



Penerapan *support vector machine* pada analisis sentimen kebijakan pembelajaran tatap muka saat pandemi 2021

Application of support vector machine in policy sentiment analysis for face-to-face learning during the 2021 pandemic

Naufal Dhianur Alam Putra^{1*}, Didi Juardi¹, Arif Sholehudin¹, Slamet Abadi²

1* Prodi Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, Jawa Barat, Indonesia

2 Prodi Agribisnis, Fakultas Pertanian, Universitas Singaperbangsa Karawang, Jawa Barat, Indonesia

INFORMASI ARTIKEL

Article History:

Submission: 14-09-2023

Revised: 17-11-2023

Accepted: 27-11-2023

Kata Kunci:

Covid; data mining; information gain; pembelajaran tatap muka; *support vector machine*

Keywords:

Covid, data mining; face-to-face learning; information gain; support vector machine

* Korespondensi:

Naufal Dhianur Alam Putra

naufal.dhianur17160@student.unsika.ac.id

ABSTRAK

Dampaknya dikeluarkannya Surat Edaran 4 Mendiknas tahun 2020 tentang kebijakan pendidikan pada masa darurat akibat penyebaran wabah Covid-19 menjadi pembelajaran daring. Sentimen yang disampaikan masyarakat mengenai kebijakan Pembelajaran Tatap Muka terbatas dapat dianalisis dan hasilnya digunakan sebagai *feedback* dari kebijakan untuk mengetahui efektifitas dari kebijakan tersebut. Tujuannya a). mengklasifikasikan sentimen pengguna sosial media twitter terhadap kebijakan pembelajaran tatap muka menggunakan *support vector machine*. b). menganalisis hasil performan dari penerapan algoritma *support vector machine* berdasarkan nilai akurasi menggunakan metode *confusion matrix*. Penelitian mempunyai 4 tahap penelitian yang dilakukan, yaitu tahap pendahuluan, tahap pengumpulan data, tahap pengolahan data, tahap analisis dan pembahasan, dan kesimpulan. Metode *Knowledge Discovery in Database* terdiri dari, *Data Selection, Preprocessing, Transformation, Data Mining, dan Evaluasi*. Data yang digunakan bisa berupa satu tweet dari aplikasi Twitter atau sebanyak 200 data, *splitting* data sebanyak 80% data digunakan untuk pelatihan dan 20% digunakan untuk pengujian sebelum tahap penemuan data. *Information Gain* menghasilkan 665 fitur yang tidak relevan. Hasil *Support Vector Machine* dan *Information Gain* yang digunakan dalam pemrosesan data dan evaluasi kinerja ini sebesar *Accuracy* sebesar 85,00%, *Precision* sebesar 40,00%, dan *recall* sebesar 66,67% untuk kernel linear sedangkan *Accuracy* sebesar 82,50%, *Precision* sebesar 0%, dan *recall* tidak bernilai untuk kernel *Radial Basis Function*

ABSTRACT

The impact is the issuance of Circular Letter 4 of the Minister of National Education in 2020 concerning education policies during the emergency period due to the spread of the Covid-19 outbreak into online learning. The sentiments expressed by the public regarding the limited



face-to-face learning policy can be analyzed and the results used as feedback on the policy to determine the effectiveness of the policy. The goal is a). classify the sentiments of Twitter social media users towards face-to-face learning policies using a support vector machine. b). analyze the performance results from the application of the support vector machine algorithm based on the accuracy value using the confusion matrix method. The research is divided into four stages: preliminary, data collecting, data processing, analysis and discussion, also conclusions. This step consists of Knowledge Discovery in Database, namely Data Selection, Preprocessing, Transformation, Data Mining, and Evaluation. The data used can be one tweet from the Twitter application or as much as 200 data. 80% of the data is used for training and 20% is used for testing before the data discovery stage. The resulting Support Vector Machine and Information Gain used in data processing and performance evaluation are Accuracy of 85.00%, Precision of 40.00%, and recall of 66.67% for linear kernels while Accuracy is 82.50%, Precision is 0%, and recall is of no value for the Radial Basis Function kernel.

