

**Penggunaan Deep Learning untuk Mengklasifikasi Hate speech
dan Good Speech Terhadap Pertamina di Platform Twitter
dengan Metode Convolutional Neural Network (CNN)**

SKRIPSI

DISUSUN OLEH

HASAN A. SITUMORANG
NPM. 2109020151



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2025

**Penggunaan Deep Learning untuk Mengklasifikasi Hate speech
dan Good Speech Terhadap Pertamina di Platform Twitter
dengan Metode Convolutional Neural Network (CNN)**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

HASAN A. SITUMORANG
NPM. 2109020151

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

LEMBAR PENGESAHAN

Judul Skripsi : PENGGUNAAN DEEP LEARNING UNTUK
MENGKLASIFIKASI HATE SPEECH DAN
GOOD SPEECH TERHADAP PERTAMINA DI
PLATFORM TWITTER DENGAN METODE
CONVOLUTIONAL NEURAL NETWORK (CNN)

Nama Mahasiswa : Hasan A. Situmorang
NPM : 2109020151
Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing

(Martiano S.Pd, S.Kom., M.Kom)

NIDN. 0128029302

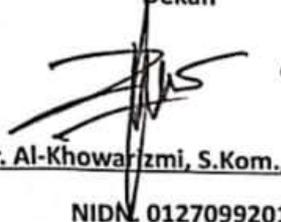
Ketua Program Studi



Fatma Sari Hutagalung, S.Kom., M.Kom.

NIDN. 0117019301

Dekan



Dr. Al-Khowarizmi, S.Kom., M.Kom.

NIDN. 0127099201

PERNYATAAN ORISINALITAS

Penggunaan Deep Learning untuk Mengklasifikasi Hate speech dan Good Speech Terhadap Pertamina di Platform Twitter dengan Metode Convolutional Neural Network (CNN)

SKRIPSI

Saya menyatakan bahwa karya tulis ini hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 16 september 2025
Yang Membuat Pernyataan



Hasan A. Situmorang
NPM. 2109020151

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademi Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan di bawah ini:

Nama : Hasan A. Situmorang
NPM : 2109020151
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi oengembangan ilmu pengetahuan, menyetujui untuk membetikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif ini, (*Non-Exclusive Royalty Free Right*) atas penelitian skripsi yang berjudul:

**Penggunaan Deep Learning untuk Mengklasifikasi Hate speech dan
Good Speech Terhadap Pertamina di Platform Twitter dengan
Metode Convolutional Neural Network (CNN)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih mediah, memformat, mengelola dalam bentuk database, merawat dan mempulikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau pemilik hak cipta.

Demikian Pernyataan ini dibuat dengan sebenarnya.

Medan, 16 september 2025
Yang Membuat Pernyataan



Hasan A. Situmorang
NPM. 2109020151

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : HASAN A. SITUMORANG
Tempat dan Tanggal Lahir : SIPITU HUTA , 29 JANUARI 2003
Alamat Rumah : Lumban Situmorang
Telepon/Faks/HP : 082276122669
E-mail : Situmoranghasan@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN 173416 POLLUNG TAMAT: 2015
SMP : MTsN DOLOKSANGGUL TAMAT: 2018
SMA : SMA NEGERI 1 POLLUNG TAMAT: 2021

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji syukur Penulis ucapkan kepada Allah SWT atas rahmat-Nya Penulis dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua yang telah memberikan dukungan moril maupun materil kepada Penulis. Dan Penulis menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak sejak masa perkuliahan hingga masa penyusunan skripsi, sulit untuk menyelesaikan skripsi ini. Oleh karena itu dengan segala kerendahan hati, Penulis ingin menyampaikan terima kasih dan rasa hormat kepada:

1. Bapak Prof. Dr. Agussani, M.AP., selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Halim Maulana, S.T., M.Kom., selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi.
4. Bapak Dr. Lutfi Basit, S.Sos., M.I.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi.
5. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom., selaku Ketua Program Studi Teknologi Informasi.

