

# Analisis Sentimen Rencana Pembelajaran Tatap Muka saat Pandemi Menggunakan *Naïve Bayes* dan *Support Vector Machine*

Syarif Hidayatulloh<sup>1</sup>, Yuni Yamasari<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika, Universitas Negeri Surabaya

<sup>1</sup>[syarif.17051204049@mhs.unesa.ac.id](mailto:syarif.17051204049@mhs.unesa.ac.id)

<sup>2</sup>[yuniyamasari@unesa.ac.id](mailto:yuniyamasari@unesa.ac.id)

**Abstrak**— Virus Coronavirus Disease 19 (Covid-19) yang menyebabkan infeksi saluran pernafasan pada akhir tahun 2019 mengakibatkan pandemi pada hampir seluruh negara di dunia. Berbagai kebijakan diambil demi kesehatan masyarakat masing-masing negara. Salah satu kebijakan yang diambil negara Indonesia adalah kebijakan sekolah dari rumah yaitu pendidikan yang dilakukan dengan sistem daring. Sehingga, siswa tidak perlu pergi ke sekolah, tetapi siswa mengakses internet dengan perangkat tertentu. Namun, muncul permasalahan baru yaitu tidak optimalnya proses pembelajaran daring yang dipicu oleh beberapa faktor, antara lain perangkat yang tidak memadai, penyediaan anggaran kuota internet dan lain sebagainya. Setelah ditemukan vaksin, pemerintah mengeluarkan rencana kebijakan pembelajaran tatap muka dengan memberikan vaksin bagi guru dan siswa. Namun, pemberian vaksin masih belum maksimal sehingga menjadi pembicaraan masyarakat di media sosial twitter. Sentimen masyarakat tentang kebijakan ini dapat diklasifikasikan menjadi positif, negatif, dan netral. Penelitian ini memperoleh 900 data yang terdiri dari 41,78% sentimen positif, 44,44% sentimen negatif, dan 13,78% sentimen netral. Hal ini menunjukkan bahwa sebagian besar masyarakat kurang setuju dengan rencana kebijakan pembelajaran tatap muka. Teknik pengujian dilakukan dengan *split data* dan *cross validation*. Hasil uji coba menunjukkan bahwa metode *Naïve Bayes* memiliki hasil lebih baik dari pada *Support Vector Machine* pada kedua jenis pengujian. Pengujian dengan hasil rata-rata akurasi tertinggi sebesar 65,30%, nilai precision 51,09%, nilai recall 70,51%, dan F1-Score sebesar 48,69% untuk metode *Naïve Bayes* dengan pengujian *split data*. Sedangkan rata-rata akurasi sebesar 64,63%, nilai precision 50,87%, nilai recall 53,81%, dan F1-Score sebesar 49,88% untuk metode *Support Vector Machine* dengan pengujian *split data*. Metode *Naïve Bayes* rata-rata akurasinya lebih tinggi 0,67% pada pengujian *split data*, dan lebih tinggi 0,33% pada pengujian *cross validation* dibanding metode *Support Vector Machine*.

**Kata Kunci**— Covid-19, Analisis Sentimen, PTM, *Naïve Bayes*, *SVM*.

## I. PENDAHULUAN

Pada akhir tahun 2019, ditemukan sebuah virus *Severe Acute Respiratory Syndrome Coronavirus (Sars-Cov-2)* atau lebih dikenal sebagai *Coronavirus Disease 19 (Covid-19)* yang menginfeksi saluran pernafasan [1]. Wabah virus ini pertama kali diidentifikasi di Wuhan, China. Virus ini kemudian

menyebarkan dengan cepat ke seluruh dunia. Dilansir dari laman resmi WHO, hingga tanggal 20 Mei 2021, secara keseluruhan terdapat lebih dari 100 juta kasus yang terkonfirmasi positif dan lebih dari 3 juta orang meninggal dunia [2]. Di Indonesia, konfirmasi kasus *Covid-19* tercatat mencapai 1.758.898 kasus dan 48.887 diantaranya menyebabkan kematian, sedangkan untuk proses vaksinasi tahap pertama sebanyak 14.369.233 dosis dan vaksinasi tahap kedua sebanyak 9.536.102 dosis [3].

Demi mencegah kenaikan angka positif *Covid-19* akibat aktivitas yang menimbulkan kerumunan, pemerintah menghimbau agar segala aktivitas dilakukan secara daring(dalam jaringan) atau *online*. Salah satu dampaknya terdapat pada bidang pendidikan, yaitu perubahan sekolah tatap muka menjadi sekolah dari rumah(*school from home*) yang dilakukan secara *online*. Namun, dalam pelaksanaannya masih ada beberapa masalah seperti keterbatasan penguasaan teknologi, sarana dan prasana, akses internet, hingga anggaran kuota internet yang disediakan [4].

Pemerintah sempat ingin membuka kembali sekolah pada awal januari 2021, tetapi dibatalkan karena jumlah kasus *Covid-19* yang masih tinggi [5]. Seiring menurunnya jumlah positif *Covid-19* dan upaya vaksinasi yang diprioritaskan bagi guru dan siswa, pemerintah memberikan pilihan untuk menggelar sekolah tatap muka dengan menerapkan protokol kesehatan pada tahun ajaran baru [6].

Kebijakan ini mendapat berbagai opini masyarakat, salah satunya di media sosial twitter, ada yang mendukung dan ada yang keberatan dengan rencana tersebut. Opini-opini di media sosial twitter ini, dapat dianalisis menggunakan metode-metode *text mining*.

Analisis sentimen juga dikenal sebagai *Opinion Mining* mengacu pada penggunaan *natural language processing*, *text analysis* and *computational linguistic* untuk mengidentifikasi dan mengekstrak informasi secara subjektif dari sumbernya [7]. Analisis ini dapat digunakan diberbagai bidang seperti bisnis, medis, pemerintahan, dan pendidikan dalam pengambilan sebuah kesimpulan atau keputusan[8]. Ada beberapa algoritma yang sering digunakan untuk melakukan analisis sentimen seperti *Naïve Bayes Classifier(NBC)*, *Support Vector Machine*, *Decision Tree*, dan *K-Nearest Neighbor(K-NN)*.

