

KOMPARASI ALGORITMA *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES* DALAM ANALISIS SENTIMEN KEBIJAKAN PPKM

Rico Pramestiawan

Institut Informatika dan Bisnis Darmajaya

pramestiawanrico@gmail.com

ABSTRACT

Corona virus is an infectious disease that attacks human respiration. This virus originated in China, more precisely in Wuhan City, Hubei Province. early March 2020 this virus entered Indonesia and has spread throughout the world. In Indonesia there are 6,721,095 people exposed to Covid-19 and 160,648 people have died. This makes the government quickly make policies, in a policy there will be pros and cons in the community, so that sentiment analysis is needed to find out the response of the community regarding policies issued by the government, in this case the policy on the Implementation of PPKM.

Keywords: *Support Vector Machine algorithm; Naïve Bayes Algorithm; RapidMiner*

ABSTRAK

Virus corona merupakan penyakit menular yang menyerang pernafasan manusia. Virus ini bermula dari Negara China, lebih tepatnya di Kota Wuhan, Provinsi Hubei. awal maret tahun 2020 virus ini masuk ke Indonesia dan telah menyebar kelseluruh dunia. Di Indonesia terdapat 6.721.095 jiwa terpapar covid 19 dan 160.648 jiwa meninggal dunia. Hal ini membuat pemerintah dengan cepat membuat kebijakan, dalam sebuah kebijakan aka nada pro dan kontra dalam masyarakat, sehingga diperlukan analisis sentiment untuk mengetahui respon masyarakat terkait kebijakan-kebijakan yang dikeluarkan pemerintah, dalam hal ini adalah kebijakan tentang Penerapan PPKM.

Kata Kunci: *Algoritma Support Vector Machine; Algoritma Naïve Bayes; RapidMiner*

1. PENDAHULUAN

Virus corona adalah penyakit yang menular dan menyerang pernafasan. Virus ini berasal dari Negara China, yaitu di Kota Wuhan, Provinsi Hubei. Virus ini masuk ke Indonesia pada awal maret tahun 2020 dan menyebar keseluruh dunia (Firmansyah and Puspitasari 2021). Penyebaran virus corona terjadi dengan cepat sehingga belum ada cara untuk mengantisipasi penyebaran virus tersebut. Penyebaran virus corona semakin cepat dan korban mulai berjatuhan data yang didapatkan dari website resmi covid19.go.id terdapat 6.721.095 jiwa terpapar covid 19 dan 160.648 jiwa meninggal dunia. Sehingga dengan cepat pemerintah berupaya mengatasi virus tersebut. (Nurlia and Enri 2021)

Upaya yang dilakukan pemerintah adalah mengeluarkan peraturan dan kebijakan membatasi kegiatan masyarakat atau yang dikenal dengan PPKM, pembatasan ini berlaku hamper disemua kegiatan seperti kegiatan sekolah yang dilakukan dari rumah, kegiatan ibadah dan juga kegiatan bekerja semua dilakukan dari rumah. Selain itu juga pelaksanaan Vaksinasi juga terus digencarkan oleh sebagai upaya dalam mencegah penyebaran virus (Astari, Dewa Gede Hendra Divayana, and Gede Indrawan 2020). Dari kebijakan yang dikeluarkan pemerintah tentang penanggulangan covid 19 ini menimbulkan pro dan kontra pada masyarakat. Setiap kebijakan publik tidak bias terlepas dari opini masyarakat. Arah opini ini terbagi menjadi tiga kategori, opini positif, opini negatif, dan opini netral (Puspitasari, Ratnawati, and Widodo 2018).

Demikian dalam penerapan kebijakan PPKM, tidak lepas dari perdebatan dimasyarakat ada yang setuju dan juga tidak setuju. Di era yang modern ini opini public bisa dengan mudah didapatkan hanya dengan melalui media social. Sebagai salah satu media sosial yang sederhana, twitter memiliki pengaruh yang tinggi terhadap proses pembentukan dan pengarahannya opini publik. Melalui twitter kita akan banyak mendapatkan opini dari pengguna yang tidak hanya bersifat positif, tetapi juga tidak sedikit publik mengungkapkan opini negatif dengan menggunakan kata-kata yang dapat menimbulkan permusuhan, penghinaan, perdebatan dan perselisihan di media sosial khususnya Twitter(Saepudin et al. 2018).

