KLASIFIKASI SENTIMEN PUBLIK TERHADAP KEBIJAKAN KENDARAAN LISTRIK MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE*

Muhammad Nouval Daffa Ramadhan, Aris Gunaryati
Sistem Informasi, Universitas Nasional
Jl. Sawo Manila No.61, RT.14/RW.7, Pejaten Bar., Ps. Minggu, Jakarta, Indonesia
muhammadnouvaldaffaramadhan2021@student.unas.ac.id

ABSTRAK

Perubahan iklim menjadi isu global yang mendesak, dengan emisi gas rumah kaca dan polusi udara berdampak buruk pada kesehatan dan lingkungan. Kendaraan listrik (EV) muncul sebagai solusi efektif untuk mengurangi emisi dan polusi, karena tidak menghasilkan emisi langsung. Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap kebijakan kendaraan listrik dengan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. *Data* yang digunakan diperoleh melalui proses *crawling* dari media sosial Twitter, yang kemudian melalui tahapan *preprocessing*, *labeling* menggunakan TextBlob, dan *weighting* menggunakan TF-IDF, model diklasifikasikan ke dalam tiga kategori sentimen positif, negatif, dan netral. Model klasifikasi diuji berdasarkan metrik akurasi, presisi, *recall*, dan F1-*score*. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* memiliki akurasi yang lebih tinggi dibandingkan *Naïve Bayes*, dengan nilai masing-masing sebesar 81% dan 76%. Temuan ini mengindikasikan bahwa *Support Vector Machine* lebih efektif dalam mengklasifikasikan sentimen publik terhadap kebijakan kendaraan listrik. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam serta meningkatkan pemahaman masyarakat mengenai sentimen publik terhadap kebijakan pemerintah terkait kendaraan listrik.

Kata kunci : *Klasifikasi Sentimen, Naïve Bayes, Support Vector Machine, Twitter, Kendaraan Listrik*

1. PENDAHULUAN

Di era modern, perubahan iklim menjadi perhatian utama di tingkat global. Emisi gas rumah kaca yang meningkat serta polusi udara telah memberikan dampak besar terhadap kesehatan manusia, keseimbangan ekosistem, dan stabilitas iklim secara keseluruhan. Untuk mengatasi permasalahan ini, berbagai upaya telah dilakukan, termasuk beralih ke energi terbarukan dan mengadopsi kendaraan yang lebih ramah lingkungan. Kendaraan listrik (EV) dipandang sebagai solusi yang efektif dalam menekan emisi gas rumah kaca serta mengurangi polusi udara. Tidak seperti kendaraan berbahan bakar fosil, EV tidak menghasilkan emisi langsung yang berdampak negatif terhadap lingkungan, sehingga dapat berkontribusi dalam meningkatkan kualitas udara dan mengurangi risiko terhadap kesehatan manusia. Selain itu, EV juga memanfaatkan sumber energi terbarukan, seperti tenaga surya dan angin, yang berkontribusi dalam mengurangi ketergantungan terhadap bahan bakar fosil sekaligus mempercepat peralihan menuju sistem energi yang lebih berkelanjutan.

Jumlah kendaraan bermotor di Indonesia terus meningkat dari tahun ke tahun, meliputi sepeda motor, kendaraan mobil seperti mobil penumpang, mobil barang, dan mobil bis [1]. Untuk mendorong percepatan adopsi kendaraan listrik, pemerintah telah mengeluarkan Perpres No. 79 Tahun 2023 sebagai revisi dari Perpres No. 55 Tahun 2019, dengan tujuan mempercepat transisi menuju energi bersih serta mengurangi ketergantungan terhadap bahan bakar fosil. Kebijakan ini mencakup insentif serta pengembangan infrastruktur seperti stasiun pengisian daya guna membangun ekosistem kendaraan listrik. Namun, efektivitas kebijakan ini masih perlu dikaji lebih lanjut, terutama dalam memahami sentimen dan respons publik terhadap regulasi tersebut.

Penelitian berjudul Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan *Naïve Bayes* Classification. Penelitian ini menganalisis sentimen publik terhadap kebijakan penanganan COVID-19 di Indonesia melalui cuitan Twitter, dengan tujuan memberikan wawasan mengenai opini masyarakat. Hasil penelitian menunjukkan mayoritas tweet bersentimen negatif, dengan evaluasi model Naïve Bayes mencapai akurasi 87,34%, sensitivitas 93,43%, dan spesifisitas 71,76%. Kata-kata yang sering muncul dalam sentimen positif adalah "demo", "Jakarta", dan "kerja", sedangkan dalam sentimen negatif adalah "Jakarta", "demo", dan "orang". Meskipun metode ini efektif dalam klasifikasi sentimen, penelitian ini memiliki keterbatasan karena tidak mempertimbangkan sentimen netral serta hanya menggunakan data Twitter berbahasa Indonesia, sehingga belum mencakup seluruh spektrum opini masyarakat terhadap kebijakan COVID-19 [2].

Analisis sentimen merupakan proses sistematis dalam mengolah data teks guna mengekstrak serta mengukur aspek afektif dan informasi subjektif yang terkandung di dalamnya. Analisis sentimen adalah studi tentang komputer yang mengubah item dan kualitasnya menjadi teks, yang mencakup opini dan perasaan public [3] .

Penelitian ini merupakan upaya untuk memahami sentimen publik terhadap kebijakan pemerintah mengenai kendaraan listrik di Indonesia melalui

analisis data tweet. Dengan membandingkan akurasi dua algoritma *machine learning*, penelitian ini diharapkan dapat memberikan pemahaman yang mendalam mengenai sentiment publik terhadap kebijakan pemerintah terkait kendaraan listrik. Selain itu, temuan dari penelitian ini diharapkan mampu meningkatkan kesadaran dan partisipasi masyarakat dalam mendukung program pemerintah menuju penggunaan kendaraan yang lebih ramah lingkungan.

