

## Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma *Support Vector Machine* dan *Naive Bayes*

Hendrik Setiawan<sup>1\*</sup>, Ema Utami<sup>2</sup>, Sudarmawan<sup>3</sup>  
<sup>1,2,3</sup>Magister Teknik Informatika, Universitas Amikom Yogyakarta  
\*email: [Hendrik.1305@students.amikom.ac.id](mailto:Hendrik.1305@students.amikom.ac.id)

DOI: <https://doi.org/10.31603/komtika.v5i1.5189>

Date received: 22-06-2021, Last Revised: 10-07-2021, Accepted: 13-07-2021

### ABSTRACT

The World Health Organization (WHO) COVID-19 is an infectious disease caused by the Coronavirus which originally came from an outbreak in the city of Wuhan, China in December 2019 which later became a pandemic that occurred in many countries around the world. This disease has caused the government to give a regional lockdown status to give students the status of "at home" for students to enforce online or online lectures, this has caused various sentiments given by students in responding to online lectures via social media twitter. For sentiment analysis, the researcher applies the naive Bayes algorithm and support vector machine (SVM) with the performance results obtained on the Bayes algorithm with an accuracy of 81.20%, time 9.00 seconds, recall 79.60% and precision 79.40% while for the SVM algorithm get an accuracy value of 85%, time 31.60 seconds, recall 84% and precision 83.60%, the performance results are obtained in the 1st iteration for naive Bayes and the 423th iteration for the SVM algorithm.

**Keywords:** Covid-19, Online Lecture, Twitter, Naïve Bayes, Support Vector Machine

### ABSTRAK

World Health Organization (WHO) COVID-19 merupakan penyakit menular yang disebabkan oleh Coronavirus yang awal mulanya berasal dari wabah di kota Wuhan, Tiongkok pada bulan Desember 2019 yang kemudian menjadi pandemi yang terjadi di banyak negara di seluruh Dunia. Penyakit ini menyebabkan pemerintah memberikan status penguncian daerah (*lockdown*) memberikan status "dirumahkan" terhadap pelajar dan mahasiswa untuk memberlakukan kuliah online atau daring, hal ini menyebabkan berbagai sentimen yang diberikan oleh mahasiswa dalam menanggapi kuliah online lewat sosial media twitter. Untuk analisis sentimen peneliti menerapkan algoritma *naïve bayes* dan *support vector machine* (SVM) dengan hasil peforma yang didapat pada algoritma *bayes* akurasi 81,20%, waktu 9,00 detik, *recall* 79,60% dan presisi 79,40% sedangkan untuk algoritma SVM mendapatkan nilai akurasi 85%, waktu 31,60 detik, *recall* 84% dan presisi 83,60%, hasil peforma tersebut diperoleh pada iterasi ke 1 untuk *naïve bayes* dan iterasi ke 423 untuk algoritma SVM.

**Kata-kata kunci:** Covid-19, Kuliah Online, Twitter, Naïve Bayes, Support Vector Machine

### PENDAHULUAN

Menurut World Health Organization (WHO) COVID-19 adalah salah satu penyakit menular yang disebabkan oleh Coronavirus yang awal mulanya berasal dari wabah di kota Wuhan, Tiongkok pada bulan Desember 2019 yang kemudian menjadi pandemi yang terjadi di banyak negara di seluruh Dunia. Pandemi COVID-19 yang dialami negara-negara di dunia memberikan dampak yang sangat besar, Penerapan status penguncian daerah (*Lockdown*) menghentikan aktifitas masyarakat, baik dari lembaga pemerintahan, perusahaan swasta, wirausaha, transportasi, pariwisata, pendidikan, dan banyak lagi sektor lain yang terkena imbasnya dari penerapan ini [1].