1. PENDAHULUAN

Pandemi Covid-19 yang merajalela sejak awal tahun 2020 telah memberikan dampak yang tidak biasa di segala aspek kehidupan. Data *World Health Organization* (WHO) per tanggal 22 Oktober 2021, perkembangan kasus *Coronavirus Disease* (Covid-19) untuk situasi global telah dikonfirmasi sebanyak 242.348.657 kasus dengan jumlah kematian sebanyak 4.927.723 kasus. Pada perkembangan kasus di Indonesia yang terkonfirmasi sebanyak 4.238.594 kasus dengan jumlah kematian 143.153 kasus per tanggal 30 Desember 2020 data dari WHO, [1]. Catatan kasus Covid-19 yang terus meningkat membuat beberapa negara termasuk Indonesia memberlakukan kebijakan *social distancing* atau *lockdown* sebagai upaya untuk melakukan pencegahan dan memutus rantai penularan virus terhadap orang lain. Dampak lainnya adalah dikeluarkannya Surat Edaran 4 Mendiknas tahun 2020 tentang kebijakan pendidikan pada masa darurat akibat penyebaran wabah Covid-19 menjadi pembelajaran daring [2]. Pemerintah terus mengupayakan solusi dalam menekan angka kasus Covid-19 dengan melakukan vaksinasi.

Menurut data Kementerian Kesehatan per 10 September 2021, proporsi penduduk yang telah menerima vaksin dosis pertama adalah 35,17% dan dosis kedua adalah 20,20% [3]. Kesadaran masyarakat tentang vaksin, pemerintah telah mengeluarkan Surat Edaran Nomor 4 Tahun 2021. Tentang penerjunan resmi terkait pelaksanaan pembelajaran tatap muka. Wilayah Jawa Timur memiliki keterbatasan pelaksanaan pembelajaran langsung (PTM), dari 4073 pada jenjang pendidikan menengah, kejuruan, dan luar biasa, 3944 lembaga atau 96,83% dari total jenjang penyelenggaraan Pembelajaran Tatap Muka (PTM) terbatas adalah 48,34% [4].

Pelaksanaan pembelajaran tatap muka menuai beberapa opini pro dan kontra dari masyarakat. Beberapa masyarakat menyampaikan opininya melalui sosial media salah satunya yaitu twitter. Terdapat masyarakat yang tidak menyetujui pembelajaran tatap muka salah satunya karena dianggap terlalu beresiko akan penyebaran Covid19 namun ada juga masyarakat yang menyetujuinya karena tidak cocok akan sistem pembelajaran daring [5]. Sentimen-sentimen yang disampaikan masyarakat mengenai kebijakan PTM terbatas dapat dianalisis dan hasilnya digunakan sebagai *feedback* dari kebijakan untuk mengetahui efektifitas dari kebijakan tersebut.

Analisis sentimen merupakan metode berkelanjutan dalam penelitian tekstual [6]. Analisis sentimen adalah mempelajari mengenai cara memecahkan masalah dari opini, sikap dan emosi yang disampaikan masyarakat terhadap suatu topik [7]. Analisis sentimen bertujuan untuk melihat opini pada sebuah masalah oleh individu, apakah cenderung beropini negatif atau positif.

Beberapa peneliti telah melakukan analisis sentimen, seperti studi perbandingan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*. Berdasarkan opini publik di *Twitter* dan menunjukkan bahwa hasil algoritma SVM 92,48% lebih akurat daripada hasil *Naïve Bayes*, khususnya 76,56% [8]. Studi lain telah membandingkan algoritma *K-Nearest Neighbor* dan SVM. Hasil algoritma SVM mempunyai akurasi lebih tinggi dibandingkan *K-Nearest Neighbor* (KNN), [9].

Berdasarkan penjelasan yang telah disampaikan, akan dilakukan penelitian mengenai analisis sentimen terkait PTM dengan menggunakan *Support Vector Machine* dan informasi yang diperoleh berupa algoritma seleksi. Penelitian dapat mengetahui bagaimana opini masyarakat tentang pembelajaran tatap muka dan sebagai masukan pemerintah sebagai selaku pembuat kebijakan.