Klasifikasi sentimen publik terhadap kebijakan kendaraan listrik menjadi aspek yang krusial dalam mendukung proses evaluasi dan pengambilan keputusan berbasis data. Dengan memahami kecenderungan opini masyarakat, baik yang bersifat positif, negatif, maupun netral, para pemangku kebijakan dapat merancang strategi komunikasi yang lebih efektif, mengidentifikasi potensi resistensi sosial, serta meningkatkan akseptabilitas terhadap implementasi kebijakan yang berorientasi pada transisi energi bersih.

2. TINJAUAN PUSTAKA

2.1. Literatur Review

Pada penelitian berjudul "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM)" bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap pemindahan Ibu Kota Negara (IKN) melalui cuitan di Twitter serta meningkatkan akurasi dalam analisis sentimen. Hasil penelitian menunjukkan bahwa metode klasifikasi menggunakan *Naïve Bayes* dan *Support Vector Machine* (SVM) menghasilkan akurasi tinggi, dengan *Naïve Bayes* mencapai 86.94% dan SVM mencapai 90.81%. Penelitian ini memberikan wawasan mendalam tentang opini publik terkait pemindahan IKN dan berkontribusi pada pemahaman isu tersebut di Indonesia [4].

Penelitian selanjutnya yang berjudul "Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform Twitter Menggunakan Metode *Naïve Bayes*", Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap kendaraan listrik ke dalam kategori positif, negatif, dan netral berdasarkan data dari Twitter. Hasil penelitian menunjukkan bahwa sentimen mayoritas cenderung positif, dengan akurasi analisis sentimen mencapai 75%. Presisi untuk sentimen positif sebesar 81%, negatif 15%, dan netral 30%, menunjukkan bahwa kendaraan listrik mendapat respons yang lebih dominan dari sisi positif [5].

Penelitian ketiga berjudul "*Sentiment Analysis of Towards Electric Cars using Naïve Bayes Classifier and Support Vector Machine Algorithm*", Penelitian ini bertujuan untuk menentukan opini publik mengenai penggunaan mobil listrik di Indonesia melalui analisis sentimen menggunakan *Naïve Bayes Classifier* dan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa sentimen mayoritas bersifat netral, dengan SVM memiliki akurasi 90%, sedikit lebih unggul dibandingkan Naïve Bayes yang mencapai 88%. Studi ini juga mengidentifikasi kata kunci yang sering muncul dalam percakapan terkait mobil listrik [6].

Penelitian keempat yang berjudul "*Twitter Sentiment Analysis of Electric Vehicle Subsidy Policy Using Naïve Bayes Algorithm*" Penelitian ini mengeksplorasi sentimen masyarakat terhadap kebijakan subsidi kendaraan listrik melalui analisis teks dan visualisasi seperti *word cloud*. Hasil penelitian menunjukkan bahwa mayoritas sentimen bersifat negatif, dengan akurasi analisis sentimen menggunakan *Naïve Bayes* mencapai 69.49%, serta *precision* 53.27% dan *recall* 74.26%. Studi ini menekankan pentingnya analisis sentimen dalam mengevaluasi serta meningkatkan efektivitas kebijakan publik [7].

Penelitian terakhir yang berjudul "*Improving Public Opinion Analysis Using Hybrid Model Based on SVM and Naïve Bayes*" mengembangkan model hibrida yang menggabungkan algoritma SVM dan NB untuk analisis sentimen, yang berhasil mencapai akurasi 87%, lebih tinggi dibandingkan dengan SVM (82%) dan NB (78%). Penggunaan fitur *unigram* dan teknik *preprocessing* teks seperti *stopword removal* dan *stemming* meningkatkan kualitas data dan akurasi model. Namun, penelitian ini belum mengatasi masalah *dataset* tidak seimbang, yang dapat mempengaruhi deteksi sentimen minoritas. Selain itu, hanya fitur *unigram* yang digunakan. Penelitian ini juga berfokus pada data berbahasa Inggris dan belum menguji model pada bahasa lain, seperti bahasa Indonesia [8].

2.2. Klasifikasi Sentimen

Klasifikasi sentimen adalah salah satu bidang dalam text mining yang termasuk dalam studi komputasi untuk mengenali opini, emosi, serta sikap individu terhadap suatu entitas [9]. Analisis sentimen sering disamakan dengan *opinion mining*, karena keduanya bertujuan untuk menilai suatu opini dalam kategori positif, netral, atau negatif.

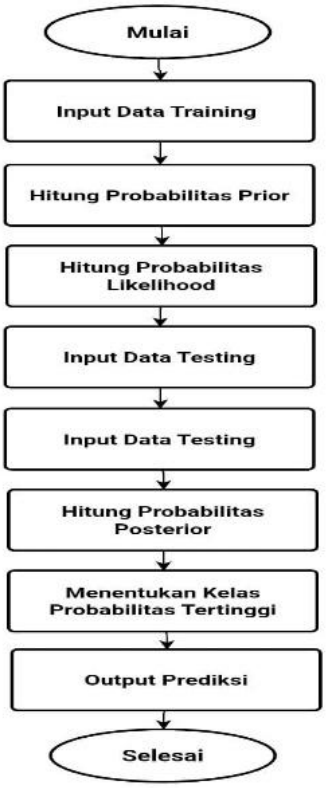
2.3. Twitter

Twitter adalah *platform* media sosial dan *microblogging* yang didirikan tahun 2006 yang memungkinkan pengguna untuk mengirim dan membaca pesan singkat hingga 280 karakter, yang dikenal sebagai "*tweet*". Sebagai salah satu *platform* media sosial yang populer, Twitter memiliki jutaan pengguna aktif di seluruh dunia. Kemudian Twitter berganti nama menjadi X. Sosial media X, merupakan *platform microblogging* menerima lebih dari 500 juta *tweet* setiap hari [10]. Twitter merupakan salah satu media sosial yang populer digunakan sebagai sumber *data* untuk penelitian analisis sentimen. Twitter memungkinkan pengguna tetap saling terhubung meskipun mereka tidak saling mengikuti.