Sulit untuk mengatakan suatu sektor merupakan sektor yang paling berat terdampak COVID-19. Hal ini disebabkan bahwa Pandemi COVID-19 hampir memberi dampak pada semua sektor. Pemberian status penguncian daerah (*lockdown*) memberikan status "dirumahkan" terhadap pelajar dan mahasiswa tentu saja hal ini memberikan dampak yang cukup berat. Sehingga untuk membantu pelajar dan mahasiswa dalam kegiatan belajar kementerian pendidikan memutuskan untuk dilakukan dengan cara jarak jauh atau yang biasa kita kenal dengan kuliah online [2].

Melalui perkembangan teknologi banyak orang yang menuangkan opini dan ekspresi mereka melalui teknologi seperti sosial media, salah satu sosial yang banyak digunakan oleh masyarakat dalam menuangkan opini dan ekspresi adalah sosial media Twitter, dimana pada Twitter masyarakat dapat menulis opini mereka dalam postingan untuk dapat dilihat banyak orang. Berdasarkan lembaga survey Twitter pada tahun 2019 menyatakan bahwa pada tahun 2018 twitter mendapatkan jumlah pengguna aktif sebanyak 126 juta dengan total pertumbuhan pengguna berkisar 9% disetiap tahunnya. Jumlah ini yang mendorong banyak masyarakat mulai menuangkan opini dan pendapat mereka pada sosial media twitter dengan tujuan agar dapat dilihat banyak orang [3].

Postingan opini dan pendapat masyarakat pada Twitter dapat dijadikan sebagai instrumen untuk mengukur tingkat kepuasan masyarakat mengenai belajar online, dimana postingan tersebut akan digunakan untuk mengukur sentimen masyarakat melalui 3 kategori kelas yaitu positif, netral dan negatif, dimana melalui kategori tersebut, dapat dijadikan untuk bahan evaluasi untuk pemerintah dalam melakukan kebijakan kedepan khususnya dalam kebijakan kuliah online [4].

Analisis sentimen merupakan salah satu metode yang dapat mengukur dan menganalisis pada suatu kasus atau objek tertentu, dimana melalui sentimen analisis tersebut dapat diambil sebuah kesimpulan dan keputusan berdasarkan teks dalam bentuk kalimat ataupun dokumen. Salah satu algoritma yang dapat digunakan untuk melakukan sentimen analisis adalah Naïve Bayes dimana pada algoritma ini sistem akan melakukan klasifikasi melalui probabilitas kemungkinan dari data yang diperoleh [3], algoritma lain yang dapat digunakan dalam teknik analisis sentimen adalah Support Vector Machine (SVM), dimana pada algoritma tersebut proses pembobotan dilakukan melalui pembentukan patern garis untuk dilakuan pemobobotan dan klasifikasi [5].

Berdasarkan latar belakang diatas maka pada penelitian ini akan mencoba melakukan analisis sentimen pada sosial media twitter untuk mengukur opini dan pendapat masyarakat mengenai kuliah online pasca covid-19 dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM).

## STUDI LITERATUR

Penelitian Deni et al (2020) yang menganalisis terkait perbandingan Algoritma *Support Vector Mechine* dengan *Navive Bayes* kepada calon gubernur jawa barat. Pada penelitian ini, penulis menggunakan data analisis sentiment pada twitter yang kemudian di kelompokkan menjadi data positif serta negative. Berdasarkan hasil pengolahan data Algoritma *Support Vector Mechine* medapatkan nilai rata-rata akurasi yang lebih tinggi dibandingkan dengan Algoritma *Naïve Bayes*, 93,03% dengan 92,85%. Dan dapat disimpulkan bahwa dalam

penelitian ini metode Algoritma *Support Vector Machine* lebih akurat dibandingkan dengan Algoritma *Naïve Bayes* [6].

Pada penelitian Yuris et al (2020) yang membahas mengenai Algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) yang berbasis *Particel Swarm Optimization* (PSO) pada analisis sentiment penghapusan ujian nasional. Penelitian dimulai dengan menggumpulkan data sentiment twitter yang kemudian diproses dengan menggunakan 4 metode yang berbeda yakni SVM tanpa menggunakan PSO, SVM dengan menggunakan PSO, NB tanpa menggunakan PSO serta NB dengan menggunakan PSO. Mendapatkan kesimpulan bahwa SVM dengan menggunakan PSO memiliki nilai akurasi tertinggi dibandingkan dengan 3 metode lainnya dengan data akurasi 92,92%, sedangkan nilai akurasi lainnya 94,81% SVM tanpa menggunakan PSO, 85,9% NB tanpa menggunakan PSO dan 86,92% NB menggunakan PSO [7].

