



Analisis Sentimen Pembelajaran Daring Era Pandemi COVID-19

Menggunakan Naive Bayes Dan Support Vector Machine

Sentiment Analysis of Online Learning in the Era of COVID-19 Pandemic Using Naive Bayes and Support Vector Machine Algorithms

Ketut Mediana Ayu Candrayani¹, I Made Agus Dwi Suarjaya², Anak Agung Ketut Agung Cahyawan Wiranatha³

^{1,2,3}Prodi Teknologi Informasi, Fakultas Teknik, Universitas Udayana

¹dianmediana5@gmail.com, ²agussuarjaya@it.unud.ac.id, ³agung.cahyawan@unud.ac.id

Abstract

The emergence of the COVID-19 virus in Indonesia has a significant impact on all sectors including education. To reduce the spread of the virus, the government created an online learning system policy that requires learning activities to be carried out from home. This drastic change has led to various responses from the public, including on Twitter. The public opinion contained in tweets is textual data that can be extracted and processed to understand the views and feelings of the public on a topic, called sentiment analysis. Sentiment analysis is an application of Big Data, the science of handling large and complex data to obtain important information, reveal hidden patterns, and help the decision-making process. This study aims to determine public sentiment toward online learning during the COVID-19 pandemic using Twitter data. The data used were 953,378 tweets from January 2020 to May 2022, which were classified into 3 classes (positive, negative, and neutral) using the Naive Bayes and SVM algorithms. The results showed that the percentage of neutral sentiment was 43.28%, negative 32.91%, and positive 23.82%. This shows that public sentiment tends to be neutral, meaning that people do not fully support or oppose online learning during the COVID-19 pandemic.

Keywords: big data; naive bayes; online learning; sentiment analysis; support vector machine; twitter

Abstrak

Kondisi pandemi COVID-19 di Indonesia memberikan pengaruh yang signifikan di seluruh sektor kehidupan, tidak terkecuali pendidikan. Dalam upaya mengurangi tingkat penyebaran virus dan memastikan bahwa kegiatan pembelajaran tetap berjalan meski di situasi pandemi, pemerintah mencetuskan kebijakan sistem pembelajaran daring yang mewajibkan kegiatan pembelajaran dijalankan secara *online* dari rumah. Perubahan drastis yang terkesan mendadak ini memunculkan beragam respons dari masyarakat termasuk dalam *platform* media sosial seperti Twitter. Opini masyarakat yang tertuang dalam tweet merupakan *textual* data yang dapat diekstrak dan diolah untuk dapat memahami pandangan dan perasaan masyarakat terhadap suatu topik, yang biasa disebut analisis sentimen. Analisis sentimen merupakan bentuk penerapan konsep *Big Data*, yaitu ilmu yang menangani kumpulan data besar dan kompleks untuk mendapatkan informasi penting, mengungkap pola tersembunyi, serta membantu pengambilan keputusan. Penelitian ini dilakukan untuk mengetahui sentimen masyarakat terhadap pembelajaran daring selama pandemi COVID-19 menggunakan data Twitter. Total data yang digunakan berjumlah 953.378 *tweet* terhitung dari bulan Januari 2020 sampai dengan Mei 2022, yang diklasifikasi menggunakan tiga kelas sentimen, yaitu negatif, positif, dan netral. Model klasifikasi dibentuk untuk mengklasifikasi data *tweet* dengan mengimplementasikan metode TF-IDF untuk pembobotan kata, serta 2 algoritma *machine learning*, yaitu *Naive Bayes* dan *Support Vector Machine* (SVM). Hasil evaluasi model menemukan bahwa persentase akurasi tertinggi mencapai 84% menggunakan SVM dan hasil penelitian menunjukkan bahwa data didominasi oleh sentimen netral dengan persentase sebesar 43.28%, sementara persentase sentimen negatif 32.91%, dan sentimen positif 23.82%. Hal ini mencerminkan bahwa sentimen masyarakat cenderung netral, dalam artian masyarakat tidak sepenuhnya mendukung maupun menentang pembelajaran daring di era pandemi COVID-19.