Ardan Fahmi dkk [9], melakukan penelitian sentimen masyarakat di media sosial *twitter* tentang aktivitas yang

dilaksanakan selama bulan ramadhan saat pandemi menghasilkan opini netral sebesar 93%, positif sebesar 6%, dan negatif sebesar 0.67%. Kata-kata positif yang sering muncul yaitu ‘ramadhan’, ‘semangat’, ‘sabar’, dan ‘promo’.

Yanuar Nurdiansyah dkk [10], menganalisis review film yang diklasifikasikan menjadi dua kategori, yaitu positif dan negatif menggunakan *Naive Bayes* dengan menggunakan jumlah data latih dan data uji coba dengan jumlah yang berbeda-beda dan kemudian diambil nilai rata-rata. Jumlah data latih secara berurutan 600, 700, 800, 900, dan 1000. Sedangkan untuk data uji secara berurutan 120, 140, 160, 180, dan 201. Dari penelitian yang dilakukan diperoleh akurasi secara berurutan 87.50%, 87.86%, 88.13%, 88,33%, dan 90.05% sehingga diperoleh rata-rata akurasi yaitu 88.37%.

Irwansyah Saputra dkk [11], membandingkan beberapa metode untuk menemukan metode terbaik dalam menganalisis sentimen masyarakat terhadap efek PSBB, data yang digunakan sebanyak 170 *tweet*, dan ekstraksi fitur yang digunakan adalah pembobotan *tf-idf*. Hasil yang didapatkan yaitu akurasi metode *Decision Tree* sebesar 81.06%, *K-NN* sebesar 82.72%, dan *Naive Bayes* sebesar 87.54%.

Lopamudra Dey dkk [12], membandingkan metode *Naive Bayes* dan *K-NN* dalam analisis sentimen review film dan hotel. Pengujian dilakukan sebanyak 10 kali dengan data latih sebanyak 100, 200, 500, 1000, 1500, 2000, 2500, 3000, 4000. Dan 4500. Hasil akurasi review film menggunakan *Naive Bayes* lebih tinggi yaitu sekitar 80% dibandingkan *K-NN* yang mendapat 70%. Sedangkan untuk review hotel, *Naive Bayes* mendapat 55% dan *K-NN* mendapat 51%.

Penelitian oleh Rahmat Hidayat dan Sekar Kinanti [13], membandingkan *Naive Bayes* dan *Support Vector Machine* untuk klasifikasi terjemahan Al-Quran dalam bahasa Inggris dan bahasa Indonesia. Akurasi yang diperoleh *Naive Bayes* lebih tinggi dalam klasifikasi dengan bahasa Inggris yaitu 78.35%, sedangkan *Support Vector Machine* lebih tinggi dalam klasifikasi dengan bahasa Indonesia yaitu 81.43%.

Penelitian oleh Elly Indrayuni [14], membandingkan metode *Naive Bayes* dan *Support Vector Machine* untuk menganalisis sentimen review film. Akurasi *Support Vector Machine* sebesar 90% dengan *AUC (Area Under Curve)* sebesar 0.982, lebih tinggi daripada *Naive Bayes* yang memiliki akurasi sebesar 84.5% dan *AUC* 0.500.

Dari beberapa penelitian di atas, metode *Naive Bayes* mempunyai keunggulan yaitu kemudahan implementasi, jumlah data yang dibutuhkan sedikit, serta hasil akurasi yang cukup tinggi untuk melakukan analisis data dalam berbagai kasus dibanding beberapa metode lainnya. Sedangkan metode *Support Vector Machine* memiliki keunggulan yaitu akurasi yang tinggi dengan *AUC* yang baik serta hasil klasifikasi yang lebih baik untuk teks berbahasa Indonesia.

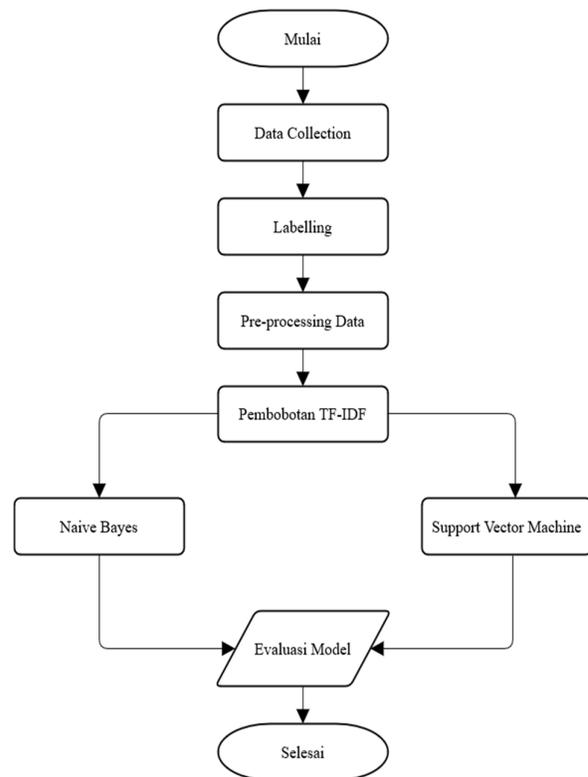
Berdasarkan keunggulan tersebut, maka dipilih kedua metode tersebut untuk membandingkan akurasinya dalam penelitian kebijakan rencana pembelajaran tatap muka yang sedang mendapat perhatian karena masih dalam situasi pandemi *Covid-19* dan jumlah vaksinasi yang dilaksanakan masih relatif kecil, dan data yang digunakan dalam penelitian ini bersumber dari pendapat masyarakat di media sosial twitter.

Mengingat pentingnya kesehatan masyarakat adalah yang utama dan pendidikan juga tidak dapat diabaikan, penulis merasa perlu adanya penelitian tentang pendapat masyarakat sebagai pertimbangan bagi pemerintah untuk pengambilan keputusan pembelajaran tatap muka saat pandemi, karena kebijakan ini akan mempengaruhi pendidikan dan kesehatan masyarakat.

