

PERBANDINGAN ANALISIS SENTIMEN ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES TERHADAP TANGGAPAN PUBLIK TENTANG PEMBELAJARAN ONLINE DI MASA PANDEMI COVID-19

Yulia Ardana^{1*}, Ridwan A. Kambau², Mustikasari³

Universitas Islam Negeri Alauddin Makassar

60200117043@uin-alauddin.ac.id¹, ridwan.kambau@uin-alauddin.ac.id²,
mustikasari@uin-alauddin.ac.id³

Abstrak

Pada awal tahun 2020 covid-19 mulai tersebar ke seluruh penjuru dunia, termasuk Indonesia. Pemerintah terus mencari cara untuk mencegah mata rantai penyebarannya, salah satunya dengan melaksanakan pembelajaran online. Penelitian ini dilatarbelakangi untuk memanfaatkan twitter dalam mengetahui respons dan sentimen publik tentang pembelajaran online selama masa pandemi covid-19. Tujuan dari penelitian ini adalah untuk mengetahui pendapat masyarakat tentang penerapan pembelajaran online dan juga untuk membandingkan tingkat performa algoritma support vector machine dan naïve bayes. Dalam melakukan penelitian ini, jenis penelitian yang digunakan adalah penelitian kualitatif agar mampu memahami dengan baik seperti apa fenomena yang dialami oleh subjek penelitian. Hasil analisis sentimen yang terbaik didapatkan dengan cara melakukan perbandingan dua algoritma klasifikasi, support vector machine dan naïve bayes. Pengujian berdasarkan k-fold cross validation bertujuan untuk memperoleh nilai accuracy, precision, dan recall. Algoritma yang terbaik akan menghasilkan output yang tepat dengan skor pengujian yang lebih tinggi.

Kata kunci: analisis sentimen, pembelajaran online, twitter, support vector machine, naïve bayes

Abstract

[COMPARISON OF SENTIMENT ANALYSIS OF SUPPORT VECTOR MACHINE AND NAÏVE BAYES ALGORITHMS ON PUBLIC RESPONSE ABOUT ONLINE LEARNING DURING THE COVID-19 PANDEMIC] At the beginning of 2020, COVID-19 began to spread throughout the world, including Indonesia. The government continues to look for ways to prevent the chain from spreading, one of which is by implementing online learning. The background of this research is to use twitter to find out the response and public sentiment about online learning during the covid-19 pandemic. The purpose of this research is to find out public opinion about the application of online learning and also to compare the performance level of support vector machine and naïve Bayes algorithms. In conducting this research, the type of research used is qualitative research in order to be able to understand well what kind of phenomena experienced by the research subjects. The best sentiment analysis results are obtained by comparing two classification algorithms, support vector machine and naïve Bayes. Testing based on k-fold cross validation aims to obtain accuracy, precision, and recall values. The best algorithm will produce the right output with a higher test score.

Keywords: sentiment analysis, online learning, twitter, support vector machine, naïve bayes

1. PENDAHULUAN

Paruh pertama 2020 *covid-19* mulai tersebar ke seluruh penjuru dunia, termasuk Indonesia. Pemerintah tak berhenti mencari cara untuk mencegah mata rantai penyebaran virus tersebut, salah satunya dengan menerapkan pembelajaran online. Pembelajaran *online* adalah kegiatan belajar mengajar yang sama halnya dengan proses pembelajaran dalam kelas, perbedaannya adalah dilakukan secara *online* menggunakan layanan internet (Firman & Rahman, 2020).

Pembelajaran *online* tentu membutuhkan perangkat yang bisa dimanfaatkan untuk mencari informasi, seperti *smartphone*, tablet, atau laptop. Teknologi *mobile* sangat berpengaruh besar dalam dunia pendidikan, termasuk dalam mencapai tujuan dari pembelajaran jarak jauh (Sadikin & Hamidah, 2020). Google Meet, Google Classroom, Zoom, dan WhatsApp, adalah beberapa contoh *platform* yang banyak dimanfaatkan dalam melaksanakan pembelajaran secara *online*.

