

Analisis Sentimen Pembelajaran Tatap Muka dengan *Apache SPARK*

Sabar Aritonang Rajagukguk*¹, Siti Mutmainah², Anas Satria³

^{1,2}Sekolah Cendekia Harapan, Bali

³Program Studi Informatika, Universitas Islam Indonesia, Yogyakarta
e-mail: *¹20917054@students.uii.ac.id, ²20917057@students.uii.ac.id,
³20917037@students.uii.ac.id

Abstrak

Mendikbud Ristek Nadiem Makarim menegaskan bahwa akan memprioritaskan kepada guru atau staf pengajar untuk melaksanakan vaksinasi, sehingga pada minggu kedua dan ketiga di bulan Juli dengan tahun ajaran baru diharapkan semua sekolah sudah dapat melakukan pembelajaran tatap muka secara terbatas dengan tetap memperhatikan protokol kesehatan. Namun dari hasil statistik pada covid19.go.id jumlah kasus terkonfirmasi kembali naik hingga kasus terkonfirmasi hingga tanggal 13 Juli 2021 sebanyak 2,5 juta kasus dan pemerintah kembali menerapkan PPKM Darurat di wilayah Jawa dan Bali. Hal ini menyebabkan berbagai respons pada masyarakat terhadap pembelajaran tatap muka, Penelitian ini bermaksud untuk menganalisis pendapat masyarakat terhadap pembelajaran tatap muka dan kasus yang kembali melonjak. Penelitian ini menggunakan Scraping Twitter API dengan fitur streaming Tweepy dan Spark untuk pengumpulan data, Spark MLlib dan Spark NLP untuk model klasifikasi dan metode klasifikasi regresi logistik. Hasil analisis yang didapatkan 47.7% sentimen "Neutral" dan 31.4% sentimen "Positive" sedangkan untuk sentimen "Positive" pada pembelajaran tatap muka 20.9%. berdasarkan hasil penelitian ini, dapat dikatakan bahwa masyarakat memiliki opini "Neutral" terhadap pembelajaran tatap muka selama pandemi. bukan berarti masyarakat tidak peduli atau kritis dalam menghadapi kasus. ditemukan bahwa beberapa masyarakat menentang dan membantah pembelajaran tatap muka. akan tetapi, terdapat beberapa masyarakat juga yang mendukung adanya pembelajaran tatap muka agar anak-anak mereka dapat kembali ke sekolah untuk menuntut pendidikan seperti semula.

Kata kunci—Analisis sentimen; klasifikasi; sekolah daring; Spark.

1. PENDAHULUAN

Covid19 merupakan penyakit menular yang disebabkan oleh jenis *coronavirus* yang baru ditemukan. virus baru dan penyakit yang disebabkan ini tidak dikenal sebelumnya mulanya wabah di kota Wuhan, Provinsi Hubei, Tiongkok, pada bulan Desember 2019 [1]. Hingga merambat ke wilayah Asia bagian tenggara, khususnya di Indonesia. Berdasarkan laporan Satuan Pengawasan Covid19 di Indonesia, jumlah kasus terkonfirmasi hingga tanggal 13 Juli 2019 sebanyak 2,5 juta, untuk konfirmasi kasus "Positive" sebanyak 367 ribu dengan korban meninggal mencapai 66.464 orang [2]. virus ini memberikan dampak yang besar di berbagai aspek kehidupan, diantaranya dalam bidang Pendidikan.

Upaya pemerintah dalam memutus rantai persebaran virus ini salah satunya dengan melakukan vaksinasi. vaksinasi merupakan suatu proses di dalam tubuh dimana seseorang menjadi kebal atau terlindungi dari suatu penyakit, sehingga apabila suatu saat terpapar oleh penyakit tersebut, dampaknya tidak terlalu berat [3]. Vaksinasi juga bermanfaat untuk memberi

perlindungan tubuh agar tidak mudah terpapar oleh virus covid19 dengan cara menimbulkan atau menstimulasikan kekebalan spesifik dalam tubuh.

Menteri Pendidikan, Kebudayaan, Riset dan Teknologi (Mendikbud Ristek), Nadiem Makarim menegaskan bahwa akan memprioritaskan kepada guru atau staf pengajar untuk melaksanakan vaksinasi, sehingga pada minggu kedua dan minggu ketiga di bulan Juli dengan tahun ajaran baru diharapkan semua sekolah sudah dapat melakukan pembelajaran tatap muka secara terbatas dengan tetap memperhatikan protokol kesehatan [4]. Pernyataan tersebut menimbulkan berbagai komentar di media sosial seperti *Twitter*. Beragam persepsi masyarakat mengenai kebijakan ini dapat dilihat pada *tweet* yang berisi komentar “*Positive*”, komentar “*Positive*” bahkan komentar yang bersifat “*Neutral*” [5].