Oleh karena itu, penulis ingin mengajukan penelitian dengan judul “Analisis Sentimen Rencana Pembelajaran Tatap Muka saat Pandemi menggunakan *Naive Bayes* dan *Support Vector Machine*” sebagai artikel ilmiah.

## II. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini yaitu dengan melakukan uji coba atau eksperimen terhadap studi kasus kebijakan pembelajaran tatap muka yang akan dilaksanakan saat pandemi. Tujuannya untuk mengetahui sentimen masyarakat terhadap kebijakan pembelajaran tatap muka, sekaligus membandingkan akurasi metode *naive bayes* dan *support vector machine* yang akan digunakan dalam pembuatan model analisis. Gbr. 1 merupakan diagram alir dari penelitian ini.

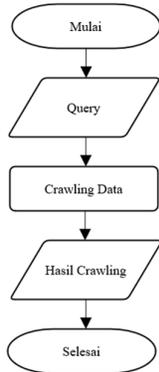


Gbr. 1 Diagram alir tahapan analisis sentimen

### A. Data Collections

*Dataset* yang digunakan pada penelitian ini berasal dari hasil *crawling tweet* pengguna media sosial *twitter* yang memiliki keterkaitan dengan topik pembelajaran tatap muka.

Proses *crawling tweet* dilakukan dengan melakukan pencarian berdasarkan *query* dengan *keyword* yang terkait pembelajaran tatap muka: pembelajaran tatap muka, sekolah tatap muka, atau sekolah *offline*. Data yang diperoleh kemudian disimpan menjadi sebuah *dataset* dalam format *csv*. Gbr. 2 menunjukkan proses *data collection*.



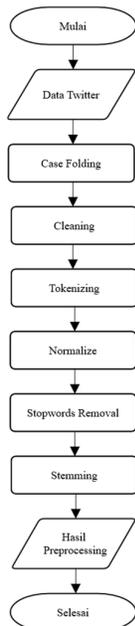
Gbr. 2 Proses data collection

### B. Labelling Data

Data yang sudah diperoleh akan diklasifikasikan menjadi tiga kategori: positif, negatif, dan netral. Tahap ini dilakukan sebelum *pre-processing* agar lebih mudah dalam menentukan sentimen *tweet*.

### C. Pre-processing Data

Sebelum pemrosesan data menggunakan metode yang dipilih, data akan melalui tahap *pre-processing*. Pada tahap ini, data mentah yang tidak terstruktur akan diolah menjadi data yang mempunyai lebih sedikit *noise* dan terstruktur, dengan harapan akurasi metode akan meningkat. Gbr. 3 adalah langkah-langkah dalam melakukan tahapan ini.



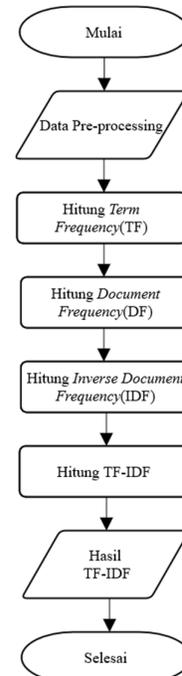
Gbr. 3 Langkah-langkah pada tahapan pre-processing

Berikut ini adalah penjelasan dari Gbr. 3:

- 1) *Case Folding*: Kata “Muka” dengan “muka”, walaupun memiliki arti yang sama, tetapi nilainya berbeda karena bahasa pemrograman python memiliki karakteristik *case sensitive*. Untuk mengatasi hal tersebut, perlu adanya penyeteraan dengan mengubah kata menjadi *uppercase* atau *lowercase*.
- 2) *Cleaning*: Proses pembersihan *tweet* dari data yang tidak digunakan. Pembersihan digunakan untuk menghilangkan *emoji*, *mention*, *hashtag*, dan simbol-simbol.
- 3) *Tokenizing*: Proses memisahkan sebuah *string* menjadi *input* tiap kata penyusunnya.
- 4) *Normalize*: Proses mengubah kata yang berupa singkatan ataupun *slang* menjadi kata yang sempurna menurut kaidah ejaan yang benar.
- 5) *Stopwords Removal*: Proses menghapus kata yang sering muncul namun tidak memiliki informasi atau makna yang penting.
- 6) *Stemming*: Proses yang bertujuan mereduksi kata berimbuhan menjadi bentuk kata dasar.

### D. Pembobotan TF-IDF

Pembobotan TF-IDF digunakan untuk memberi *score* pada tiap *term* atau kata berdasarkan frekuensi kemunculannya dalam dokumen. *Score* tersebut berfungsi untuk menyeleksi fitur yang akan digunakan pada metode pilihan. Gbr. 4 adalah Langkah-langkahnya.



Gbr. 4 Tahapan TF-IDF

Rumus menghitung TF-IDF adalah sebagai berikut:

$$TF - IDF = TF \times IDF \quad (1)$$

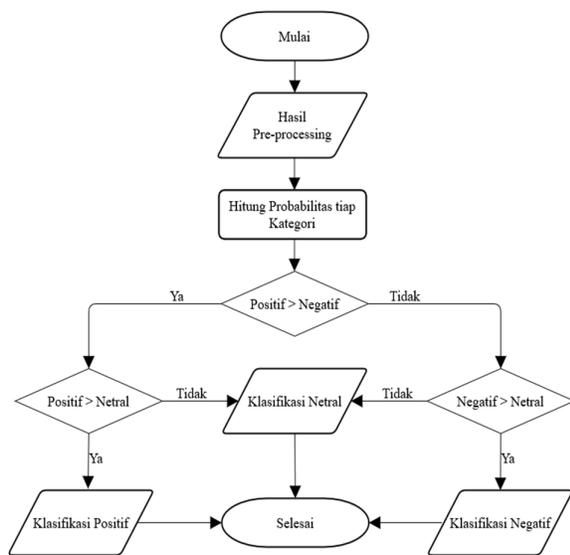
dimana,

$$TF = \frac{\text{total munculnya term dalam satu dokumen}}{\text{total term dalam dokumen}} \quad (2)$$

$$DF = \frac{\text{total seluruh dokumen}}{\text{total dokumen dimana term muncul}} \quad (3)$$

### E. Naïve Bayes

Pengujian dengan metode *naïve bayes* yaitu dengan melakukan perhitungan probabilitas bagi setiap kelasnya dalam menentukan klasifikasi. Probabilitas positif akan dibandingkan dengan negatif, kemudian dibandingkan dengan probabilitas netral. Nilai tertinggi dari perbandingan probabilitas akan menjadi hasil klasifikasi dari data tersebut. Gbr. 5 menunjukkan bagaimana *naïve bayes* bekerja dalam penentuan kelas.