2.4. *Naïve Bayes*

Naïve Bayes memanfaatkan perhitungan probabilitas berdasarkan statistik dari teori *bayes*, yang mengasumsikan adanya atau tidak adanya suatu kelas

[11]. Metode *Naïve Bayes* memiliki beberapa keunggulan, antara lain kemudahan dalam implementasi, kecepatan dalam proses klasifikasi data, dan tingkat akurasi yang cukup tinggi. *Naïve Bayes* memiliki beberapa varian utama yang disesuaikan dengan jenis data yang digunakan seperti *GaussianNB. MultinomiaNB, BernoulliNB, ComplementNB, CategoricalNB.*

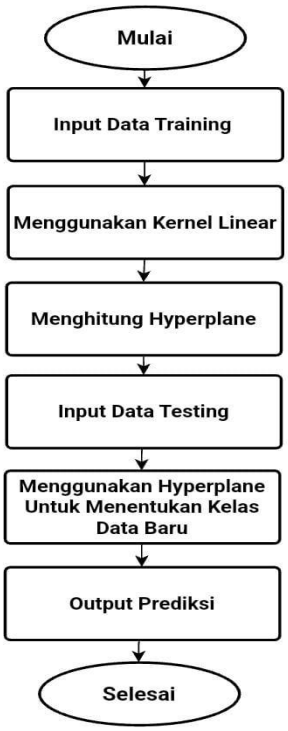


Gambar 1. Flowchart *Naïve Bayes*

Gambar 1 menampilkan proses klasifikasi menggunakan algoritma *Naïve Bayes* dimulai dengan memasukkan data pelatihan yang telah diberi label untuk menghitung probabilitas prior setiap kelas. Selanjutnya, dihitung probabilitas *likelihood* untuk setiap fitur terhadap kelas tertentu. Setelah model terbentuk, data uji dimasukkan, dan probabilitas posterior dihitung menggunakan *Teorema Bayes.* Kelas dengan probabilitas tertinggi kemudian dipilih sebagai hasil prediksi, dan proses klasifikasi selesai.

2.5. *Support Vector Machine*

Support Vector Machine adalah algoritma *machine learning* yang digunakan untuk klasifikasi dan regresi dengan mencari *hyperplane* (garis/permukaan pemisah) terbaik yang memaksimalkan margin antara kelas-kelas data. SVM sering digunakan untuk klasifikasi teks, pengenalan wajah, kategorisasi gambar, bioinformatika, prediksi finansial dan banyak aplikasi lainnya. Nilai yang dihasilkan oleh algoritma ini berupa sebuah garis pemisah, atau yang dikenal sebagai *hyperplane*, yang akan digunakan untuk membedakan *tweet* berdasarkan labelnya, seperti positif, netral, dan negatif.



Gambar 2. Flowchart SVM

Gambar 2 menampilkan proses klasifikasi menggunakan algoritma SVM dimulai dengan memasukkan data pelatihan untuk membangun model. Selanjutnya, pemilihan kernel dilakukan untuk menentukan fungsi pemisah menggunakan kernel linear. Setelah kernel dipilih, SVM menghitung *hyperplane* optimal yang memaksimalkan *margin* antara kelas-kelas data. Setelah model terbentuk, data uji dimasukkan, dan *hyperplane* digunakan untuk menentukan kelas dari data uji. Kelas dengan posisi paling dekat dengan hyperplane dipilih sebagai hasil prediksi, dan proses klasifikasi selesai.

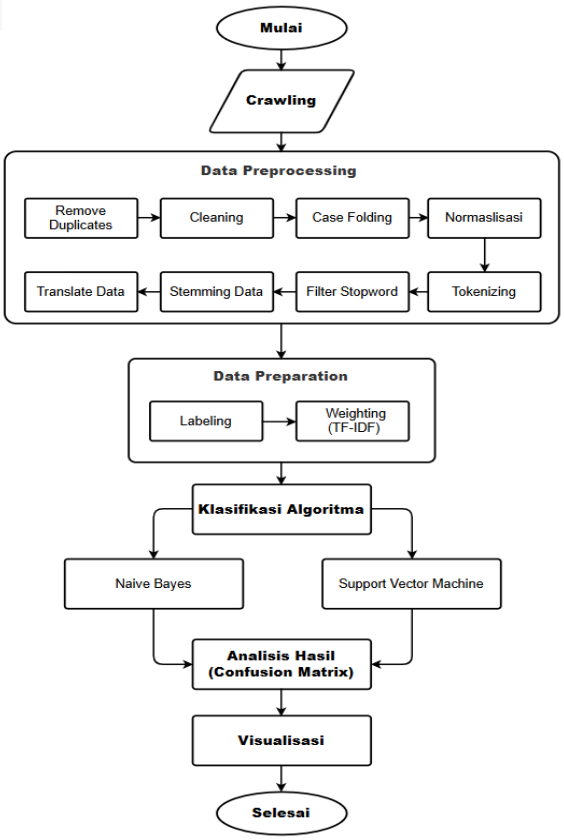
2.6. *Confusion Matrix*

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap nilai aktual. *Confusion Matrix* menyajikan informasi secara rinci mengenai jumlah prediksi yang tepat maupun keliru dalam setiap kategori, sehingga mempermudah analisis efektivitas model.

Dalam *confusion matrix* memiliki metrik evaluasi seperti presisi, *recall*, dan F1-*score* yang berfungsi dalam menilai kinerja model klasifikasi. Presisi mengukur seberapa akurat model dalam memprediksi kelas, yaitu seberapa banyak prediksi kelas yang benar dari semua prediksi kelas yang dihasilkan. Sementara itu, *recall* menilai kemampuan model dalam mendeteksi semua contoh kelas yang ada, dengan memperhatikan seberapa banyak kelas yang berhasil diprediksi dengan benar. F1-*score* adalah rata-rata harmonis antara presisi dan *recall*,

yang memberikan gambaran keseimbangan antara keduanya. Ketiga metrik ini digunakan untuk memastikan bahwa model tidak hanya menghasilkan akurasi yang tinggi, tetapi juga mampu meminimalkan kesalahan dalam memprediksi kelas yang relevan, terutama pada dataset yang tidak seimbang.

