

Analisis Sentimen Terhadap Kebijakan Penggunaan Kendaraan Listrik di Media Youtube menggunakan Metode *Convolutional Neural Network* (CNN)

Riska Amilia Putri¹, Muhsi², Moh Aminullah Hamzah³

¹Teknik Informatika, Universitas Islam Madura

²Teknik Informatika, Fakultas Teknik, Universitas Islam Madura

¹riskaameliaputri2509@gmail.com, ²mail2@mail.org

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan penggunaan kendaraan listrik di Indonesia melalui komentar pada platform youtube. Data dikumpulkan dari video yang membahas kebijakan kendaraan listrik, kemudian dilakukan tahap preprocessing teks seperti pembersihan karakter, normalisasi, tokenisasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah Convolution Neural Network (CNN). Yang terbukti efektif dalam menangani data teks dalam bentuk sekuens. Model CNN dilatih untuk mengklasifikasikan komentar menjadi sentiment positif dan negative, atau netral. Hasil evaluasi menunjukkan 97% bahwa model mampu mengklasifikasikan sentiment dengan tingkat akurasi yang baik, sehingga dapat digunakan untuk memahami respons publik terhadap kebijakan pemerintah. Penelitian ini diharapkan dapat menjadi referensi bagi pengambil kebijakan dalam memahami opini masyarakat serta meningkatkan pendekatan komunikasi publik terkait transisi menuju kendaraan listrik.

Kata kunci: Analisis sentiment kendaran listrik, CNN.

ABSTRACT

This study aims to analyze public sentiment towards the policy of using electric vehicles in Indonesia through comments on the YouTube platform. Data were collected from videos discussing electric vehicle policies, then text preprocessing stages were carried out such as character cleaning, normalization, tokenization. The classification method used in this study is the Convolution Neural Network (CNN). Which has proven effective in handling text data in the form of sequences. The CNN model was trained to classify comments into positive and negative sentiments, or neutral. The evaluation results showed 97% that the model was able to classify sentiment with a good level of accuracy, so that it can be used to understand the public response to government policies. This study is expected to be a reference for policy makers in understanding public opinion and improving the public communication approach related to the transition to electric vehicles.

Keywords: Electric vehicle SentimentAnalysis, CNN.

1. PENDAHULUAN

Di era digitalisasi yang berkembang pesat, masyarakat semakin aktif menyampaikan opini melalui berbagai platform media social. Salah satu media social yang populer adalah youtube. Yang tidak hanya digunakan untuk hiburan, tetapi juga sebagai sarana penyampaian opini

publik terhadap isu-isu aktual, seperti kebijakan pemerintah mengenai kendaraan listrik, masyarakat memanfaatkan fitur komentar di youtube untuk menyuarakan pendapatnya secara terbuka dan langsung dengan besarnya volume data opini ini, di

perlu metode yang efektif untuk memahami sentimen publik[1].

Kendaraan listrik adalah alat transportasi yang dapat menggabungkan manfaat dari segi kesehatan dan ramah lingkungan dengan kenyamanan berkendara yang mirip dengan kendaraan bermotor [2]. Penerbitan peraturan pemerintahan (PP) pada Nomor 55 tahun 2019 lalu mengenai percepatan program kendaraan listrik berbasis baterai untuk angkutan jalan menjadi sorotan utama. Beragam tanggapan dari masyarakat melalui media sosial, Termasuk Youtube [3]. Seiring dengan pertumbuhan ekonomi dan peningkatan mobilitas penduduk, permintaan akan kendaraan bermotor di Indonesia semakin meningkat. Namun, penggunaan kendaraan bermotor yang masih didominasi oleh bahan bakar fosil menyebabkan Indonesia mengalami masalah polusi udara dan kontribusi besar terhadap emisi gas rumah kaca. Untuk mengatasi masalah tersebut, pemerintah Indonesia telah menetapkan target penggunaan kendaraan listrik sebesar 2,2 juta unit pada tahun 2030 sebagai bagian dari komitmen Indonesia dalam Paris Agreement. Kendaraan listrik menjadi salah satu solusi potensial dalam mengurangi dampak lingkungan akibat emisi karbon dari kendaraan berbahan bakar fosil [4]. Namun implementasi kebijakan ini tidak lepas dari berbagai tantangan, termasuk infrastruktur yang belum memadai dan opini publik yang beragam. Sentimen masyarakat terhadap kebijakan ini menjadi indikator penting untuk mengukur efektivitas penerapannya dan merancang kebijakan yang lebih baik di masa depan [5].