Gbr. 5 penentuan klasifikasi oleh *naïve bayes*

### F. Support Vector Machine(SVM)

Klasifikasi yang dilakukan oleh SVM yaitu mengubah *text* menjadi data vektor, pada penelitian ini vektor adalah bobot(TF-IDF) dan dimensi(*id word*). Setelah itu dimensi akan diberi bobot berdasarkan seberapa penting keberadaannya dalam dokumen. Metode SVM akan mencoba menemukan garis terbaik untuk membagi kelas dan mengklasifikasikannya berdasarkan posisi kemunculan terhadap garis. *Kernel* yang digunakan pada penelitian ini adalah *kernel linear*.

### G. Pengujian Data

Pengujian data dilakukan dengan dua cara, yaitu *split data* dan *cross validation*. Kedua pengujian tersebut akan dilihat *accuracy*, *precision*, *recall*, dan *f1-score*.

### H. Evaluasi Model

Model yang menggunakan kedua metode akan dibandingkan berdasarkan hasil pengujian dari *split data* dan *cross validation*, sehingga dapat ditemukan metode yang terbaik dalam analisis data terhadap kebijakan pembelajaran tatap muka.

## III. HASIL DAN PEMBAHASAN

Penelitian analisis sentimen yang dilakukan terhadap kebijakan pembelajaran tatap muka saat pandemi ini menggunakan dua metode, yaitu *Naïve Bayes* dan *Support Vector Machine*. Untuk *Support Vector Machine*, *kernel* yang akan digunakan adalah *kernel linear* yang akan membagi kelas dengan garis lurus.

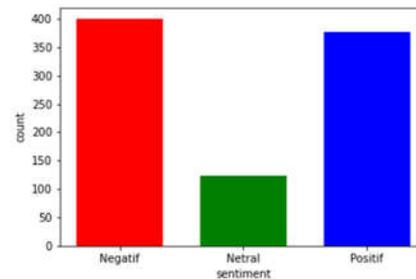
Penelitian ini bertujuan untuk mengetahui sentimen masyarakat terhadap kebijakan yang akan diambil pemerintah, serta membandingkan hasil akurasi dari kedua metode yang dipilih untuk melakukan analisis.

*Dataset* yang akan dianalisis berasal dari proses *crawling* terhadap *tweet* yang memiliki keterkaitan dengan pembelajaran tatap muka.

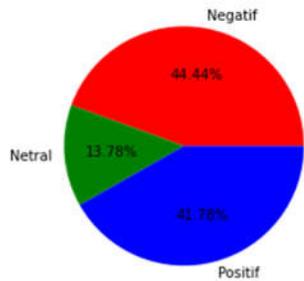
### A. Hasil Data Collection dan Labelling

*Dataset* yang didapat dari proses *crawling tweet* dengan *keyword* yang memiliki keterkaitan dengan pembelajaran tatap muka berjumlah 900 data yang belum memiliki label.

*Dataset* ini kemudian dilakukan proses *labelling*, yaitu pemberian label kepada masing-masing *tweet* agar dapat diklasifikasikan menjadi sentimen positif, negatif, ataupun netral. Setelah *dataset* diberi label, diperoleh label positif sebanyak 376 data, data negatif sebanyak 400 data, dan data netral sebanyak 124 data. Visualisasi dari klasifikasi data tersebut dapat dilihat pada Gbr. 6. Jika dipersentasekan, data positif sebesar 41,78%, data negatif sebesar 44,44%, dan data netral sebesar 13,78% seperti yang terdapat pada Gbr. 7.



Gbr. 6 Visualisasi pembagian data



Gbr. 7 persentase pembagian data

**B. Hasil Pre-processing Data**

Pada *pre-processing* data, data yang sudah diberi label akan melalui beberapa tahapan sebelum siap digunakan sebagai *input* dari metode *naïve bayes* dan *support vector machine*. Contoh dari *pre-processing* terdapat pada Tabel I.

TABEL I  
TAHAPAN PADA *PRE-PROCESSING*

Tweet asli	
Kalau pembelajaran tatap muka jadi digelar, Gw takut soalnya corona masih naik. Walaupun udh vaksin, tetep aja bahaya.	
Tahap <i>Pre-processing</i>	Hasil
<i>Case Folding</i>	kalau pembelajaran tatap muka jadi digelar, gw takut soalnya corona masih naik. walaupun udh vaksin, tetep aja bahaya.
<i>Cleaning</i>	kalau pembelajaran tatap muka jadi digelar gw takut soalnya corona masih naik walaupun udh vaksin, tetep aja bahaya
<i>Tokenizing</i>	[kalau, pembelajaran, tatap, muka, jadi, digelar, gw, takut, soalnya, corona, masih, naik, walaupun, udh, vaksin, tetep, aja, bahaya]
<i>Normalize</i>	[kalau, pembelajaran, tatap, muka, jadi, digelar, saya, takut, soalnya, corona, masih, naik, walaupun, udah, vaksin, tetep, saja, bahaya]
<i>Stopwords Removal</i>	[pembelajaran, tatap, muka, digelar, takut, soalnya, corona, naik, udah, vaksin, tetep, bahaya]
<i>Stemming</i>	belajar tatap, muka, gelar, takut, soal, corona, naik, udah, vaksin, tetep, bahaya

Hasil Akhir
belajar tatap muka gelar takut soal corona naik udah vaksin tetep bahaya

**C. Wordcloud**

Hasil dari *pre-processing*, dapat divisualisasikan ke dalam bentuk *wordcloud*. *Wordcloud* akan menampilkan kata-kata berdasarkan frekuensi kemunculan kata tersebut, semakin besar visualisasi kata maka semakin sering kata tersebut muncul.