Kata kunci: analisis sentimen; big data; naive bayes; pembelajaran daring; support vector machine; twitter

1. Pendahuluan

Pada 11 Maret 2020 World Health Organisation resmi mendeklarasikan virus COVID-19 sebagai pandemi

global sejak ditemukan pertama kali di Wuhan pada Desember 2019 [1]. Di Indonesia angka kematian akibat virus ini terus bertambah sejak diumumkan kasus positif pertama pada 2 Maret 2020 [2]. Munculnya virus COVID-19 ini mengakibatkan Indonesia mengalami keadaan darurat nasional yang berdampak pada seluruh sektor kehidupan, tidak terkecuali pendidikan. Dalam dunia pendidikan, pemerintah Indonesia mencetuskan kebijakan sistem pembelajaran daring yang mewajibkan kegiatan pembelajaran dilaksanakan secara daring (*online*) dari rumah dengan tujuan meminimalkan penyebaran virus dan memastikan bahwa KBM tetap terlaksana meski di kondisi pandemi. Pembelajaran daring dalam pelaksanaannya tidak terlepas dari peranan teknologi seperti perangkat *mobile*, jaringan internet, dan media belajar kelas virtual untuk mendukung proses pembelajaran secara daring [3].

Perubahan drastis yang terkesan mendadak ini seolah 'memaksa' seluruh jenjang pendidikan untuk mampu bertransformasi dan beradaptasi dengan sistem pembelajaran daring. Penguasaan teknologi yang masih rendah menjadi salah satu kendala dari pelaksanaan pembelajaran daring selama pandemi. Kondisi ini akhirnya memunculkan beragam respons dari masyarakat termasuk dalam *social media platforms* seperti Twitter. Opini masyarakat pada media sosial Twitter yang dituangkan dalam bentuk *tweet* merupakan *textual* data yang diekstrak dan diolah untuk dapat memahami pandangan dan perasaan masyarakat terhadap suatu topik, atau biasa disebut analisis sentimen. Analisis sentimen merupakan salah satu bentuk penerapan konsep *Big Data*, yaitu ilmu yang menangani kumpulan data besar dan kompleks dengan mengekstraksi dan menganalisis data untuk mendapatkan informasi penting yang terkandung didalamnya, mengungkap pola tersembunyi, dan membantu dalam proses pengambilan keputusan [4]. Dalam praktiknya, konsep *Big Data* tidak terlepas dari implementasi algoritma pembelajaran mesin (*machine learning*) untuk membantu mengidentifikasi pola, tren, dan informasi berguna yang terkandung dalam data.

Seiring perkembangan teknologi yang kian canggih, penelitian terkait analisis sentimen saat ini telah berkembang luas dikalangan peneliti. Penelitian terkait analisis sentimen banyak diterapkan dalam berbagai bidang, contohnya di bidang kesehatan seperti pada tahun 2021 analisis sentimen dilakukan terhadap vaksin Sinovac dengan menerapkan metode *Naive Bayes*. Hasil penelitian menunjukkan bahwa sentimen masyarakat didominasi oleh kelas positif dengan persentase 86% [5]. Pada bidang politik, penelitian terkait *sentiment analysis* dilakukan dengan mengangkat topik terkait pemilihan Presiden Amerika dengan metode *Naive Bayes* dan *SVM*, serta pembobotan kata TF-IDF. Hasil evaluasi model menggunakan *cross-validation* menunjukkan bahwa

akurasi tertinggi mencapai 82% menggunakan *SVM* kernel linear. Penelitian tersebut juga menemukan bahwa sentimen didominasi oleh sentimen netral, baik untuk Donald Trump maupun Joe Biden [6]. Pada bidang olahraga, penelitian analisis sentimen mengangkat topik keberhasilan Indonesia dalam Thomas Cup 2020 dengan algoritma *Naive Bayes* dan *Decision Tree*. Hasil penelitian tersebut menemukan bahwa sentimen masyarakat cenderung netral, dengan akurasi tertinggi mencapai 83% menggunakan metode *Naive Bayes* [7].

Pembelajaran daring menjadi salah satu alternatif dalam keberlangsungan pendidikan di tengah kondisi pandemi COVID-19. Sistem ini memberikan pengaruh yang besar pada kualitas pendidikan dan keberhasilan peserta didik dalam proses belajar. Dengan demikian perlu dilakukan analisis untuk mengetahui bagaimana pandangan atau persepsi publik terhadap pembelajaran daring selama pandemi. Penelitian ini akan memberikan informasi yang berguna untuk mengetahui kecenderungan serta perkembangan sentimen masyarakat terhadap pembelajaran daring selama pandemi COVID-19. Informasi yang dihasilkan dapat membantu pemerintah, lembaga pendidikan, dan masyarakat untuk mengevaluasi proses pembelajaran, meningkatkan partisipasi masyarakat dalam memperkuat sistem pendidikan, dan menentukan strategi yang tepat dalam upaya peningkatan kualitas pembelajaran, baik dari segi pengembangan infrastruktur maupun fasilitas untuk mendukung proses pembelajaran diberbagai wilayah termasuk daerah pelosok yang memiliki keterbatasan akses internet maupun dukungan teknologi yang memadai.