Analisis sentiment adalah sebuah metode analisis pada data pada teks untuk mengklasifikasi teks berbentuk opini publik berdasarkan sentiment masyarakat. Analisis sentiment berguna untuk mengklasifikasi teks berdasarkan kalimat atau fitur yang memiliki sifat positif dan negatif. Analisis sentiment melibatkan komputasi dalam memahami sentiment yang tersirat dalam sebuah teks. *Natural Language Processing* (NLP)

merupakan untuk membuat program khusus dalam bidang linguistik yang memungkinkan mesin untuk memahami bahasa manusia. NLP dapat memproses dan menerjemahkan ke Bahasa alami manusia dalam format yang dapat dimengerti oleh mesin [6]. Dalam penelitian ini, pendekatan yang digunakan adalah metode *Convolution Neural Network* (CNN). CNN memiliki keunggulan dalam mengenali pola dalam teks dan telah terbukti efektif untuk klasifikasi sentiment dalam Bahasa Indonesia. Metode ini banyak digunakan karena akurasi nya yang tinggi serta kemampuannya untuk menangani data tidak terstruktur dari media sosial.

Terdapat beberapa penelitian yang telah dilakukan sebelumnya menggunakan *Convolution Neural Network* (CNN). Pada penelitian (George et al, 2022) mengimplementasikan metode CNN untuk analisis sentiment dengan mendapatkan hasil sebesar 97% dimana hasil tersebut termasuk dalam kategori sangat baik. Penelitian yang dilakukan oleh (Dany & Aulia, 2024) yang berjudul “Implementasi algoritma *Convolutional Neural Network* Untuk analisis sentiment bacapres 2024 pada kolom komentar youtube mata najwa” dengan mendapatkan hasil klasifikasi sentiment akurasi sebesar 91%.

Penelitian ini dilakukan karena penting nya pemahaman opini publik terhadap kebijakan penggunaan kendaraan listrik yang memiliki dampak besar terhadap lingkungan dan ekonomi nasional. Dengan memahami sentiment publik melalui CNN, diharapkan dapat memberikan masukan bagi pemerintah dan pemangku kepentingan dalam menyusun kebijakan yang lebih responsif. Dan adaptif. Penelitian ini difokuskan untuk mengklasifikasi sentiment publik menjadi kategori *Positive*, *negative* dan *neutral* berdasarkan komentar-komentar yang terdapat di youtube menggunakan metode CNN.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan data sekunder berupa dengan pendekatan data mining yang di fokuskan pada analisis sentiment terhadap komentar pengguna youtube mengenai kebijakan kendaraan listrik indonesia. Tujuan dari penelitian ini untuk mengetahui reaksi dan persepsi publik terhadap kebijakan tersebut melalui analisis komentar pada video yang relevan. pengumpulan data diperoleh dari platform youtube dengan menggunakan Teknik *web scraping*, dengan kriteria video yang dipilih antara lain membahas isu kendaraan listrik, tahap pemodelan dilakukan menggunakan perangkat lunak google colab berbasis python dan penggunaan library Tensorflow pada model CNN . Untuk memberikan gambaran yang jelas mengenai proses analisis sentimen yang akan dilakukan, penelitian ini membagi tahapan pembangunan model ke dalam beberapa langkah yang terstruktur.



Gambar 1. Tahapan Penelitian

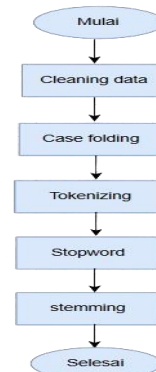
2.1 Crawling Data

Langkah pertama yang dilakukan dalam analisis sentiment adalah *crawling* data atau disebut pengumpulan data. Pada saat ini banyak metode yang bisa dilakukan. Terdapat *crawling* data yang menggunakan pemrograman python untuk mengambil dan metode web Crawler.

Data di kumpulkan dengan memanfaatkan google collabs untuk melakukan *crawling* data komentar youtube dengan pencarian “Kendaraan listrik”. Dibutuhkan youtube Data API V3 dan link URL tentang kendaraan listrik untuk melakukan *crawling*. Youtube API di sediakan oleh Youtube untuk para pengembang untuk membuat application yang dapat berinteraksi dengan resource video-video yang ada di youtube [7].

2.2 Pre-processing

Tujuan dari preprocessing data ini adalah untuk menghilangkan dan mengatasi noise data supaya hasil perhitungannya lebih optimal [8]. Selain itu, preprocessing ini dilakukan pada beberapa tahapan dalam pre-processing yakni sebagai berikut :



Gambar 2. Tahapan penelitian

Penelitian ini melakukan klasifikasi terlebih dahulu menggunakan algoritman CNN pada data training. Setelah itu, hasil dari klasifikasi data latih, diterapkan pada data testing. Setelah data testing terklasifikasi, Selanjutnya dilakukan evaluasi kinerja [9].