3. METODE PENELITIAN



Gambar 3. Tahapan penelitian

Gambar 3 memperlihatkan bahwa penelitian ini terdiri dari enam tahapan, yaitu:

a. Crawling

Mengumpulkan opini masyarakat terkait kebijakan kendaraan Listrik di Twitter, menggunakan teknik *web scrapping* dengan *library TweetHarvest.*

b. *Data Pre-processing*

Melakukan tahapan *data pre-processing* pada opini masyarakat, sehingga data mentah dapat dipersiapkan untuk pemrosesan lebih lanjut oleh algoritma.

c. Data Preparation

Tahap ini terdiri dari dua proses utama, yaitu *labeling*, dan *weighting*. Proses *labeling* dilakukan menggunakan *TextBlob*, *TextBlob* menggunakan pendekatan berbasis leksikon untuk menilai sentimen suatu teks *Weighting* menggunakan TF-IDF yang bertujuan untuk menentukan tingkat kepentingan kata-kata kunci dalam *tweet.*

d. Klasifikasi Algoritma

Klasifikasi dilakukan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. Sebelum proses klasifikasi, *data* terlebih dahulu dibagi melalui *splitting*, dengan proporsi 80% sebagai *data* latih dan 20% sebagai *data* uji.

e. Analisis Hasil

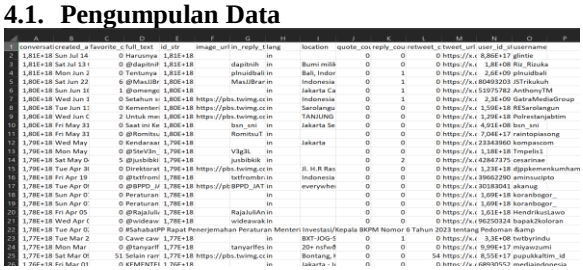
Kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* dianalisis menggunakan *confusion matrix*, dengan metrik akurasi, presisi, *recall*, dan F1-*score* sebagai tolok ukur, kemudian hasilnya dibandingkan untuk menilai performa masing-masing model.

f. Visualisasi

Pada tahap akhir, dilakukan visualisasi data *tweet* dengan *dashboard* sederhana yang membahas kebijakan kendaraan listrik guna mempermudah pemahaman terhadap persepsi publik. Proses visualisasi menggunakan Streamlit, yang memungkinkan penyajian data secara interaktif dan mudah dipahami.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data



Gambar 4. Hasil *crawling*

Gambar 4 menampilkan distribusi *data* yang berhasil dikumpulkan melalui *crawling* menggunakan *TweetHarvest* berjumlah 4.234 data tweet.

4.2. Pre-processing Data

a. *Remove Duplicates*

Remove duplicates berfungsi untuk menghapus *data* yang berulang dalam *dataset* untuk memastikan keunikan setiap data, proses ini menyebabkan pengurangan sebanyak 1.098 *data* dari jumlah awal *data*.

b. *Cleaning*

Cleaning adalah langkah penting dalam klasifikasi sentimen Twitter untuk menghilangkan elemen-elemen tidak relevan, seperti URL, *username*, *retweet*, dan simbol. Proses ini memastikan model dapat lebih fokus pada teks utama *tweet* yang mengandung informasi penting.

full_text	cleaning
Harusnya transum diperbagus dan ditambah modan...	Harusnya transum diperbagus dan ditambah modan...
@dapitnih Rakyatnya lah yg dipenjara karena ti...	Rakyatnya lah yg dipenjara karena tidak bayar ...
Tentunya kegiatan ini sejalan dengan Peraturan...	Tentunya kegiatan ini sejalan dengan Peraturan...
@MasJJBram @ngurahsaka Salah tujuan EV adalah ...	Salah tujuan EV adalah mengurangi polusi udara...
@omengoren @jellypastaa Kalo masih mematuhi pe...	Kalo masih mematuhi peraturan yg ada masih bai...

Gambar 5. Hasil *cleaning*

Pada gambar 5 menampilkan hasil *cleaning* yang menghilangkan simbol dan *username* dalam *dataset*

c. *Case Folding*

Case Folding merupakan proses konversi seluruh huruf pada data tweet menjadi huruf kecil. Langkah ini bertujuan untuk menyederhanakan klasifikasi serta mempermudah pemrosesan teks.

cleaning	case_folding
Harusnya transum diperbagus dan ditambah modan...	harusnya transum diperbagus dan ditambah modan...
Rakyatnya lah yg dipenjara karena tidak bayar ...	rakyatnya lah yg dipenjara karena tidak bayar ...
Tentunya kegiatan ini sejalan dengan Peraturan...	tentunya kegiatan ini sejalan dengan peraturan...
Salah tujuan EV adalah mengurangi polusi udara...	salah tujuan ev adalah mengurangi polusi udara...
Kalo masih mematuhi peraturan yg ada masih bai...	kalo masih mematuhi peraturan yg ada masih bai...

Gambar 6. Hasil case folding

Pada gambar 6 menampilkan data *tweet* telah dikonversi menjadi huruf kecil.

d. Normalisasi

Normalisasi bertujuan untuk menyederhanakan dan menyelaraskan kata dalam teks agar lebih seragam dan konsisten. Proses ini mencakup penghapusan atau penggantian ejaan tidak baku, bahasa gaul, dan variasi kata dalam teks informal, sehingga mengurangi ambiguitas dan mempermudah pemrosesan lebih lanjut.

case_folding	normalisasi
harusnya transum diperbagus dan ditambah modan...	harusnya transum diperbagus dan ditambah modan...
rakyatnya lah yg dipenjara karena tidak bayar ...	rakyatnya lah yang dipenjara karena tidak baya...
tentunya kegiatan ini sejalan dengan peraturan...	tentunya kegiatan ini sejalan dengan peraturan...
salah tujuan ev adalah mengurangi polusi udara...	salah tujuan ev adalah mengurangi polusi udara...
kalo masih mematuhi peraturan yg ada masih bai...	kalau masih mematuhi peraturan yang ada masih ...