6. Bapak Mhd. Basri, S.Si., M.Kom., selaku Sekretaris Program Studi Teknologi Informasi.
7. Bapak Martiano, S.Pd, S.Kom., M.Kom., selaku Dosen Pembimbing yang telah banyak membantu penulis baik dukungan, saran, dan bimbingan dalam penyusunan penelitian ini.
8. Bapak Fiter Situmorang dan Alm. Ibu Raya Lumban Gaol, S.Pd, selaku Kedua Orang Tua penulis yang selalu mendoakan dan memberi dukungan baik mental maupun finansial selama ini.
9. Terimakasih juga kepada Safryal Situmorang S.Pd, Vinny M. Situmorang AMd.Keb., Rahmad R. Situmorang S.T., Iyen H. Situmorang S.I.Kom, Istiq H. Situmorang, Fitri O. Situmorang, selaku saudara kandung yang berpartisipasi dari segi moral dan dalam menyupport finansial.
10. Terima kasih sebesar-besarnya untuk seluruh teman dan rekan seperjuangan. Dukungan dan masukan dari kalian semua sangat berarti hingga penulis bisa sampai di titik ini.
11. Kepada diri sendiri yang sudah berusaha keras dan berjuang sejauh ini. Mampu mengendalikan diri dari berbagai tekanan di luar keadaan dan tak pernah menyerah sesulit apapun. Mampu menguatkan dan meyakinkan bahwa semuanya akan selesai pada waktunya.

Medan, 4 Agustus 2025

Penulis
Hasan A. Situmorang

Penggunaan *Deep Learning* untuk Mengklasifikasi *Hate speech* dan *Good Speech* Terhadap Pertamina di Platform Twitter dengan Metode *Convolutional Neural Network (CNN)*

ABSTRAK

Perkembangan teknologi digital telah mengubah cara masyarakat menyampaikan pendapat, salah satunya melalui media sosial Twitter. Dalam konteks sosial dan korporat, Pertamina sebagai BUMN di sektor energi kerap menjadi sasaran opini publik, baik positif (ujaran baik) maupun negatif (ujaran kebencian). Untuk mengelola informasi tersebut, dibutuhkan sistem yang mampu mengklasifikasikan twit secara otomatis dan akurat. Penelitian ini bertujuan untuk membangun model klasifikasi teks berbasis Deep Learning, khususnya menggunakan metode Convolutional Neural Network (CNN), untuk mengidentifikasi twit yang memuat ujaran kebencian dan ujaran positif terhadap Pertamina. Data diambil dari platform Twitter dengan kata kunci terkait, kemudian melalui proses preprocessing dan pelabelan manual. Dataset yang telah dibersihkan kemudian dibagi menjadi data latih dan data uji untuk dilatih dengan arsitektur CNN. Hasil penelitian menunjukkan bahwa model CNN memiliki kinerja klasifikasi yang sangat baik, dengan akurasi validasi mencapai 98,08% dan akurasi pengujian sebesar 97,66%. Evaluasi dengan confusion matrix menunjukkan presisi dan recall yang tinggi, serta f1-score mencapai 0,98. Temuan ini menunjukkan bahwa metode CNN mampu mengidentifikasi ujaran kebencian dan ujaran positif secara akurat dalam data teks berbahasa Indonesia, terutama terkait isu-isu terkait Pertamina. Penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem pemantauan media sosial dan manajemen opini publik otomatis.

Kata Kunci: Deep Learning , CNN, Hate Speech, Good Speech, Twitter, Pertamina, Klasifikasi Teks

**Using Deep Learning to Classify Hate Speech and Good Speech Against
Pertamina on the Twitter Platform Using the Convolutional Neural Network
(CNN) Method**

ABSTRACT

Advances in digital technology have changed the way people express their opinions, particularly through social media platforms like Twitter. In social and corporate contexts, Pertamina, a state-owned energy company in Indonesia, is frequently the subject of public discourse, both in the form of positive (good speech) and negative (hate speech) expressions. To manage this information, a system capable of accurately and automatically classifying tweets is crucial. This study aims to develop a Deep Learning-based text classification model, specifically using the Convolutional Neural Network (CNN) method, to identify tweets containing hate speech and good speech related to Pertamina. Data was collected from Twitter using relevant keywords, followed by manual preprocessing and labeling. The cleaned dataset was then divided into training and testing data for processing using a CNN architecture. The results showed that the CNN model performed very well in the classification task, achieving a validation accuracy of 98.08% and a testing accuracy of 97.66%. Evaluation using a confusion matrix also showed high precision and recall values, with an f1 score of 0.98. These findings demonstrate that the CNN method is effective in accurately identifying hate speech and positive speech in Indonesian text data, particularly regarding issues related to Pertamina. This research is expected to contribute to the development of automated social media monitoring systems and public opinion management tools..

