

PENERAPAN SENTIMEN ANALISIS TWITTER TERHADAP COVID-19 MENGGUNAKAN SUPPORT VECTOR MACHINE

Ari Syafri¹, Azhar Alfaini², Gilang Dhiya Ulhaq³, Mochammad Gymnastiar⁴, Hidayat⁵

^{1,2,3,4} Teknik Informatika, Universitas Komputer Indonesia

⁵ Teknik Komputer, Universitas Komputer Indonesia

Jl. Dipati Ukur No.112-116, Lebakgede, Kecamatan Coblong, Kota Bandung, Jawa Barat 40132

E-mail: hidayat@email.unikom.ac.id⁵

Abstrak

Pemanfaatan media sosial seperti Twitter sangat masif digunakan oleh masyarakat di dunia internet, sehingga informasi yang sedang viral dapat diperoleh dengan cepat. Kondisi pandemi Covid-19 yang telah melanda dunia tentunya dapat menyebabkan perubahan-perubahan pada tatanan kehidupan masyarakat saat ini khususnya di daerah DKI Jakarta. Penelitian ini bertujuan untuk mengetahui respon masyarakat terhadap Covid-19 melalui analisis sentimen menggunakan perbandingan model *decision tree*, regresi logistik, *random forest*, dan *support vector machine* (SVM). Melalui pemanfaatan teknik *Text Mining*, metode klasifikasi akan menghasilkan sentimen bernilai positif, netral atau negatif. Hasil pengujian menunjukkan bahwa model SVM menjadi model utama dengan rata-rata akurasi sebesar 0,6094 dan standar deviasi terendah sebesar 0,0378. Analisis sentimen diuji dengan 5230 data *tweet*. Hasil analisis menunjukkan bahwa prosentase sentimen positif (31,64%) lebih besar daripada sentimen negatif (20,48%), walaupun sentimen netral mendominasi dengan prosentase sebesar 47,88%.

Kata kunci : Analisis Sentimen, Covid-19, SVM, Twitter.

Abstract

The utilization of social media such as Twitter is massively embraced by the global internet community, enabling quick access to viral information. The Covid-19 pandemic has undoubtedly led to significant changes in the societal landscape, particularly in the DKI Jakarta region. This research aims to discern the public response towards Covid-19 through sentiment analysis utilizing a comparative approach involving decision tree, logistic regression, random forest, and support vector machine (SVM) models. Leveraging Text Mining techniques, the classification method will yield sentiments categorized as positive, neutral, or negative. Test results reveal that the SVM model emerges as the primary model with an average accuracy of 0.6094 and the lowest standard deviation of 0.0378. Sentiment analysis is conducted on 5230 tweet data. The analysis demonstrates that the percentage of positive sentiment (31.64%) outweighs the negative sentiment (20.48%), although neutral sentiment predominates at 47.88%.

Keywords : Sentiment analysis, Covid-19, SVM, Twitter.

1. PENDAHULUAN

Pandemi Covid-19 telah menjadi tantangan global yang melanda dunia sejak awal tahun 2020. Dalam menghadapi pandemi ini, berbagai negara dan wilayah di seluruh dunia berjuang untuk memahami, menangani, dan mengendalikan penyebaran virus yang mematikan ini. Salah satu upaya penting dalam penanganan pandemi adalah pemantauan dan pemahaman terhadap persepsi serta reaksi masyarakat terhadap situasi yang berkembang.

DKI Jakarta, sebagai ibu kota Indonesia, adalah salah satu wilayah yang terdampak parah oleh pandemi Covid-19. Menghadapi situasi ini, pemerintah dan berbagai pihak terkait telah melakukan upaya dalam melaksanakan kebijakan, mengkomunikasikan informasi, dan menggerakkan berbagai langkah penanganan Covid-19. Namun, untuk mendapatkan pemahaman yang komprehensif mengenai persepsi, pendapat, serta sentimen masyarakat terhadap pandemi di DKI Jakarta, diperlukan alat analisis yang efektif.