Gambar 7. Hasil normalisasi

Gambar 7 menunjukkan hasil normalisasi yang mengganti ejaan tidak baku menjadi baku.

e. *Tokenizing*

Proses *tokenizing* dalam klasifikasi sentimen Twitter berfungsi untuk memecah teks tweet menjadi unit yang terstruktur, seperti kata, frasa, atau simbol.

normalisasi	tokenize
harusnya transum diperbagus dan ditambah modan...	[harusnya, transum, diperbagus, dan, ditambah,...
rakyatnya lah yang dipenjara karena tidak baya...	[rakyatnya, lah, yang, dipenjara, karena, tida...
tentunya kegiatan ini sejalan dengan peraturan...	[tentunya, kegiatan, ini, sejalan, dengan, per...
salah tujuan ev adalah mengurangi polusi udara...	[salah, tujuan, ev, adalah, mengurangi, polusi...
kalau masih mematuhi peraturan yang ada masih ...	[kalau, masih, mematuhi, peraturan, yang, ada,...

Gambar 8. Hasil *tokenizing*

Gambar 8 menampilkan hasil *tokenizing*, di mana setiap data *tweet* dipecah menjadi unit-unit kecil yang berupa token kata.

f. *Stopword Removal*

Stopword berfungsi untuk menghapus kata-kata umum yang tidak memiliki arti signifikan dari teks cuitan, dengan demikian model klasifikasi sentimen dapat fokus pada kata-kata yang lebih penting dan relevan dalam menentukan sentimen.

tokenize	stopword removal
[harusnya, transum, diperbagus, dan, ditambah,...	[transum, diperbagus, ditambah, modanya, perat...
[rakyatnya, lah, yang, dipenjara, karena, tida...	[rakyatnya, dipenjara, bayar, pajak, bpjs, kes...
[tentunya, kegiatan, ini, sejalan, dengan, per...	[kegiatan, sejalan, peraturan, gubernur, no, p...
[salah, tujuan, ev, adalah, mengurangi, polusi...	[salah, tujuan, ev, mengurangi, polusi, udara,...
[kalau, masih, mematuhi, peraturan, yang, ada,...	[mematuhi, peraturan, anggap, dimasalahkan, li...

Gambar 9. Hasil *stopword removal*

Pada gambar 9 menampilkan hasil *stopword* removal pada *data tweet*, di mana kata-kata umum yang tidak memiliki makna signifikan, seperti "dan", "lah", "ini", dan "adalah", telah dihapus.

g. *Stemming*

Stemming merupakan proses mengonversi kata ke bentuk dasar. Dalam *data* Twitter, yang sering mengandung slang dan kata turunan, *stemming* membantu menyederhanakan teks agar kata-kata dengan makna serupa diperlakukan secara konsisten dalam klasifikasi sentimen.

stopword removal	stemming_data
[transum, diperbagus, ditambah, modanya, perat...	transum bagus tambah modanya atur guna kendara...
[rakyatnya, dipenjara, bayar, pajak, bpjs, kes...	rakyat penjara bayar pajak bpjs sehat beli ken...
[kegiatan, sejalan, peraturan, gubernur, no, p...	giat jalan atur gubernur no guna kendara motor...
[salah, tujuan, ev, mengurangi, polusi, udara,...	salah tuju ev kurang polusi udara negara atur ...
[mematuhi, peraturan, anggap, dimasalahkan, li...	patuh atur anggap masalah lingkung pemuda mala...

Gambar 10. Hasil *stemming*

Gambar 10 menampilkan hasil *stemming*, di mana kata-kata diubah ke bentuk dasarnya, seperti "diperbagus" menjadi "bagus".

h. *Translate*

Translate data bertujuan untuk meningkatkan akurasi dalam penentuan sentimen, terutama saat menggunakan model berbasis leksikon seperti *TextBlob*, yang lebih optimal dalam bahasa Inggris.

Proses translate data dilakukan menggunakan fungsi *deep_translator* sebagai langkah awal sebelum masuk ke tahap *labeling*.

stemming_data	translate
transum bagus tambah modanya atur guna kendara...	transum good add mode set for personal vehicle...
rakyat penjara bayar pajak bpjs sehat beli ken...	people in prison pay tax bpjs healthy buy elec...
giat jalan atur gubernur no guna kendara motor...	active road governor regulates no use of elect...
salah tuju ev kurang polusi udara negara atur ...	wrong direction ev less air pollution countrie...
patuh atur anggap masalah lingkung pemuda mala...	obey the rules consider the environmental prob...

Gambar 11. Hasil *translate*

Gambar 11 menampilkan data *tweet* yang sudah melalui proses *translate* dari bahasa Indonesia ke bahasa Inggris.

4.3. Data Preparation

a. *Labeling*

Proses *labeling* dalam klasifikasi sentimen Twitter bertujuan memberikan label sentimen (positif, negatif, atau netral) pada setiap *tweet*. Labeling dalam penelitian ini dilakukan dengan menggunakan TextBlob, sebuah library open-source berbasis Python yang memanfaatkan teknik *natural language processing* untuk menganalisis teks secara sederhana.

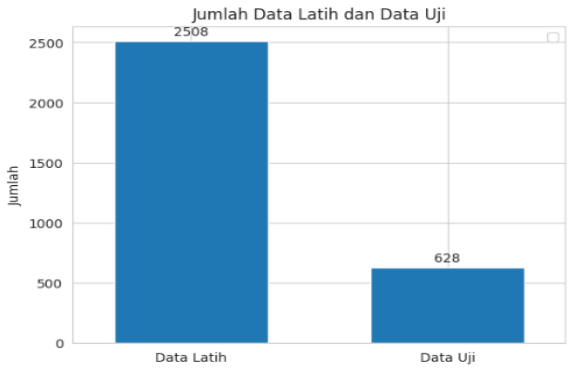
stemming_data	translate	sentimen
transum bagus tambah modanya atur guna kendara...	transum good add mode set for personal vehicle...	positive
rakyat penjara bayar pajak bpjs sehat beli ken...	people in prison pay tax bpjs healthy buy elec...	positive
giat jalan atur gubernur no guna kendara motor...	active road governor regulates no use of elect...	negative
salah tuju ev kurang polusi udara negara atur ...	wrong direction ev less air pollution countrie...	negative

Gambar 12. Hasil labeling

Gambar 12 menampilkan data *tweet* yang sudah melewati proses *labeling*, sentimen diklasifikasikan menjadi positif, netral, dan negatif berdasarkan analisis polaritas dari *TextBlob*.

b. *Splitting Data*

Splitting data bertujuan untuk mengevaluasi kinerja model secara objektif. *Data* latih digunakan dalam proses pelatihan model, sedangkan *data* uji berfungsi untuk mengukur performa model pada *data* yang belum pernah dilihat sebelumnya. Pemisahan ini membantu mencegah *overfitting* dan memastikan model dapat menggeneralisasi dengan baik.