Penelitian ini berupaya melakukan analisis sentimen untuk mengetahui pandangan masyarakat terhadap pembelajaran daring selama pandemi COVID-19 menggunakan data Twitter. Data yang digunakan berjumlah 953.378 *tweet* terhitung dari Januari 2020 sampai Mei 2022, yang diklasifikasi menggunakan tiga kelas data, meliputi kelas positif, negatif, dan netral dengan mengimplementasikan 2 algoritma *machine learning*, yaitu *Naive Bayes* dan *Support Vector Machine* (*SVM*).

Naive Bayes dipilih karena kemampuannya yang baik dalam klasifikasi data walau hanya membutuhkan sedikit data latih [8]. Selain itu, algoritma ini memiliki kesederhanaan dalam perhitungan dan kecepatan pemrosesan yang tinggi pada *database* dengan data yang besar [9]. Berdasarkan penelitian, *Naive Bayes Classifier* sering digunakan pada penelitian di bidang *text mining* karena dinilai memiliki kinerja yang cukup baik serta memiliki tingkat akurasi yang tinggi [10]. Sementara keunggulan yang dimiliki *SVM* ialah kemampuan generalisasi yang tinggi dan dinilai dapat melakukan klasifikasi data dengan baik walaupun dilatih dengan dengan jumlah data yang sedikit [11].

Walaupun memiliki waktu pelatihan yang cenderung lebih lambat, tetapi dengan kemampuannya yang mampu menangani data non linier kompleks membuat metode ini memiliki nilai akurasi yang lebih tinggi [12].

Penelitian terkait analisis sentimen sebelumnya pernah dilakukan pada Maret 2023 lalu terhadap pembukaan pariwisata di masa pandemi COVID-19 menggunakan K-NN dan *Naive Bayes*. Hasil evaluasi model menggunakan *confusion matrix* menunjukkan bahwa akurasi tertinggi dicapai dengan persentase 75.53% menggunakan metode *Naive Bayes* pada percobaan *split* data dengan rasio 80:20. Pada penelitian tersebut penulis juga menyarankan pengembangan sistem dengan menggunakan metode perbaikan kata baku untuk dapat meningkatkan akurasi model [13].

Penelitian [14] melakukan analisis sentimen terhadap aplikasi Ruang Guru menggunakan metode *Decision Tree* dan SVM. Data diklasifikasi menggunakan 2 kelas yaitu positif dan negatif. Hasil klasifikasi menunjukkan bahwa sentimen didominasi oleh kelas positif. Penelitian ini menemukan bahwa akurasi tertinggi diraih dengan menggunakan algoritma SVM kernel linear dengan perbandingan *data training* dan *data testing* 70:30.

2. Metode Penelitian

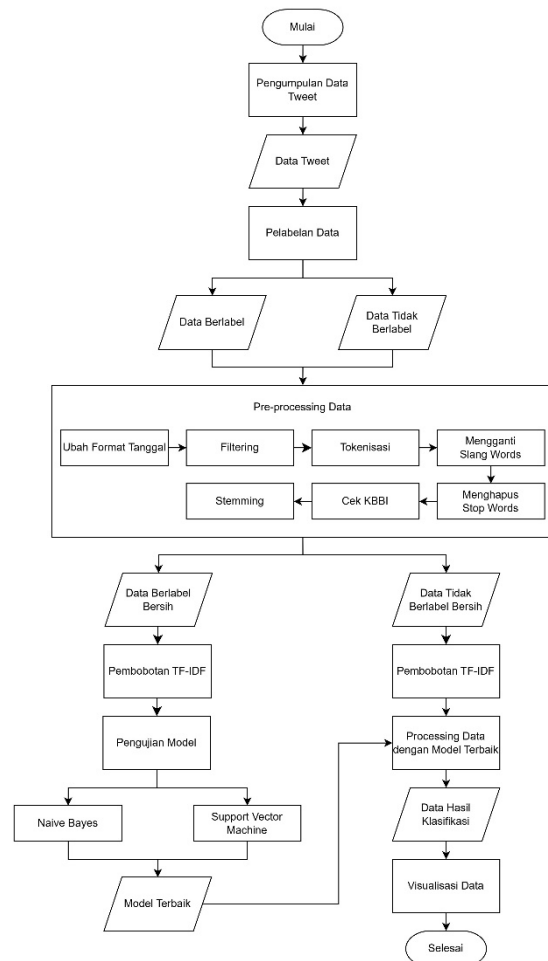
Metode penelitian terdiri dari 6 tahapan penelitian yang diawali dengan pengumpulan data *tweet*. Data *tweet* diambil dari bulan Januari 2020 sampai Mei 2022. Gambar 1 menggambarkan *flowchart* metode penelitian.