Pada penelitian Samsir et al (2021) yang membahas tentang Pembelajaran Daring pada Masa Pandemi. Pada penelitian ini penulis menggunakan analisis *text mining* pada data yang dihasilkan dari analisis sentiment twitter, kemudian mengklasifikasikan data menggunakan metode *Naïve Bayes* dengan tahapan *Data Crawling, Processing, Extrasi Fitur, dan Naïve Bayes Classifier*. Dari perhitungan data didapatkan 8,9 ribu twit bernilai negative, 2,8 ribu twit bernilai positif dan 134 ribu twit bernilai netral, kemudian dari data tersebut diklasifikasi kembali sehingga mendapatkan data akurasi yang dapat dijadikan kesimpulan pada penelitian. 3830 atau 30 % untuk indifikasi nilai positif, 8943 atau 69% indifikasi dari nilai Negatif dan 134 atau 1% indifikasi nilai netral dari 12906 data yang ada. Dapat disimpulkan bahwa terdapat ketidak puasan terhadap pembelajaran secara daring menurut hasil dari penelitian yang telah dilakukan [8].

Penelitian Efid dan Iqbal (2021) membahas analisis sentiment warga China disaat Pandemi. Penelitian ini menggunakan Algoritma *Term Frequency – Invers Document Frequency* dan juga *Support Vector Machine*. Tahapan penelitian yang dilakukan masih sama seperti penelitian lainnya dengan menggunakan data *sentiment* yaitu dimulai dengan pengumpulan data, pemelahan data kemudian *text processing*. Dari hasil pengolahan data ditemukan prediksi positif sebesar 342 data dan prediksi *negative* 370 data dalam perhitungan evaluasi akurasi. Kemudian tidak hanya selesai disitu, untuk lebih memastikan keakuratan data atau mengukur tingkat akurasi tahapan selanjutnya yaitu dengan melakukan perhitungan hasil *recall* dan *precision*. Dengan hasil akhir 89.04% sentiment positif dan 92% *sentiment negative*. Pada penelitian ini dapat di tarik kesimpulan bahawa menganalisis menggunakan sentiment twitter yang kemudian diproses lebih lanjut menggunakan SVM dapat memperoleh hasil atau kesimpulan yang lebih akurat, akan tetapi. Kekurangan pada penelitian ini penulis tidak menyebutkan berapa data awal yang digunakan [9].

Penelitian Sodik dan Iqbal (2021) mengenai tanggapan masyarakat Indonesia mengenai Covid-19 dengan menggunakan metode SVM, *Naïve Bayes* dan KKN melalui analisis sentimen media Twitter. Dari 10000 data yang di dapatkan 6128 sentimen positif dan 3875 *sentiment negative*, setelah mendapatkan data kemudian dilakukan perhitungan dengan menggunakan SVM, *Naïve Bayes* dan KKN dengan menggunakan validasi model 10-fold cross validation. Berdasarkan perhitungan dari masing – masing model mendapatkan kesimpulan SVM 90,1%, *Naïve Bayes* 72,9% dan KKN 62,1%. Dari hasil 3 metode tersebut