Keywords: Deep Learning, CNN, Hate Speech, Good Speech, Twitter, Pertamina, Text Classification

DAFTAR ISI

LEMBAR PENGESAHAN.....	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP.....	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
BAB II.....	6
LANDASAN TEORI.....	6
Tabel 2.1 Penelitian Terdahulu.....	6
2.2. Media Sosial Twitter.....	9
2.3. Hate Speech dan Good Speech.....	9
2.4. Deep Learning.....	11
2.5. Convolutional Neural Network.....	12
2.6. Python.....	13
2.7. Text Classification.....	13
BAB III.....	15
METODOLOGI PENELITIAN.....	15
3.1. Metode Penelitian.....	15
3.2. Tahap Penelitian.....	16
3.3 Pengumpulan Data.....	17

3.4 Alur Pengolahan Data.....	18
BAB IV	22
HASIL DAN PEMBAHASAN.....	22
4.1. Hasil Pengumpulan Data	22
4.2. Import Library.....	23
4.3. Import Dataset.....	24
4.4. Preprocessing & Labelling.....	24
4.5. Tokenizer dan Padding	27
4.6. Split Data (Training dan Validasi).....	28
4.7. Training Model CNN.....	29
4.8. Evaluasi Model di Test Set	32
4.9. Classification Report & Confusion Matrix.....	33
4.10 Plot Akurasi dan Loss	34
4.11 Hasil Klasifikasi.....	37
BAB V	38
PENUTUP	38
5.1 Kesimpulan	38
5.2 Saran	39
DAFTAR PUSTAKA	40

DAFTAR TABEL

Table 2.1 Peneliti Terdahulu	6
------------------------------------	---

DAFTAR GAMBAR

3.1 Gambar Tahapan Penelitian.....	17
3.2 Gambar Diagram Pengolahan Data	21
4.1 Gambar Hasil Pengumpulan Data	22
4.2 Gambar Hasil Akurasi dan Loss	36
4.3 Gambar HasilKlasifikasi	37

BAB I

PENDAHULUAN

1.1 Latar Belakang

Platform media sosial dikenal karena berbagai kelebihannya, termasuk memfasilitasi komunikasi antar individu, memungkinkan pertukaran informasi, dan mendorong pengembangan bisnis. Twitter, salah satu platform media sosial yang populer, telah mengalami peningkatan signifikan dalam jumlah pengguna aktif dan telah menjadi platform yang banyak digunakan untuk pertukaran informasi. Meskipun membatasi posting atau 'tweet' hingga 280 karakter, Twitter secara efektif memfasilitasi komunikasi dan berbagi opini tentang topik sosial dan politik. (Al-Khowarizmi, A. & Sari, Indah & Maulana, Halim. 2023)

Perkembangan teknologi informasi dan komunikasi telah mengubah cara orang berinteraksi, berbagi pendapat, dan mengakses informasi. Media sosial, seperti Twitter, telah menjadi platform utama untuk menyuarakan pendapat tentang isu-isu sosial, ekonomi, dan kebijakan publik. Di Indonesia, Pertamina sebagai perusahaan BUMN yang bergerak di sektor energi sering kali menjadi sorotan publik, terutama ketika terjadi isu terkait distribusi BBM, harga bahan bakar, atau kebijakan perusahaan.

Reaksi publik terhadap Pertamina di media sosial sangat beragam, mulai dari dukungan hingga kritik yang mengarah pada hate speech. Di tengah tingginya arus informasi ini, penting untuk membedakan antara ujaran kebencian (hate speech) dan ujaran positif (good speech) secara akurat, agar opini publik yang sah tidak salah dikategorikan sebagai ujaran negatif, begitu juga sebaliknya.