Tahap pelabelan data dilakukan oleh penulis secara manual pada 3.000 data *tweet*, terdiri dari 1000 data untuk setiap kelas (positif, negatif, dan netral). Setelah dilakukan pelabelan data, antara kelompok data berlabel dan data tidak berlabel dipisahkan. Data berlabel nantinya akan digunakan untuk menguji performa model.

Tahap ketiga merupakan *pre-processing* data yang dilakukan untuk membuat data menjadi lebih bersih sehingga proses klasifikasi oleh algoritma *machine learning* dapat memberikan hasil yang lebih optimal. Tahap ini dilakukan pada data berlabel maupun data tidak berlabel.

Tahap *pre-processing* pada penelitian ini meliputi proses ubah format tanggal *tweet*, *filtering* yaitu menghilangkan karakter-karakter yang tidak diperlukan (*hashtag*, *mention*, *punctuation*, angka, URL, dan karakter lainnya) termasuk *lowering case*, dilanjutkan dengan tokenisasi yaitu memecah struktur kalimat menjadi kumpulan kata/token, mengganti *slang words* yaitu memperbaiki ejaan kata menjadi bentuk baku, menghapus *stop words*, cek KBBI yaitu menghapus kata yang tidak masuk dalam Kamus Besar Bahasa Indonesia, dan terakhir adalah *stemming* yang akan

mengubah bentuk kata turunan (berimbuhan) menjadi kata dasar. Proses ini menghasilkan kelompok data berlabel bersih dan kelompok data tidak berlabel bersih



Gambar 1. Metode Penelitian

Tahap pengujian model menggunakan data berlabel bersih dengan mengimplementasikan metode TF-IDF untuk menghitung bobot kata (*term*) serta algoritma *Naive Bayes* dan SVM [15]. *Naive Bayes* adalah metode klasifikasi probabilitas sederhana dengan menerapkan teori *Teorema Bayes* yang independent pada setiap kondisi [16]. Formula 1 merupakan perhitungan *Teorema Bayes*.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A)$ menunjukkan prob. prior dari peristiwa A, sementara $P(B)$ merupakan probabilitas prior peristiwa B. $P(A|B)$ merupakan perobabilitas posterior dari peristiwa A dengan asumsi B terjadi, dan $P(B|A)$ sebaliknya [17].

Support Vector Machine (SVM) adalah metode klasifikasi *supervised learning*, yang dapat mengatasi permasalahan baik secara linier maupun non-linier [18]. SVM menggunakan bantuan kernel untuk menghubungkan *training data input* ke dimensi yang

lebih besar dan mengidentifikasi *hyperplane* sebagai ruang pemisah [19]. Pada penelitian ini menggunakan SVM *kernel* RBF dengan persamaan seperti yang ditunjukkan pada formula 2 [20].

$$K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2) \quad (2)$$

Variabel X_i dan X_j yang masing-masing merupakan dua titik data dalam ruang fitur. Pengujian model dilakukan untuk mengetahui performa setiap algoritma *machine learning* yang digunakan. Hasil pengujian model meliputi nilai *accuracy* (akurasi), *precision* (presisi), *recall* (sensitivitas), dan *f1 score*. Nilai akurasi menunjukkan seberapa akurat hasil prediksi benar yang dihasilkan dari keseluruhan data [21]. Pada tahap ini model dengan nilai akurasi tertinggi dianggap sebagai model terbaik, yang selanjutnya akan digunakan pada tahap *processing* data.

Pada tahap *processing* data, dilakukan klasifikasi menggunakan data tidak berlabel bersih dengan mengimplementasikan model terbaik yang sebelumnya telah diuji. Tahap ini juga diterapkan metode pembobotan kata TF-IDF. Proses ini akan menghasilkan *output* berupa data hasil klasifikasi.

Tahap terakhir yaitu visualisasi data, merupakan tahap yang merepresentasikan data hasil klasifikasi dari proses *processing* ke bentuk visual menggunakan berbagai jenis *chart* untuk memudahkan dalam pengambilan informasi terkait sentimen masyarakat terhadap topik yang diangkat.