Gambar 13. Hasil splitting data

Gambar 13 menampilkan visualisasi proporsi pembagian *data*, di mana 80% dari total *dataset* digunakan sebagai *data* latih sebanyak 2.508 *data*, sementara 20% sisanya digunakan sebagai *data* uji sebanyak 626 *data*.

c. *Weighting*

Weighting TF-IDF digunakan untuk menilai tingkat kepentingan suatu kata dalam dokumen dibandingkan dengan seluruh koleksi dokumen. TF menghitung frekuensi kemunculan kata dalam satu dokumen, sedangkan IDF mengukur seberapa jarang kata tersebut muncul di seluruh dokumen. Metode ini membantu model fokus pada kata-kata yang lebih relevan dan mengurangi dampak kata umum, sehingga meningkatkan akurasi dalam analisis teks.

```
TF-IDF scores for document 0:
add: 0.2200902700131938
car: 0.15640541002180955
electric: 0.04371014915266301
for: 0.09293961582670185
free: 0.18735930340738408
gage: 0.6466310974301955
good: 0.18982407610730578
mode: 0.27436647141551296
personal: 0.21232267134386662
review: 0.2606867558435679
set: 0.21700096432374869
transum: 0.19889626686023426
tricked: 0.33980025904057953
using: 0.17915731634668314
vehicle: 0.06895390646359958
```

Gambar 14. Hasil TF-IDF

Gambar 14 menampilkan hasil pembobotan kata menggunakan TF-IDF.

4.4. Implementasi Algoritma

a. Klasifikasi *Naïve Bayes*

Classification Report Naive Bayes:

	precision	recall	f1-score	support
negative	0.88	0.54	0.67	124
neutral	0.81	0.68	0.74	210
positive	0.71	0.90	0.79	294
accuracy			0.76	628
macro avg	0.80	0.71	0.73	628
weighted avg	0.78	0.76	0.75	628

Gambar 15. Classification report NB

Gambar 15 menampilkan *classification report* untuk model *ComplementNB* yang digunakan dalam klasifikasi sentimen. Kelas negatif memiliki *precision* sebesar 0.88, *recall* 0.54, dan f1-*score* 0.67, yang menunjukkan model cukup akurat dalam mengenali *data* negatif, tetapi sering melewatkan banyak sampel yang seharusnya diklasifikasikan sebagai negatif. Kelas netral memiliki *precision* 0.81, *recall* 0.68, dan f1-*score* 0.74, yang menandakan bahwa model memiliki keseimbangan yang cukup baik dalam mengenali dan mengklasifikasikan *data* netral, meskipun masih terdapat beberapa kesalahan. Sementara itu, kelas positif memiliki *precision* 0.71, *recall* 0.90, dan f1-*score* 0.79, yang berarti model sangat baik dalam mendeteksi data positif dengan *recall* tinggi, meskipun terkadang masih salah mengklasifikasikan *data* dari kelas lain sebagai positif. Secara keseluruhan, model ini mencapai akurasi 76%, dengan *macro average* f1-*score* sebesar 0.73 dan *weighted average* f1-*score* sebesar 0.75.

b. Klasifikasi *Support Vector Machine*

Classification Report Support Vector Machine:

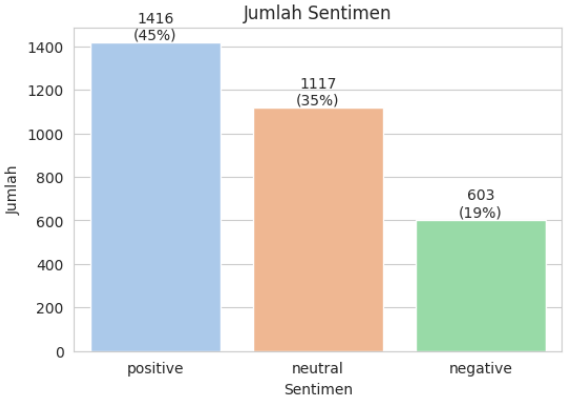
	precision	recall	f1-score	support
negative	0.77	0.69	0.73	124
neutral	0.79	0.84	0.82	210
positive	0.84	0.84	0.84	294
accuracy			0.81	628
macro avg	0.80	0.79	0.80	628
weighted avg	0.81	0.81	0.81	628

Gambar 16. Classification report SVM

Gambar 16 menampilkan *Classification Report* untuk model *Support Vector Machine* menggunakan *hyperparameter* (C: 10, kernel: 'linear') yang digunakan dalam klasifikasi sentimen. Kelas negatif memiliki *precision* sebesar 0.77, *recall* 0.69, dan f1-*score* 0.73, yang menunjukkan bahwa model cukup baik dalam mengidentifikasi *tweet* negatif, meskipun masih sering melewatkan beberapa sampel. Kelas netral memiliki *precision* 0.79, *recall* 0.84, dan f1-*score* 0.82, menandakan bahwa model mampu mengenali *tweet* dengan sentimen netral dengan cukup akurat. Sementara itu, kelas positif memiliki *precision* dan *recall* sebesar 0.84, menghasilkan f1-*score* 0.84, menunjukkan performa yang kuat dan seimbang dalam mengenali *tweet* dengan sentimen positif. Secara keseluruhan, model ini mencapai akurasi sebesar 81%, dengan *macro average* f1-*score* sebesar 0.80 dan *weighted average* f1-*score* sebesar 0.81.