Komentar-komentar tersebut dapat dijadikan bahan untuk menyerap aspirasi masyarakat untuk menangkap persoalan yang ada pada masyarakat [6]. permasalahannya, komentar tersebut begitu banyak dan dalam jumlah ragam kata, singkatan kata bahasa tidak formal, maupun simbol, sehingga menimbulkan kesulitan ketika harus memilih kata atau kelompok kata yang dapat menghasilkan pernyataan “*Negative*”, “*Positive*” atau “*Neutral*” [7]. Berdasarkan pernyataan yang diperoleh, dapat dilakukan klasifikasi dengan melakukan analisis sentimen.

Analisis sentimen adalah studi yang berhubungan dengan komputasi dikaitkan dengan pendapat, emosi, dan komentar seseorang atau masyarakat ketika memberikan respon terhadap suatu peristiwa yang dapat dinyatakan dalam bentuk teks [8]. Analisis sentimen juga dapat mengubah data yang awalnya tidak terstruktur menjadi data yang memiliki struktur yang jelas [9]. Penelitian dilakukan untuk mengetahui hasil dari analisis sentimen mengenai pembelajaran tatap muka yang segera dilaksanakan kembali oleh pemerintah. penelitian analisis sentimen ini menggunakan metode regresi logistik untuk melakukan klasifikasi berdasarkan dari tanggapan masyarakat di media sosial *Twitter*. Hasil analisis sentimen ini dapat menjadi acuan pemerintah untuk mengetahui tanggapan atau komentar masyarakat terhadap kebijakan yang telah dibuat dan dapat menjadi bahan pertimbangan untuk mengambil kebijakan lainnya di masa yang akan datang dalam bidang pendidikan.

2. METODE PENELITIAN

Analisis sentimen pembelajaran tatap muka di masa pandemi covid19 ini untuk mengetahui tanggapan masyarakat pada kasus ini. tanggapan dapat berisi sentimen “*Positive*”, “*Positive*” dan “*Neutral*” di media sosial *Twitter* dengan menggunakan metode klasifikasi regresi logistik. pada penelitian terdahulu, banyak penelitian yang menggunakan beberapa metode *machine learning* dalam melakukan analisis sentimen. metode yang digunakan seperti *Logistic Regression*, *Support Vector Machine (SVM)*, *Naive Bayes* dan *K-Nearest Neighbors* [10].

Beberapa penelitian telah melakukan analisis sentimen menggunakan regresi logistik, pada literatur [11] dilakukan analisis sentimen sikap masyarakat terhadap kebijakan sekolah daring pada media sosial *Twitter*. Penelitian ini melakukan klasifikasi sentimen “*Positive*”, “*Positive*”, dan “*Neutral*”. dalam penelitian ini diperoleh nilai akurasi sebesar 87%. Penggunaan kata dalam sentimen “*Positive*” yang paling banyak digunakan yaitu “bahagia”, dengan angka kemunculan sebanyak 30 kali, diikuti dengan kata “rajin” sebanyak 35 kali kemudian kata “senang” sebanyak 20 kali. penggunaan kata dalam sentimen “*Positive*” yang memiliki frekuensi paling banyak adalah kata “capek”, “muak” dan “bosan”. untuk tren penggunaan kata yang memiliki sentimen “*Neutral*” yang paling sering digunakan yaitu kata “bingung” dengan frekuensi penggunaan diatas 17 kali, kata “paham” dengan frekuensi penggunaan diatas 5 kali, dan kata “kasih” dengan frekuensi penggunaan diatas 5 kali [11].

Perbandingan model klasifikasi menggunakan berbagai metode ditunjukkan pada literatur [12] yang dilakukan analisis sentimen terhadap kebijakan pemerintah dalam penerimaan siswa baru menggunakan sistem zonasi. Perbandingan dilakukan dengan menggunakan metode klasifikasi *Support Vector Machine*. Berdasarkan model yang telah dikembangkan menggunakan SVM diperoleh akurasi sebesar 92,93% dan Naive Bayes sebesar

79,86%. hasil analisis sentimen yang didapatkan cenderung pada sentimen “*Positive*” terhadap kebijakan zonasi [12].