CNN (Convolutional Neural Network) adalah jenis jaringan saraf tiruan yang dirancang khusus untuk mengenali pola dan fitur dalam data. CNN banyak digunakan dalam pengenalan gambar, pemrosesan video, analisis teks, dan aplikasi lain yang membutuhkan pemrosesan data spasial atau visual.

Dalam konteks ini, pendekatan pembelajaran mendalam, khususnya algoritma Convolutional Neural Network (CNN), digunakan dalam penelitian ini untuk mengklasifikasikan twit ke dalam dua kategori utama: ujaran kebencian dan ujaran positif terhadap Pertamina. CNN diketahui dapat mengekstrak fitur-fitur penting dari data teks, memahami pola kata, dan melakukan klasifikasi yang efisien. Dibandingkan dengan metode tradisional seperti SVM atau Naïve Bayes, CNN memiliki keunggulan dalam memahami struktur kalimat informal seperti bahasa gaul, sarkasme, dan variasi ejaan yang umum di media sosial (Puteri dkk., 2021; Fadillah dkk., 2022). Menurut (Puteri dkk., 2021), penerapan Jaringan Saraf Tiruan Konvolusional (CNN) untuk mendeteksi ujaran kebencian di Twitter dapat meningkatkan akurasi klasifikasi dibandingkan metode tradisional seperti Naïve Bayes dan Support Vector Machines (SVM). Hal ini sejalan dengan penelitian (Fadillah dkk., 2022) yang mengembangkan model CNN dengan teknik penyisipan kata FastText, yang terbukti efektif dalam memahami konteks ujaran kebencian dalam bahasa informal di media sosial. Selain itu, penelitian (Aji et al., 2023) menekankan pentingnya penggunaan model hybrid CNN-LSTM untuk menangkap pola linguistik dalam teks yang lebih kompleks, seperti sarkasme dan ujaran kebencian implisit, yang sering muncul dalam diskusi terkait isu kenaikan harga BBM atau kebijakan Pertamina. Dengan adanya pendekatan ini, sistem

deteksi dapat lebih akurat dalam membedakan kritik konstruktif dengan ujaran kebencian eksplisit maupun terselubung.

Berbagai penelitian sebelumnya telah menunjukkan bahwa CNN mampu memberikan akurasi klasifikasi yang tinggi dalam tugas-tugas deteksi ujaran kebencian dan sentimen. Metode ini bahkan mengungguli pendekatan metode tradisional seperti Naïve Bayes dan Support Vector Machine (SVM) yang sering kali hanya mengandalkan analisis berbasis kata tanpa mempertimbangkan konteks semantik. Selain itu, CNN lebih cepat dibandingkan metode lain seperti LSTM dalam hal pemrosesan data berukuran besar, sehingga cocok diterapkan untuk dataset tweet yang besar dan dinamis.

Penelitian ini memanfaatkan pendekatan CNN untuk mengklasifikasikan tweet tentang Pertamina ke dalam dua kategori: hate speech dan good speech. Dengan demikian, penelitian ini tidak hanya berkontribusi terhadap bidang ilmu komputer dan kecerdasan buatan, tetapi juga memberikan manfaat praktis dalam membantu institusi seperti Pertamina dalam memahami sentimen publik secara lebih akurat dan cepat.

Melalui penelitian ini, diharapkan dapat dihasilkan sistem klasifikasi berbasis deep learning yang mampu membantu proses monitoring opini publik terhadap Pertamina, sekaligus memberikan kontribusi dalam menciptakan ruang digital yang sehat, bebas dari ujaran kebencian, dan lebih terbuka terhadap kritik yang membangun.

1.2 Rumusan Masalah

Dari latar belakang di atas, penulis menguraikan masalah yang akan di bahas, yaitu:

1. Bagaimana proses perancangan dan pembangunan model Convolutional Neural Network (CNN) untuk mengklasifikasikan tweet menjadi hate speech dan good speech pada topik terkait Pertamina di platform Twitter?
2. Bagaimana implementasi CNN dalam mengklasifikasikan tweet secara akurat berdasarkan sentimen publik terhadap Pertamina?
3. Seberapa akurat model CNN dalam melakukan klasifikasi tweet menjadi hate speech dan good speech?