*Wordcloud* tersebut ditampilkan berdasarkan label yang dimiliki, sehingga menghasilkan tiga *wordcloud*. Gbr. 8 menunjukkan *wordcloud* sentimen positif, Gbr. 9 menunjukkan *wordcloud* sentimen netral, sedangkan Gbr. 10 menunjukkan *wordcloud* sentimen negatif.



Gbr. 8 wordcloud sentimen positif



Gbr. 9 wordcloud sentimen netral



Gbr. 10 wordcloud sentimen negatif

**D. Hasil Pengujian Split Data**

Pengujian pada metode *naïve bayes* dan *support vector machine* akan menggunakan pengujian *split data*, dimana data akan dibagi menjadi data *train* yang digunakan untuk melatih model dan data *test* yang akan digunakan untuk menguji

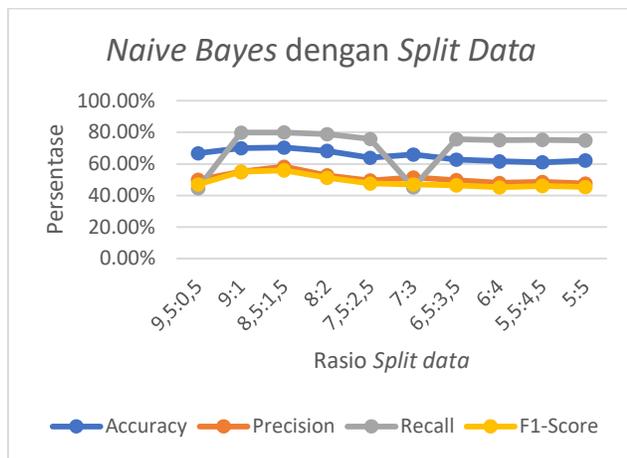
performance dari model. Pada pengujian ini, rasio *split data* untuk data *train* dan data *test* ditunjukkan pada Tabel II.

Hasil pengujian metode *naive bayes* dengan rasio *split data* akan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* seperti pada Tabel II.

TABEL II  
HASIL PENGUJIAN NAIVE BAYES DENGAN *SPLIT DATA*

Rasio	Accuracy	Precision	Recall	F1-Score
9,5:0,5	66,67%	49,96%	44,53%	47,03%
9:1	70,00%	55,03%	79,78%	54,99%
8,5:1,5	70,37%	58,26%	79,97%	55,96%
8:2	68,33%	52,74%	78,92%	51,19%
7,5:2,5	64,00%	49,42%	75,89%	47,54%
7:3	65,93%	51,47%	45,09%	47,12%
6,5:3,5	62,86%	49,77%	75,68%	46,51%
6:4	61,67%	48,03%	75,04%	45,18%
5,5:4,5	61,00%	48,48%	75,32%	46,17%
5:5	62,22%	47,70%	74,88%	45,25%

Berdasarkan hasil pengujian pada Tabel II, diperoleh hasil akurasi tertinggi pada pengujian dengan rasio data *train* dan data *test* yaitu 8,5:1,5, yang artinya 85% data akan digunakan sebagai data *train* dan 15% data akan digunakan sebagai data *test*. Akurasi tertinggi yang diperoleh dari pengujian yaitu sebesar 70,37%, dengan nilai *precision* 58,26%, nilai *recall* 79,97%, dan *F1-Score* sebesar 55,96%. Grafik pada Gbr. 11 merepresentasikan hasil metode *naive bayes* dengan pengujian *split data*.



Gbr. 11 grafik metode *naive bayes* dengan pengujian *split data*

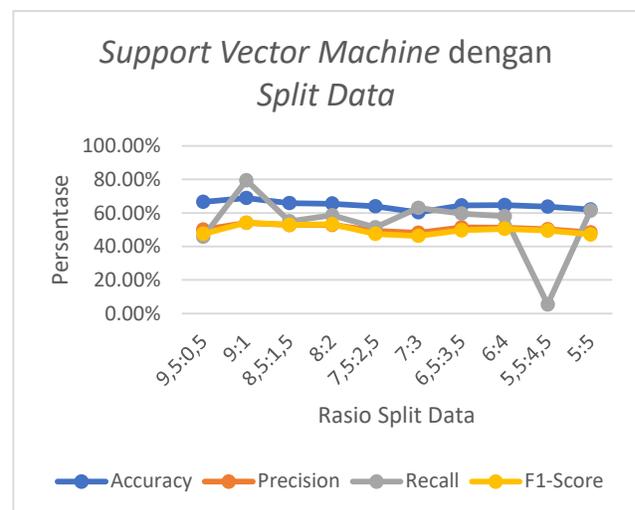
Hasil pengujian metode *support vector machine* dengan rasio *split data* akan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* seperti pada Tabel III.

TABEL III  
HASIL PENGUJIAN SUPPORT VECTOR MACHINE DENGAN *SPLIT DATA*

Rasio	Accuracy	Precision	Recall	F1-Score
9,5:0,5	66,67%	49,96%	45,69%	47,62%

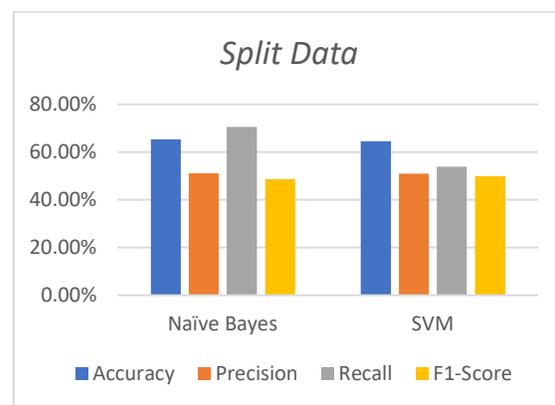
9:1	68,89%	54,09%	79,46%	54,22%
8,5:1,5	65,93%	52,96%	55,12%	52,90%
8:2	65,56%	52,94%	58,61%	53,29%
7,5:2,5	64,00%	49,25%	51,47%	47,57%
7:3	60,37%	48,26%	63,02%	46,30%
6,5:3,5	64,44%	51,43%	59,76%	49,74%
6:4	64,72%	51,25%	58,08%	50,56%
5,5:4,5	63,70%	50,15%	5,41%	49,47%
5:5	62,00%	48,41%	61,53%	47,12%