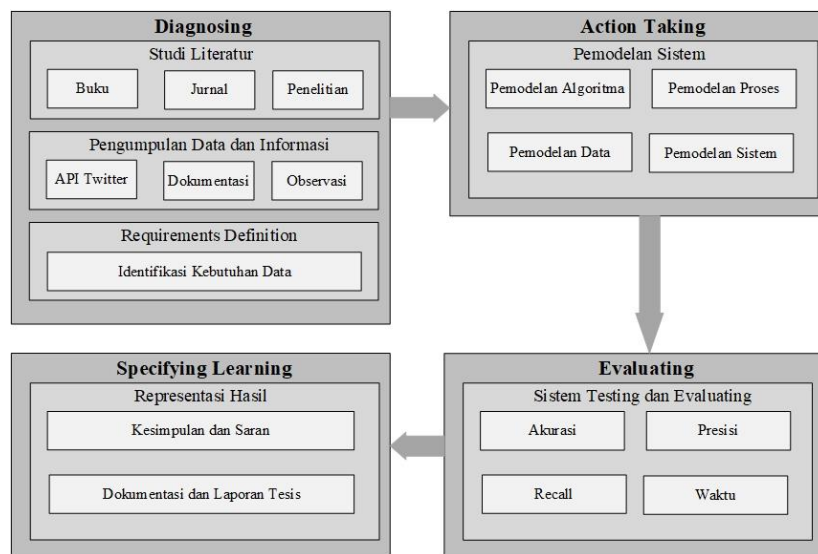
dapat disimpulkan bahawa dalam penelitian ini metode SVM merupakan metode yang paling akurat diantara 2 metode lainnya [10].

## METODE

Metode penelitian ini dilakukan pada beberapa proses yaitu diagnosing, action taking, specifying learning dan evaluating, proses ini akan dilakukan untuk mendapatkan hasil terbaik dalam melakukan sentimen analisis pada data twitter melalui kata kunci kuliah online pasca covid-19, dengan detail sebagai berikut:

### Alur Penelitian

Alur penelitian mencakup proses yang saling terkait. Alur sistem sentimen analisis pada data twitter melalui kata kunci kuliah online pasca covid-19. Metode atau alur penelitian seperti pada Gambar 1.



Gambar 1. Metode Penelitian

#### 1. *Diagnosing*

*Diagnosing* dilakukan melalui proses melakukan rievew pada peneliaian sebelumnya baik dalam buku ataupun jurnal, selanjutnya dilakukan proses pengumlan data untuk pembentukan informasi melalui observasi pada dokumentasi API Twitter dan kemudian dilanjutkan proses pendefinisian kebutuhan dari sistem yang akan dibuat.

#### 2. *Action Taking*

*Action taking* merupakan proses perencanaan penelitian yang akan dibuat, perencanaan akan dilakukan melalui pemodelan sistem ataupun pemodelan algoritma dalam bentuk detail poses dari mulai hingga selesai.

#### 3. *Evaluating*

*Evaluating* merupakan proses untuk mengukur peforma dari setiap algoritma untuk dibandingkan melalui parameter waktu, recall, presisi dan akurasi ke dalam bentuk visual.

