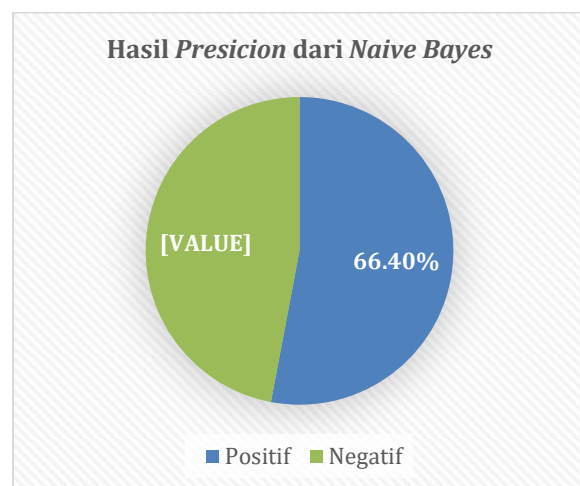
1. Nilai *accuracy*, sejatinya nilai *accuracy* adalah nilai tingkat kedekatan antara nilai prediksi dengan nilai *actual* (Antinasari et al., 2017), didapatkan dari pengujian *Naive Bayes* 63.21% sedangkan KNN 58.10%.
2. Nilai *precision*, yakni tingkat ketepatan prediksi dari suatu sistem, dengan menghitung prediksi positif dari total data yang diprediksi sistem, termasuk prediksi salah (Rahutomo et al., 2018). didapatkan *Naive Bayes* dengan nilai 59.11% : KNN 53.10%.
3. Nilai *recall* yaitu, nilai tingkat keberhasilan dalam mengenali suatu kelas yang harus dikenali (Hakiem & Fauzi, 2019), dari data diatas didapatkan *Naive Bayes* 56.80%: KNN 50.40%.
4. Nilai *F-Measure*, merupakan nilai gabungan antara nilai *precision* dan *recall*, atau nilai yang mewakili keseluruhan kinerja dari suatu sistem (Habibi et al., 2016), didapatkan *Naive Bayes* 57.96% : KNN 51.75% dan
5. Nilai AUC (*Area under Curve*), yaitu digunakan untuk mengukur perbedaan performansi yang

telah dihitung (Faisal et al., 2020), didapatkan *Naive Bayes* 0.862 sedangkan KNN 0.822.

Sehingga dapat ditarik benang merah bahwa model klasifikasi *Naive Bayes* memiliki nilai tertinggi dari semua pengujian, dan lebih akurat untuk mengklasifikasikan data yang ada, dibandingkan dengan model klasifikasi KNN.

6. Menentukan Kecenderungan Data Opini yang Diuji

Untuk mengetahui kecenderungan apakah opini yang diuji bersifat positif atau negatif, dapat dilihat dari banyaknya opini yang ada, yaitu positif 610 opini, sedangkan negatif sebanyak 488 opini, untuk memperkuat data kecenderungan opini yang diuji dapat dilihat pada hasil pengujian model *Naive Bayes* (model yang paling akurat), pada bagian *presicion*, karena *presicion* menurut Indrayuni, dapat memprediksi ketepatan data dari suatu sistem (Indrayuni, 2019). Diketahui *presicion* positif lebih banyak persentasenya dibanding *presicion* negative, yaitu 66.40% : 58.94%, yang dapat dilihat pada Gambar 8 bar dibawah ini:



Sumber: (Syarifuddin, 2020b)

Gambar 8. Kecenderungan Positif Data Opini yang Diuji

Sehingga dapat ditarik inti sari bahwa opini atau *tweets* atau kicauan masyarakat umum di

twitter tentang COVID-19 saat ini, cenderung pada opini positif. Sehingga masyarakat tidak perlu berpikir bahwa COVID-19 hanya memberi dampak negatif, tetapi juga ada dampak positif.

KESIMPULAN

Berdasarkan hasil penelitian, dapat ditarik beberapa kesimpulan diantaranya adalah COVID-19 merupakan penyakit yang tidak sedikit meresahkan masyarakat umum, apalagi ditambah dengan kicauan di media social tentang COVID-19 yang semakin merabak, salah satunya adalah *twitter* yang menambah masyarakat ketakutan. Namun, disisi lain terdapat dampak positif yang masyarakat perlu tahu, hal tersebut didapatkan dari hasil pegujian menggunakan metode Naïve Bayes dan KNN, yang menunjukkan opini masyarakat cenderung positif. Dari kedua metode tersebut metode Naïve Bayes merupakan model klasifikasi yang paling akurat dibandingkan dengan KNN dengan nilai accuracy sebesar 63.21%. Sehingga, penulis berharap penelitian ini dapat menambah wawasan masyarakat umum tentang COVID-19.

REFERENSI

- Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1733–1741.
- Azizah, K. N. (2020). Heboh Karena Virus Corona, Memangnya Siapa Sih yang Wajib Pakai Masker_. *Health.Detik.Com*, Berita Detikhealth. <https://health.detik.com/berita-detikhealth/d-4902084/heboh-karena-virus-corona-memangnya-siapa-sih-yang-wajib-pakai-masker>
- Faisal, A., Alkhalifi, Y., Rifai, A., & Gata, W. (2020). Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization. *JOINTECS (Journal of Information Technology and Computer Science)*, 5(2), 61. <https://doi.org/10.31328/jointecs.v5i2.1362>
- Fink, C. R., Chou, D. S., Kopecky, J. J., & Llorens, A. J. (2011). Coarse- and fine-grained sentiment analysis of social media text. *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 30(1), 22–30.
- Habibi, R., Setyohadi, D. B., & Wati, E. (2016). Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation. *Jurnal Informatika*, 12(1), 103–109. <https://doi.org/10.21460/inf.2016.121.462>
- Hakim, M., & Fauzi, M. A. (2019). *Klasifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Naïve Bayes Berbasis N-Gram Dengan Seleksi Fitur Information Gain*. 3(3), 2443–2451.
- Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. *Jurnal Khatulistiwa Informatika*, 7(1), 29–36. <https://doi.org/10.31294/jki.v7i1.1>
- Merdeka. (2020). Data Terkini Jumlah Korban Virus Corona di Indonesia. *Merdeka.Com*, Peristiwa. <https://www.merdeka.com/peristiwa/data-terkini-jumlah-korban-virus-corona-di-indonesia.html>
- Negara, E. S., Andryani, R., & Saksono, P. H. (2016). Analisis Data Twitter: Ekstraksi dan Analisis Data Geospasial. *Jurnal INKOM*, 10(1), 27. <https://doi.org/10.14203/j.inkom.433>
- Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, KNN dan SVM. *Inti Nusa Mandiri*, 14(1), 27–32.
- Rahutomo, F., Saputra, P. Y., & Fidyawan, M. A. (2018). Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. *Jurnal Informatika Polinema*, 4(2), 93. <https://doi.org/10.33795/jip.v4i2.152>
- Suryono, S., Utami, E., & Luthfi, E. T. (2018). Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naive Bayes Classifier. *Seminar Nasional Geotik 2018*, 9–15.
- Syarifuddin, M. (2020a). Hasil Pencarian Pada Twitter Dengan Kata Kunci “COVID-19.” *Twitter.Com*. https://twitter.com/search?q=COVID-19&src=typed_query
- Syarifuddin, M. (2020b). *Laporan Akhir Penelitian Mandiri: Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn*.
- Wibowo, A., & Winarko, E. (2014). Paper Review: Data Mining Twitter. *Maintaining Cultural Heritage Through Information Technology for a Smart Future, November 2014*.