Berdasarkan hasil pengujian pada Tabel III, diperoleh hasil akurasi tertinggi pada pengujian dengan rasio data *train* dan data *test* yaitu 9:1. Akurasi tertinggi yang diperoleh dari pengujian yaitu sebesar 68,89%, dengan nilai *precision* 54,09%, nilai *recall* 79,46%, dan *F1-Score* sebesar 54,22%. Grafik pada Gbr. 12 merepresentasikan hasil metode *support vector machine* dengan pengujian *split data*.



Gbr. 12 grafik metode *support vector machine* dengan pengujian *split data*

Apabila kedua metode yang diuji dengan *split data* dibandingkan rata-ratanya, maka hasilnya akan seperti pada Gbr. 13.



Gbr. 13 perbandingan rata-rata *naive bayes* dan *svm* dengan *split data*

E. Hasil Pengujian Cross Validation

Pengujian berikutnya adalah dengan *cross validation*, yaitu metode pengujian model dengan mebagik ukuran data berdasarkan *K-fold*. Data akan memiliki K bagian, dimana satu bagian akan bertindak sebagai data *validation* dan sisanya sebagai data *train*. setiap bagian K, akan dicoba secara bergantian sebagai data *train* dan data *validation*, sehingga setiap pengujian dengan K bagian akan memiliki sebanyak K skenario hasil. Hasil tersebut kemudian akan diambil rata-ratanya. Pada pengujian ini akan menguji dengan K=2 hingga K=20.

Hasil pengujian metode *naive bayes* dengan *cross validation* akan menghasilkan nilai *accuracy*, *precicion*, *recall*, dan *f1-score* seperti pada Tabel IV.

TABEL IV  
HASIL PENGUJIAN NAIVE BAYES DENGAN CROSS VALIDATION

K-fold	Accuracy	Precision	Recall	F1-Score
2	59,00%	39,55%	45,72%	42,23%
3	59,89%	40,81%	46,36%	43,29%
4	59,56%	39,84%	46,12%	42,62%
5	59,56%	39,79%	46,12%	42,63%
6	62,11%	41,49%	48,08%	44,45%
7	60,79%	41,19%	47,03%	43,72%
8	60,80%	40,64%	47,04%	43,51%
9	61,33%	40,93%	47,43%	43,74%
10	62,22%	41,63%	48,17%	44,52%
11	62,12%	47,51%	48,64%	45,67%
12	62,89%	42,18%	48,67%	44,98%
13	62,25%	41,47%	48,10%	44,22%
14	62,15%	41,35%	48,01%	44,18%
15	62,11%	41,56%	48,01%	44,33%
16	61,81%	41,40%	47,84%	44,24%
17	62,56%	45,73%	48,96%	45,90%
18	61,44%	40,82%	47,42%	43,63%
19	62,59%	45,27%	48,91%	45,76%
20	62,11%	41,58%	48,10%	44,44%

Berdasarkan hasil pengujian pada Tabel IV, diperoleh hasil akurasi tertinggi pada pengujian dengan K=12. Akurasi

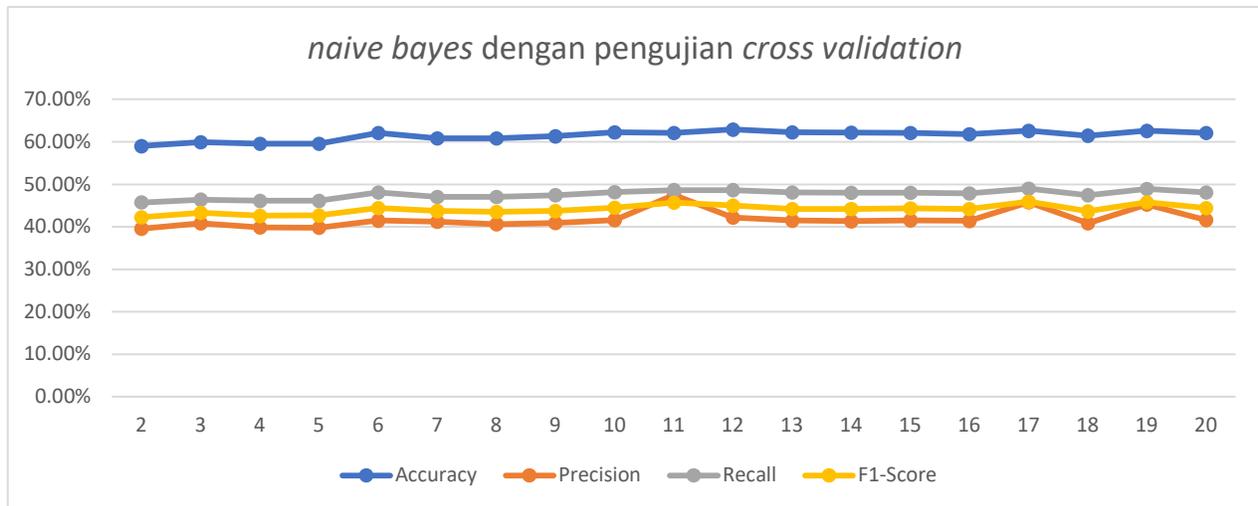
tertinggi yang diperoleh dari pengujian yaitu sebesar 62,89%, dengan nilai *precision* 42,18%, nilai *recall* 48,67%, dan *F1-Score* sebesar 44,98%. Gbr. 14 merepresentasikan hasil dari tabel IV.

Hasil pengujian metode *support vector machine* dengan *cross validation* akan menghasilkan nilai *accuracy*, *precicion*, *recall*, dan *f1-score* seperti pada Tabel V.