Dalam era digital seperti sekarang ini, media sosial telah menjadi *platform* utama bagi individu untuk berbagi berita, opini, dan pengalaman mereka. Twitter khususnya telah menjadi salah satu sumber informasi dan interaksi sosial yang penting [1], [2]. Pengguna Twitter secara aktif berbagi pemikiran mereka tentang berbagai topik, termasuk pandemi Covid-19. Dalam konteks ini, analisis sentimen dapat menjadi pendekatan yang efektif untuk memperoleh wawasan tentang persepsi dan sikap masyarakat.

Berdasarkan pertimbangan-pertimbangan tersebut, penelitian ini dilakukan untuk menerapkan analisis sentimen terhadap tweet-tweet terkait Covid-19 di DKI Jakarta. Penelitian ini bertujuan untuk menggali informasi yang berharga mengenai sentimen masyarakat terhadap pandemi Covid-19 di wilayah tersebut, yang dapat digunakan sebagai dasar dalam perumusan kebijakan dan strategi penanganan yang lebih tepat sasaran [3]–[5].

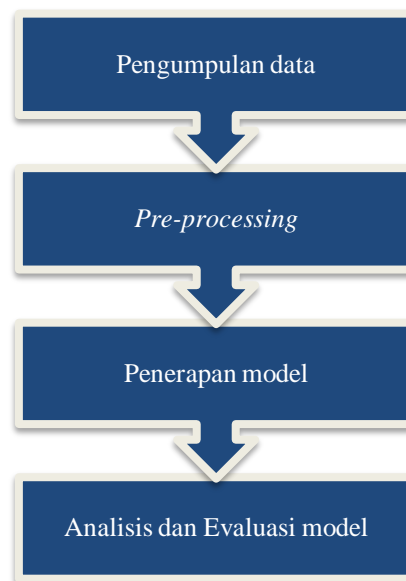
Dengan melakukan analisis terhadap *tweet-tweet* yang mengandung pendapat, perasaan, dan keyakinan masyarakat, diharapkan hasil penelitian ini dapat memberikan gambaran yang lebih mendalam tentang sikap dan persepsi masyarakat DKI Jakarta terhadap pandemi Covid-19 [6]–[9]. Hasil analisis sentimen ini diharapkan dapat memberikan panduan berharga bagi para pembuat kebijakan dan pihak-pihak terkait dalam merancang strategi penanganan dan upaya komunikasi yang lebih efektif.

Dalam konteks ini, laporan penelitian ini memiliki peran penting dalam memberikan gambaran yang lebih lengkap dan mendalam mengenai pandangan masyarakat di DKI Jakarta tentang pandemi Covid-19. Diharapkan laporan ini dapat memberikan sumbangan yang signifikan dalam memahami dinamika sosial dan kebutuhan masyarakat dalam menghadapi tantangan yang ditimbulkan oleh pandemi.

Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan dan rekomendasi yang berharga serta menjadi kontribusi nyata dalam upaya penanggulangan Covid-19 dan kesejahteraan masyarakat DKI Jakarta.

2. METODOLOGI

Tahapan penelitian ditampilkan pada Gambar 1. Tahapan penelitian terdiri atas: pemilihan data, *pre-processing*, ekstraksi fitur, pemilihan model, pelatihan model, evaluasi model, penggunaan model.



Gambar 1. Tahapan dalam penelitian

2.1. Pengumpulan data

Tahapan ini merupakan tahapan penting dalam menentukan sumber data yang akan dianalisis. Beragam sumber informasi dapat dimanfaatkan untuk mengetahui sentimen masyarakat berdasarkan opini-opini yang tersebar melalui media sosial, salah satunya adalah media sosial Twitter.