2. METODE

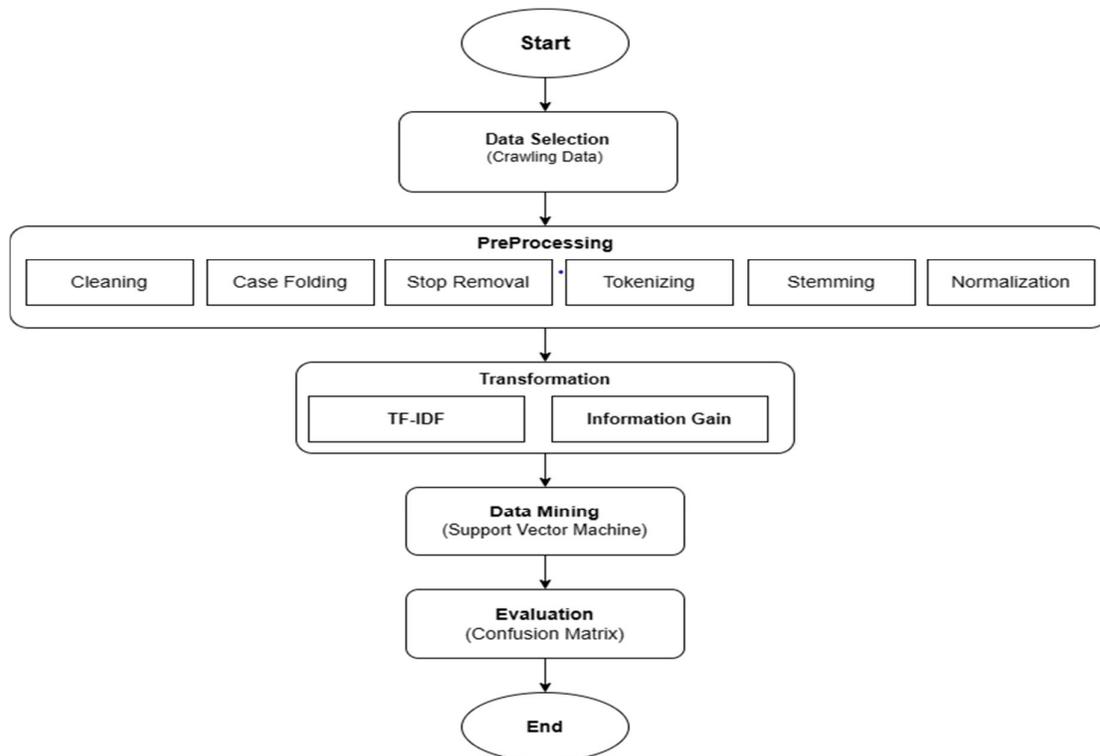
Penelitian yang dilakukan adalah jenis penelitian eksperimen, bertujuan untuk mendapatkan informasi perihal keadaan suatu fenomena pada saat dilakukannya penelitian. Penelitian mempunyai 4 tahap yaitu tahap pendahuluan, tahap pengumpulan data, tahap pengolahan data, tahap analisis dan pembahasan serta tahap penarikan kesimpulan.

2.1 Objek penelitian

Objek penelitian ini berupa data postingan pada aplikasi *Twitter* mengenai tanggapan mengenai kebijakan sekolah tatap muka saat pandemi di Indonesia. Data diambil dengan menggunakan kombinasi dari kata kunci sekolah tatap muka. Kata kunci tersebut dipilih karena termasuk ke dalam ruang lingkup topic yang dibahas. Proses pengambilan data berasal dari aplikasi *Twitter* diakses menggunakan *tool twitnt* pada Bahasa pemrograman python.

2.2 Rancangan penelitian

Langkah-langkah yang dilakukan untuk menyelesaikan penelitian ini menggunakan *Knowledge Discovery in Database* (KDD), seperti terlihat pada gambar 1.



Gambar 1. Bagan alur penelitian

2.2.1 Data selection

Pengumpulan data merupakan salah satu langkah dalam proses metode *Knowledge Discovery in Database* (KDD), menggunakan kumpulan data yang ada untuk mengidentifikasi pola atau hubungan dalam kumpulan data yang dapat digunakan untuk menginformasikan keputusan di masa depan [10]. Metode Langkah-langkah ini digunakan untuk mengumpulkan dan mengkaji data penelitian. Ada berbagai tahapan dalam penemuan *Knowledge Discovery in Database* (KDD). Langkah-langkah yang tercantum termasuk mengumpulkan data, membersihkannya, memprosesnya, mengujinya [11].

Seleksi Data untuk penelitian ini dikumpulkan dengan teknik *crawling* dengan bahasa pemrograman *python* menggunakan *tool google colab*. Kata kunci yang digunakan adalah “pembelajaran tatap muka” dan “sekolah *offline* pandemi”. Setelah menerima data dari tweet tersebut, maka data tersebut akan diklasifikasikan menjadi 2 kategori yaitu positif dan negatif. Selama proses penandaan ini, kami akan meminta bantuan ahli bahasa untuk meminimalkan kesalahan penandaan karena ahli bahasa dapat lebih memahami makna teks [12].