#### 4. Specifying Learning

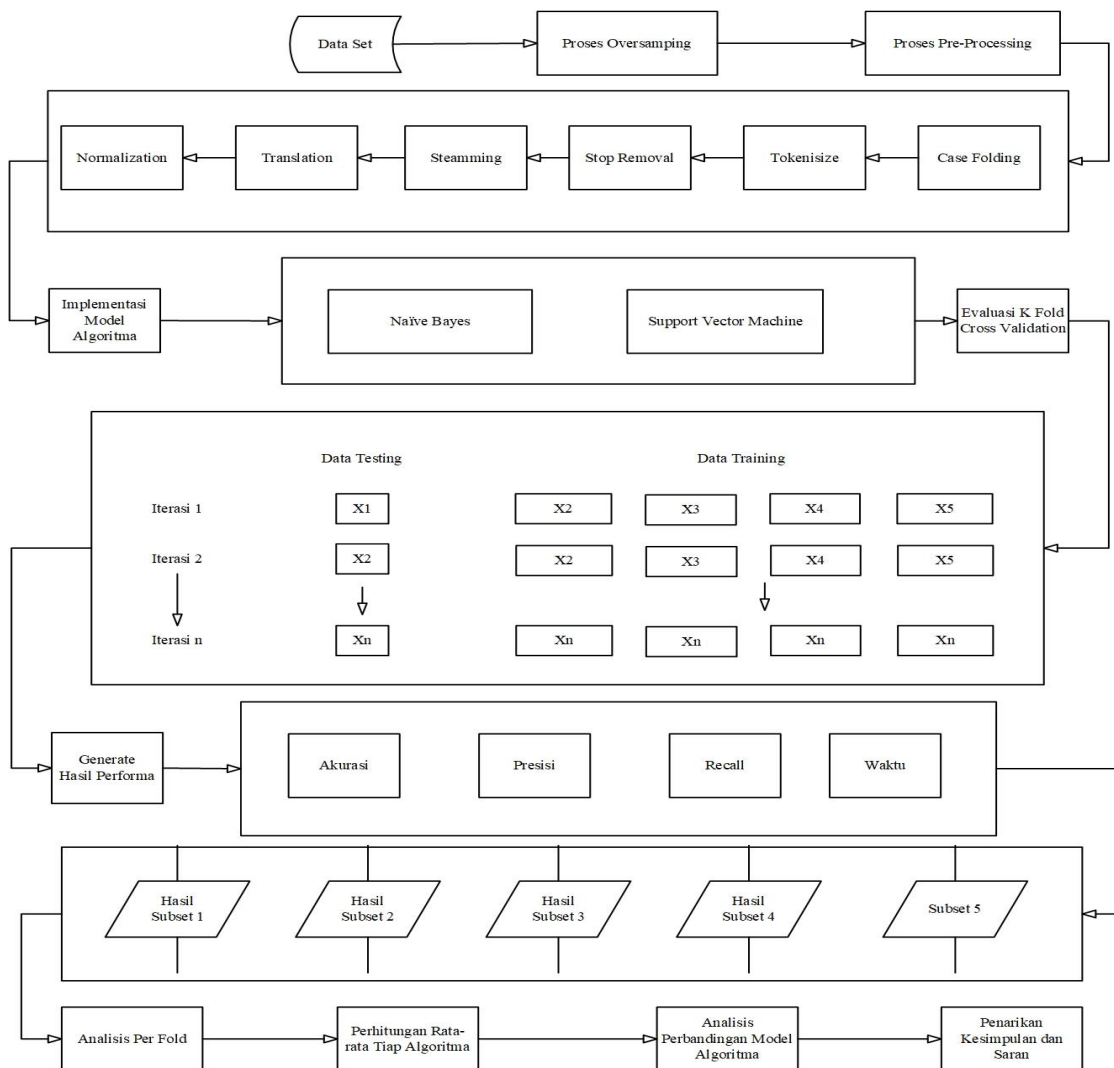
Specifying learning merupakan proses membuat dokumentasi penelitian unntuk dapat dipublikasikan.

#### Metode Pengumpulan Data

Metode pengumpulan data dilakukan melalui proses scaraping data, dimana scaraping data dilakukan melalui *Application Programming Integration* (API) twitter dengan mekanisme pencarian berdasarkan keyword yaitu kuliah online. Teknik scaraping data dilakukan dengan menggunakan bahasa pemrograman PHP melalui proses Twitter authentication untuk mendapatkan akses ke *twitter*, *proses authentication* dilakukan dengan melakukan konfigurasi *API secret*, *API key*, *access token secret* dan *access token* untuk proses perizinannya.

#### Metode Analisis Data

Analisi data akan dilakukan melalui beberapa proses yang akan dilakukan dengan detail seperti pada Gambar 2.



Gambar 2. Alur Analisis data pada Algoritma

Detail penjelasan analisis data pada Gambar 2 adalah sebagai berikut:

1. *Oversampling*

Pada tahap ini akan dilakukan pendefinisian dari data, dimana data didapatkan melalui API Twitter dengan dengan parameter kuliah online dengan pembagian kelas positif, netral dan negatif. Data yang didapatkan pada proses ini adalah 350 data.

2. *Pre-processing*

Tahap ini akan dilakukan proses pembersihan data dari kata-kata yang tidak valid untuk digunakan melalui proses *case folding* untuk membentuk standar kalimat, *tokenize* memisahkan kalimat menjadi satu per satu kata, *stop removal* untuk membuang kata yang tidak digunakan, *stemming* untuk mendapatkan kata dasar, *transformation* untuk membentuk data kedalam format *database* dan *normalization* untuk menormalkan data.