2.3 Labelling

Labeling adalah proses memberikan kategori sentimen seperti positif, negatif, atau netral pada data, yang penting untuk melatih model CNN agar dapat mengenali pola sentimen. Keakuratan labeling sangat berpengaruh terhadap kinerja model, sebagaimana dibuktikan oleh penelitian Susanti et al. (2023) yang menunjukkan peningkatan akurasi CNN dalam klasifikasi data dengan label yang jelas [10].

2.4 Pembobotan

Tahapan selanjutnya adalah pembobotan dalam penelitian. Pembobotan ini menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF ini digunakan untuk melakukan pembobotan

fitur dari data teks. TF-IDF adalah Teknik statistic yang men[11]. gukur pentingnya sebuah kata dalam sebuah dokumen relatif terhadap seluruh korpus. Pembobotan ini membantu memfokuskan model hanya pada kata-kata yang relevan dengan isi dokumen. Berikut adalah rumus TF-IDF:

$$\text{TF}(t,d) = \frac{\text{number of times } t \text{ appears in } d}{\text{Total number of terms in } d}$$

$$\text{IDF}(t) = \log \frac{N}{1+df} \quad (1)$$

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) * \text{IDF}(t)$$

Dimana:

TFidFT = bobot dari term t

Ftd = frekuensi munculnya termt pada dokumen d

N = jumlah kumpul dokumen

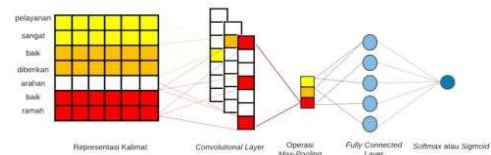
Dft = jumlah dokumen yang mengandung term t

CNN akan mempelajari pola-pola dari representasi vector ini untuk melakukan klasifikasi sentiment positif, negative, atau netral, Penggunaan TF-IDF sebagai metode ekstrasi fitur sebelum CNN telah terbukti efektif dalam meningkatkan performa analisis sentiment, sebagaimana diterapkan dalam berbagai penelitian yang menggunakan pendekatan serupa untuk mengolah teks sebelum klasifikasi [12].

2.5 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu algoritma *deep learning* yang cara kerjanya melakukan training data pada kumpulan data yang besar. Metode ini sudah tidak asing di ranah deep learning karena metode ini memiliki performa yang lebih baik dalam menganalisis sentiment [13]. Metode ini juga memiliki kelebihan daripada dengan metode Deep Learning lainnya, CNN memiliki lebih sedikit koneksi, parameter, dan mudah untuk dilakukan proses training [14]. Berikut

ini adalah gambar arsitektur *Convolution Neural Network* :



Gambar 3. Arsitektur Convolutional Neural Network

Rumus perhitungan CNN untuk analisis sentiment :

1. Embedding

Embedding layer mengubah kata dalam teks menjadi vektor angka berdimensi tetap yang bermakna secara semantik, memungkinkan CNN memahami konteks dan emosi, lebih efektif dibanding metode seperti TF-IDF.

Embedding layer memiliki rumus :

$$E(w_i)=v_i \in \mathbb{R}^d \quad (2)$$

Rumus ini menunjukkan bahwa kata w_i direpresentasikan sebagai vektor v_i berdimensi d melalui *embedding layer*.

2. Convolutional layer

Convolutional layer pada CNN berfungsi mengekstraksi pola penting seperti n-gram emosional dari vektor teks, sehingga membantu mengidentifikasi ciri khas sentimen positif, negatif, atau netral dalam komentar.

Rumus konvolusi umum adalah :

$$y_i = f\left(\sum_{j=0}^{k-1} w_j \cdot x_i + j + b\right) \quad (3)$$

Rumus tersebut menjelaskan cara CNN menghitung fitur lokal dengan mengalikan input dan filter, menjumlahkannya, menambahkan bias, lalu melewatkannya ke fungsi aktivasi untuk menghasilkan output y_i .

3. Max pooling

Max Pooling adalah proses dalam CNN yang digunakan untuk mereduksi dimensi data sambil mempertahankan fitur penting dari hasil konvolusi.

Rumus max pooling adalah :

$$z = \max(y_1, y_2, \dots, y_n) \quad (4)$$

Rumus tersebut menunjukkan bahwa max pooling mengambil nilai terbesar dari sekelompok output, tujuannya untuk menyederhanakan data dan mempertahankan informasi paling kuat.

4 . *Fully Connected layer*

Fully Connected Layer adalah lapisan terakhir dalam CNN yang menghubungkan semua neuron dari lapisan sebelumnya untuk menghasilkan output akhir.