2.2. Pre-processing

Dalam proses *text mining*, terdapat beberapa tahapan yang bertujuan untuk mengolah teks agar menjadi lebih terstruktur. Salah satu tahapan kunci dalam *text mining* adalah *pre-processing* agar data dapat disiapkan untuk dianalisis. Tahap *pre-processing* mencakup beberapa langkah, seperti *tokenizing* (pemisahan teks menjadi unit-unit tertentu), *cleansing* (pembersihan dari karakter yang tidak relevan), *case folding* (mengubah semua huruf menjadi huruf kecil), perbaikan kata tidak baku, *filtering* (penghapusan kata-kata tidak relevan), dan *stemming* (pengubahan kata ke bentuk dasarnya).

2.2.1. Perbaikan kata tidak baku

Langkah perbaikan kata tidak baku atau normalisasi bahasa bertujuan untuk mengubah kata-kata yang tidak baku menjadi kata baku berdasarkan Kamus Besar Bahasa Indonesia (KBBI). Pada penelitian ini, perbaikan kata tidak baku dilakukan untuk mengatasi masalah kata-kata tidak baku yang berasal dari bahasa modern atau slang, penyingkatan kata, dan kesalahan ejaan. Proses perbaikan kata tidak baku dilakukan melalui normalisasi, yaitu dengan memeriksa token-token dalam kalimat dan membandingkannya dengan kamus_katabaku. Jika terdapat kata-kata yang tidak baku, maka kata-kata tersebut akan diubah menjadi bentuk kata baku sesuai dengan kamus. Hasil akhir proses ini adalah kalimat yang menggunakan kata-kata baku.

2.2.2. Pembobotan TF-IDF

TF-IDF adalah suatu metode yang digunakan untuk menghitung bobot atau nilai penting dari setiap kata yang telah diekstrak dari teks. Penggunaan metode ini umumnya bertujuan untuk menilai kata-kata umum dalam *information retrieval*. Model pembobotan TF-IDF menggabungkan dua komponen, yaitu *term frequency (tf)* dan *inverse document frequency (idf)*. *tf* menghitung berapa kali suatu kata muncul dalam satu dokumen, sedangkan *idf* digunakan untuk menghitung seberapa umum kata tersebut muncul di berbagai dokumen atau komentar. Kata-kata yang sering muncul dalam satu dokumen tetapi jarang muncul di seluruh koleksi dokumen dianggap memiliki bobot lebih tinggi, sedangkan kata-kata yang sering muncul di banyak dokumen dianggap memiliki bobot lebih rendah dan dianggap kurang penting dalam analisis informasi[9].

Tahapan pembobotan dengan TF-IDF adalah:

1. Hitung *term frequency*: tft, d
2. Hitung *weighting term frequency*: $Wtft, d = \{1 + \log 10 tft, d, \text{if } tft, d > 0\}$ (1)
3. Hitung *document frequency (df)*
4. Hitung bobot *inverse document frequency (idf)*: $idft = \log 10 N df$ (2)
5. Hitung nilai bobot TF-IDF: $Wt, d = Wtft, d \times idft$ (3)

Keterangan:

- tft, d = frekuensi term
 $Wtft, d$ = bobot frekuensi term
 df = jumlah frekuensi dokumen yang mengandung term
 N = jumlah total dokumen
 Wt, d = bobot TF-IDF

2.3. Penerapan Model

Tahapan penerapan model ini melibatkan beberapa teknik *Machine Learning* yang telah terbukti efektif dalam analisis sentimen. Beberapa model yang digunakan, yaitu *Decision Tree*, *Logistic Regression*, *Random Forest Classifier*, dan *Support Vector Machine (SVM)*[3], [4], [10]–[13]. Penggunaan kombinasi teknik-teknik ini akan memberikan pendekatan yang komprehensif dalam menganalisis sentimen *tweet* terkait Covid-19 di DKI Jakarta.