3. Implementasi algoritma

Pada tahap ini algoritma *naïve bayes* dan SVM akan diimplementasikan secara bergantian, akan tetapi walaupun proses implementasi dilakukan bergantian, kedua algoritma akan mengalami proses yang sama untuk mekanisme pengujiannya.

4. Evaluasi *K-Fold Cross Validation*

*K-Fold Cross Validation* akan membagi data menjadi menjadi 5 fold atau 5 mekanisme pengujian.

5. Analisis Hasil

Setelah melewati hasil validasi pengujian menggunakan *K-Fold cross validation* maka selanjutnya akan diambil kesimpulan perbandingan performa setiap algoritma melalui parameter presisi, waktu, *recall* dan akurasi.

## HASIL DAN PEMBAHASAN

Pada bagian ini hasil percobaan dari implementasi *Naive Bayes* serta *Support Vector Machine* akan dilakukan perbandingan. Hasil yang akan dibandingkan merupakan rata-rata terbaik dari keseluruhan percobaan pengujian dengan parameter Akurasi, *Presisi*, *Recall* dan waktu training yang telah dilakukan pada sub bab sebelumnya.

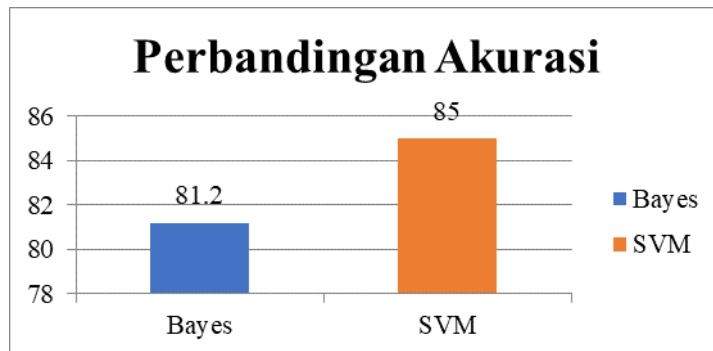
Sebelum melakukan proses pengujian terlebih dahulu dilakukan beberapa proses yang *cleansing* data, dimana pada tahap ini data yang didapat sebanyak 350 data akan dilakukan pembersihan untuk menghindari *error* atau *missing* pada proses pengujian, proses *cleansing* dilakukan dengan mengecek *missing value* pada data yang diperoleh, dan setelah dilakukan pada 350 data terdapat 8 data yang mengalami *missing value* yaitu terjadi kerusakan yang menjadikan data tersebut tidak valid, sehingga diperoleh 342 data untuk dilakukan pengujian.

Setelah dilakukan *cleansing* data dan diperoleh data final sebanyak 342 data maka langkah selanjutnya adalah menguji performa kedua algoritma melalui data tersebut. Hasil terbaik dari algoritma bayes terdapat pada 1 kali iterasi perulangan dikarenakan algoritma tersebut melakukan proses pembobotan berdasarkan probabilitas data yang ada serta algoritma SVM mendapatkan hasil terbaik pada iterasi ke 423 dikarenakan algoritma tersebut memerlukan proses training data untuk mendapatkan nilai performa terbaik dengan detail perbandingan seperti pada Tabel 1.

Tabel 1. Hasil Perbandingan Tingkat Performa

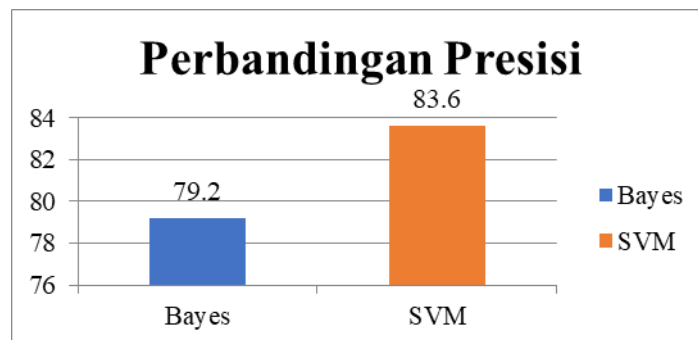
No	Algoritma	Akurasi (%)	Presisi(%)	Recall (%)	Waktu (Detik)
1	Bayes	81,20	79,40	79,60	9,00
2	SVM	85	83,60	84	40,60
	Selisih	3,80	4,20	4,40	31,60