3. Hasil dan Pembahasan

3.1. Pengumpulan dan Pelabelan Data

Proses pengumpulan data Twitter menghasilkan data *tweet* sejumlah 953.378 data. Filter data diterapkan untuk mengambil data *tweet* berbahasa Indonesia dari rentang waktu tertentu serta menggunakan beberapa kata kunci yang berkaitan dengan topik penelitian.

Proses pelabelan data menghasilkan data berlabel berjumlah 3000 *tweet* yang terbagi menjadi tiga kelas sentimen, meliputi negatif, positif, dan netral. Tabel 1 menunjukkan contoh hasil pelabelan data.

Tabel 1 Contoh Hasil Pelabelan Data Tweet

Label	Tweet
Positif	Online, udah nyaman aja. Sekolah offline menurut aku ribet kalo lagi pandemi giniðŸ- https://t.co/23yAWcHD32
	@lilkookiest Online lebih santuy dan nyaman di rumah ya kawan :3 gk usah capek2 pergi ke sekolah atau kuliah :v
Negatif	@schfess Aku males udah sekolah online pake kelompokan pula, ga enak bgt sumpah ga efektif kesimpulan 2 semester kuliah online; sangatlah tidak efektif bagi kemampuan otak ini. UNTUNG aja udah selesai nih 7 semester
Netral	Bosen online, tapi malah jadi ga siap kuliah offline. Serba salah emang

Label	Tweet
	kadang pingin pol kuliah offline, tapi kalo udah nyaman online gmnðŸ~ç

3.2. Pre-processing Data

Proses *pre-processing* data didukung oleh beberapa *library* untuk membantu dalam pembersihan data. Beberapa *library* yang digunakan antarlain seperti RegEx untuk proses *filtering*, NLTK untuk tokenisasi dan *remove stop words*, dan Sastrawi untuk proses *stemming* berbahasa Indonesia.

Setelah melewati tahap *pre-processing*, ditemukan bahwa pada kelompok data tidak berlabel bersih terdapat penurunan jumlah data akibat adanya data yang kosong (*null*). Tabel 2 menggambarkan tahapan-tahapan dalam *pre-processing*.

Tabel 2. Contoh Hasil Pre-processing Data

Sebelum Pre-processing	
Tanggal	2021-10-03T11:58:55+00:00
Tweet	Gw kapan kuliah offline? Tapi takut ketemu orang-orang tapi gak nyaman juga kalo terusÀan online hmm...
Pre-processing	
Ubah Format	2021-10-03
Tanggal	
Filtering	gw kapan kuliah offline tapi takut ketemu orang orang tapi gak nyaman juga kalo terusan online hmm
Tokenisasi	['gw', 'kapan', 'kuliah', 'offline', 'tapi', 'takut', 'ketemu', 'orang', 'orang', 'tapi', 'gak', 'nyaman', 'juga', 'kalo', 'terusan', 'online', 'hmm']
Replace Slang Words	['aku', 'kapan', 'kuliah', 'luring', 'tapi', 'takut', 'ketemu', 'orang', 'orang', 'tapi', 'tidak', 'nyaman', 'juga', 'kalau', 'terusan', 'daring', 'hmm']
Remove Stop Words	['aku', 'kuliah', 'luring', 'tapi', 'takut', 'ketemu', 'orang', 'orang', 'tapi', 'tidak', 'nyaman', 'terusan', 'daring', 'hmm']
Cek KBBI	['aku', 'kuliah', 'luring', 'tapi', 'takut', 'ketemu', 'orang', 'orang', 'tapi', 'tidak', 'nyaman', 'terusan', 'daring']
Stemming	aku kuliah luring tapi takut ketemu orang orang tapi tidak nyaman terus daring

3.3. Pengujian Model

Skenario pengujian model dilakukan dengan menerapkan 3 rasio *split* data yang berbeda pada setiap algoritma. Hasil terbaik didapatkan pada skenario menggunakan rasio *split* data *training* dan *testing* 90:10, dengan akurasi 79.33% untuk *Naive Bayes* dan 84% untuk SVM.