Pada literatur [13], analisis sentimen mengenai netralitas berita dalam membahas pemilu presiden 2019 menggunakan Spark. pemanfaatan spark ini berdasarkan data yang digunakan dalam jumlah besar untuk klasifikasi sentimen terhadap berita pemilu yang beredar. metode klasifikasi yang digunakan yaitu *Naive Bayes*. diperoleh hasil akurasi pada model sebesar 80%. kemudian dilakukan evaluasi performa dari model menggunakan *K-Fold Cross Validation*. Komparasi antara model menggunakan spark dan tanpa spark. hasil komparasi menunjukkan bahwa waktu *runtime* (waktu eksekusi program) dengan menggunakan spark lebih cepat dibandingkan tanpa menggunakan spark. Proses pengolahan data menggunakan spark hanya membutuhkan waktu 5,915 detik sedangkan model tanpa spark membutuhkan waktu lebih lama sekitar 202,185 detik [13].

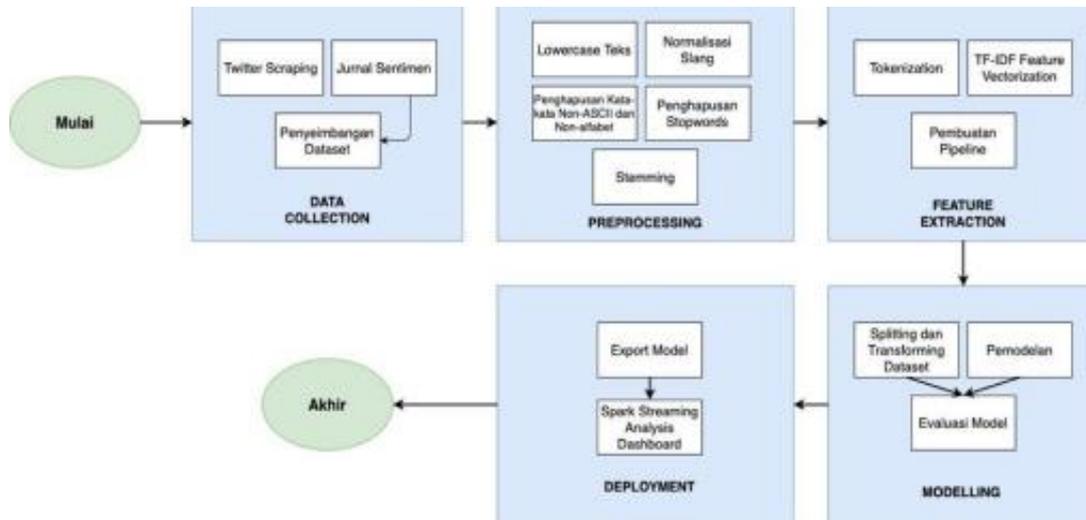
Penelitian terkait lain, dilakukan analisis sentimen terhadap penerapan pembatasan sosial berskala besar (PSBB) di wilayah ibukota Jakarta. pada literatur [14] dilakukan *scraping* data melalui media sosial *Twitter*. *Scraping* data dilakukan mulai tanggal 8 September 2020 hingga 9 Oktober 2020. kata kunci “*PSB_Jakarta*” digunakan untuk pengambilan data dengan menggunakan aplikasi *Orange* dan *Anaconda* dibantu dengan tools *TwitterScraper*. diperoleh hasil dari *Twitter* sebanyak 10.070 *tweets*. kemudian dilakukan tahapan *prapemrosesan teks* untuk melakukan *filtering* data menggunakan aplikasi *Orange* untuk mengetahui data mana yang termasuk dalam sentimen “*Positive*” dan sentimen “*Negative*”. sehingga hasil olahan analisis sentimen menunjukkan bahwa sebanyak 672 *tweet* masyarakat di *Twitter* memiliki opini yang “*Positive*” sedangkan untuk opini “*Negative*” diperoleh sebanyak 573 *tweet*.

Selain melihat sentimen pada kasus PSBB, pada literatur [14] juga melakukan analisis dampak sentimen pada Indeks Harga Saham Gabungan (IHSG) dengan menggunakan beberapa pemodelan. dilakukan komparasi model dengan menggunakan metode *Logistic Regression*, *K-Nearest Neighbors*, *Naive Bayes* dan *Random Forest*. Berdasarkan model yang telah dibuat, diperoleh hasil bahwa metode *logistic regression* memiliki tingkat akurasi paling tinggi dibandingkan model lainnya dengan tingkat akurasi sebesar 75%, kemudian *K-Nearest Neighbors* dan *Naive Bayes* memperoleh tingkat akurasi sebesar 70%. sedangkan metode *Random Forest* memiliki tingkat akurasi paling rendah dengan tingkat akurasi sebesar 60% [14].