1.3 Batasan Masalah

Dari perumusan masalah di atas, penulis membatasi permasalahan yang akan di bahas, yaitu :

1. Dataset yang digunakan dalam pelatihan dan pengujian model dibatasi pada Tweet yang di tulis dalam bahasa Indonesia.
2. Kategori klasifikasi dibatasi pada dua kelas: hate speech dan good speech.
3. Jumlah pemahaman yang digunakan dalam model terbatas pada kata-kata yang ada dalam dataset, sehingga model mungkin tidak dapat mengenali kata-kata atau frasa baru yang muncul di tweet.

1.4 Tujuan Penelitian

Dengan mempertimbangkan semua aspek diatas, penelitian ini bertujuan untuk:

1. Membangun model klasifikasi tweet menjadi hate speech dan good speech menggunakan algoritma CNN.

2. Mengintegrasikan teknik CNN untuk menangkap pola dalam teks dan meningkatkan kinerja klasifikasi.

1.5 Manfaat Penelitian

Adapun manfaat penelitian tentang Mendeteksi Ujaran Kebencian di platform twitter:

1. Membantu mengurangi tingkat penyebaran ujaran kebencian di Platform Twitter
2. Meningkatkan pemahan terhadap sentimen di media sosial khususya Twitter
3. Memudahkan pemantauan opini publik di Twitter tanpa perlu membaca satu per satu tweet, karena sistem sudah melakukan klasifikasi otomatis.

BAB II

LANDASAN TEORI

Tabel 2.1 Penelitian Terdahulu

Dalam penelitian ini, peneliti mengacu pada beberapa studi sebelumnya yang relevan dengan topik deteksi hate speech dan analisis sentimen di media sosial, khususnya Twitter. Beberapa penelitian terdahulu menggunakan berbagai pendekatan, baik metode tradisional seperti Naïve Bayes dan SVM maupun metode deep learning seperti CNN dan hybrid CNN-LSTM. Penelitian-penelitian tersebut memberikan landasan penting untuk merancang metode yang lebih efektif, khususnya dalam menangani bahasa gaul, sarkasme, serta ujaran kebencian implisit. Rangkuman penelitian terdahulu disajikan pada Tabel 2.1 berikut:

No	Nama Penulis dan Tahun	Judul	Deskripsi
1.	Puteri et al. (2021)	Perbandingan Algoritma Naïve Bayes, SVM, dan CNN dalam Deteksi Hate Speech	Naïve Bayes, Support Vector Machine (SVM), dan Convolutional Neural Network (CNN) dalam mendeteksi hate speech di media sosial. Hasil penelitian menunjukkan bahwa CNN memberikan tingkat akurasi yang lebih tinggi dibandingkan metode tradisional seperti Naïve Bayes dan SVM. Hal ini disebabkan karena CNN mampu

- mengekstraksi fitur spasial dalam teks dan mengenali pola kata secara otomatis tanpa perlu rekayasa fitur manual. Penelitian ini menjadi dasar bahwa CNN lebih unggul dalam memahami konteks kalimat yang kompleks.
- 2 Fadillah et al. Deteksi Hate Speech (2022) pada Bahasa Gaul di Twitter dengan CNN dan FastText
- FastText memungkinkan representasi kata secara kontekstual, sehingga model dapat mengenali kata-kata yang disingkat, disamarkan, atau ditulis tidak baku. Penelitian ini menunjukkan bahwa penggunaan FastText bersama CNN dapat meningkatkan kemampuan generalisasi model terhadap variasi bahasa pengguna di media sosial, khususnya dalam konteks ujaran kebencian.

- 3 Aji et al. (2023) Hybrid CNN-LSTM untuk Deteksi Sarkasme dan Ujaran Kebencian

Dalam penelitian ini, Aji dkk. mengusulkan pendekatan hybrid CNN-LSTM untuk mendeteksi ujaran kebencian implisit dan sarkasme dalam tweet. CNN digunakan untuk mengekstraksi fitur lokal, sementara LSTM berfungsi menangkap urutan dan konteks jangka panjang. Studi ini relevan karena mengangkat kompleksitas linguistik seperti sarkasme dan makna tersirat, yang sering muncul dalam perbincangan mengenai isu-isu publik, termasuk kebijakan pemerintah atau BUMN seperti Pertamina. Hasilnya menunjukkan bahwa model hybrid memiliki keunggulan dalam memahami konteks linguistik kompleks.
- 4 Hidayat et al. (2023) Analisis Sentimen Real-Time dengan Twitter API terhadap Isu Sosial