Rumus dari fully connected layer adalah :

$$h = f(Wx + b) \quad (5)$$

5. *Dropout layer*

Dropout layer dalam CNN adalah teknik untuk mencegah overfitting dengan mengabaikan secara acak sejumlah neuron saat pelatihan, sehingga model belajar representasi yang lebih umum dan tidak bergantung pada neuron tertentu.

Berikut ini adalah rumus Dropout layer :

$$\hat{h}_i = r_i \cdot h_i \text{ dengan } r_i \sim \text{Bernoulli}(p) \quad (6)$$

menunjukkan bahwa dropout akan menonaktifkan neuron h_i secara acak selama training, dengan r_i bernilai 0 atau 1 berdasarkan distribusi Bernoulli. Ini dilakukan untuk mencegah overfitting dan membuat model lebih robust.

6. *Softmax*

Softmax adalah fungsi aktivasi yang digunakan di output layer CNN untuk mengubah nilai logits menjadi probabilitas antar kelas. Fungsi ini memastikan total

probabilitas semua kelas berjumlah 1, sehingga cocok untuk tugas klasifikasi.

Proses *softmax* dilakukan perhitungan pada nilai probabilitas dari hasil *Max pooling* yang matriksnya digabung dengan persamaan berikut:

$$P(y = c | x) = \frac{e^{z_c}}{\sum_{i=1}^C e^{z_i}} \quad (7)$$

fungsi softmax yang mengubah output CNN menjadi probabilitas. Nilai z_c adalah skor untuk kelas c , dan rumus ini memastikan semua skor diubah menjadi nilai antara 0 dan 1, serta totalnya menjadi 1. Biasanya digunakan pada output layer untuk klasifikasi multi-kelas.

2.6 Pengujian

Proses pengujian ini dilakukan untuk mengevaluasi kinerja metode *Convolutional Neural Network* (CNN) dalam mengklasifikasikan sentimen komentar masyarakat terhadap kebijakan penggunaan kendaraan listrik yang diekspresikan di platform youtube. Pengujian ini menggunakan confusion metrik seperti *accuracy*, *presisi*, *recall*, dan *F1-score* yang penting untuk mengukur sentimen nya dalam sisi positif, negatif, dan netral. Berikut ini adalah rumus *confusion matrix* :

Accuracy :

Mengukur seberapa banyak prediksi model yang benar dari keseluruhan data uji.

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar}}{\text{Total Data}}$$

Precision :

Mengukur proporsi prediksi positif yang benar-benar relevan.

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

Recall:

Mengukur kemampuan model untuk menemukan semua instance relevan dari suatu kategori.

Recall= $\frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$

True Positive (TP) + False Negative (FN)

F1-Score :

Rata-rata harmonis antara precision dan recall, memberikan keseimbangan antara keduanya

F-Measure= $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis sentimen kendaraan listrik di media youtube menggunakan metode Convolutional Neural Network (CNN). Model ini di evaluasi berdasarkan confusion matrix, accuracy, Precision, Recall, dan F1-Score. Data komentar kendaraan listrik, proses pengambilan nya dengan cara crawling data di media social yaitu youtube.

3.1 Data Kendaraan listrik

Pengumpulan Data dalam penelitian ini di lakukan dengan crawling data. Data di kumpulkan dengan memanfaatkan *Google Colaboratory* untuk melakukan crawling data komentar youtube melalui channel motor listrik.com [15]. Di butuhkan API V3 dan link URL channel youtube motor listrik.com untuk melakukan crawling. data kendaraan listrik di dapatkan 3000 data. Kemudian data tersebut di olah, sehingga didapatkan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*. *Crawling data* yang di hasilkan dapat terlihat pada tabel di bawah ini:

Tabel 1. Hasil crawling data

No	publishedAt	authorDisplayName	textDisplay
1	2025-06-23T10:17:21Z	@oeibenho7164	Sampai kapan promo ini?
2	2025-06-23T10:24:55Z	@motorlistrikcom	Selama pameran PRJ s/d 13 juli 2025
3	2025-06-23T09:23:36Z	@Kaisami2	Bener gak sich klo yadea sparepart nya susah termasuk bannya
4	2025-06-23T11:09:40Z	@Dablong90	Aman om.kita udh 2 tahun pakeo Y,
5	2025-06-23T06:49:06Z	@djuraganalnaz2532	gk recommended motor listrik pake sla atau aki.. baik itu graphene atau sla biasa
6	2025-06-23T05:17:01Z	@MustaqimMustaqim-b5q3o	Ombu kmn bang kok gk ada d yadea
7	2025-06-23T06:42:22Z	@motorlistrikcom	Lagi ke eropah kali bang
8	2025-06-23T05:00:31Z	@cakemet3056	Yadea Velax dapat potongan 7jt gak nich? Kl ada potongan lgsg meluncur.....serius oY™