2.4.1. Decision Tree

Decision Tree (Pohon Keputusan) adalah model pembelajaran mesin yang menggunakan struktur pohon untuk mengambil keputusan berdasarkan serangkaian kondisi dan aturan. Setiap simpul di dalam pohon mewakili fitur atau atribut. Setiap cabang menunjukkan nilai atau rentang nilai yang mungkin untuk atribut tersebut. Pada simpul daun, keputusan akhir atau label kelas akan ditentukan [10]–[12].

2.4.2. Regresi Logistik

Logistic Regression (Regresi Logistik) adalah algoritma pembelajaran mesin yang digunakan untuk memodelkan hubungan antara variabel independen (fitur) dengan variabel dependen (label kelas) yang bersifat biner. Regresi Logistik menghasilkan probabilitas prediksi dalam rentang 0 hingga 1 dengan menggunakan fungsi logistik atau sigmoid. Jika probabilitas tersebut melebihi ambang batas tertentu, data diklasifikasikan ke dalam kelas positif. Sebaliknya, jika probabilitasnya di bawah ambang batas, data diklasifikasikan ke dalam kelas negatif [11].

2.4.3. Random Forest Classifier

Random Forest Classifier adalah algoritma pembelajaran mesin yang menggabungkan banyak pohon keputusan menjadi sebuah model yang lebih kuat [14]. Setiap pohon dalam hutan diproses secara

independen, dan prediksi akhir didasarkan pada mayoritas suara dari pohon-pohon tersebut. *Random Forest* menghasilkan hasil prediksi dengan menggabungkan prediksi dari beberapa pohon, sehingga lebih tahan terhadap *overfitting* dan memiliki kemampuan untuk menangani data dengan banyak fitur [4], [11].

2.4.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi atau regresi. SVM mencari *hyperplane* (bidang pemisah) yang optimal dalam ruang fitur untuk memisahkan data dari kelas yang berbeda. *Hyperplane* ini dipilih agar jarak antara *hyperplane* dengan titik-titik data terdekat (yang disebut *support vectors*) maksimum [3], [13]. SVM juga dapat menggunakan fungsi kernel untuk mengubah ruang fitur ke ruang dimensi yang lebih tinggi, agar data yang tidak linear terpisah dengan lebih baik [15], [16]. Gambar 2 menampilkan ilustrasi penerapan *hyperplane* pada SVM.

Konsep klasifikasi dengan SVM adalah mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data. SVM mampu bekerja pada dataset yang berdimensi tinggi dengan menggunakan kernel trik. SVM menggunakan beberapa titik data terpilih yang berkontribusi (*support vector*) untuk membentuk model yang akan digunakan dalam proses klasifikasi [8]. Persamaan SVM yang digunakan adalah sebagai berikut:

$$f(x) = w \cdot x + b \tag{1}$$

atau

$$f(x) = \sum a_i y_i K(x, x_i) + b \quad m \ i = 1 \tag{2}$$

Keterangan :

w : parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik *support vector*)

x : titik data masukan *Support Vector Machine*

a_i : nilai bobot setiap titik data

K(x, x_i) : fungsi kernel

b : parameter *hyperplane* yang dicari (nilai bias)

Penelitian ini menggunakan kernel *polynomial* dengan persamaan berikut:

$$(K(x, y) = (x \cdot y + c) \ d \tag{3}$$

Keterangan:

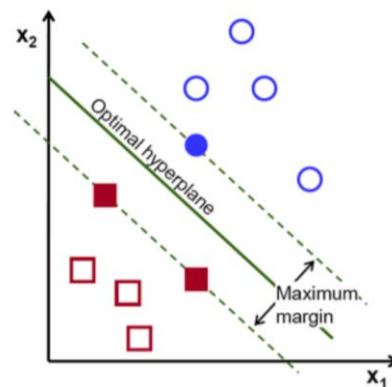
K(x, y) : Nilai kernel dari data x dan data y

x : fitur data 1

y : fitur data 2

c : nilai konstanta

d : nilai *degree*



Gambar 2. Ilustrasi SVM

2.5. Analisis dan Evaluasi Model

Tahapan ini dilakukan untuk mengetahui klasifikasi sentimen masyarakat terhadap pandemi Covid-19 melalui opini-opini (*tweet*) masyarakat melalui media sosial Twitter. Analisis sentimen pada *tweet-tweet* Covid-19 memungkinkan pengguna untuk memahami tanggapan dan pandangan yang terungkap dalam media sosial.