Penggunaan twitter untuk mencari opini *public* ini dilakukan guna mencari tanggapan masyarakat terkait kebijakan public ataupun isu-isu yang sedang diperbincangkan masyarakat dan selanjutnya di analisis, kegiatan ini disebut dengan analisis sentiment atau *opinion mining*(Nugroho et al. 2021). Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis agar makna dari opini tersebut bias di identifikasi kedalam kelas positif, negative atau netral, agar mendapatkan informasi sentimen dari kalimat opini tersebut.

Dalam melakukan klasifikasi terdapat beberapa teknik yang bias digunakan diantaranya *Naïve Bayes* (NB), *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (k-NN) (Saepudin et al. 2018). Terdapat beberapa penelitian sebelumnya yang membahas analisis sentimen dokumen teks diantaranya adalah penelitian yang dilakukan oleh Taofik Krisdiyanto penelitian dilakukan dengan menggunakan metode *Naïve bayes* Clasifiers dengan jumlah dataset 1000 data ulasan, diperoleh hasil yaitu berupa sentimen positif sebanyak 99% dan sentimen negatif sebanyak 1%(Krisdiyanto 2021).

Selanjutnya penelitian yang dilakukan oleh aldiansyah et.al yang menggunakan Algoritma Suport Vector Machine untuk analisis sentiment kebijakan PPKM, dengan dataset yang digunakan berjumlah 2000 data dan didapatkan hasil akurasi sebesar 64% dengan kernel Linear dan penalti regularisasi $C = 1$. Dimana dari hasil prediksi terhadap data tweet non label, diketahui persebarannya adalah 57%-60% berlabel positif dan sisanya berlabel negatif(Putra et al. 2021).

Berdasarkan latar belakang tersebut dan beberapa penelitian terdahulu, penelitian ini bertujuan untuk mengelompokkan opini masyarakat tentang penerapan PPKM menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes*, sehingga dapat membantu para pemangku kebijakan dalam melakukan evaluasi terhadap program yang telah terealisasi dengan metode yang efektif dan akurat tanpa harus melakukan secara manual.(Dharmawan 2021)

2. KERANGKA TEORI

2.1. Teks Mining

Menurut Jiawei Han dalam Atang Saepudin, 2018 bahwa *Text mining* atau *text analytics* (penambangan teks) adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan data mining dimana data mining mengolah data yang sifatnya terstruktur(Saepudin et al. 2018). Pada dasarnya, *text mining* merupakan bidang interdisiplin yang mengacu pada perolehan informasi (*information retrieval*), *data mining*, pembelajaran mesin (*machine learning*), statistik, dan *komputasi linguistic*.

Tujuan utama *Text Mining* adalah mendapatkan informasi yang berguna dari data yang di olah. Permasalahan yang sering di temui dalam teks mining adalah data dalam jumlah besar, berdimensi tinggi, data yang berubah serta

noise dan hal ini merupakan tantangan dalam pengerjaan *text mining*. Didalam *text mining* terdapat fitur pendukung yang sering digunakan antara lain: (Astari, Dewa Gede Hendra Divayana, and Gede Indrawan 2020).

- a. **Character**, komponen individual yang ada dalam bagian *text mining* seperti huruf, angka, karakter special dan spasi, dan merupakan level paling tinggi dalam pembentukan *semantic feature*.
- b. **Word**, diartikan kata-kata yang dipilih secara langsung dari dokumen asli yang menjadi dasar atau tingkatan dasar, tetapi kadang *fitur word* terdapat dalam dokumen asli itu sendiri.
- c. **Term**, diartikan sebagai *single word* dan frasa multi word yang terpilih secara langsung dari korpus. *Representasi term-based* dari dokumen tersusun dari *subset term* dalam dokumen.
- d. **Concept**, merupakan *feature* yang degenerate dari sebuah dokumen secara manual, *rule-based*, atau metodologi lain.