9	2025-06-23T06:42:38Z	@motorlistrikcom	Gass
10	2025-06-23T02:53:40Z	@dimaschannel4843	Beli yadea vexel harga 22 jta, dapet hadiah sepeda listrik apa?
11	2025-06-23T03:20:07Z	@Dablong90	Gc 10.harga 5.900
12	2025-06-22T15:24:08Z	@henriardyanzah9212	Pendek banget 30km minimal 60 km lah
13	2025-06-23T02:57:53Z	@bangajieh	Kan sepeda, buat apa jauh2, oY™, oY™
14	2025-06-23T03:29:18Z	@henriardyanzah9212	Lu pikir sendiri lah bang layak gak? @bangajieh
15	2025-06-23T03:33:27Z	@bangajieh	Layak donk, kan sepeda listrik, bukan motor listrik oY™, a? @henriardyanzah9212
16			
3000	2025-05-11T05:30:55Z	@linguahumanistic	Gkk papa om. Molis setahun pakai jg gak pernah minta servis ini saya pakai. 1 thn motor bensin butuh berapa x servis. Motor bensin harus bnyak tmpat servis, karena baru beli aja tiap bulan wajib servis. Jgn mutlak ditelan mentah2 ya

3.2 Hasil Pre-processing

Tahapan pre-processing ini menggunakan lima langkah, yaitu clean data, Normalisasi, Stopword removal, stemming, tokenisasi, dan case folding. Untuk output yang dihasilkan dapat terlihat pada tabel di bawah ini:

1. Cleaning

Digunakan untuk Menghapus karakter-karakter yang tidak perlu seperti angka, tanda baca, simbol, emotikon, HTML tag, dan URL.

Tabel 2. Hasil cleaning

No	Text_Display	Cleaning
1.	Sampai kapan promo ini?	sampai kapan promo ini
2.	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya	banyak belajar yaa bos harus lebih dedail lagi kalo nge review bos. misalnya dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya
3.	Siap. Makasih om...	siap makasih om

2. Case folding

Digunakan untuk Mengubah semua huruf menjadi huruf kecil (lowercase).

Tabel 3. Hasil Case folding

No	Text_Display	Casefolding
1	Sampai kapan promo ini?	sampai kapan promo ini?
2	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya	banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya

3	Siap. Makasih om..	siap. makasih om..
---	--------------------	--------------------

3. Tokenizing

Digunakan untuk Memecah kalimat atau paragraf menjadi potongan-potongan kata atau unit dasar (token). Contoh : sampai kapan promo ini → ['sampai', 'kapan', 'promo', 'ini']. Dengan tujuan agar bisa di analisis lebih mudah dan di gunakan dalam model pembelajaran mesin atau deep learning seperti CNN.

Tabel 4. Hasil Tokenizing

	Text_Display	tokenizing
1	Sampai kapan promo ini?	['sampai', 'kapan', 'promo', 'ini']
2	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya -dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya	['banyak', 'belajar', 'yaa', 'bosbrharus', 'lebih', 'dedail', 'lagi', 'kalo', 'nge', 'review', 'bos', 'misalnya', 'brdinamanya', 'berapa', 'wattbrnge', 'cas', 'nya', 'berapa', 'kwh', 'yg', 'dibutuhkan', 'da', 'lainnya']
3	Siap. Makasih om...	['siap', 'makasih', 'om']

4. Stopword removal

Stopword removal dalam analisis sentimen digunakan untuk membersihkan teks dari kata-kata umum yang tidak memiliki makna penting atau relevan terhadap analisis sentiment. seperti: *banyak, lebih, lagi* dan sebagainya.

Tabel 5. Hasil Stopword removal

No	Text_Display	stopword_removal
1	Sampai kapan promo ini?	sampai kapan promo
2	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya -dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya	belajar yaa bosbrharus dedail kalo nge review bos misalnya brdinamanya berapa wattbrnge cas nya berapa kwh yg dibutuhkan da lainnya
3	Siap. Makasih om..	siap makasih om

5. Stemming

Stemming Dalam analisis sentimen digunakan untuk mengubah kata ke bentuk dasarnya (akar kata) agar kata-kata yang memiliki makna sama tetapi bentuk berbeda bisa dikenali sebagai satu entitas.