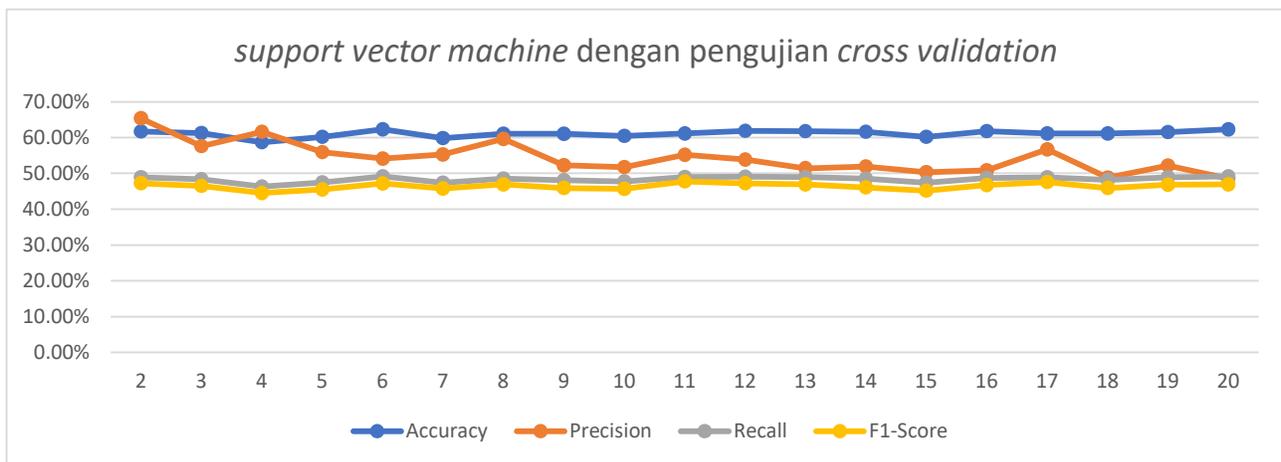
TABEL V  
HASIL PENGUJIAN SUPPORT VECTOR MACHINE DENGAN CROSS VALIDATION

K-fold	Accuracy	Precision	Recall	F1-Score
2	61,67%	65,35%	48,91%	47,24%
3	61,33%	57,64%	48,34%	46,52%
4	58,67%	61,62%	46,32%	44,53%
5	60,22%	55,96%	47,50%	45,50%
6	62,33%	54,11%	49,15%	47,17%
7	59,90%	55,30%	47,44%	45,79%
8	61,03%	59,66%	48,48%	46,93%
9	61,00%	52,26%	48,09%	45,96%
10	60,44%	51,78%	47,74%	45,66%
11	61,12%	55,18%	49,00%	47,73%
12	61,89%	53,84%	49,07%	47,25%
13	61,80%	51,42%	49,02%	46,92%
14	61,59%	51,90%	48,48%	46,13%
15	60,22%	50,31%	47,47%	45,17%
16	61,81%	50,82%	48,80%	46,73%
17	61,12%	56,70%	48,94%	47,62%
18	61,11%	48,85%	48,21%	45,92%
19	61,51%	52,15%	48,94%	46,87%
20	62,33%	48,57%	49,19%	46,96%

Berdasarkan hasil pengujian pada Tabel V, diperoleh hasil akurasi tertinggi *SVM* pada pengujian dengan K=6 dan K=20. Pengujian K=6 memiliki akurasi sebesar 62,33%, nilai *precision* 54,11%, nilai *recall* 49,15%, dan *F1-Score* sebesar 47,17%. Sedangkan pengujian K=20 memiliki akurasi sebesar 62,33%, nilai *precision* 48,57%, nilai *recall* 49,19%, dan *F1-Score* sebesar 46,96%. Gbr. 15 merepresentasikan hasil dari tabel V.

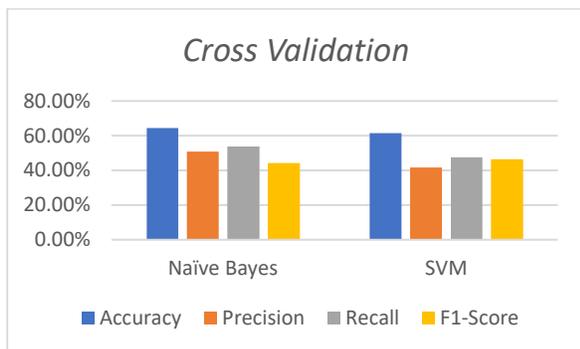


Gbr. 14 grafik metode *naive bayes* dengan pengujian *cross validation*



Gbr. 15 grafik metode *support vector machine* dengan pengujian *cross validation*

Apabila kedua metode yang diuji dengan *cross validation* dibandingkan rata-ratanya, maka hasilnya akan seperti pada Gbr. 12.



Gbr. 16 perbandingan rata-rata *naive bayes* dan *svm* dengan *cross validation*

#### F. Hasil Perbandingan Split Data dan Cross Validation

Hasil metode *naive bayes* dan *support vector machine* dari kedua pengujian akan dibandingkan agar dapat menentukan metode dengan akurasi yang lebih baik. Tabel VI menunjukkan perbandingan dari kedua pengujian yang telah dilakukan.

TABEL VI  
 PERBANDINGAN RATA-RATA DARI KESELURUHAN PENGUJIAN

Pengujian	Metode	Accuracy	Precision	Recall	F1-Score
Split Data	Naive Bayes	65,30%	51,09%	70,51%	48,69%
	SVM	64,63%	50,87%	53,81%	49,88%
Cross validation	Naive Bayes	61,44%	41,83%	47,62%	44,11%
	SVM	61,11%	54,39%	48,37%	46,45%

Dari Tabel VI, dapat dilihat bahwa pada pengujian model dengan *split data*, level akurasi tertinggi diperoleh dari model

yang menggunakan metode *naïve bayes* dengan rata-rata akurasi sebesar 65,30%, nilai *precision* 51,09%, nilai *recall* 70,51%, dan *F1-Score* sebesar 48,69%. Sedangkan metode *support vector machine* masih memiliki rata-rata akurasi lebih rendah dengan akurasi sebesar 64,63%, nilai *precision* 50,87%, nilai *recall* 53,81%, dan *F1-Score* sebesar 49,88%.