Peneliti bermaksud melakukan analisis sentiment menggunakan dua *classifier*, yaitu *support vector machine* dan *naïve bayes*. Hasil analisis sentimen yang terbaik bisa diperoleh dengan melakukan perbandingan algoritma. Keduanya merupakan *classifier* yang dapat melakukan pengelompokan data ke dalam kategori positif atau negatif, dan populer untuk kasus *text mining* khususnya klasifikasi teks.

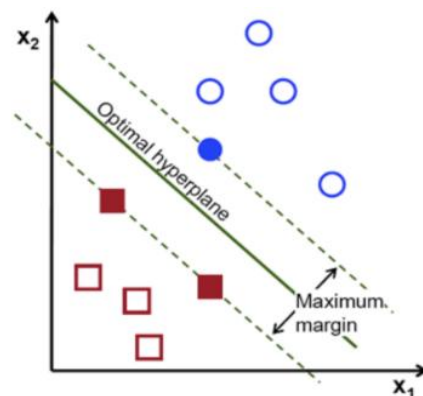
Algoritma *support vector machine* dipilih dengan alasan, yang pertama proses komputasi yang cepat karena menentukan jarak menggunakan *support vector* (Octaviani dkk., 2014). Yang kedua, kemampuannya meminimalkan *error* dalam *data training*. Sedangkan algoritma *naïve bayes* dipilih dengan alasan, yang pertama selain bisa dimanfaatkan untuk klasifikasi *biner*, juga bisa dimanfaatkan untuk klasifikasi *multiclass* (Binus University, 2019). Yang kedua, proses eksekusi yang cepat jadi cocok digunakan jika memiliki banyak data. Yang ketiga,

paling efektif dan efisien untuk *machine learning*. Kedua algoritma tersebut juga sama-sama sederhana dalam memproses data teks sehingga eksekusi data dapat selesai dengan cepat, karena untuk memproses tanggapan publik tentu ada banyak data dan memang dibutuhkan proses eksekusi yang cepat.

2. METODE

Analisis sentimen ialah teknik untuk mendapatkan informasi yang bersifat positif atau negatif tentang pendapat dan penilaian pihak lain. Penelitian tentang analisis sentimen jauh lebih berkembang berkat pengaruh dan manfaat dari analisis sentimen itu sendiri. Adapun manfaat dari analisis sentimen dalam dunia pendidikan seperti untuk mengetahui bagaimana sentimen dan pendapat masyarakat tentang keefektifan pembelajaran *online*, sehingga bisa dilakukan langkah-langkah perbaikan.

Support vector machine bertujuan untuk melakukan klasifikasi dan analisis data. Karena *support vector machine* adalah sebuah *classifier*, maka diberikan himpunan latih yang masing-masing dimiliki oleh salah satu dari dua kategori. *Support vector machine* bekerja dengan mendapatkan *hyperplane* terbaik yang menjadi pemisah dua *class*. Klasifikasi *linear* adalah prinsip dasar dari *support vector machine*, kemudian dikembangkan lagi supaya mampu beroperasi pada masalah *non-linear* (Octaviani dkk., 2014: 812).



Gambar 1. Linear Support Vector Machine

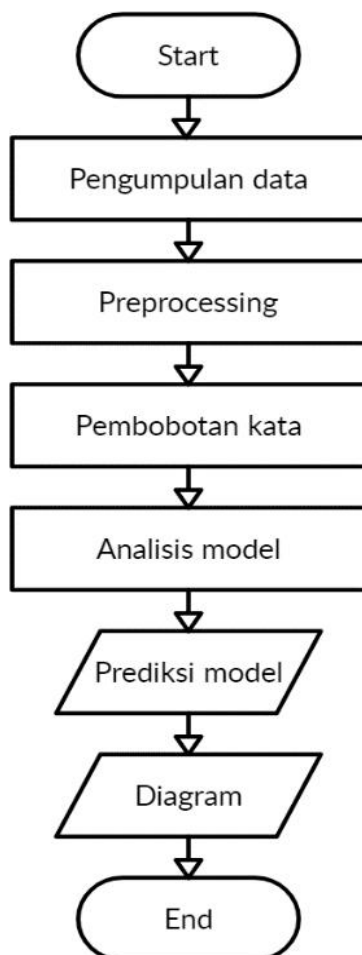
Naïve bayes adalah *classifier* yang bekerja dengan mencari nilai probabilitas tertinggi supaya data *testing* atau data uji terklasifikasi ke dalam kategori yang seharusnya. *Naïve bayes* beranggapan bahwa data tidak berhubungan satu sama lain. Teknik penggunaannya sangat cepat dan sederhana dengan penggunaan probabilitas.