3. HASIL DAN PEMBAHASAN

3.1. Pemodelan Data

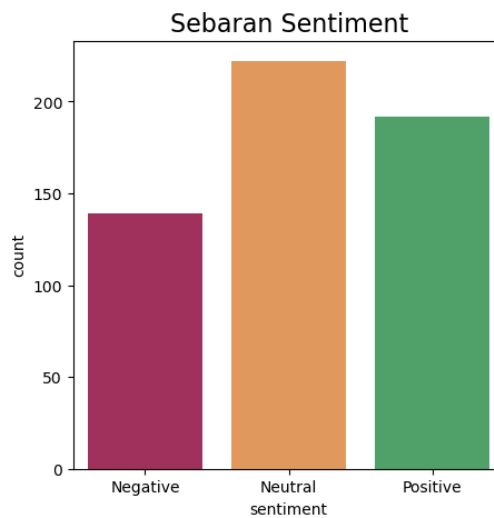
Pengujian diawali dengan mengolah dataset *train* yang digunakan. Pengujian menggunakan *miniconda* dengan *jupyter notebook* sebagai alat bantu pengolahan data yang akan dilakukan analisis. Pada tahap ini pemanggilan data *train* diolah dengan melakukan tahap *preprocessing*, *casefolding*, menghapus *whitespace*, menerjemahkan emoji menjadi kata-kata, normalisasi karakter *unicode* string ke *ascii*, menghapus *html* dan *url*, mengganti *word elongation*, mengganti kata gaul (*slang*) ke bentuk yang lebih formal, menghapus tanda baca dan angka, dan *stemming*. Gambar 3 menampilkan contoh hasil *preprocessing*.

Out[4]:

	tweets	sentiment	cleaned
0	Salut! Polri mendapat apresiasi dari DPR RI da...	2	salut polri dapat apresiasi dari dpr ri dalam ...
1	Sebelum booming corona biasanya ke pasar nongk...	0	belum booming corona biasa ke pasar nongkrong ...
2	Covid 19 plus 30 ribu narapidana https://t.co/...	0	covid plus ribu narapidana
3	@faruqmd Lu corona ora, muntaber iya.	0	faruqmd lu corona ora muntaber iya
4	3. Jenazah belum dapat dipastikan apakah posit...	1	jenazah belum dapat pasti apakah positif infek...

Gambar 3. Hasil *preprocessing*

Langkah selanjutnya adalah memodelkan data *train* dalam bentuk grafik batang, dan membagi data menjadi tiga kelas, yaitu positif, negatif, dan netral. Gambar 4 menampilkan grafik distribusi data ke dalam tiga kelas berdasarkan hasil *preprocessing*. Grafik tersebut menunjukkan bahwa distribusi setiap sentimen pada data *sample* tidak menunjukkan *imbalance* pada setiap kelasnya. Walaupun data didominasi oleh sentimen netral tetapi jumlah data setiap sentimen tidak terlalu berbeda jauh.



Gambar 4. Model Data Train

3.2. Membandingkan Model

Langkah selanjutnya adalah membandingkan setiap model yang digunakan, yaitu model *decision tree*, *logistic regression*, *random forest classifier*, *support vector machine*. Performa model-model tersebut dibandingkan dengan melihat *metric accuracy* agar dapat dipilih model terbaiknya. *Fitting* model-model dilakukan dengan menggunakan *StratifiedKfold* sebanyak 5 *split*. Kemudian model-model tersebut diukur keakuratannya dengan menghitung rata-rata akurasi. Hasil akurasi yang diperoleh diperlihatkan pada Gambar 5.