Tahap pertama, data berupa *tweet* dikumpulkan melalui teknik pengumpulan data bahasa pemrograman *Python* pada aplikasi *Twitter*. Periode publikasi kata kunci penemuan data adalah mulai 1 Januari hingga 30 Desember 2021 sebanyak 881 data.

2.2.2 Preprocessing

Langkah preprocessing adalah proses pelatihan dan pembersihan data sebelum klasifikasi, proses dari langkah *preprocessing*:

a. Case Folding

Langkah pertama dalam proses ini adalah case folding, yang bertujuan untuk mengubah teks menjadi huruf kecil. Contohnya adalah tweet “BISSMILAH SEMUA YANG IKUT PCR NEGATIF BIAR SENEN PTM” maka dilakukan perubahan menjadi “bismillah semua yang ikut pcr negatif biar senen ptm”.

b. Cleaning Proses

Cleaning yaitu pembersihan data, termasuk menghapus emoticon, hashtag, URL, username, atau simbol yang tidak perlu. Contohnya pada *tweet* “bismillah semua yang ikut pcr negatif biar senen ptm <https://t.co/42zqxkarlu>” diubah menjadi “bismillah semua yang ikut pcr negatif biar senen ptm”.

c. Tokenizing

Langkah selanjutnya adalah tokenizing. Langkah ini bertujuan untuk memotong kalimat kata demi kata. Contohnya seperti pada *tweet* “kangen sekolah tatap muka” menjadi “kangen” “sekolah” “tatap” “muka”.

d. Stemming Proses

Stemming adalah proses melepas imbuhan. Setelah melakukan tokenizing maka akan dihapus dari token 19 yang dihasilkan. Contohnya “para” “siswa” “bisa” “Kembali” “menikmati” “sekolah” “offline” saat dilakukan stemming maka imbuhan yang terdapat pada token akan dihilangkan menjadi “para” “siswa” “bisa” “kembali” “nikmat” “sekolah” “offline”.

e. Stopword Removal

Stopword Removal dengan daftar 32 stopwords list akan menghapus kata-kata dalam data yang berisi kata-kata dari daftar. *stopwords list* berisi kata-kata yang tidak mempunyai peran atau nilai dalam proses klasifikasi.

f. Normalization

Dalam data akan terdapat berbagai kata yang memiliki kesalahan ketik maupun mengandung bahasa slang. Untuk itu dilakukan normalization untuk memperbaiki kata yang memiliki kesalahan. *Preprocessing* merupakan upaya untuk mengurangi kenyaringan kata, menghilangkan noise, dan membuat jenis kata menjadi lebih homogen. Langkah-langkah pelipatan kasus, pembersihan, *stemming*, tokenisasi, dan penghapusan *stopword* kini terlibat.

2.2.3 Transformation

Pada titik ini, transformasi dilakukan dengan menggunakan *Information Gain* setelah memberikan nilai pada setiap kata menggunakan TF-IDF. Jumlah informasi yang diperlukan untuk membuat pilihan klasifikasi yang tepat dalam kategori apa pun yang memengaruhi hadir atau tidaknya suatu kata dapat diukur menggunakan teknik pemilihan fitur yang dikenal sebagai perolehan informasi [13]. Persamaan untuk memperoleh informasi tersebut dapat dilihat disajikan dalam rumus 1.

$$\text{info}(D) = -\sum_{i=1}^m (P_i) \log_2 (P_i) \quad (1)$$

Di mana

D = merupakan himpunan kasus,

m = jumlah kelas partisi D,

P_i = proporsi dari sampel acak D_i terhadap D dalam kelas m

Sedangkan p_i merupakan probabilitas sebuah tuple pada D yang masuk kedalam kelas D_i . Fungsi log dalam hal ini digunakan log berbasis 2 karena informasi dikodekan berbasis bit. Sebelumnya data dibagi persentase data training 80% dan data uji 20%. Pembagian data sebesar 80:20 mengacu pada penelitian [14], di mana perbandingan analisis sentimen dengan algoritma *Support Vector Machine* memberikan akurasi yang tinggi. Langkah berikutnya, pelaksanaan dengan proses pemilihan fitur dengan memakai algoritma *Information Gain* agar fitur tersebut digunakan dalam proses klasifikasi lebih sesuai.

2.2.4 Data mining

Data mining digunakan untuk diimplementasikan algoritma SVM untuk mengklasifikasikan data memakai Kernel Linear dan *Gaussian Radial Basis Function* (RBF). Penggunaan Kernel Linear dan *Gaussian Radial Basis Function* (RBF) mengacu pada hasil penelitian sebelumnya yang keduanya memiliki akurasi yang baik dan nilai yang relatif sama dalam analisis sentiment [15].