Berdasarkan analisis sentimen pada dua kasus terakhir, bahwa masyarakat mendukung kebijakan pemerintah dalam penerapan PSBB berdasarkan jumlah tanggapan “*Positive*” terhadap kebijakan pemerintah pada media sosial *Twitter*. sedangkan masyarakat yang bersentimen “*Positive*” menilai bahwa kebijakan PSBB yang dilakukan gagal karena masih banyak yang ditemui pelanggaran-pelanggaran terkait dengan kebijakan yang dibuat [14]. Kemudian dalam kasus IHSG, model *logistic regression* yang telah dibuat akan digunakan sebagai prediksi kedepannya. jika model dapat memprediksi lebih banyak sentimen “*Positive*” maka kemungkinan besar IHSG cenderung akan menurun begitu juga sebaliknya yang terjadi pada sentimen “*Positive*”. Berbanding lurus dalam dunia nyata jika mayoritas opini buruk yang dihasilkan oleh masyarakat maka nilai IHSG akan menurun karena para investor akan menarik saham pada pasar modal [15].

Dalam penelitian ini menggunakan Spark sebagai arsitektur utama [16]. Spark telah banyak digunakan dalam penelitian lainnya karena kemampuan dalam toleransi kesalahan dan pemrosesan data secara *real-time* [17]. Implementasi spark dalam pembuatan model untuk klasifikasi sentimen terhadap pembelajaran tatap muka menggunakan Spark MLlib dan Spark NLP. sehingga terciptanya model *logistic regression* dengan tingkat akurasi sebesar 90,04%. berdasarkan penelitian yang telah dilakukan diperoleh hasil analisis sentimen didominasi oleh sentimen yang bersifat “*Neutral*” dengan sentimen sebesar 47,7%, kemudian diikuti oleh sentimen “*Positive*” sebesar 31,4% dan sentimen “*Negative*” memperoleh hasil sebesar 20,9%.

Analisis sentimen dilakukan untuk memeriksa tanggapan masyarakat pada pelaksanaan pembelajaran tatap muka di masa pandemic. Analisis menggunakan metode klasifikasi pembelajaran mesin, yakni regresi logistik dengan bantuan alat *Apache Spark*. Tahapan dalam melakukan sentimen analysis pada penelitian yaitu: *Data Collection*, *Prapemrosesan*, *Feature Extraction*, *Modelling*, *Deployment*. Gambar 1, menunjukkan diagram alir tahapan yang dilakukan pada penelitian.



Gambar 1 Diagram Alir Penelitian

2.1 Data Collection

Pengumpulan data tweets pada media social *Twitter* dengan melakukan *Scraping Twitter* API menggunakan fitur streaming Tweepy dan Spark. Spark streaming dipakai karena memiliki kelebihan dalam menangani komputasi yang berat dalam pengumpulan data tweet. Ekstraksi data pada 1 Januari 2020 hingga bulan 7 Juli 2021 menggunakan kata kunci: *sekolah tatap muka*, *sekolah offline*, *pembelajaran tatap muka*, *sekolah offline*, *sekolah dibuka*. Dataset yang didapatkan dari ekstraksi berjumlah 2368 dengan 6 variabel: *tanggal*, *id*, *teks*, *rt*, *fav*, *coord*, yang disimpan dalam bentuk .csv.

2.2 Data Collection

Prapemrosesan memahami data dan meningkatkan kualitas dari data sehingga dapat memaksimalkan kinerja metode machine learning dalam penambangan data [10], Proses prapemrosesan yang dilakukan pada penelitian ini yaitu:

1. *Case Folding*, keseluruhan teks data dikonversi menjadi huruf kecil atau lowercase untuk menyamakan huruf dan memiliki bobot yang sama.
2. *Penghapusan Stopwords*, pada proses ini menghilangkan kata kata umum dan tidak memiliki makna. Frekuensi kemunculan yang tinggi dapat menyebabkan bias pada model. Pada Gambar 2 contoh distribusi kata yang sering muncul dan dilakukan penghapusan pada tahap *stopwords*.
3. *Text Cleaning*, pembersihan data seperti penghapusan tanda baca, karakter non-ASCII, situs, angka, hexadecimal, username *Twitter*, dan sebagainya untuk mengurangi *redundancy*.
4. Normalisasi slang, pada saat pemodelan machine learning akan menginterpretasikan kata slang dan formal berbeda padahal memiliki arti yang sama, proses ini menyamakan kata-kata slang menjadi kata-kata yang lebih formal untuk mengurangi *redundancy*.
5. *Stemming*, proses ini mereduksi dan mengelompokkan kata pada data ke dalam bentuk dasar.
6. *Tokenization*, memisahkan kalimat pada tweets menjadi kata-kata, ini dilakukan untuk pembobotan kata pada kalimat dalam data.