SVM terbukti memiliki performa yang lebih unggul dibandingkan *Naive Bayes* berdasarkan hasil dari pengujian model yang dilakukan. Hal ini didukung oleh kemampuan SVM yang mampu mengolah data berdimensi besar dan mampu menangani model non linier yang lebih kompleks, oleh karena itu nilai akurasi yang dihasilkan dapat lebih tinggi dan dikatakan bahwa SVM kurang rentan mengalami *overfitting* jika dibandingkan dengan algoritma lainnya [12]. Dengan

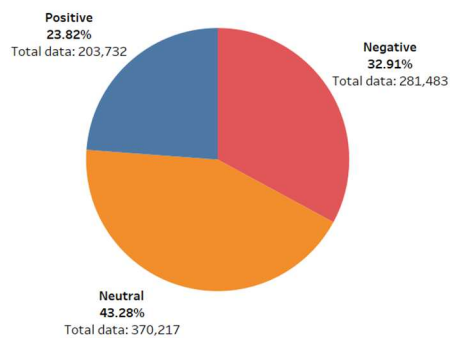
demikian SVM ditetapkan sebagai model terbaik yang akan digunakan selanjutnya pada tahap *processing* data. Tabel 3 menunjukkan perbandingan hasil pengujian model.

Tabel 3. Perbandingan Hasil Pengujian Model

Train:Test		70:30 (%)	80:20 (%)	90:10 (%)
Naïve Bayes	Accuracy	76.0	76.17	79.33
	Recall	76.0	76.17	79.33
	Precision	76.43	76.48	79.91
	F1 Score	75.92	76.04	79.25
SVM	Accuracy	82.33	82.5	84.0
	Recall	82.33	82.5	84.0
	Precision	82.67	82.67	84.15
	F1 Score	82.31	82.46	84.03

3.4. Processing dan Visualisasi Data

Tahapan *processing* data menghasilkan *output* berupa data hasil klasifikasi yang selanjutnya digunakan pada tahap terakhir, yaitu visualisasi data untuk memudahkan dalam pengambilan informasi. Pada penelitian ini visualisasi digambarkan dalam bentuk *pie chart*, *line chart*, visualisasi *map*, dan *word cloud*.



Gambar 2. Pie Chart Jumlah Tweet Per Kelas

Gambar 2 merupakan visualisasi *pie chart* yang menunjukkan perbandingan persentase serta total data dari kelas negatif, positif, dan netral. Kelas netral yang memiliki persentase dan jumlah data terbanyak menunjukkan bahwa sentimen masyarakat terhadap pembelajaran daring selama pandemi cenderung netral.



Gambar 3. Line Chart Tren Sentimen Per Bulan

Gambar 3 merupakan visualisasi *line chart* yang menunjukkan tren sentimen kelas positif dan negatif per bulan. Kelas negatif ditandai dengan garis berwarna merah sementara kelas positif adalah garis berwarna

biru. Terjadi peningkatan yang cukup signifikan pada sentimen negatif di bulan Juli 2020 dengan selisih 8.234 menjadi 21.793 *tweet*. Salah satu faktor yang mempengaruhi peningkatan jumlah sentimen negatif tersebut adalah karena pada 13 Juli 2020 ditetapkan sebagai awal dimulainya Tahun Ajaran baru 2020/2021. Hal serupa terjadi pada bulan Desember 2020, dimana terdapat peningkatan jumlah sentimen negatif cukup signifikan yang diikuti dengan penurunan jumlah sentimen positif meski tidak signifikan. Hal tersebut dipengaruhi oleh keputusan bersama 4 menteri terkait kegiatan pembelajaran di perguruan tinggi yang dapat dilaksanakan secara *hybrid*, daring, dan luring, sebagaimana tercantum dalam Surat Edaran Nomor 6 Tahun 2020 tentang Penyelenggaraan Pembelajaran pada Semester Genap Tahun Akademik 2020/2021. Dapat diketahui bahwa sentimen masyarakat didominasi oleh kelas negatif, namun pada bulan Januari-Februari 2022 terdapat peningkatan jumlah sentimen positif yang membuatnya sedikit lebih unggul dari jumlah sentimen negatif. Kondisi ini dipengaruhi oleh Surat Edaran Mendikbudristek Nomor 2 Tahun 2022 tentang Diskresi Pelaksanaan Keputusan Bersama Empat Menteri tentang Panduan Penyelenggaraan Pembelajaran di Masa Pandemi COVID-19, yang menyebutkan bahwa pembelajaran luring dapat dilaksanakan pada daerah dengan status PPKM level 2 dengan jumlah siswa yang terbatas.

Perolehan jumlah data *tweet* terbanyak adalah pada bulan Maret 2020 yang mencapai 90.799 untuk total keseluruhan data. Kondisi ini sejalan dengan fakta bahwa sistem pembelajaran daring mulai diterapkan di Indonesia pada bulan Maret 2020 [22].