Rata - rata akurasi dari Decision Tree Classifier : 0.4881 - std : 0.0552
 Rata - rata akurasi dari Logistic Regression : 0.6040 - std : 0.0514
 Rata - rata akurasi dari Random Forest Classifier : 0.5696 - std : 0.0387
 Rata - rata akurasi dari SVM : 0.6094 - std : 0.0378

Gambar 5. Hasil Perbandingan Model

Hasil rata-rata setiap model menunjukkan bahwa SVM memiliki performa yang lebih baik dibandingkan dengan model-model lainnya. SVM memiliki akurasi rata-rata tertinggi (0,6094) dengan tingkat standar deviasi yang paling rendah (0,0378). Hasil ini mengindikasikan bahwa SVM memiliki kinerja yang sangat baik dan konsisten. Oleh karena itu, pada penelitian ini SVM digunakan sebagai model utama untuk mengklasifikasikan sentimen.

3.3. Hyperparameter Tunning

Proses penyetelan *hyperparameter* dilakukan pada model SVM dengan tujuan untuk mendapatkan model terbaik. Pendekatan yang digunakan adalah *gridsearch* dengan menguji berbagai kombinasi parameter. Perbandingan rata-rata akurasi dari setiap kombinasi parameter dilakukan dengan metode *StratifiedKfold*. Hasil proses *tunning parameter* ditunjukkan pada Gambar 6.

Out[8]:

	param_svm_C	param_svm_kernel	param_tf-idf_max_features	param_svm_coef0	params	mean_test_score	std_test_score	rank_test_score
44	0.9	linear	3000	NaN	{'svm_C': 0.9000000000000001, 'svm_kernel': ...}	0.611155	0.044532	1
37	0.8	linear	2000	NaN	{'svm_C': 0.8000000000000003, 'svm_kernel': ...}	0.609386	0.039866	2
38	0.8	linear	3000	NaN	{'svm_C': 0.8000000000000003, 'svm_kernel': ...}	0.609369	0.042560	3
50	1.0	linear	3000	NaN	{'svm_C': 1.0000000000000002, 'svm_kernel': ...}	0.609353	0.037839	4
31	0.7	linear	2000	NaN	{'svm_C': 0.7000000000000002, 'svm_kernel': ...}	0.607568	0.035870	5
70	1.3	rbf	2000	NaN	{'svm_C': 1.3000000000000003, 'svm_kernel': ...}	0.605782	0.050123	6
64	1.2	rbf	2000	NaN	{'svm_C': 1.2000000000000004, 'svm_kernel': ...}	0.605749	0.051658	7
83	1.5	rbf	3000	NaN	{'svm_C': 1.5000000000000004, 'svm_kernel': ...}	0.605749	0.052283	7

Gambar 6. Hasil Tunning

Hasil penyetelan *hyperparameter* menunjukkan bahwa model SVM dengan parameter C=0,9 dan kernel="linear" memiliki rata-rata skor akurasi tertinggi dibandingkan dengan model lainnya. Namun, model tersebut memiliki nilai standar deviasi yang tinggi sebesar 0,044532. Hal ini menunjukkan kurangnya konsistensi dalam performa model. Sebaliknya, model SVM dengan parameter C=1 dan kernel="linear" memiliki skor akurasi yang lebih rendah dibandingkan model sebelumnya, tetapi memiliki nilai standar deviasi yang jauh lebih kecil. Hal ini menandakan konsistensi yang lebih baik. Oleh karena itu, model dengan parameter C=1 dan kernel="linear" digunakan sebagai model utama.