3. HASIL DAN PEMBAHASAN

a. Flowchart

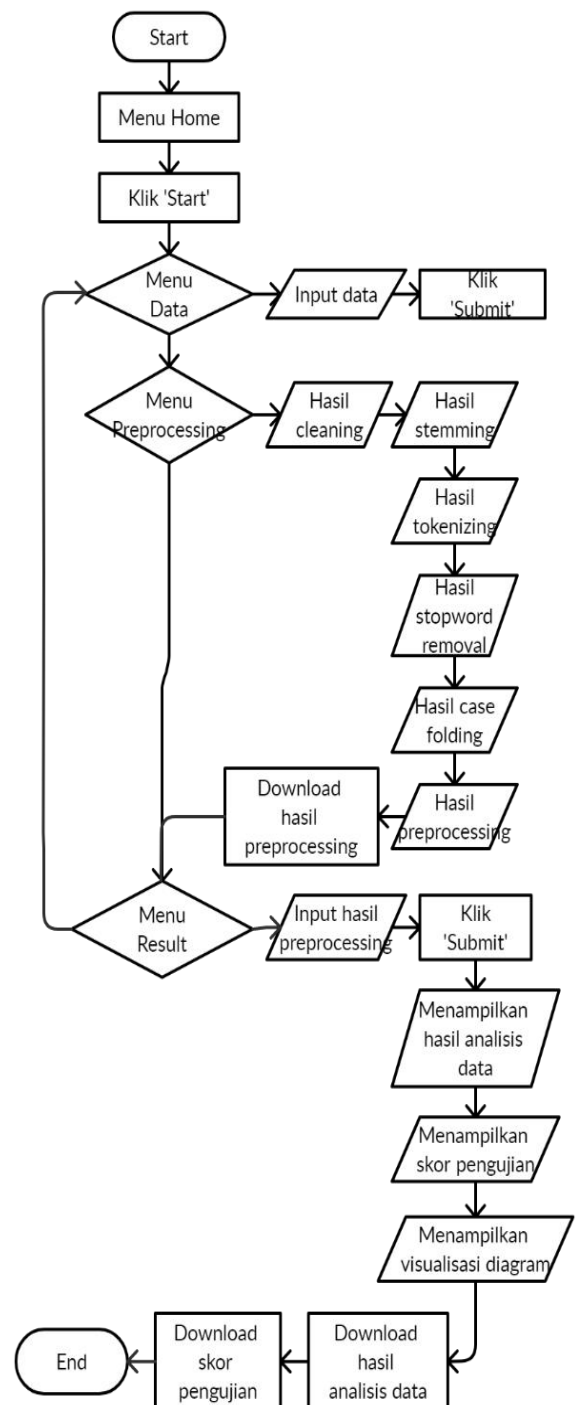
Flowchart memberikan gambaran alur kerja atau proses dari sistem yang memiliki simbol tertentu dan urutannya ditandai dengan arah panah.

1) Flowchart Pengembangan Sistem



Gambar 2. Flowchart Pengembangan Sistem

2) Flowchart Penggunaan Sistem



Gambar 3. Flowchart Penggunaan Sistem

b. Implementasi

1) Data

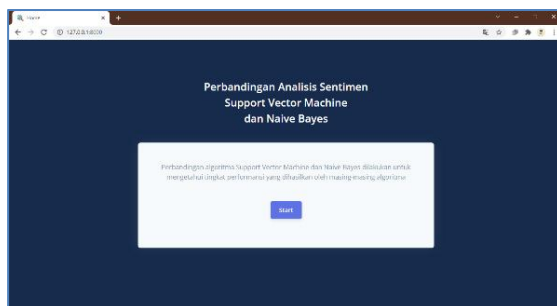
Berikut contoh data twitter hasil *crawling* (data mentah) yang telah diberi *label* dan siap melalui proses *preprocessing*.