Tabel 1 Prapemrosesan

	Sebelum	Sesudah
<i>Case Folding</i>	Haha kasian adek Gue sekolah tatap muka september, Sekolah masuk cuman setahun mana dia yg pilih sendiri pengen masuk ke situ. Ini semua gara" corona please get out right now b*tch ð	haha kasian adek gue sekolah tatap muka september, sekolah masuk cuman setahun mana dia yg pilih sendiri pengen masuk ke situ. ini semua gara" coronaplease get out right now b*tch ð
<i>Tokenization</i>	amin lebih bagus lagi moga covid segera lalu biar anak bisa sekolah tatap muka seperti biasa	amin, lebih, bagus, lagi, moga, covid, segera, lalu, biar, anak, bisa, sekolah, tatap, muka, seperti, biasa
<i>Penghapusan Stopwords</i>	haha kasian adek gue sekolah tatap muka september, sekolah masuk cuman setahun mana dia yg pilih sendiri pengen masuk ke situ. ini semua gara" corona please get out right now b*tch ð	haha kasian adek gue sekolah tatap muka september, sekolah masuk cuman setahun pilih sendiri pengen masuk situ. semua gara" corona please get out right now b*tch ð
<i>Text Cleaning</i>	presiden @jokowi: tolak sekolah tatap muka! batalkan skb 4 menteri! – tanda tangani petisi! https://t.co/kmeygBw3mz lewat @ChangeOrg_ID	presiden tolak sekolah tatap muka batalkan skb menteri tanda tangani petisi lewat
<i>Normalisasi slang</i>	mls bgt sekolah tatap muka temen gue kayak setan semua	malas banget sekolah tatap muka temen saya seperti setan semua
<i>Stemming</i>	besok anak anak sudah masuk kelas baru tahun ajar kemarin aku pikir banget kapan anak anak bisa masuk tatap sama	besok anak anak sudah masuk kelas baru tahun ajar kemarin aku pikir banget kapan anak anak bisa masuk tatap sama

Gambar 2. Distribusi Kata *Stopwords*

2.3 Features Extraction

Tahap selanjutnya sebelum modelling yaitu feature extraction, pada tahap ini menggunakan Spark NLP dan fitur N-gram. Metode ini diterapkan untuk mengubah kata yang telah di tokenisasi pada teks menjadi vektor yang akan diproses oleh klasifikasi dengan bobot yang berbeda. Pemodelan N-gram digunakan untuk memeriksa fitur pada postingan dan pembelajaran teks. TF-IDF yaitu konversi ke bentuk biner atau memberikan bobot term pada semua kata yang ada pada data, penelitian ini menggunakan HashTF untuk menghasilkan vektor dan IDF sebagai skala pada setiap vektor. Setelah menerapkan features extraction selanjutnya membuat pipelines untuk menggabungkan fitur transformer dan estimator agar menyesuaikan dengan keperluan proses machine learning.

2.4 Modelling

Pada tahap modelling menggunakan 19.194 data yang telah dilakukan labeling, data tersebut dibagi menjadi dua bagian untuk keperluan training (80%) dan testing (20%), data *training* akan digunakan untuk membuat model sedangkan data *testing* untuk keperluan validasi model yang telah dibuat. Modelling dibuat menggunakan variabel *Label* dan *Features* (vektor yang didapatkan pada *tweets*). Modelling dibuat menggunakan model Spark MLlib, Spark NLP dan metode klasifikasi regresi logistic yang akan melakukan klasifikasi dengan menyaring inputan numerik sebagai fiturnya, metode ini akan merubah vektor kata menjadi vektor yang merata pada semua nilai dalam urutan.

2.5 Deployment

Setelah melakukan *modelling* selanjutnya melakukan *deployment* untuk analisis sentimen menggunakan data *Scraping Twitter API* sebanyak 2368. *Deployment* model menggunakan Spark MLlib Logistic Regression dapat digeneralisasi menjadi multinomial atau klasifikasi multi kelas, dalam kasus ini yaitu prediksi sentimen “*Positive*”, “*Neutral*” dan “*Negative*”.