3.4. Fitting Main Model

Fitting salah satu tahap penting dalam pengembangan model untuk membuat prediksi atau mengklasifikasikan data baru. Proses *fitting* dilakukan pada model utama yang akan digunakan, yaitu sebuah model bertipe *Pipeline*. Dalam model ini, berbagai langkah dalam analisis teks digabungkan secara terstruktur dalam satu alur kerja. Langkah-langkah tersebut mencakup penggunaan metode TF-IDF dengan parameter *max_features*=3000, yang bertujuan untuk mengubah teks menjadi representasi numerik dengan mempertimbangkan frekuensi kata-kata dan kemunculan kata-kata yang paling penting. Selain itu, model juga menggunakan metode *Support Vector Machine* (SVM) dengan parameter C=1 dan kernel="linear", yang bertujuan untuk melakukan klasifikasi teks berdasarkan fitur-fitur yang telah diekstraksi. Dengan kombinasi langkah-langkah ini dalam sebuah pipeline, model dapat dilatih dan dievaluasi secara efisien untuk membuat prediksi sentimen atau klasifikasi teks yang akurat. Hasil proses *Fitting main model* ditampilkan pada Gambar 7.

Out[9]:

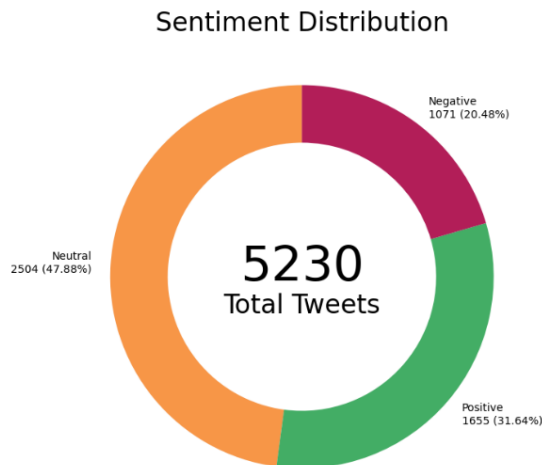
	precision	recall	f1-score	support
0	0.666667	0.500000	0.571429	28.000000
1	0.640000	0.711111	0.673684	45.000000
2	0.525000	0.552632	0.538462	38.000000
accuracy	0.603604	0.603604	0.603604	0.603604
macro avg	0.610556	0.587914	0.594525	111.000000
weighted avg	0.607357	0.603604	0.601598	111.000000

Gambar 7. Fitting Main Model

Berdasarkan hasil proses *fitting main model* pada Gambar 6 dapat disimpulkan bahwa model memiliki kinerja yang relatif seimbang, namun masih ada ruang untuk perbaikan. Meskipun nilai *precision*, *recall*, dan *F1-score* untuk setiap kelas (negatif (0), netral (1), dan positif (2)) menunjukkan variasi, nilai-nilai tersebut relatif serupa, dengan nilai tertinggi terdapat pada kelas netral (1). Akurasi keseluruhan model adalah 0,603604. Akurasi dapat menjadi indikator yang kurang akurat jika distribusi kelas dalam dataset tidak seimbang. Secara keseluruhan, model ini dapat dianggap sebagai alat yang dapat memberikan wawasan dalam menganalisis sentimen dalam teks.

3.5. Visualisasi Data

Tahapan selanjutnya adalah menerapkan model yang telah diperoleh untuk melakukan prediksi pada dataset *twitter* mengenai Covid-19 di DKI Jakarta [17]. Hasil pengolahan data divisualisasikan seperti pada Gambar 8.



Gambar 8. Visualisasi data

Berdasarkan hasil analisis sentimen yang dilakukan pada 5230 *tweet* yang terdapat pada *dataset* ditemukan bahwa mayoritas respon masyarakat terhadap Covid-19 cenderung netral, dengan persentase sebesar 47,88%. Namun, penting untuk dicatat bahwa terdapat juga bagian yang signifikan dari masyarakat yang menunjukkan sentimen positif sebesar 31,4%, sementara 20,48% mengekspresikan sentimen negatif. Data hasil analisis ini dapat disimpulkan bahwa meskipun mayoritas respon masyarakat cenderung netral, masih terdapat sebagian yang signifikan menunjukkan sentimen positif dan negatif terhadap Covid-19. Hal ini menunjukkan kompleksitas persepsi dan reaksi masyarakat terhadap situasi pandemi, yang membutuhkan pemahaman yang lebih mendalam dan respons yang tepat dari berbagai pihak terkait.