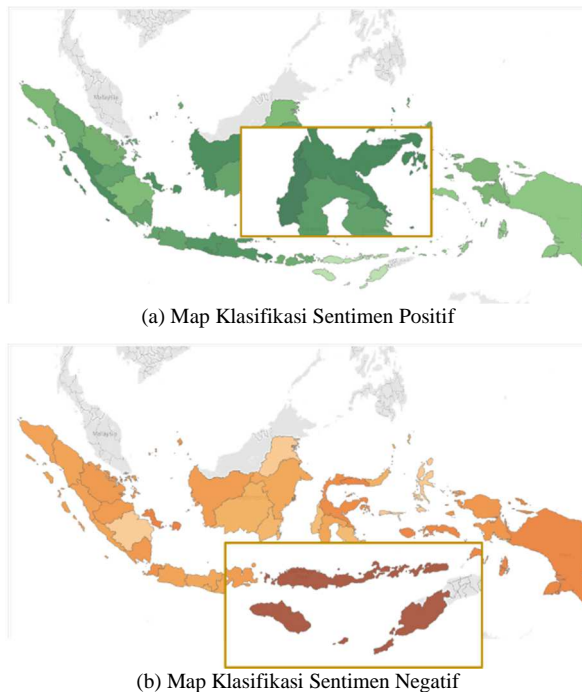
(a) Word Cloud Kelas Positif



(b) Word Cloud Kelas Negatif

Gambar 4. Word Cloud Kelas Positif dan Negatif

Gambar 4 merupakan visualisasi *word cloud* dari sentimen kelas positif dan negatif berdasarkan data hasil klasifikasi. Beberapa kata yang sering muncul dalam data yang diklasifikasi sebagai kelas positif, seperti kata “semangat”, “anak”, “tugas”, “tidur”, “nyaman”, dan lainnya mencerminkan bahwa masyarakat yang mendukung pembelajaran daring merasa nyaman dengan sistem pembelajaran daring yang berlangsung selama pandemi. Sementara kumpulan kata yang sering muncul dalam data yang diklasifikasi sebagai kelas negatif, seperti kata “tugas”, “ponsel”, “kuota”, “susah”, “capek”, dan lainnya menunjukkan bahwa masyarakat yang menentang pembelajaran daring masih menemukan banyak kendala dalam pelaksanaannya, contohnya seperti keterbatasan ponsel maupun kuota, merasa kesusahan, baik itu susah sinyal, susah beradaptasi, maupun alasan lainnya.



Gambar 5. Map Klasifikasi Sentimen Positif dan Negatif

Gambar 5 merupakan visualisasi *map* yang menampilkan sebaran data pada 34 Provinsi di Indonesia berdasarkan label klasifikasi positif dan negatif terhitung dari bulan Januari 2020 sampai dengan Mei 2022. Persentase *tweet* dengan label positif tertinggi ditemukan di Provinsi Sulawesi Barat yaitu sebesar 28.95%, sementara persentase *tweet* dengan label negatif tertinggi ditemukan di Nusa Tenggara Timur sebesar 51.97%.

4. Kesimpulan

Analisis sentimen dilakukan menggunakan data Twitter dengan jumlah data yang berhasil dikumpulkan sebanyak 953,378 *tweet*. Hasil pengujian model yang mengimplementasikan algoritma *Naive Bayes* dan

SVM ditemukan bahwa akurasi tertinggi mencapai 84% menggunakan SVM. Berdasarkan hasil klasifikasi data menggunakan model yang telah diuji, dapat disimpulkan bahwa sentimen masyarakat terhadap pembelajaran daring selama pandemi didominasi oleh sentimen netral dengan persentase sebesar 43.28% atau sebanyak 370.217 data. Proses pembersihan data pada tahap *pre-processing* dapat mempengaruhi tingkat akurasi model. Akurasi model juga dipengaruhi oleh *split* data *training* dan *testing* yang digunakan dalam pengujian. Tren sentimen masyarakat dipengaruhi oleh beberapa faktor, seperti dimulainya tahun ajaran baru, surat edaran, atau surat keputusan bersama yang diterbitkan oleh Kemendikbud maupun instansi pemerintah lainnya. Seperti pada bulan Juli 2020 terjadi peningkatan jumlah sentimen negatif yang cukup signifikan karena pada 13 Juli 2020 ditetapkan sebagai awal dimulainya Tahun Ajaran baru 2020/2021.