Pengaplikasian kernel linier pada penelitiannya mengacu pada hasil penelitian sebelumnya yang dilakukan oleh [16], bahwa kernel linier mendapatkan nilai ketepatan terbaik dibandingkan dengan RBF dan perkalian polinomial partikel. Persamaan untuk *Support Vector Machine* untuk kernel linier dinyatakan sebagai berikut.

$$K(x \cdot x') = x \cdot x' \quad (2)$$

Di mana:

x = matriks data uji (data latih/training)

x' = matriks transpose x

K = nilai Kernel Linier

Untuk Kernel Gaussian Radial Basis Function (RBF) sebagai berikut,

$$K(x \cdot x') = \exp\left(-\gamma \|x \cdot x'\|^2\right) \quad (3)$$

di mana

x = matriks data uji (data latih/training)

x' = matriks transpose x

γ = koefisien fungsi eksponen

K = nilai Kernel Gaussian Radial Basis Function

Parameter yang digunakan pada proses klasifikasi *Support Vector Machine* dengan parameter $\gamma = 0.001$.

2.2.5 Evaluasi

Pada langkah ini, pemeriksaan akan dilakukan dengan perhitungan model *Confusion Matrix*. Matriks ini digunakan untuk mengukur kinerja *classifier/supervised learning* dengan setiap kolom matriks mewakili kelas prediksi dan setiap baris matriks menyatakan kelas sesungguhnya [17]. Dalam pengklasifikasian data yang digunakan dalam matriks antara lain adalah *True Positive*, *True Negative*, *False Positif*, dan *False Negative*, untuk dipakai untuk mewakili luaran klasifikasi [18]. Dalam proses penyeleksian *benchmark* ini luarannya akan mempunyai tipe berbentuk *accuracy*, *recall*, dan *precision* [19]. Berdasarkan nilai tersebut akan dilihat seberapa baik dan akurat proses klasifikasi.

3. HASIL DAN PEMBAHASAN

Hasil penelitian dan pembahasan ini meliputi sub bahasan sebagai berikut, *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation*.

3.1 Data selection

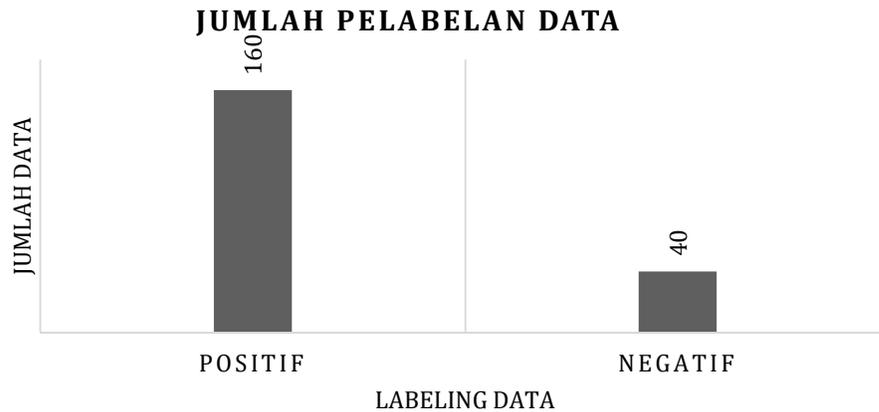
Pengumpulan data menemukan 881 baris dan 12 kolom yang digunakan, dimana hanya 2 kolom yang merupakan nama pengguna dan *tweet*, serta 10 kolom yang dihapus dan tidak digunakan. Langkah selanjutnya adalah proses penyortiran data dengan menghilangkan *tweet* yang selain berbahasa Indonesia dan *tweet* yang tidak berhubungan dengan topik pembelajaran di masa pandemi (*out of topic*).