4. PENUTUP

Penelitian ini menegaskan pentingnya pemahaman tentang persepsi dan sentimen masyarakat terkait pandemi Covid-19 di DKI Jakarta. Pemilihan model untuk melakukan analisis pada sentimen masyarakat

terkait pandemi ini sangat menentukan hasil prediksinya. Penggunaan metode SVM sebagai metode utama dalam menganalisis sentimen pada penelitian ini telah menghasilkan prediksi dengan komposisi data sentimen 47,88% pada sentimen netral, 31,4% untuk sentimen positif, dan 20,48% untuk sentimen negatif. Temuan dan hasil analisis penelitian didapatkan bahwa meskipun kecenderungan masyarakat dalam merespon pandemi ini didominasi dengan sentimen netral namun prosentase sentimen positif kesadaran masyarakat dalam menyikapi pandemi masih lebih besar dibanding dengan sentimen negatif. Melalui penerapan analisis sentimen terhadap *tweet-tweet* terkait Covid-19 -19, diharapkan hasil ini dapat memberikan kontribusi yang berarti dalam pemahaman tentang respons dan sikap masyarakat di media sosial.

Penelitian ini juga memiliki keterbatasan sehingga masih perlu pengembangan dengan melibatkan teknik dan algoritma yang memiliki tingkat akurasi yang tinggi sehingga memerlukan kajian dengan data yang lebih banyak dan relevan. Data dapat dikembangkan dengan memanfaatkan data dari *tweet-tweet* yang dipilih dengan metode tertentu agar, sehingga tidak mencakup seluruh sudut pandang dan pengalaman masyarakat.

DAFTAR PUSTAKA

- [1] A. Rossi, T. Lestari, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dan Pembobotan Emoji," vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] P. Antinasari, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," vol. 1, no. 12, pp. 1733–1741, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen analisis publik terhadap kebijakan lockdown Pemerintah Jakarta menggunakan algoritma SVM," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhmfMjtXw>
- [4] G. Khanvilkar and D. Vora, "Sentiment analysis for product recommendation using random forest," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 3, pp. 87–89, 2018, doi: 10.14419/ijet.v7i3.3.14492.
- [5] D. Alita and A. Rahman, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *Jurnal Komputasi*, vol. 8, no. 2, 2020.
- [6] 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, doi: 10.1016/j.procs.2019.05.008.
- [7] 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-Nearest Neighbor dan Pembobotan Jumlah Retweet," vol. 1, no. 12, pp. 1750–1757, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] U. Rofiqoh, R. Setya Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features," vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis," *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, vol. 1, pp. 525–531, 2019, doi: 10.5220/0008364105250531.
- [11] T. Pranckevičius and V. Marcinkevičius, "Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, 2017, doi: 10.22364/bjmc.2017.5.2.05.
- [12] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naive Bayes, decision tree, and random forest algorithm," *Procedia Comput Sci*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.

-
- [13] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment Analysis using SVM: A Systematic Literature Review," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, 2018, [Online]. Available: www.ijacsa.thesai.org
- [14] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [15] Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Comput Sci*, vol. 127, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
- [16] A. Nayak and S. Natarajan, "Comparative study of Naïve Bayes, Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter feeds," *International Journal of Advanced Studies in Computer Science and Engineering*, vol. 5, no. 1, 2016.
- [17] Wahyu Setianto, " Analisis Sentimen Corona DKI Jakarta ." <https://github.com/Hyuto/Analisis-Sentimen-Corona-DKI-Jakarta/> (accessed Jul. 20, 2023).