4.5. Analisis Hasil dan Evaluasi

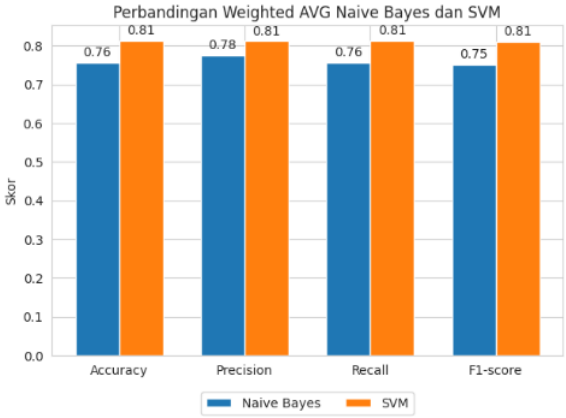
a. Distribusi Sentimen



Gambar 17. Distribusi Sentimen

Gambar 17 menampilkan visualisasi distribusi jumlah *tweet* berdasarkan analisis sentimen. Sentimen positif mendominasi dengan total 1.416 *tweet*, diikuti oleh sentimen netral sebanyak 1.117 *tweet*, sementara sentimen negatif tercatat sebanyak 603 *tweet*.

b. Perbandingan Kinerja Algoritma

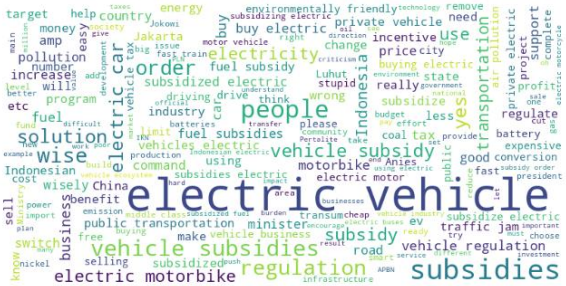


Gambar 18. Perbandingan algoritma

Gambar 18 menampilkan visualisasi perbandingan metrik evaluasi antara model *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam klasifikasi sentimen berdasarkan *weighted avg* menunjukkan bahwa SVM memiliki kinerja lebih unggul di semua metrik evaluasi. Model SVM mencapai akurasi sebesar 0.81, lebih tinggi dibandingkan *Naïve Bayes* yang hanya memperoleh 0.76. Dari segi *precision*, SVM juga lebih baik dengan skor 0.81 dibandingkan *Naïve Bayes* yang mendapatkan 0.78, yang berarti SVM lebih efektif dalam mengurangi kesalahan klasifikasi positif palsu. Pada metrik *recall*, SVM kembali menunjukkan keunggulan dengan skor 0.81, lebih tinggi dibandingkan *Naïve Bayes* yang memiliki 0.76, menunjukkan bahwa SVM lebih andal dalam mengenali kelas yang sebenarnya. Selain itu, F1-*score*, yang merupakan keseimbangan antara *precision* dan

recall, juga lebih tinggi pada SVM dengan nilai 0.81 dibandingkan *Naïve Bayes* yang hanya mencapai 0.75. Secara keseluruhan, hasil ini mengindikasikan bahwa SVM lebih optimal dalam klasifikasi sentimen dibandingkan *Naïve Bayes*, terutama dalam aspek *precision* dan *recall* yang lebih konsisten.

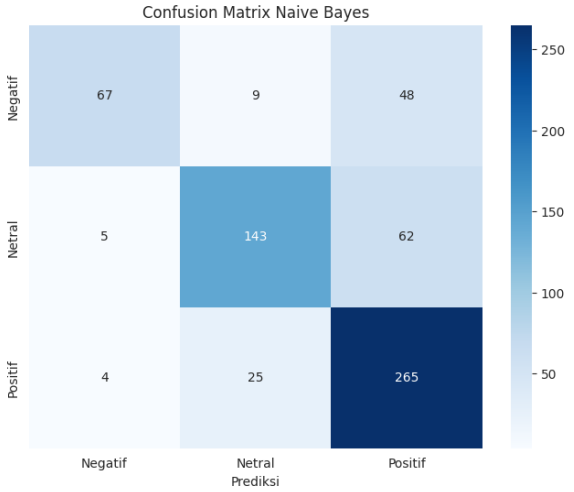
c. Word Cloud



Gambar 19. *Word Cloud*

Gambar 19 menyajikan visualisasi *Word Cloud*, yang menampilkan kata-kata paling sering muncul dalam *dataset* terkait kebijakan kendaraan listrik. Kata-kata yang paling dominan, seperti "*electic*", "*vehicle*", dan "*subsidies*" memiliki ukuran lebih besar, menunjukkan frekuensi kemunculannya yang tinggi

d. Confusion Matrix Naïve Bayes

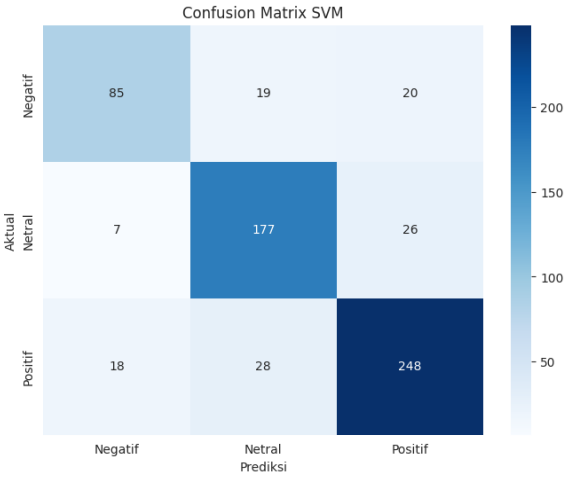


Gambar 20. Confussion Matrix NB

Gambar 20 menunjukkan performa model *Naïve Bayes* dalam mengklasifikasikan *data* ke dalam tiga kelas. Berikut adalah penjelasannya:

- Negatif: Model mengklasifikasikan 67 sampel dengan benar, tetapi salah mengklasifikasikan 48 sebagai positif dan 9 sampel sebagai netral.
- Netral: Model mengklasifikasikan 143 sampel dengan benar, tetapi salah mengklasifikasikan 5 sampel sebagai negatf dan 62 sampel sebagai positif.
- Positif: Model mengklasifikasikan 265 sampel dengan benar, tetapi salah mengklasifikasikan 4 sampel sebagai negatif dan 25 sampel sebagai netral..