2.2. Analisis Sentimen

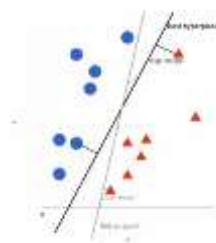
Analisis sentimen adalah sebuah proses untuk menentukan sentimen atau opini dari seseorang yang diwujudkan dalam bentuk teks dan bisa dikategorikan sebagai sentimen positif atau negative (Kevin et al. 2020), analisis sentiment bertujuan untuk menganalisa pendapat, penilaian dan emosi seseorang berkenaan dengan suatu topic atau produk maupun kegiatan tertentu. Analisis sentiment adalah proses pengelompokan teks yang ada pada sebuah kalimat atau dokumen kedalam sifat yang positif atau negatif.

Opinion mining atau analisis sentimen merupakan suatu bidang ilmu dari data mining yang berguna untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu (Faisal et al. 2020). Analisis ini bertujuan untuk mendapatkan sebuah informasi dari suatu himpunan data. Analisis sentimen adalah penelitian yang baru pada *Natural Language Processing* (NLP) dan bertujuan menemukan subjektivitas dalam teks maupun mengekstraksi dan menjalankan klasifikasi sentimen pada opini (Kevin et al. 2020).

2.3. Algoritma Support Vector Machine

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory* (Giovani et al. 2020). Konsep dasar metode SVM sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang telah ada pada tahun sebelumnya, seperti margin *hyperplane* (Dyda dan Hart, 1973; Cover, 1965; Vapnik, 1964), kernel diperkenalkan oleh Aronszajn tahun 1950, Lagrange Multiplier yang ditemukan oleh Joseph Louis Lagrange pada tahun 1766, dan demikian juga dengan konsep-konsep pendukung lain (Normah et al. 2022).

Support Vector Machine bekerja dengan mencari dan menentukan nilai *hyperplane* (fungsi pemisah kelas) terbaik. Nilai *hyperplane* terbaik adalah nilai yang terletak di tengah-tengah antara dua set obyek dari dua kelas. Mencari *hyperplane* terbaik berbanding lurus dengan memaksimalkan margin. Margin adalah jarak tegak lurus antara *hyperplane* dengan obyek terdekat sebagaimana ditunjukkan pada Gambar 1 (Nawawi, Purnama, and Hikmah 2019).



Gambar 1. *Support Vector Machine*

2.4 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan metode yang dapat digunakan untuk mengklasifikasikan sekumpulan data. Algoritma ini memanfaatkan metode probabilitas dan Statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya (Fuad Nur Hasan 2018).

Naïve Bayes merupakan *machine learning* yang menggunakan perhitungan probabilitas yang menggunakan konsep pendekatan *Bayesian*. Kata *Naïve*, yang terkesan merendahkan, berasal dari asumsi *independensi* pengaruh nilai suatu atribut dari probabilitas pada kelas yang diberikan terhadap nilai atribut lainnya. Penggunaan teorema Bayes pada algoritma *Naïve Bayes* yaitu dengan mengkombinasikan prior probability dan probabilitas bersyarat dalam sebuah rumus yang bisa digunakan untuk menghitung probabilitas tiap klasifikasi yang mungkin.

Rumus *Naïve Bayes* adalah (Utomo and Mesran 2020):

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

Keterangan:

X = Data dengan class yang belum diketahui (bukti)

H = Hipotesis data X merupakan suatu class spesifikasi

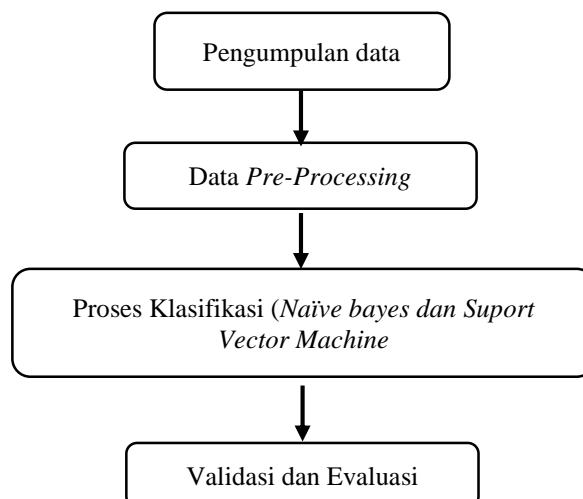
$P(H|X)$ = Probabilitas hipotesis H benar untuk kondisi X

$P(H)$ = Probabilitas hipotesis H

PX = Probabilitas prior bukti X

3. METODOLOGI

Dalam sebuah penelitian, terdapat model metode penelitian. Model Metode Penelitian pada penelitian ini dibuat sebagai alur perjalanan dari penelitian ini yang selanjutnya akan diolah untuk memudahkan dalam melakukan penelitian. Model metode penelitian dapat dilihat pada Gambar 2.