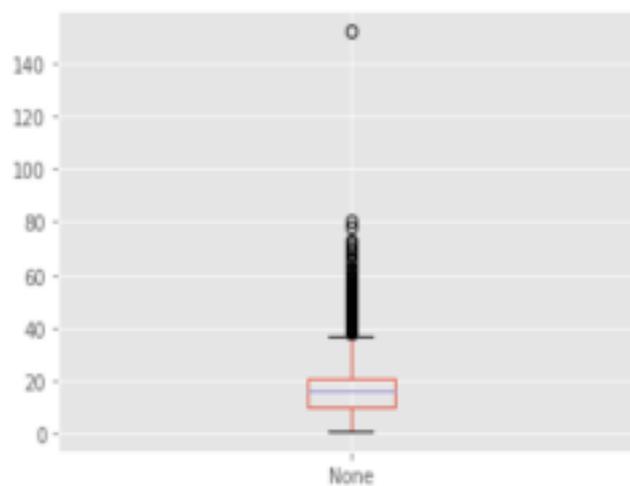
3. HASIL DAN PEMBAHASAN

3.1 Analisis

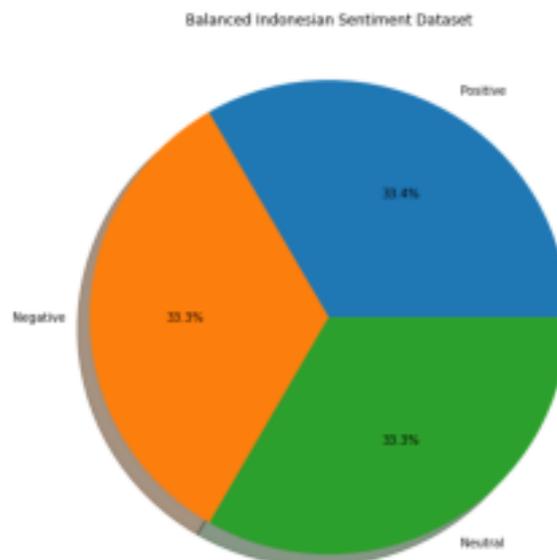
Analisis dilakukan untuk mengeksplorasi data, hal ini dilakukan sebelum tahap prapemrosesan dan pemodelan. Eksplorasi data digunakan untuk mengetahui distribusi, persebaran data dan panjang data pada proses training serta prapemrosesan data.

3.1.1 Distribusi Data

Eksplorasi ini bermaksud untuk mengetahui distribusi data *modelling*, Pada Gambar 3 menyajikan boxplot, boxplot digunakan untuk menggambarkan distribusi data lebih detail pada tweets, Gambar 3 menunjukan adanya *imbalance* pada data yang digunakan untuk *modelling*, oleh karena itu dilakukan balancing data agar prediksi oleh pembelajaran mesin tidak condong pada class mayoritas.



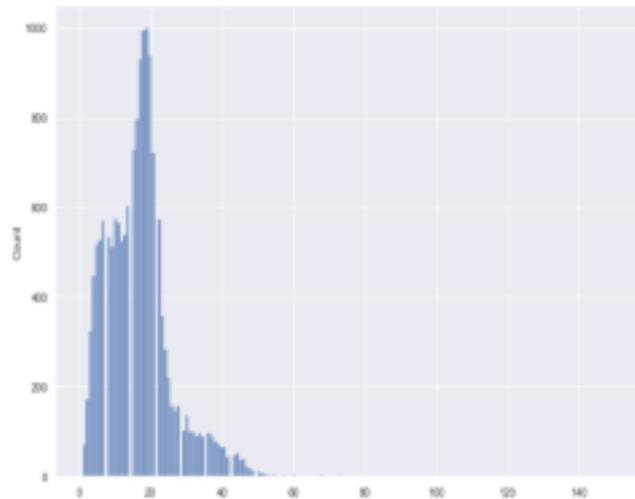
Gambar 3 Distribusi Data



Gambar 4 Distribusi Modelling

3.1.2 Panjang Proses Training

Eksplorasi digunakan untuk mengukur lama data pada proses training yang dilakukan untuk *modelling*, Diagram pada Gambar 5 menunjukkan lama proses training data yang akan dilakukan. Gambar 5 menunjukkan semakin lama proses training akan semakin pendek waktu training.



Gambar 5 Panjang Data

3.1.3 Distribusi Kata

Eksplorasi dan visualisasi ini untuk mengetahui kata yang sering muncul dan frekuensi kemunculannya pada teks, hasil dan frekuensi kata akan bermanfaat pada saat prapemrosesan yaitu penghapusan *stopwords* dan pembersihan teks serta tokenisasi. Kata dan frekuensi kemunculannya disajikan pada Tabel 2. Kata yang disajikan merupakan kata yang sering muncul pada data yang akan di *deployment* dan dilakukan analisis sentimen. Gambar 6 dan Gambar 7 merupakan *wordcloud* dari kata sebelum dan sesudah prapemrosesan. Pada penelitian ini melakukan sentimen analisis terhadap pembelajaran tatap muka menggunakan metode klasifikasi regresi logistik.