Berdasarkan hasil pada Tabel 1 dapat ditarik suatu kesimpulan bahwa nilai performa untuk akurasi model algoritma *Support Vector Machine* memiliki tingkat akurasi yang lebih baik dengan nilai akurasi sebesar 85% dengan selisih 3,80% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 81,20% pada konfigurasi dataset sebanyak 342 data. Adapun grafik dari perbandingan nilai akurasi pada Gambar 3.



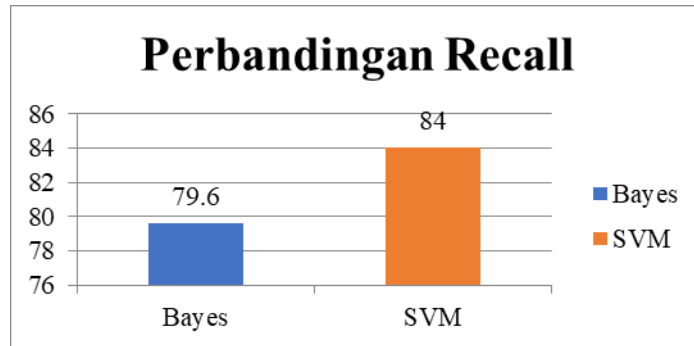
Gambar 3. Perbandingan Performa Akurasi

Kemudian untuk nilai tingkat performa presisi *Support Vector Machine* memiliki lebih baik dengan nilai presisi sebesar 83,60% dengan selisih 4,20% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 79,40% pada konfigurasi dataset sebanyak 342 data. Adapun grafik dari perbandingan nilai presisi pada Gambar 4.



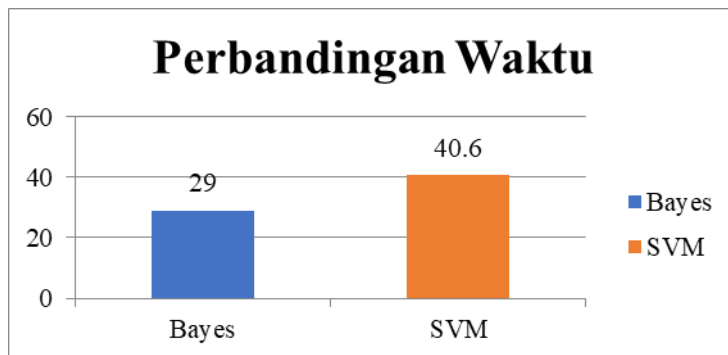
Gambar 4. Perbandingan Performa Presisi

Kemudian untuk nilai tingkat performa *recall Support Vector Machine* memiliki lebih baik dengan nilai presisi sebesar 84% dengan selisih 4,40% jika dibandingkan dengan model algoritma *Naive Bayes* dengan nilai yang lebih rendah sebesar 79,60% pada konfigurasi dataset sebanyak 342 data. Adapun grafik dari perbandingan nilai *recall* pada Gambar 5.



Gambar 5. Perbandingan Performa Recall

Kemudian untuk nilai tingkat performa waktu training *Naive Bayes* memiliki lebih baik dengan nilai waktu sebesar 9,00 detik dengan selisih 31,60 detik jika dibandingkan dengan model algoritma *Support Vector Machine* dengan nilai yang lebih rendah sebesar 40,60 detik pada konfigurasi dataset sebanyak 342 data. Adapun grafik dari perbandingan nilai waktu training pada Gambar 6.