Tabel 6. Hasil Stemming

No	Text_Display	stemming
1	Sampai kapan promo ini?	sampai kapan promo
2	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya -dinamanya berapa watt -nge cas nya berapa	ajar yaa bosbrharus dedail kalo nge review bos misal brdinamanya berapa wattbrnge cas nya berapa kwh yg butuh da lain
3	Siap. Makasih om..	siap makasih om

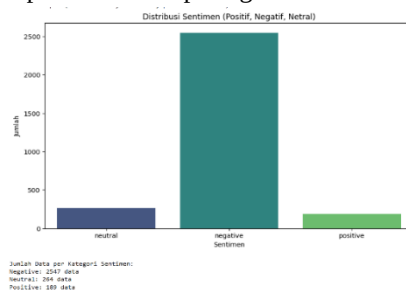
3.3 Hasil labelling

Setelah melakukan *pre-processing* atau pembersihan. Maka akan melakukan tahapan labelling. Untuk mengetahui seberapa banyak sentiment pada positif, negative, dan Neutral. Hasil pelabellan pada penelitian ini dapat terlihat pada di bawah ini :

Tabel 7. Hasil labelling

No	Clean Text	Sentiment
1	Sampai kapan promo ini?	Neutral
2	Banyak belajar yaa bos... harus lebih dedail lagi kalo nge review bos. misalnya -dinamanya berapa watt -nge cas nya berapa kwh yg dibutuhkan da lain2nya	Negative
3	Siap. Makasih om..	Positive

Pada Analisa perlu diketahui persentase terkait komentar atau opini dari masyarakat. Didapatkan hasil seperti gambar di bawah ini.



Gambar 4. Diagram labelling

Dari gambar di atas di simpulkan bahwa komentar atau opini masyarakat tentang Kendaraan listrik Negative masih yang paling tinggi disbanding dengan netral dan positive.

3.4 Hasil Pembobotan

Metode ini mengukur pentingnya kata dalam dokumen dengan memperhitungkan frekuensi kemunculannya dalam dokumen tersebut serta beberapa sering kata itu muncul diseluruh koleksi dokumen. Kata yang muncul sering

dalam dokumen tetapi jarang di dokumen lain akan diberi bobot tinggi, karena di anggap lebih relevan dan informatif. Teknik ini membantu dalam menyoroti kata-kata kunci yang benar-bener menggambarkan isi dokumen.

1. Pembobotan Term Frequency (TF)

Pembobotan *Term Frequency* mengukur seberapa sering sebuah kata muncul dalam dokumen. Kata yang muncul lebih sering di anggap lebih penting dalam konteks dokumen tersebut dan di beri bobot yang lebih tinggi.

Term	TF
ada	0.07142857142857142
nomor	0.07142857142857142
salesnya	0.07142857142857142
gak	0.07142857142857142
om	0.07142857142857142
saya	0.07142857142857142
mau	0.07142857142857142
tanya	0.07142857142857142
tentang	0.07142857142857142
promo	0.07142857142857142
prjmau	0.07142857142857142
beli	0.07142857142857142
roda	0.07142857142857142
🚗	0.07142857142857142

Gambar 5. Hasil pembobotan TF

Gambar tersebut menunjukkan daftar kata dengan nilai TF. Setiap kata memiliki nilai sebesar 0.0714, yang berarti setiap kata muncul dengan frekuensi yang sama dalam dokumen tersebut. Ini merupakan bagian dari proses pra-pemrosesan data untuk analisis sentiment, dimana TF digunakan untuk mengukur seberapa sering suatu kata muncul.

2. Pembobotan Tf-idf

term	TF	TF-IDF
ada	0.07142857142857142	0.15616318837842033
nomor	0.07142857142857142	0.36951101597100217
salesnya	0.07142857142857142	0.4728623718950254
gak	0.07142857142857142	0.2001686343266751
om	0.07142857142857142	0.11701288437855388
saya	0.07142857142857142	0.22757757157465783
mau	0.07142857142857142	0.23295967423478547
tanya	0.07142857142857142	0.32894072757057796
tentang	0.07142857142857142	0.4728623718950254
promo	0.07142857142857142	0.39439006556158895
prjmau	0.07142857142857142	0.5223728847921644
beli	0.07142857142857142	0.21117939717147927
roda	0.07142857142857142	0.38337930271678483
🚗	0.07142857142857142	0.36542827212529155

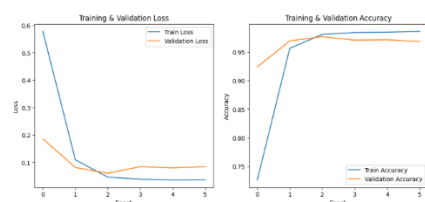
Gambar 6. Hasil pembobotan TF-IDF

Tabel tersebut menunjukkan bahwa meskipun setiap kata memiliki frekuensi yang sama (TF) sebesar 0.0714, nilai TF-IDF-nya berbeda-beda. Hal ini disebabkan karena TF-IDF tidak hanya melihat frekuensi kata dalam satu dokumen, tetapi juga mempertimbangkan

seberapa sering kata tersebut muncul di seluruh korpus dokumen. Kata seperti “prjmau”, “salesnya”, dan “tentang” memiliki nilai TF-IDF yang tinggi karena dianggap lebih spesifik atau jarang muncul di dokumen lain, sehingga lebih penting untuk analisis. Sementara itu, kata-kata umum seperti “ada”, “om”, dan “gak” memiliki nilai TF-IDF lebih rendah karena sering muncul di berbagai dokumen, sehingga kurang memberikan nilai informatif. TF-IDF digunakan untuk menekankan kata-kata yang paling relevan dalam proses analisis teks seperti klasifikasi atau analisis sentimen.