Tabel 1. Hasil data selection

No.	username	tweet	value
0	NaN	@ABSetyono @PlateJohnny @kemkominfo @jokowi @K...	positif
1	NaN	@Mikhamikha21 @florielicious @Dien_Drew @Wong...	positif
2	NaN	@dinahsholikhah @suhermannasu @emangakunfake @o...	positif
3	NaN	@awanmalangit Alhamdulillah ikut senang Denga...	positif
4	NaN	@helloiammello @punyaleefelix @convomf pelajar...	positif
...
195	beardrblestar	Ini masih covid kan, KENAPA MALAH SEKOLAH GUE ...	negatif
196	mamvianawati	Benar sekolah akan di lakukan daring lahi ? ...	positif
197	ayenedede	KENAPA COVID HARUS NAIK HAHH??? SEKOLAH GUE KE ...	positif
198	profesorzubairi	Learning loss or life loss?? Ketika angka covi ...	negatif
199	dosamrabbat	Sekolah offline kangen banget sekolah offline ...	positif
200 rows x 3 columns			

Tabel 1 hasil seleksi setelah dilakukan seleksi data akan menghasilkan sebanyak 200 data. Sebanyak 200 data dievaluasi secara positif dan negatif dengan bantuan ahli bahasa Indonesia.

Gambar 2 data berlabel negatif adalah data yang tidak setuju atau mengeluhkan kebijakan pembelajaran langsung pada masa masa pandemi Covid-19. Pelabelan data menunjukkan sentimen positif sebanyak 160 (80,00%) data melebihi sentimen negatif sebanyak 40 (20,00%) data.



Gambar 2. Perbandingan sentimen

3.2 Preprocessing

Data yang tidak berharga dan banyak *noise*, data yang diperoleh tidak akan langsung digunakan dalam proses penambangan data. Pemrosesan awal diperlukan untuk membersihkan data dan memilih atribut yang tidak akan mengganggu klasifikasi. Tabel 2 tahapan *preprocessing* menggunakan bahasa pemrograman Python ditunjukkan sebagai berikut. Prosedur *Cleaning* pertama, prosedur *Case Folding* kedua, dan prosedur *Normalisasi* ketiga.

Tabel 2. Hasil tahapan *preprocessing*

Tahap Preprocessing	Hasil
<i>Dataset</i>	Maksa ptm di sikon kek gini tuh kek tinggal menunggu kabar covid aja ga si :@ https://t.co/F6auA1mpol
<i>Cleaning</i>	Maksa ptm di sikon kek gini tuh kek tinggal menunggu kabar covid aja ga si
<i>Case Folding</i>	maksa ptm di sikon kek gini tuh kek tinggal menunggu kabar covid aja ga si
<i>Stemming</i>	Maksa ptm di sikon kek gini tuh kek tinggal tunggu kabar covid aja ga si
<i>Stopword Removal</i>	Maksa ptm sikon kek gin tuh kek tinggal tunggu kabar covid aja ga si
<i>Tokenizing</i>	maksa,ptm,sikon,kek,gin,tuh,kek,tinggal,tunggu,kabar,covid,aja,ga,si
<i>Normalization</i>	paksa,pembelajaran tatap muka ,kondisi,kayak,gini,itu,kayak,tinggal,tunggu ,kabar,covid,saja,tidak,si

Preprocessing memiliki dampak yang cukup besar dalam meningkatkan kinerja sistem. Studi ini akan berkonsentrasi pada penerapan beberapa metode *preprocessing* untuk memastikan bagaimana pra-pemrosesan mempengaruhi kinerja analisis sentimen. Kombinasi metode *preprocessing* manakah yang memberikan hasil analisis sentimen terbaik untuk komentar Twitter. *Preprocessing* yang digunakan dalam penelitian ini mencakup *case folding*, *cleaning*, *stopword removal*, *stemming*, dan penambahan kata untuk menyiasati penggunaan istilah-istilah tidak umum yang sering muncul ketika menganalisis data Twitter [20].

3.3 Transformation

Konversi dilakukan dengan metode TF-IDF untuk menyusun struktur data. TF-IDF memproses tweet sebagai teks untuk mendapatkan suatu nilai. Langkah awal, dilakukan pengelompokan sebanyak 200 data menjadi persentase perbandingan, seperti yang ditunjukkan pada Tabel 3.

Tabel 3. Jumlah data training dan data testing dari 200 data

Split Data (%)	Jumlah Data	
	Training	Testing
80 : 20	160	40

Setelah *splitting* data, hitung istilah untuk setiap tweet dan lakukan pembobotan TF-IDF. **Tabel 4** menunjukkan luaran pembobotan TF-IDF dengan menggunakan bahasa pemrograman Python. Setiap term mempunyai nilai TF-IDF untuk mengetahui seberapa sering kata tersebut muncul. Sebuah kata yang jarang muncul akan memiliki pengaruh yang lebih kecil pada proses klasifikasi dan oleh karena itu akan diseleksi ulang oleh algoritma pemilihan fitur.