Tabel 1. Contoh Data

No.	<i>Tweet</i>	<i>Class</i>
1	aku tim belajar harus dijelaskan langsung sama guru pls jangan sampe online lagið~aku kalo online tuh bawaannya males mulu gaada semangat, mana udah kelas 12, sebel bangetð	positif
2	aku sangat bosan hampir 2 tahun belajar dalam jaringan, materinya tidak ada yang masuk..	negatif
3	Media Pembelajaran Online yang Efektif di Era New Normal https://t.co/wtU1ggtiO1	positif
4	Apapun kekurangannya pembelajaran online harus tetap diadakan. Apa artinya generasi yang pintar jika fisik mereka sakit? Maaf hanya sekedar usul. Terima kasih.	positif
5	Saya Menolak Pembelajaran Online!!!	negatif

2) Antarmuka Halaman Home

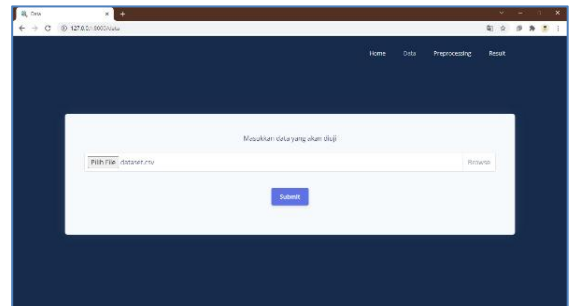
Halaman *home* merupakan halaman pembuka dalam sistem. Pada halaman tersebut dicantumkan nama beserta fungsi dari sistem yang akan digunakan. Selanjutnya klik *start* untuk memulai dan masuk ke halaman *data*.



Gambar 4. Halaman *Home*

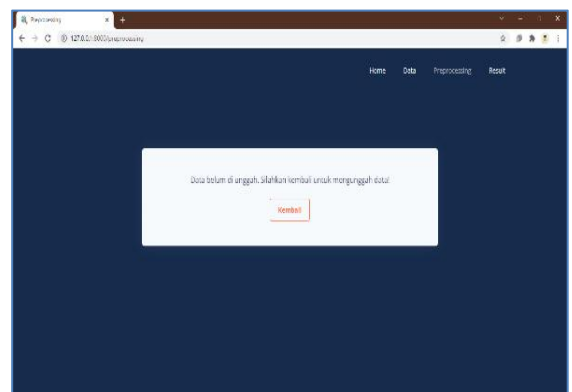
3) Antarmuka Halaman Data

Pada halaman data, *dataset* akan di *input*. Lalu klik *submit* untuk memulai *preprocessing*.



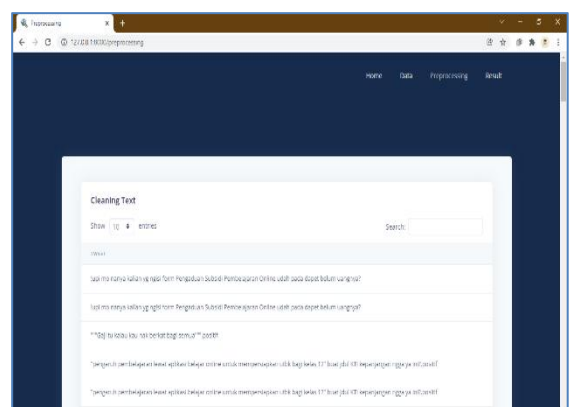
Gambar 5. Halaman Data

4) Antarmuka Halaman Preprocessing
Halaman *preprocessing* yang diklik sebelum melakukan *input* data akan memunculkan pesan seperti **Gambar 6**.