**Gambar 2.** Tahapan Penelitian

3.1 Pengumpulan Data

Pengumpulan data penelitian dilakukan terhadap komentar pengguna media sosial Twitter dengan tagar atau #PPKM yang diambil dari https://www.kaggle.com/code/mochkholil/ppkm-sentiment-classification/data?select=ppkm_test.csv data bersifat public.

3.2 Pengolahan Awal Data

Tahapan yang dilakukan terhadap dokumen selanjutnya adalah text preprocessing. Hal ini dilakukan guna menormalisasi data yang digunakan dalam proses analisis sentimen. Proses pertama yang dilakukan yaitu:

- Proses *Case Folding*. Pada proses ini akan dilakukan penghapusan atau penghilangan karakter-karakter pada dokumen yang tidak dibutuhkan, yang mana hal itu dapat menimbulkan noise seperti misalnya emotikon, tanda baca dan lain sebagainya. Selain itu pada proses ini akan dilakukan pengubahan huruf dalam dokumen menjadi huruf kecil.
- Proses *Tokenizing*. Pada proses ini dilakukan pemotongan/pemisahan setiap kata dalam teks yang disebut sebagai token.
- Stopword Removal* dimana pada tahap ini akan dihilangkan kata-kata yang tidak penting seperti kata “di”, “dan”, “karena” “oleh” dan lain sebagainya. Tahap ini dilakukan agar dapat memperbesar nilai akurasi.
- Stemming*. Pada proses ini akan dilakukan pengubahan kata yang ada dalam dokumen menjadi bentuk kata dasar. Sebagai contoh, pengubahan kata “himbauan” menjadi “himbau”, “larangan” menjadi “larang”, “penyebaran” menjadi “sebar” dan lain sebagainya

4. HASIL DAN PEMBAHASAN

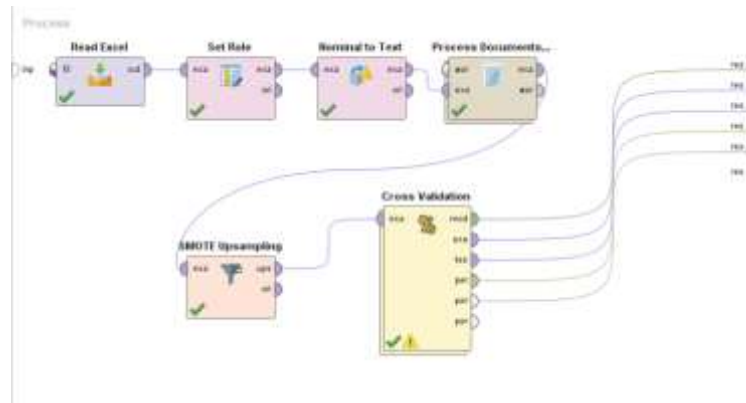
Dalam bab ini penulis akan membahas mengenai hasil dari penelitian, data tersebut akan dihitung menggunakan algoritma *Support Vector Machine* dan Algoritma Nive bayes yang kemudian akan diuji menggunakan *CrosValidation* menggunakan algoritma *Decision Support Vector Machine* dan Algoritma Naive bayes Dataset yang digunakan merupakan dataset publik yang berasal dari https://www.kaggle.com/code/mochkholil/ppkm-sentiment-classification/data?select=ppkm_test.csv Dataset dapat dilihat pada Gambar 3.

	A	B
1	Teks	Label
2	Kami siap laksanakan Instruksi pak	Positif
3	.Siap melaksanakan intruksi pak	Positif
4	.Siap dukung dan sukseskan	Positif
5	.Langkah 3M ini sudah sukses di Bali memutus penyebaran...	Positif
6	.Siap amankan seluruh kebijakan kementerian Desa PDTT	Positif
7	.Siap utk di sosialisasikan ke setiap TPP yg ada di wilayah ka...	Positif
8	.Mendukung kebijakan Gus Menteri dalam upaya pencapaian ...	Positif
9	.Mari bersama cegah penyebaran covid-19	Positif
10	.Mari kita sukseskan pphm di masyarakat	Positif
11	.Siap kawal dan amankan kebijakan kemendes palaksanaan...	Positif
12	.Semoga pak gubernur terus diberikan kekuatan untuk lanjut...	Positif
13	.Semoga bapak selalu diberi kesehatan amin... salam dari ko...	Positif