Tabel 4. Hasil term frequency - inverse document frequency (TF-IDF)

No	index	abdi	ada	adek	agama	ah	ajar	akh	akowkawok	aku	alah	alas	alhamdulillah	alias
0	0.0	0.0	0.0	0.0	0.0	0.10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	2	0.0	0.0	0.0	0.0	0.0	0.19	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	3	0.0	0.0	0.0	0.0	0.0	0.16	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	4	0.0	0.0	0.0	0.0	0.0	0.11	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
160	0.0	0.0	0.0	0.0	0.0	0.09	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
161	0.0	0.0	0.0	0.0	0.18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
162	0.0	0.0	0.0	0.0	0.0	0.11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
163	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
164	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

165 rows x 769 columns

Analisis dampak fitur dua level dari TF-IDF dan N-Gram pada dataset SS-Tweet dari analisis sentimen [21]. Menggunakan level kata TF-IDF (Terminal Frekuensi Invers Frekuensi), kinerja analisis sentimen 3-4% lebih baik dibandingkan menggunakan fitur N-gram, analisis dilakukan menggunakan enam jenis algoritma analisis (*Decision Tree, Support vector Machine, K-Nearest Neighbour, Random Forest, Logistic Regression, Naive Bayes*) dan mempertimbangkan parameter kinerja *F-Score, Accuracy, Precision, dan Recall*.

Tabel 5. Hasil information gain

Threshold: 0.0005.

Previos number of features: 665

Number of features information gain: 665

Threshold sebesar 0,0005 digunakan untuk ujicoba proses *threshold* sebesar 0,0005 yang menghasilkan luaran *features after*. Jumlah information *gain* sebanyak 665, yang tersaji dalam **Tabel 5**, kemudian dilakukan seleksi fitur untuk data tersebut yang tidak terlalu berpengaruh.

3.4 Data mining dan evaluation

Langkah proses ini akan menghasilkan klasifikasi sentimen untuk pelaksanaan kebijakan pembelajaran tatap muka 2021 menggunakan *Support Vector Machine* (SVM). Kernel yang digunakan yaitu kernel linear dan RBF. Akurasi kernel linier untuk 80% : 20% menghasilkan prosentase Kernel Linear sebesar 85,00%, dengan confusion matriks $\begin{bmatrix} 2 & 5 \\ 1 & 32 \end{bmatrix}$. Sesuai dengan hasil perhitungan diperoleh nilai presisi 40%, akurasi 85%, dan recall 66,67%. Selain itu, 2 data

kelas positif dan 32 data kelas negatif telah diprediksi dengan tepat. Akurasi kernel untuk 80% : 20% menghasilkan prosentase Kernel RBF sebesar 82,50%, dengan confusion matriks $\begin{bmatrix} 0 & 7 \\ 0 & 33 \end{bmatrix}$.

Hasil perhitungan menunjukkan bahwa tidak terdapat data kelas positif yang diprediksi tepat dan terdapat total 33 data kelas negatif yang diprediksi tepat. Precision 0%, Accuracy 82,5%, dan Recall tidak ada nilainya. Hasil presentasi *accuracy* dari Kernel Linear sebesar 86% dan Kernel RBF sebesar 82%, di mana Kernel Linear lebih besar kemiripan dari Kernel RBF. Data ini menunjukkan penggunaan model Kernel Linear lebih baik dari Kernel RBF.

4. SIMPULAN

Pembahasan Analisis Sentimen Twitter Terhadap Kebijakan pembelajaran tatap muka Tahun 2021 selama pandemi memakai model Algoritma *Support Vector Machine* dengan *Information Gain*. Pengkategorian sentimen pengguna Twitter terhadap kebijakan Pembelajaran Tatap Muka 2021 di masa pandemi menggunakan *Information Gain* menghasilkan total 665 fitur. Berdasarkan jumlah kelas positif yang lebih banyak dibandingkan kelas negatif, diketahui mayoritas masyarakat setuju dengan kegiatan pembelajaran tatap muka 2021 di masa pandemi. Evaluasi dan kinerja proses pada data mining yang memakai model *Support Vector Machine* dan *Information Gain* dihasilkan sebesar *Accuracy* sebesar 85,00%, *Precision* sebesar 40,00%, dan *recall* sebesar 66,67% untuk kernel linear sedangkan *Accuracy* sebesar 82,50%, *Precision* sebesar 0%, dan *recall* tidak mempunyai nilai untuk kernel *Radial Basis Function* (RBF). *Confussion Matrix* yang menggunakan *Radial Basis Function* (RBF) untuk *recall* tidak mempunyai nilai maka model tersebut menggambarkan bahwa keberhasilan model menemukan kembali sebuah informasi tidak berhasil. Hal tersebut menunjukkan bahwa model ini lebih cenderung model kernel linear. Penelitian selanjutnya diharapkan supaya data yang akan dipakai dapat diolah sebanyak mungkin dan kuantitas kelas negatif dan positif dibuat seimbang serta membandingkan beberapa model perbandingan sebaran data antara data data uji dan latih.