3.5 Hasil pemodelan CNN

Setelah data komentar YouTube dikumpulkan dan diproses melalui tahap preprocessing (cleaning, normalisasi, tokenisasi, stopword removal, dan stemming), data dibagi menjadi 80% data latih dan 20% data uji. Model CNN dibangun dengan tahapan sebagai berikut: pertama, *Embedding layer* mengubah kata menjadi vektor numerik bermakna; kemudian, *Convolutional layer* mengekstraksi fitur penting dari teks; dilanjutkan dengan *Max Pooling* untuk mereduksi dimensi dan mempertahankan fitur dominan. Setelah itu, *Dropout layer* digunakan untuk mencegah overfitting, dan diakhiri dengan *Fully Connected layer* serta Softmax untuk menghasilkan klasifikasi ke dalam tiga kategori sentimen (positif, negatif, netral). Model dilatih selama beberapa epoch, dengan hasil training loss menurun stabil, dan akurasi validasi mencapai 97. Evaluasi menggunakan confusion matrix menunjukkan performa sangat baik dengan akurasi 97,71%, precision, recall, dan F1-score masing-masing 0,9769, menandakan model mampu mengenali pola sentimen secara akurat dan seimbang.



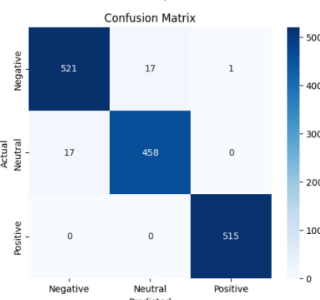
Gambar 7. Hasil gambar loss

Grafik menunjukkan *training loss* turun drastis dan stabil, sedangkan *validation loss* rendah dengan sedikit fluktuasi. *Training accuracy* mencapai hampir 99% dan *validation accuracy* stabil di kisaran 96–97%. Hasil ini menandakan model bekerja sangat baik dengan generalisasi yang baik tanpa indikasi *overfitting*.

Tabel 8. Hasil accuracy

Metrik	Nilai
Accuracy	97.71%
Precision	0.9769
Recall	0.9769
F1-Score	0.9769

Tabel ini menunjukkan bahwa model CNN mencapai kinerja sangat tinggi dalam analisis sentiment. Nilai-nilai ini menunjukkan bahwa model mampu mengklasifikasikan sentiment secara konsisten dan akurat, serta berhasil menangkap hampir semua komentar yang relevan dengan sangat baik tanpa banyak kesalahan klasifikasi.



Gambar 8. Hasil confusion matrix

Gambar di atas menunjukkan *confusion matrix* hasil evaluasi model terhadap tiga kelas sentimen: Negative, Neutral, dan Positive. Model menunjukkan performa yang sangat baik, dengan prediksi benar sebanyak 521 untuk kelas Negative, 458 untuk kelas Neutral, dan 515 untuk kelas Positive. Kesalahan klasifikasi terjadi pada kelas Negative, yaitu 17 diklasifikasikan sebagai Neutral dan 1 sebagai Positive. Pada kelas Neutral, sebanyak 17 data diklasifikasikan salah sebagai Negative. Sementara itu, kelas Positive diklasifikasikan

dengan akurasi 100% tanpa kesalahan. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan ketiga kelas sentimen tersebut.

4. KESIMPULAN

Penelitian analisis sentiment pada komentar youtube terkait kebijakan kendaraan listrik di Indonesia menggunakan metode *Convolution Neural network* (CNN) menunjukkan bahwa model mampu mengklasifikasikan sentiment positif, negatif, dan netral dengan akurasi 97%. Hasil ini membuktikan bahwa CNN cukup efektif dalam menangkap pola Bahasa masyarakat terhadap isu kebijakan publik. Kesimpulannya, CNN merupakan metode yang baik untuk analisis sentiment di media social. Namun, kualitas pelabelan manual dan tahapan preprocessing seperti penghilangan stopwords dan normalisasi kata perlu di perhatikan untuk meningkatkan akurasi. Penelitian lanjutan disarankan untuk membandingkan CNN dengan model lain seperti LSTM atau transformer guna memperoleh hasil yang lebih optimal.