e. Confusion Matrix Support Vector Machine



Gambar 21. Confussion Matrix SVM

Gambar 21 menampilkan kinerja Support Vector Machine dalam mengklasifikasikan data ke dalam 3 kelas. Berikut adalah penjelasannya:

- Negatif: Model mengklasifikasikan 85 sampel dengan benar, tetapi salah mengklasifikasikan 19 sampel sebagai netral dan 20 sampel sebagai positif.
- Netral: Model mengklasifikasikan 177 sampel dengan benar, tetapi salah mengklasifikasikan 7 sampel sebagai negatif dan 26 sampel sebagai positif.
- Positif: Model mengklasifikasikan 248 sampel dengan benar, tetapi salah mengklasifikasikan 18 sampel sebagai negatif dan 28 sampel sebagai netral.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa klasifikasi sentimen publik terhadap kebijakan kendaraan listrik dapat dilakukan dengan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. *Data* yang diperoleh dari media sosial Twitter telah melewati berbagai tahap *preprocessing*, pemberian label, pembobotan, dan proses klasifikasi, sehingga memungkinkan identifikasi sentimen ke dalam tiga kategori utama, yaitu positif, netral, dan negatif. *Dataset* yang diperoleh mengenai kebijakan kendaraan listrik menghasilkan distribusi sentimen yang beragam. Dari 3.136 *data* yang dianalisis, sebanyak 1.416 *tweet* menunjukkan sentimen positif, 603 *tweet* mengandung sentimen negatif, 1.117 *tweet* bersentimen netral. Dari hasil evaluasi, *Support Vector Machine* menunjukkan performa yang lebih unggul dibandingkan dengan *Naïve Bayes*, dengan akurasi mencapai 81%, sedangkan *Naïve Bayes* hanya mencapai 76%. Berdasarkan hasil penelitian, SVM lebih direkomendasikan untuk klasifikasi sentimen publik terhadap kebijakan kendaraan listrik, terutama jika tujuan utama adalah memperoleh hasil yang lebih akurat dan mampu menangani kompleksitas data dengan lebih baik. Namun, jika diperlukan metode yang lebih ringan dan cepat dalam proses pelatihan, *Naïve Bayes* tetap dapat menjadi pilihan alternatif.

Untuk penelitian selanjutnya, beberapa saran dapat dipertimbangkan agar hasil yang diperoleh lebih optimal dan akurat. Penelitian ini hanya menggunakan *data* dari media sosial Twitter (X), sehingga disarankan untuk memperluas sumber *data*, seperti komentar di *platform* berita, forum diskusi, atau survei *online*, guna memperoleh perspektif yang lebih luas. Selain itu, peningkatan jumlah *dataset* diperlukan agar model dapat menangkap lebih banyak variasi opini publik. Proses *preprocessing* juga perlu dioptimalkan, terutama pada teknik *stemming* dan *stopword removal*, agar model tidak kehilangan informasi penting dalam teks. Meskipun *Naïve Bayes* dan SVM menunjukkan kinerja yang baik, penelitian mendatang dapat mengeksplorasi algoritma yang lebih kompleks, seperti *Random Forest*, *XGBoost*, atau metode *deep learning* seperti LSTM dan BERT, guna meningkatkan akurasi dan performa model. Selain klasifikasi sentimen, penelitian selanjutnya dapat menganalisis faktor-faktor yang memengaruhi opini publik terhadap kebijakan kendaraan listrik, sehingga memberikan wawasan lebih mendalam mengenai alasan di balik sentimen yang muncul.

DAFTAR PUSTAKA

[1] A. Darmawan Sidik, A. Ansawarman, K. Kunci, J. Kendaraan Bermotor, M. Regresi, and F. Jalan, "Prediksi Jumlah Kendaraan Bermotor Menggunakan Machine Learning," *Formosa Journal of Multidisciplinary Research (FJMR)*, vol. 1, no. 3, pp. 559–568, 2022, doi: 10.55927.

[2] N. Putu Gita Naraswati, D. Cindy Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, and R. Nooraeni, "SISTEMASI: Jurnal Sistem Informasi Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification." 2021, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[3] N. Ananda, "Klasifikasi Sentimen Tweet Masyarakat Terhadap Kendaraan Listrik Menggunakan Support Vector Machine," 2023, doi: 10.32493/informatika.v8i4.36754.

[4] A. M. Siregar, "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *Faktor Exacta*, vol. 16, no. 3, Oct. 2023, doi: 10.30998/faktorexacta.v16i3.16703.

[5] F. Nuriy Alin, M. Hafid Totohendarto, and M. Rafi Muttaqin, "Analisis Sentimen Terhadap Kendaraan Listrik Pada Platform Twitter Menggunakan Metode Naive Bayes," *Informatics for Educators And Professionals : Journal of Informatics*, vol. 8, no. 2, pp. 96–107, 2023.

[6] M. Fauzi Fayyad, D. Takratama Savra, V. Kurniawan, and B. Hilmi Estanto, "Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm," vol. 1, no. 1, pp. 1–9, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/predatecs/article/view/814

[7] A. S. Wicaksono, "TWITTER SENTIMENT ANALYSIS OF ELECTRIC VEHICLE SUBSIDY POLICY USING NAÏVE BAYES ALGORITHM," 2023. [Online]. Available: https://jurnalmipa.unri.ac.id/jsmds

[8] H. R. Farahani, E. Mahdavi, and A. J. Ebrahimi, "Improving public opinion analysis using hybrid model based on SVM and Naïve Bayes," Journal of Artificial Intelligence Research and Applications, vol. 12, no. 3, pp. 123–134, 2023.

[9] J. Elektronika and D. Komputer, "Perbandingan Naïve Bayes dan KNN Dalam Klasifikasi Tweet BBM Subsidi," vol. 16, no. 1, pp. 35–44, 2023, [Online]. Available: https://journal.stekom.ac.id/index.php/elkom■page35

[10] S. Singh and P. Suri, "Sentiment Analysis and Opinion Mining: A Survey of Techniques and Tools," International Journal of Computer Applications, vol. 177, no. 3, pp. 1–7, 2021.

[11] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," 2021