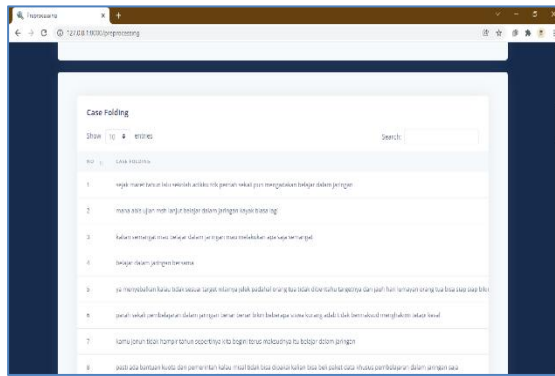
Gambar 6. Halaman *Preprocessing 1*

Proses *preprocessing* selesai, lalu secara otomatis akan masuk ke halaman *preprocessing*. Di bawah ini merupakan dokumentasi hasil *cleaning text*.



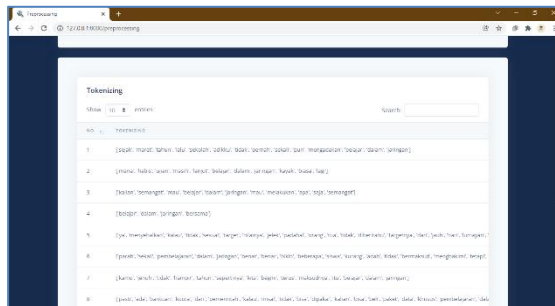
Gambar 7. Halaman *Preprocessing 2*

Gambar 8 merupakan dokumentasi hasil *case folding*. Pada bagian *case folding* seluruh *upper case* berubah menjadi *lower case*.



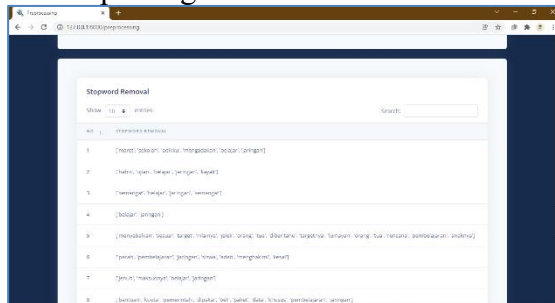
Gambar 8. Halaman *Preprocessing* 3

Pada bagian *tokenizing* akan memisahkan kalimat menjadi kumpulan kata-kata.



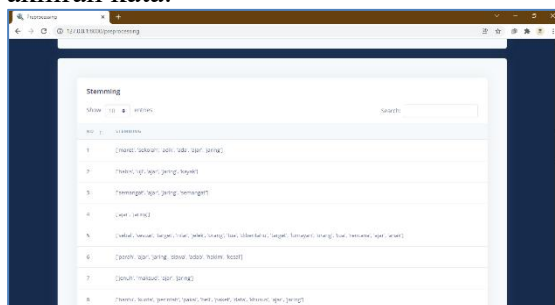
Gambar 9. Halaman *Preprocessing* 4

Bagian *stopword removal* akan menghilangkan kata-kata yang tidak punya makna penting.



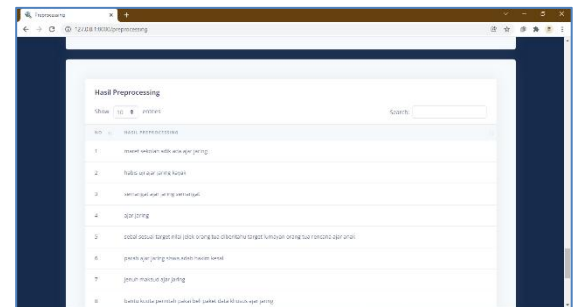
Gambar 10. Halaman *Preprocessing* 5

Dokumentasi hasil *stemming* dapat dilihat pada **Gambar 11**. Tahap *stemming* menghapus imbuhan berupa awalan dan akhiran kata.



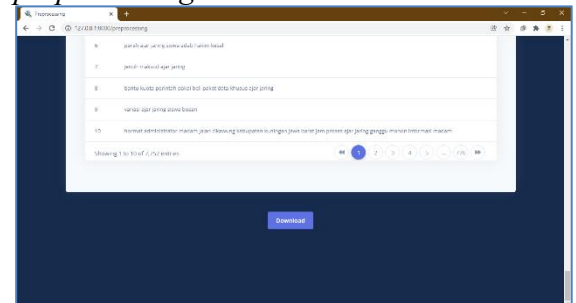
Gambar 11. Halaman *Preprocessing* 6

Gambar di bawah merupakan hasil *preprocessing* secara keseluruhan setelah melalui beberapa tahap tadi.