Tabel 2 Top 10 Kata

No	Teks	Jumlah
1	sekolah	2061
2	offline	976
3	tatap	860
4	muka	846
5	di	482
6	enggak	472

7	ajar	389
8	yang	351
9	sudah	330
10	buka	249



Gambar 6 Wordcloud Sebelum Prapemrosesan



Gambar 7 Wordcloud Setelah Prapemrosesan

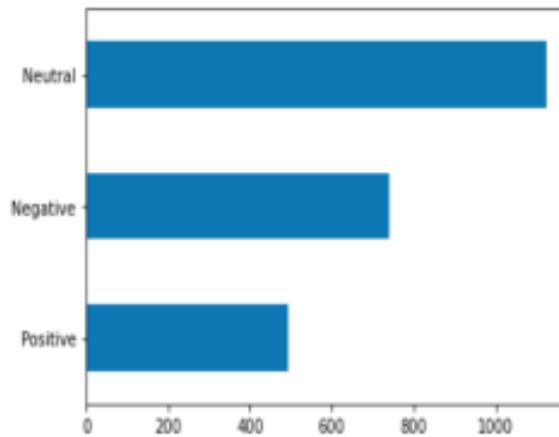
Data yang telah dilakukan prapemrosesan dibagi menjadi dua data untuk *modelling*, proporsi pembagian data ada pada Tabel 3. Performa yang dihasilkan oleh model menggunakan metode regresi logistik adalah 90.04%.

Tabel 3 Distribusi Dataset

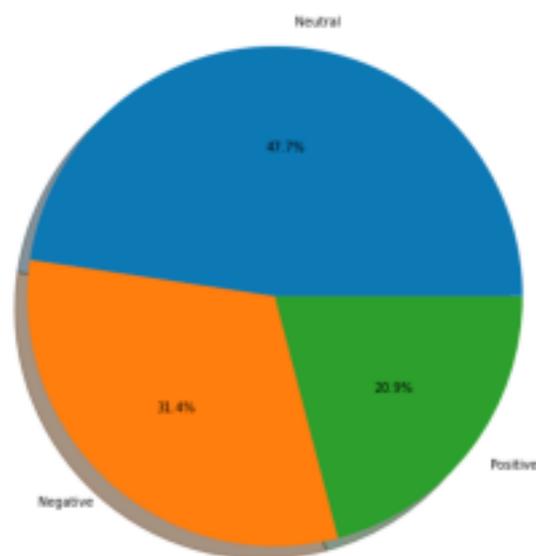
Data	Jumlah
Train set	15.989
Test set	3205

3.2 Hasil

Analisis sentimen dilakukan menggunakan data *Scraping Twitter* API sebanyak 2368 untuk prediksi sentimen pada masyarakat dengan menggunakan model yang telah dibangun. Gambar 8 merupakan hasil yang didapatkan oleh model pada sentimen analisis pembelajaran tatap muka. Masyarakat bersikap “*Neutral*” terhadap pembelajaran tatap muka, untuk melihat proporsi yang dihasilkan masing masing sentimen ada pada Gambar 9 diagram lingkaran.



Gambar 8 Sentimen Tweet



Gambar 9 Proporsi Sentimen

Pada Gambar 9 merupakan diagram lingkaran proporsi masyarakat yang bersikap “*Neutral*” sebesar 47.7% dan “*Positive*” sebesar 31.4% sedangkan untuk bersikap “*Positive*” pada pembelajaran tatap muka 20.9%.

4. KESIMPULAN

Hasil analisis terhadap sentimen masyarakat pada pembelajaran tatap muka pada tweet media sosial *Twitter* didapatkan, beberapa kata paling sering muncul pada tweet masyarakat seperti “sekolah”, “offline” dan “tatap”, “muka”. hasil sentimen masyarakat bersikap “*Neutral*” pada pembelajaran tatap muka sebesar 47.7%. Sentimen “*Neutral*” dan 31.4% sentimen “*Positive*” sedangkan untuk sentimen “*Positive*” pada pembelajaran tatap muka 20.9%.

Sentimen masyarakat dari data yang didapatkan bahwa kebanyakan dari masyarakat cenderung bersikap “*Neutral*” dan “*Positive*” terhadap pembelajaran tatap muka, hal ini dapat disebabkan oleh kasus terinfeksi yang terus meningkat. masyarakat menginginkan segera pembelajaran namun masih masyarakat khawatir terhadap pembelajaran tatap muka.