UCAPAN TERIMA KASIH

Bagian ini berisi ucapan terima kasih kepada bapak Muhsi, MT dan bapak Moh Aminullah Hamzah, M.Kom selaku dosen pembimbing saya, dan kepada kedua orangtua atas segala doa, pesan dan harapannya yang membuat saya termotivasi untuk tetap semangat menyelesaikan Tugas Akhir ini.

DAFTAR PUSTAKA

- [1] Diaz Tiyasya Putra and Erwin Budi Setiawan, "Sentiment Analysis on Social Media with Glove Using Combination CNN and RoBERTa," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 457–563, 2023, doi: 10.29207/resti.v7i3.4892.
- [2] D. D. Suranto, S. Anwar, M. Nuruddin, A. Rofi'i, and A. T. Zain, "Analisa Perancangan dan Pengujian Kendaraan

- Listrik Roda Dua dengan Variasi Pembebanan,” *J-Proteksion J. Kaji. Ilm. dan Teknol. Tek. Mesin*, vol. 7, no. 2, pp. 47–53, 2023, doi: 10.32528/jp.v7i2.8927.
- [3] G. H. Kusuma, I. Permana, F. N. Salisah, M. Afdal, M. Jazman, and A. Marsal, “Pendekatan Machine Learning: Analisis Sentimen Masyarakat Terhadap Kendaraan Listrik Pada Sosial Media X,” *JUSIFO (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 65–76, 2023, doi: 10.19109/jusifo.v9i2.21354.
- [4] H. Dian Fatikno and O. Wicaksono Adi, “Evaluasi Persebaran Infrastruktur Pengisian Kendaraan Listrik Di Indonesia Dengan Memperhatikan Indeks Kualitas Udara Untuk Mendorong Penjualan Kendaraan Listrik,” *J. Ilmu Manaj.*, vol. 8, no. 2, pp. 119–128, 2023, doi: 10.33474/jimmu.v8i2.19667.
- [5] S. Fariz Ramadhan, M. Sa’dan Dardiri, R. Agusta Arifin, and B. Mardikawati, “Ramah Lingkungan Sebagai Alasan Tertinggi Masyarakat Berminat Pada Kendaraan Konversi Hasil Kuesioner,” *Berk. Forum Stud. Transp. antar Perguru. Tinggi*, vol. 2, no. 1, pp. 22–29, 2024, doi: 10.19184/berkalafstpt.v2i1.913.
- [6] B. A. Yuniarossy, K. M. Hindrayani, and A. T. Damaliana, “Analisis Sentimen Terhadap Isu Feminisme Di Twitter Menggunakan Model Convolutional Neural Network (Cnn),” *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 1, pp. 477–491, 2024, doi: 10.46306/lb.v5i1.585.
- [7] D. E. Saputra and A. R. Isnain, “IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK UNTUK ANALISIS SENTIMEN BACAPRES 2024 PADA KOLOM KOMENTAR YOUTUBE MATA NAJWA,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1431–1441, Aug. 2024, doi: 10.29100/jupi.v9i3.5420.
- [8] D. Rifaldi, Abdul Fadlil, and Herman, “Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet ‘Mental Health,’” *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.
- [9] S. Abrori and Z. Fatah, “Gudang Jurnal Multidisiplin Ilmu Implementasi Metode CNN Untuk Klasifikasi Status Stunting Pada Balita,” vol. 2, pp. 380–385, 2024.
- [10] L. A. Susanti, A. M. Soleh, and B. Sartono, “Deep Learning Image Classification Rontgen Dada pada Kasus Covid-19 Menggunakan Algoritma Convolutional Neural Network,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 5, pp. 973–982, 2023, doi: 10.25126/jtiik.20231057142.
- [11] A. M. Priyatno and L. Ningsih, “TF-IDF Weighting to Detect Spammer Accounts on Twitter based on Tweets and Retweet Representation of Tweets,” *Sistemasi*, vol. 11, no. 3, p. 614, 2022, doi: 10.32520/stmsi.v11i3.1995.
- [12] Y. Yuliska, D. H. Qudsi, J. H. Lubis, K. U. Syaliman, and N. F. Najwa, “Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, p. 1067, 2021, doi: 10.25126/jtiik.2021854842.
- [13] A. Permana Putra and A. Farrah Syafira, “Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn),” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 20, pp. 36–41, 2023.
- [14] C. M. Bachri and W. Gunawan, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Deteksi Email Spam menggunakan Algoritma Convolutional Neural Network (CNN),” *Edukasi dan Penelit. Inform.*, vol. 10, no. 1, pp. 88–94, 2024.
- [15] M. Listrik.com, “motor Listrik.com,” Kendaraan listrik. [Online]. Available: <https://youtube.com/@motorlistrikcom?si=IyLzcdwhbQVujpg0>