Gambar 12. Halaman *Preprocessing* 7

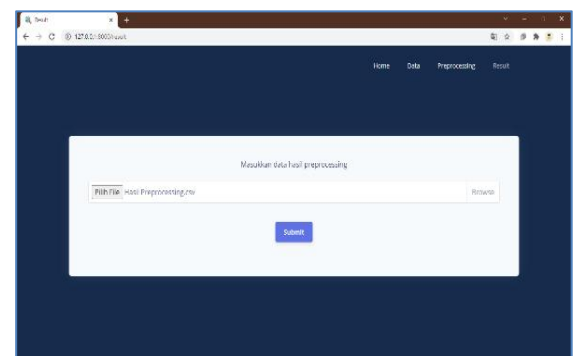
Tersedia fitur *download* data hasil *preprocessing* dalam format csv.



Gambar 13. Halaman *Preprocessing* 8

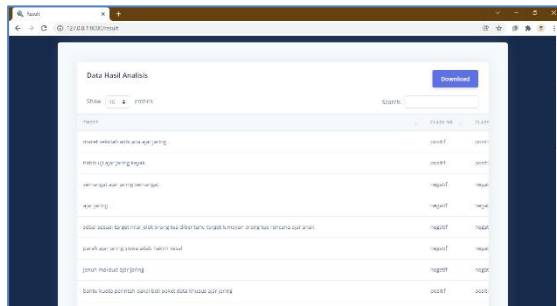
5) Antarmuka Halaman Result

Proses *preprocessing* telah selesai, selanjutnya klik halaman *result*. Masukkan data hasil *preprocessing* yang telah di *download*, kemudian klik *submit*.



Gambar 14. Halaman *Result* 1

Gambar di bawah menunjukkan tampilan data hasil analisis yang terdiri dari hasil *preprocessing* beserta *class* masing-masing data yang menggunakan algoritma *support vector machine* dan *naïve bayes*. Selain itu, data tersebut juga dapat di unduh dalam format csv.



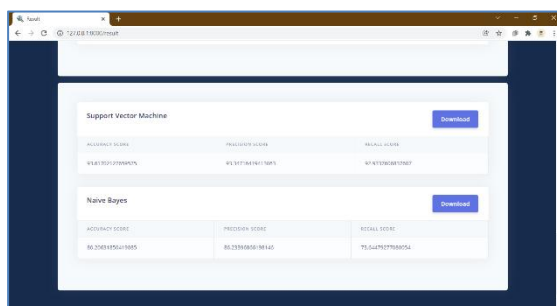
Gambar 15. Halaman *Result 2*

Tersedia juga *diagram* untuk menampilkan kata-kata yang paling dominan atau paling banyak ditemukan dalam data yang masuk kategori negatif.



Gambar 16. Halaman *Result 3*

Gambar 17 menampilkan skor pengujian kedua algoritma yang berupa *accuracy*, *precision*, dan *recall* untuk mengetahui performa algoritma.



Gambar 17. Halaman *Result* 4

c. Pengujian Sistem

Cross validation ialah metode untuk melakukan evaluasi seberapa akurat kinerja dari algoritma. Dari 7752 *dataset* yang dimiliki akan dibagi menjadi *data training* dan *data testing*. *K-fold cross validation* dipilih dengan alasan mampu meminimalisir penggunaan waktu dalam memproses, namun keakuratannya masih terjaga. Misal nilai k yang digunakan adalah 10, maka data dipecah menjadi 10 bagian dengan jumlah yang seimbang,

sehingga untuk melakukan evaluasi model terdapat 10 subset. *Cross-validation* memakai 1 *fold* untuk data uji dan 9 *fold* untuk data latih di setiap subsetnya. (MTI Binus University, 2017)

Tahapan ini akan dilakukan pengujian sistem dan algoritma untuk mengetahui kemampuannya. Pengujian sistem memanfaatkan *k-fold cross validation* dengan menetapkan 5, 7, dan 10 sebagai nilai k.