REFERENSI

- [1] WHO, "Dashboard Corona Virue Case," New York, 2021.
- [2] Kemendikbud-a, "Penyelenggaraan Pembelajaran Tatap Muka Tahun Akademik 2021/2022.," *Kementeri. Pendidik. dan Kebudayaan.-*, 2021.
- [3] Kemenkes, "Vaksinasi COVID-19 Nasional," 2021.
- [4] Kemendikbud-b, "Serba-Serbi Pembelajaran Tatap Muka Terbatas di Wilayah PPKM Level 3.," *Kementeri. Pendidik. dan Kebudayaan.-b*, 2021.
- [5] Fernando Perdana Putra, Iskandar, and N. Nurkholis, "Rancang bangun sistem pembelajaran sekolah berbasis web menggunakan framework codeigniter 4.0 (studi kasus SMP Muhammadiyah 1 Cileungsi)," *INFOTECH J. Inform. Teknol.*, vol. 3, no. 2, pp. 92–103, 2022, doi: 10.37373/infotech.v3i2.360.
- [6] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 157–163, 2021.
- [7] Reza Fadhilah, Pria Sukamto, and Nurkholis, "Analisis sistem ujian berbasis mobile di SMK Muhammadiyah 1 Cileungsi," *INFOTECH J. Inform. Teknol.*, vol. 3, no. 1, pp. 38–45, 2022, doi: 10.37373/infotech.v3i1.180.
- [8] J. P. Singh, Y. K. Dwivedi, N. P. Rana, A. Kumar, and K. K. Kapoor, "Event classification and location prediction from tweets during disasters," *Ann. Oper. Res.*, vol. 283, pp. 737–757, 2019.
- [9] S. Efendi and P. Sihombing, "Sentiment Analysis of Food Order Tweets to Find Out Demographic Customer Profile Using SVM," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 583–594, 2022.
- [10] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi

- Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, pp. 56–62, 2020.
- [11] A. V. Sudiantoro *et al.*, "Analisis Sentimen Twitter Menggunakan Text Mining Dengan," vol. 10, no. 2, pp. 398–401, 2018.
- [12] W. E. Nurjanah, R. Setya Perdana, and M. A. Fauzi, "Analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial twitter menggunakan metode k-kearest neighbor dan pembobotan jumlah retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1750–1757, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [13] A. B. P. Negara, H. Muhandi, and I. M. Putri, "Analisis sentimen maskapai penerbangan menggunakan metode naive bayes dan seleksi fitur information gain," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, 2020.
- [14] S. M. ARFINA, "ANALISIS SENTIMEN OPINI MASYARAKATPENGGUNA TWITTER TERHADAPPARIWISATA LAMPUNGMEGGUNAKAN SUPPORT VECTOR MACHINEDANNAIVE BAYES," 2022.
- [15] J. Li, G. Lu, B. Zhang, J. You, and D. Zhang, "Shared linear encoder-based multikernel Gaussian process latent variable model for visual classification," *IEEE Trans. Cybern.*, vol. 51, no. 2, pp. 534–547, 2019.
- [16] F. F. Irfani, M. Triyanto, and A. D. Hartanto, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, p. 258, 2020, doi 10.26487/jbmi.v16i3.8607, 2020.
- [17] A. Kulsumarwati, I. Purnamasari, and B. A. Darmawan, "Penerapan SVM dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, pp. 101–109, 2021.
- [18] N. Amalina, "Uji akurasi aplikasi Augmented Reality pembelajaran huruf alfabet Bahasa Isyaratindonesia (BISINDO) pada Vuforia menggunakan Confusion Matrix," Universitas Islam Negeri Maulana Malik Ibrahim, 2019.
- [19] R. Melita, V. Amrizal, H. B. Suseno, T. Dirjam, T. Informatika, and F. Sains, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam)," *J. Tek. Inf.*, vol. 11, no. 2, pp. 149–164, 2018.
- [20] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 406–414, 2021.
- [21] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019.