Gambar 3. Dataset PPKM

4.1 Pengujian Algoritma *Support Vector Machine*

Pengujian data yang digunakan dalam algoritma *Support Vector Machine* adalah dataset ppkm yang didapatkan dari Crawling Twiter, selanjutnya data diproses dengan menggunakan Software Rapid Miner. Secara umum Proses pengujian menggunakan algoritma *Support Vector Machine* dijelaskan pada gambar berikut ini:



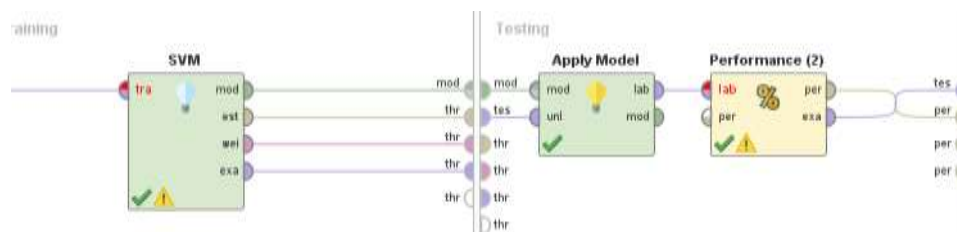
Gambar 4. Proses

Akan tetapi sebelum data di uji menggunakan Algoritma *Support Vector Machine* terlebih dahulu data dilakukan *Preprocessing* data, proses *Preprocessing* data dijelaskan pada Gambar 5.



Gambar 5. Tahapan *Preprocessing* Data

Pengimplementasian algoritma *Support Vector Machine* pada Rapidminer untuk klasifikasi dengan menggunakan *cross validation* dengan nilai akurasi, nilai *recall*, presisi dan nilai AUC dapat dilihat pada Gambar 6.



Gambar 6. Implementasi Algoritma *Support Vector Machine*

accuracy: 86.00% +/- 7.75% (micro average: 86.00%)

	true Positif	true Negatif	class precision
pred. Positif	96	24	80.00%
pred. Negatif	4	76	95.00%
class recall	96.00%	76.00%	

Gambar 7. Tabel Akurasi, Recall, dan Presisi



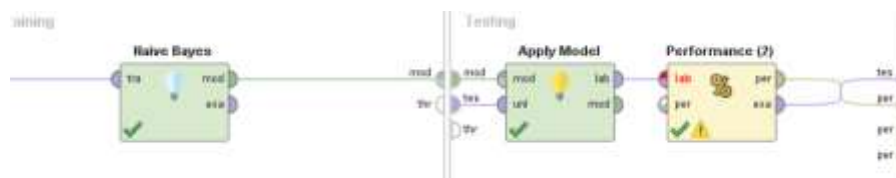
Gambar 8. Grafik ROC-AUC

4.2 Pengujian Algoritma *Naïve Bayes*

Pengujian data selanjutnya yaitu menggunakan algoritma *Naïve Bayes*. Secara umum Proses pengujian menggunakan algoritma *Naïve Bayes* dijelaskan pada Gambar 9.



Gambar 9. Tahapan *Preporocessing* Data



Gambar 10. Implementasi Algoritma *Naïve Bayes*

accuracy: 90.50% +/- 7.25% (micro average: 90.50%)

	true Positif	true Negatif	class precision
pred. Positif	91	10	90.10%
pred. Negatif	9	90	90.91%
class recall	91.00%	90.00%	

Gambar 11. Tabel Akurasi, *Recall*, dan Presisi**Gambar 12.** Grafik ROC-AUC

4.3 Perbandingan Hasil Algoritma *Support Vector Machine* dan *Naive Bayes*

Setelah dilakukan pengujian terhadap masing-masing algoritma, selanjutnya yaitu membandingkan hasil akurasi dari masing-masing algoritma yang dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Akurasi

No	Algoritma	Akurasi	AUC
1	<i>Support Vector Machine</i>	86,00%	0,966
2	<i>Naive Bayes</i>	90,50%	0,771

Dari hasil pengujian pada penelitian ini, dapat dilihat bahwa performansi akurasi algoritma *Naive Bayes* lebih tinggi jika dibandingkan dengan algoritma *Support Vector Machine*. Dari hasil yang didapatkan masing-masing algoritma menunjukkan performa yang baik dan tingkat akurasi yang tinggi dalam melakukan analisis *sentiment*.