5. SARAN

Adapun beberapa kekurangan dalam penelitian ini yang dapat dipertimbangkan sebagai pengembangan penelitian dikemudian hari. Penelitian ini telah menganalisis *tweets* atau opini publik dalam platform media sosial *Twitter* dari tahun 1 Januari 2020 hingga 7 Juli 2021. Tetapi, analisis sentimen publik pada pembelajaran tatap muka ini belum memperhitungkan opini setelah pandemi mulai dapat ditangani dan mereda, yakni 2021 akhir hingga 2022. Penelitian kedepannya dapat menganalisis sentimen publik pada tiga periode *pre-pandemic* (sebelum pandemi), *pandemic* (pandemi), dan *post-pandemic* (setelah pandemic mereda) agar dapat dianalisis lebih dalam bagaimana pandemi mempengaruhi sentimen public terhadap pembelajaran tatap muka. Selain itu, berbagai alternatif metode lain seperti *deep learning* dapat digunakan untuk menganalisis sentimen masyarakat dengan lebih akurat.

DAFTAR PUSTAKA

- [1] Samsir, Ambiyar, U. Verawardina, F. Edi, and R. Watrionthos, “Analisis Sentimen Pembelajaran Daring Pada *Twitter* di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [2] G. T. P. P. Covid19, “Peta Sebaran,” 2019. [Online]. Available: <https://covid19.go.id/peta-sebaran-covid19>.
- [3] I. S. Adhi, “4 Manfaat Vaksin Covid-19 yang Perlu Dipahami,” 2021. [Online]. Available: <https://health.kompas.com/read/2021/06/27/120400768/4-manfaat-vaksin-covid-19-yang-perludipahami?page=all>.
- [4] Y. Makdori, “Kasus Covid-19 Melonjak, Nadiem Makarim Diminta Tunda Pembelajaran Tatap Muka Terbatas,” 2021. [Online]. Available: <https://www.liputan6.com/news/read/4590185/kasus-covid-19-melonjak-nadiem-makarim-diminta-tunda-pembelajaran-tatap-muka-terbatas>.
- [5] P. Aliandu, “*Twitter* Used by Indonesian President : An Sentimen Analysis of Timeline,” *Inf. Syst. Int. Conf.*, no. December, pp. 2–4, 2013.

- [6] O. Somantri and D. Apriliani, "Support Vector Machine Berbasis Feature Selection Untuk Sentimen Analisis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 537, 2018, doi:10.25126/jtiik.201855867.
- [7] N. Ramdhani and R. H. Al-fadillah, "ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP BELAJAR DARING SELAMA PANDEMI COVID-19 DENGAN DEEP LEARNING Kata Kunci : Analisis sentimen , *Twitter* , belajar daring , COVID-19 , Deep Learning," 2021.
- [8] A. P. Natasuwarna, "Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes," *Semin. Nas. Sist. Inf. dan Tek. Inform.*, pp. 47–53, 2019.
- [9] A. P. Kirilenko, S. O. Stepchenkova, and H. Kim, "Automated Sentimen Analysis in Tourism: Comparison of Approaches," 2017, doi: <https://doi.org/10.1177/0047287517729757>.
- [10] E. Mona Cindo, Dian Palupi Rini, "Literatur Review : Metode Klasifikasi Pada Sentimen Analisis," *Semin. Nas. Teknol. Komput. Sains*, pp. 66–70, 2019.
- [11] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada *Twitter* Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [12] R. A. Ariyanto and N. Chamidah, "Sentimen Analysis for Zoning System Admission Policy Using Support Vector Machine and Naive Bayes Methods," *J. Phys. Conf. Ser.*, vol. 1776, no. 1, 2021, doi: 10.1088/1742-6596/1776/1/012058.
- [13] R. A. Fauzi, I. Cholissodin, and B. Rahayudi, "Pemanfaatan Spark untuk Analisis Sentimen Mengenai *Neutralitas* Berita dalam Membahas Pemilu Presiden 2019 Menggunakan Metode Naive Bayes Classifier," vol. 5, no. 3, pp. 1070–1077, 2021.
- [14] N. C. Lengkong, O. Safitri, S. Machus, Y. R. Putra, A. Syahadati, and R. Nooraeni, "Analisis Sentimen Penerapan Psbb Di Dki Jakarta Dan Dampaknya Terhadap Pergerakan Ihsg," *J. Teknoinfo*, vol. 15, no. 1, p. 20, 2021, doi: 10.33365/jti.v15i1.866.
- [15] G. A. K. R. S. Dewi and D. P. Vijaya, *Investasi dan Pasar Modal*. 2018.
- [16] R. R. Aryanto and D. H. Fudholi, "Analyzing Public's Reaction towards Black Lives Matter Campaign using Machine learning-based Approach through Spark," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1077, no. 1, p. 012005, 2021, doi: 10.1088/1757-899x/1077/1/012005.
- [17] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Streaming big data analysis for real-time sentimen based targeted advertising," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 1, pp. 402–407, 2017, doi: 10.11591/ijece.v7i1.pp402-407.
- [18] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data "Edisi Revisi."* Bandung: Informatika, 2019.
-