Tabel 2. Pengujian *Support Vector Machine*

k	Score		
	Accuracy	Precision	Recall
5	92,97%	92,79%	92,15%
7	93,61%	93,34%	92,93%
10	94,39%	93,72%	93,91%

Dari hasil pengujian algoritma *support vector machine* di atas, bisa disimpulkan bahwa skor pengujian paling tinggi ada pada nilai $k=10$, dengan skor pengujian *accuracy* 94,39%, *precision* 93,72%, dan *recall* 93,91%.

Tabel 3. Pengujian *Naïve Bayes*

k	Score		
	Accuracy	Precision	Recall
5	83,43%	87,82%	77,36%
7	80,20%	86,23%	73,64%
10	82,84%	88,81%	76,65%

Dari hasil pengujian algoritma *naïve bayes* di atas, bisa disimpulkan bahwa skor pengujian paling tinggi ada pada nilai $k=5$, dengan skor pengujian *accuracy* 83,43%, *precision* 87.82%, dan *recall* 77,36%.

d. Kuesioner

Kuesioner terdiri dari 10 buah pertanyaan yang disebarakan kepada 21 orang responden. Berdasarkan data yang dihasilkan dari kuesioner, dilakukan perhitungan menggunakan *skala likert* dengan skala 1 - 5. Hasil dari responden

sebanyak 6 orang kemudian dapat kita lihat sebagai berikut:

Keterangan:

81% - 100% = Sangat Setuju

61% - 80% = Setuju

41% - 60% = Cukup Setuju

21% - 40% = Tidak Setuju

< 21% = Sangat Tidak Setuju

Adapun nilai rata-rata dari hasil kuesioner adalah 84,77%. Maka bisa disimpulkan bahwa kelayakan sistem tergolong ‘Sangat Setuju’, serta sistem yang telah dibuat mampu menghasilkan data hasil analisis sentimen dengan mudah dan akurat.

4. KESIMPULAN

Hasil dari penelitian “Perbandingan Analisis Sentimen Algoritma *Support Vector Machine* dan *Naïve Bayes* Terhadap Tanggapan Publik Tentang Pembelajaran *Online* di Masa Pandemi *Covid-19*” bisa diambil beberapa kesimpulan, di antaranya:

- a. Algoritma *support vector machine* dan *naïve bayes* dapat menganalisis sentimen berupa teks, seperti data dari twitter.
- b. Semakin banyak jumlah data yang digunakan, maka performa dari algoritma akan semakin baik.
- c. Dari hasil pengujian *k-fold*, untuk *support vector machine* diketahui skor pengujian paling tinggi ada pada nilai $k=10$, dengan skor pengujian *accuracy* 94,39%, *precision* 93,72%, dan *recall* 93,91%. Sedangkan untuk *naïve bayes* diketahui skor pengujian paling tinggi ada pada nilai $k=5$, dengan skor pengujian *accuracy* 83,43%, *precision* 87,82%, dan *recall* 77,36%.
- d. Performa *support vector machine* lebih unggul dari *naïve bayes* dalam kasus ini berdasarkan hasil penelitian.

5. DAFTAR PUSTAKA

- [1] Firman dan Sari Rahayu Rahman. 2020. “Pembelajaran *Online* di Tengah Pandemi *Covid-19*”. *Indonesian Journal of Educational Science (IJES)* 2, no. 2: h. 81–89.
- [2] Sadikin, Ali, dan Afreni Hamidah. 2020. “Pembelajaran Daring di Tengah Wabah *Covid-19*”. *Biodik: Jurnal Ilmiah Pendidikan Biologi* 6, no. 2: h. 214–224. <https://online-journal.unja.ac.id/biodik/article/view/9759/5665>
- [3] Octaviani, Pusphita Anna, Yuciana Wilandari, dan Dwi Ispriyanti. 2014. “Penerapan Metode Klasifikasi *Support Vector Machine* (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang”. *Jurnal Gaussian* 3, no. 4: h. 811–820. <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/8092>
- [4] Binus University. 2019. Algoritma *Naïve Bayes*. <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/> (23 Mei 2021)
- [5] MTI Binus University. 2017. 10 *Fold-Cross Validation*. <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/> (24 Mei 2021)