Sedangkan pada pengujian *cross validation*, level akurasi tertinggi diperoleh dari model yang menggunakan metode *naïve bayes* dengan rata-rata akurasi sebesar 61,44%, nilai *precision* 41,83%, nilai *recall* 47,62%, dan *F1-Score* sebesar 44,11%. Sedangkan metode *support vector machine* masih memiliki rata-rata akurasi lebih rendah dengan akurasi sebesar 61,11%, nilai *precision* 54,39%, nilai *recall* 48,37%, dan *F1-Score* sebesar 46,45%.

#### IV. KESIMPULAN

Berdasarkan pengujian analisis sentimen terhadap kebijakan pembelajaran tatap muka saat pandemi menggunakan metode *naïve bayes* dan *support vector machine*, kesimpulan yang dapat diperoleh yaitu dari 900 data, 41,78% memiliki sentimen positif, 44,44% memiliki sentimen negatif, dan sebesar 13,78% memiliki sentimen netral. Hal ini menunjukkan bahwa sebagian besar masyarakat masih kurang setuju dengan kebijakan pembelajaran tatap muka saat pandemi.

Pada kedua pengujian, metode *naïve bayes* memiliki rata-rata akurasi lebih tinggi 0,67% dibanding metode *support vector machine* pada pengujian *split data* dan lebih tinggi 0,33% pada pengujian *cross validation*. Hal ini menunjukkan bahwa metode *naïve bayes* lebih baik daripada *support vector machine* pada kedua jenis pengujian dalam analisis sentimen terhadap pembelajaran tatap muka.

#### V. SARAN

Pada penelitian ini, data yang digunakan adalah *tweet* yang tidak menggunakan bahasa baku. Pada penelitian selanjutnya disarankan untuk menggunakan data yang menggunakan bahasa baku, seperti formulir atau *survey* bagi siswa, orangtua siswa, dan tenaga pendidik. Selain itu, perlu penambahan kosakata pada tahapan *normalize* saat *pre-processing*. Penelitian selanjutnya juga dapat menggunakan metode lainnya.

#### UCAPAN TERIMA KASIH

Puji syukur kepada Allah SWT yang memberikan kemampuan untuk menyelesaikan penelitian ini. Terima kasih kepada orangtua yang memberikan dukungan dan do'a sehingga dipermudah segala hal yang berkaitan dengan penelitian ini, serta dosen pembimbing dan pihak-pihak yang memberikan dukungan pada jalannya penelitian ini.

#### REFERENSI

- [1] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir and R. Siddique, "COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses," *Journal of Advanced Research*, vol. 24, pp. 91-98, 2020.
- [2] (2021) World Health Organization, "WHO Coronavirus (COVID-19) Dashboard," [Online], <https://covid19.who.int/>, tanggal akses: 20 Mei 2021.
- [3] (2021) Komite Penanganan Covid-19 dan Pemulihan Ekonomi Nasional, "Data Vaksinasi COVID-19 (Update per 20 Mei 2021)," [Online], <https://covid19.go.id/berita/data-vaksinasi-covid-19-update-20-mei-2021>, tanggal akses: 20 Mei 2021.
- [4] R. H. S. Aji, "Dampak Covid-19 pada Pendidikan di Indonesia: Sekolah, Keterampilan, dan Proses Pembelajaran," *Jurnal Sosial & Budaya Syar-i*, vol. 7, pp. 395-402, 2020.
- [5] (2021) G. S. Sandro, "Kemendikbud Resmi Batalkan Belajar Tatap Muka Tahun Ajaran 2020-2021," *Economic Zone*, 4 January 2021. [Online], <https://www.economiczone.id/read/1741/kemendikbud-resmi-batalkan-belajar-tatap-muka-tahun-ajaran-2020-2021>, tanggal akses: 21 Mei 2021.
- [6] (2021) A. P. Kasih, "Kemendikbudristek: Sekolah Tatap Muka Juli, Orangtua Berhak Memilih," *Kompas.com*, [Online], <https://www.kompas.com/edu/read/2021/05/06/150000171/kemendikbudristek--sekolah-tatap-muka-juli-orangtua-berhak-memilih?page=all>, tanggal akses: 21 Mei 2021].
- [7] S. Jusoh and H. M. Alfaraweh, "Techniques, applications and challenging issue in text mining," *International Journal of Computer Science Issues (IJCSI)*, pp. 431-436, 2012.
- [8] W. Fan, L. Wallace, S. Rich and Z. Zhang, "Tapping into the Power of Text Mining," *Communications of the ACM*, 2006.
- [9] A. Fahmi, I. Ramadhan and A. , "Analisis Sentiment Masyarakat Selama Bulan Ramadhan dalam Menghadapi Pandemi Covid-19," *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, pp. 608-617, 2020.
- [10] Y. Nurdiansyah, S. Bukhori and R. Hidayat, "Sentiment Analysis System for Movie Review in Bahasa Indonesia using Naïve Bayes Classifier Method," *Journal of Physics: Conference Series*, 2018.
- [11] I. Saputra, J. A. Halomoan, A. B. Raharjo and C. R. A. Syavira, "SENTIMENT ANALYSIS ON TWITTER OF PSBB EFFECT USING MACHINE LEARNING," *Jurnal Techno Nusa Mandiri*, Vols. 143-150, 2020.
- [12] L. Dey, S. Chakraborty, B. Bose and A. Biswas, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *International Journal of Information Engineering and Electronic Business*, pp. 54-62, 2016.
- [13] R. Hidayat and S. Minati, "Comparative Analysis of Text Mining Classification Algorithms for English and Indonesian Qur'an Translation," *International Journal on Informatics for Development*, pp. 47-51, 2019.
- [14] E. Indrayuni, "Komparasi Algoritma Naive Bayes dan Support Vector Machine untuk Analisa Sentimen Review Film," *Jurnal PILARNusa Mandiri*, vol. 14, pp. 175-180, 2018.