5. KESIMPULAN

Dari pengukuran kinerja dengan melakukan komparasi dua algoritma yang telah dilakukan berdasarkan jumlah data maka dapat disimpulkan bahwa algoritma *Naive Bayes* memiliki kemampuan dalam melakukan analisis sentimen. Adapun hasil penelitian dari percobaan yang telah dilakukan dapat disimpulkan bahwa:

1. Algoritma *Support Vector Machine* memiliki tingkat akurasi sebesar 86,00% sedangkan *Naive Bayes* sebesar 90,50%. Berdasarkan hasil penelitian ditunjukkan bahwa algoritma *Naive Bayes* memiliki tingkat akurasi lebih baik dari algoritma *Support Vector Machine*.
2. Model algoritma *Support Vector Machine* memiliki AUC sebesar 0,966 dan algoritma *Naive Bayes* sebesar 0,771.

DAFTAR PUSTAKA

Astari, Ni Made Ayu Juli, Dewa Gede Hendra Divayana, and Gede Indrawan. 2020. "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier." *Jurnal Sistem dan*

Informatika (JSI) 15(1): 27–29.

Dharmawan, Weiskhy Steven. 2021. “I N F O R M a T I K a Dalam Prediksi Penyakit Jantung.” *Jurnal Informatika, Manajemen dan Komputer* 13(2): 31–41.

Faisal, Anas, Yuris Alkhalifi, Achmad Rifai, and Windu Gata. 2020. “Analisis Sentimen Dewan Perwakilan Rakyat Dengan Algoritma Klasifikasi Berbasis Particle Swarm Optimization.” *JOINTECS (Journal of Information Technology and Computer Science)* 5(2): 61.

Firmansyah, Zulfikar, and Nila Feby Puspitasari. 2021. “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes.” *Jurnal Teknik Informatika* 14(2): 171–78.

Fuad Nur Hasan. 2018. “No Title.” 3(November).

Giovani, Angelina Puput et al. 2020. “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi.” *Jurnal Teknoinfo* 14(2): 115.

Kevin, Valentino, Sitanayah Que, Ade Iriani, and Hindriyanto Dwi Purnomo. 2020. “Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization).” 9(2): 162–70.

Krisdiyanto, Taofik. 2021. “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers.” *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi* 7(1): 32.

Nawawi, Hendri Mahmud, Jajang Jaya Purnama, and Agung Baitul Hikmah. 2019. “Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung.” *Jurnal Pilar Nusa Mandiri* 15(2): 189–94.

Normah, Bakhtiar Rifai, Satrio Vambudi, and Rifki Maulana. 2022. “Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE.” *Jurnal Teknik Komputer AMIK BSI* 8(2): 174–80. <https://ejournal.bsi.ac.id/ejurnal/index.php/jtk/article/view/13041/pdf>.

Nugroho, Dian Ari et al. 2021. “Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan SVM.” *Seminar Nasional Teknologi Fakultas Teknik Universitas Krisnadwipayana* 3(1): 1–11.

Nurlia, Elin, and Ultach Enri. 2021. “Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5.” *Jurnal Teknik Informatika Musirawasa Elin Nurlia* 6(1): 42.

Puspitasari, Ana Mariyam, Dian Eka Ratnawati, and Agus Wahyu Widodo. 2018. “Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine.” *J-Ptiik* 2(2): 802–10.

Putra, Aldiansyah, Dede Haeirudin, Hasna Khairunnisa, and Retnani Latifah. 2021. “Analisis Sentimen Masyarakat

Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma Svm.” *Seminar Nasional Sains dan Teknologi 2021* (November): 1–6.

Saepudin, Atang et al. 2018. “KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR BERBASIS PARTICLE SWARM OPTIMIZATION PADA ANALISIS.”

Utomo, Dito Putro, and Mesran Mesran. 2020. “Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung.” *Jurnal Media Informatika Budidarma* 4(2): 437.