Daftar Rujukan

- [1] E. M. Rahmawati, Putri, “Learning From Home dalam Perspektif Persepsi Mahasiswa Era Pandemi Covid-19,” *Pros. Semin. Nas. Hardiknas*, vol. 1, pp. 17–24, 2020, [Online]. Available: <http://proceedings.ideaspublishing.co.id/index.php/hardiknas/article/view/3>.
- [2] R. N. Putri, “Indonesia dalam Menghadapi Pandemi Covid-19,” *J. Ilm. Univ. Batanghari Jambi*, vol. 20, no. 2, p. 705, 2020, doi: 10.33087/jiubj.v20i2.1010.
- [3] F. Firman and S. Rahayu, “Pembelajaran Online di Tengah Pandemi Covid-19,” *Indones. J. Educ. Sci.*, vol. 2, no. 2, pp. 81–89, 2020, doi: 10.31605/ijes.v2i2.659.
- [4] I. N. K. Bayu, I. M. A. D. Suarjaya, and P. W. Buana, “Classification of Indonesian Population’s Level Happiness on Twitter Data Using N-Gram, Naïve Bayes, and Big Data Technology,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 5, pp. 1944–1949, 2022, doi: 10.18517/ijaseit.12.5.14387.
- [5] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naïve Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 163–170, 2021.
- [6] G. Nugroho, D. T. Murdiansyah, and K. M. Lhaksana, “Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine,” vol. 8, no. 5, pp. 10106–10115, 2021.
- [7] F. Fitriana, E. Utami, and H. Al Fatta, “Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
- [8] A. Y. P. Yusuf and R. Sari, “Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Pemahaman Program MBKM Bagi Mahasiswa,” *J. Inform. Inf. Secur.*, vol. 3, no. 2, pp. 171–180, 2022, doi: 10.31599/jiforty.v3i2.1713.
- [9] H. P. Doloksaribu and Y. T. Samuel, “Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi,” *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022, doi: 10.47111/jti.v16i1.3747.
- [10] E. R. Kaburuan, Y. Sartika, and I. Agustina, “Sentiment Analysis on Product Reviews from Shopee Marketplace using the Naïve Bayes Classifier,” vol. 13, no. 3, pp. 150–159, 2022, doi: 10.24843/LKJITI.2022.v13.i03.p02.
- [11] Y. A. Singgalen, “Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646, 2022, doi:

- 10.47065/bits.v4i3.2766.
- [12] S. S. N. N. Ferdy Maylani, "Implementasi Metode Data Mining Untuk Memprediksi Warna Anak Kucing Pada Proses Pengembangbiakan Kucing Ras Menggunakan Algoritma Support Vector Machine (SVM)," *Pros. Semin. Nas. Dharmajaya*, pp. 114–125, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2921>.
- [13] J. S. Komputer, D. Era, S. Andryana, A. Rubhasy, and U. Nasional, "Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19," vol. 7, pp. 263–272, 2023, doi: 10.30645/j-sakti.v7i1.590.
- [14] I. N. Hassanah, S. Faisal, and A. M. Siregar, "PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DENGAN DECISION TREE PADA APLIKASI RUANG GURU," pp. 39–50, doi: 10.20527/klik.v10i1.602.
- [15] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *J. Ilm. Educ.*, vol. 7, no. 1, pp. 1–11, 2020.
- [16] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019, doi: 10.24176/simet.v10i2.3487.
- [17] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [18] T. S. Sabrila, Y. Azhar, and C. S. K. Aditya, "Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 1, pp. 10–19, 2022, doi: 10.14421/jiska.2022.7.1.10-19.
- [19] N. W. E. Rosiana Dewi, I. G. Aris Gunadi, and G. Indrawan, "Detection of Class Regularity with Support Vector Machine methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 20, 2020, doi: 10.24843/lkjiti.2020.v11i1.p03.
- [20] N. A. Utami, W. Maharani, and I. Atastina, "Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 177–184, 2021, doi: 10.1016/j.procs.2020.12.023.
- [21] N. I. Purnayasa, I. M. A. D. Suarjaya, and I. P. A. Dharmadi, "Analysis of Cyberbullying Level using Support Vector Machine Method," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 10, no. 2, pp. 81–90, 2022, doi: 10.24843/JIM.2022.v10i02.p01.
- [22] F. N. Arifa, "Tantangan Pelaksanaan Kebijakan Belajar Dari Rumah Dalam Masa Darurat Covid-19," 2020.