Gambar 6. Perbandingan Performa Waktu

Berdasarkan hasil pada Tabel 1 dapat disimpulkan bahwa algoritma SVM lebih unggul dalam hal presisi, recall dan akurasi dikarenakan pada proses training algoritma melakukan perulangan sebanyak 423 iterasi untuk mendapatkan pembobotan yang optimal, sedangkan algoritma *naïve bayes* melakukan proses pembobotan sebanyak 1 iterasi sehingga hasil yang didapat kurang optimal jika dibandingkan dengan algoritma SVM, akan tetapi dikarenakan proses iterasi tersebut algoritma *naïve bayes* mendapatkan hasil lebih cepat dalam segi waktu jika dibandingkan dengan algoritma SVM.

## KESIMPULAN

Kesimpulan pada penelitian ini adalah hasil terbaik algoritma *naïve bayes* didapatkan pada iterasi ke 1 dikarenakan pada algoritma *naïve bayes* melakukan proses pencarian bobot dengan melakukan perhitungan probabilitas pada data dan mendapatkan hasil akurasi 81,20%, waktu 9,00 detik, *recall* 79,60% dan presisi 79,40%. Untuk hasil terbaik pada algoritma SVM didapatkan pada iterasi ke 423 dikarenakan proses training data pada algoritma SVM untuk mendapatkan hasil optimal dengan hasil akurasi 85%, waktu 31,60 detik, *recall* 84% dan presisi 83,60%.



## DAFTAR PUSTAKA

- [1] WHO. 2020. Coronavirus disease 2019 (COVID-19) situation report-94. WHO.
- [2] Permenkes 9 tahun (2020). 2020. Pedoman PSBB dalam rangka Percepatan Penanganan COVID-19.
- [3] Ibrahim Moge Noor, Metin Turan. 2020. Sentiment Analysis using Twitter Dataset. IJID (International Journal on Informatics for Development) Vol. 8 No. 2 (2020).
- [4] Ghulam Asrofi Buntoro. 2018. Sentiment Analysis to Prediction DKI Jakarta Governor 2017 on Indonesian Twitter. International Journal of Science, Engineering and Information Technology (IJSEIT) Vol 2, No 02 (2018)
- [5] Siti Rahmawati, Muhammad Habibi. 2020. Public Sentiments Analysis about Indonesian Social Insurance Administration Organization on Twitter. IJID (International Journal on Informatics for Development) VOL. 9 NO. 2 (2020).
- [6] Deni Gunawan, Dwiza Riana, Dian Ardiansyah, Fajar Akbar, Salman Alfarizi, 2020, Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023, Jurnal Teknik Komputer AMIK BSI. Volume VI No.1 Januari 2020.
- [7] Yuris Alkhalifi, Windu Gata, Arfhan Prasetyo, Imam Budiawan4, 2020, Analisis Sentimen Penghapusan Ujian Nasional pada Twitter Menggunakan Support Vector Machine dan Naïve Bayes berbasis Particle Swarm Optimization 2018-2023, Jurnal CoreIT, Vol.6, No.2, Desember 2020.
- [8] Samsir, Ambiyar, Unung Verawardina, Firman Edi, Ronal Watrianthos, 2021, Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes, Jurnal Sosial & Budaya Syar-i, JURNAL MEDIA INFORMATIKA BUDIDARMA Volume 5, Nomor 1, Januari 2021, Page 157-163.
- [9] Efid Dwi Agustono, Daniel Sianturi, Andi Taufik, Windu Gata. 2020., Analisis Sentimen Terhadap Warga China Saat Pandemi Dengan algoritma term Frequency-Inverse Document Frequency Dan Support Vector Machine, Jurnal Sosial & Budaya Syar-i, JURNAL MEDIA INFORMATIKA BUDIDARMA JIRE (Jurnal Informatika & Rekayasa Elektronika) Volume 3, No 2, November 2020.
- [10] Fajar Sodik Pamungkasa, Iqbal Kharisudin, 2021, Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter, Jurnal Sosial & Budaya Syar-i, PRISMA, Prosiding Seminar Nasional Matematika. PRISMA 4 (2021): 628-634.