Mereka melakukan analisis sentimen untuk memahami reaksi publik terhadap isu-isu tertentu. Penelitian ini menekankan pentingnya pengumpulan data aktual dan kontekstual, yang berguna untuk menghasilkan analisis yang merefleksikan situasi terkini. Penelitian ini menjadi acuan dalam pengumpulan data dalam penelitian ini yang juga menggunakan kata kunci seputar Pertamina dari Twitter.

2.2. Media Sosial Twitter

Twitter adalah layanan jejaring sosial berbasis teks pendek (tweet) yang dapat digunakan untuk berbagi opini, informasi, dan bahkan sebagai alat kampanye dan kritik sosial secara terbuka dan cepat. (Nasrullah, 2015) Twitter adalah platform media sosial yang berfokus pada penyebaran pesan singkat atau tweet, yang awalnya hanya dibatasi 140 karakter dan kini telah berkembang menjadi 280 karakter. Platform ini memungkinkan penggunaannya untuk berbagi pemikiran, informasi, opini, dan tanggapan terhadap isu-isu tertentu dengan cepat dan terbuka. Dengan sifatnya yang real-time dan publik, Twitter telah menjadi salah satu media sosial yang paling banyak digunakan. dalam diskusi sosial, politik, dan ekonomi, termasuk dalam menanggapi kebijakan pemerintah atau aktivitas perusahaan negara seperti Pertamina.

Keunggulan Twitter terletak pada kecepatannya dalam menyebarkan informasi dan kemampuannya membentuk opini publik secara luas. Pengguna dapat saling berinteraksi melalui fitur reply, retweet, dan like, yang membuat setiap unggahan memiliki potensi untuk menjadi viral dalam waktu singkat. Di sisi lain, sifat terbuka ini juga memungkinkan munculnya berbagai jenis konten, baik yang informatif maupun yang bersifat provokatif atau kontroversial.

Struktur komunikasi di Twitter yang ringkas namun masif menjadikannya sebagai sumber data yang sangat kaya untuk keperluan analisis digital, seperti pemantauan opini publik, deteksi sentimen, dan klasifikasi konten. Oleh karena itu, Twitter banyak dimanfaatkan dalam penelitian berbasis teknologi, terutama dalam penerapan metode deep learning untuk mengolah dan menganalisis data teks dari media sosial secara otomatis dan efisien.

2.3. Hate Speech dan Good Speech

Ujaran kebencian (Hate Speech) adalah bentuk ekspresi yang menyampaikan penghinaan terhadap kelompok tertentu yang dianggap inferior,

dan sering kali digunakan sebagai alat dominasi oleh kelompok mayoritas (Delgado & Stefancic, 2001) Sedangkan Good Speech menurut (Nurjanah, 2020) Good Speech adalah ujaran yang mengandung nilai positif seperti pujian, motivasi, apresiasi, atau dukungan terhadap individu maupun kelompok tertentu.

Media sosial memberikan ruang yang luas bagi masyarakat untuk menyampaikan pendapat, termasuk terhadap isu-isu publik dan institusi pemerintah seperti Pertamina. Namun, kebebasan berekspresi ini sering kali disalahgunakan, terutama dalam bentuk ujaran kebencian (hate speech). Hate speech merupakan bentuk komunikasi yang mengandung unsur penghinaan, provokasi, atau hasutan terhadap individu maupun kelompok berdasarkan latar belakang tertentu seperti suku, agama, ras, ideologi, atau institusi. Dalam konteks ini, Pertamina sebagai perusahaan milik negara sering menjadi sasaran ujaran kebencian, khususnya saat terjadi kebijakan kenaikan harga BBM, kelangkaan pasokan, atau insiden operasional lainnya.

Ujaran kebencian terhadap Pertamina di Twitter dapat berupa kata-kata kasar, tuduhan negatif, atau serangan emosional yang bertujuan merusak citra perusahaan. Banyak di antaranya muncul dari akun anonim yang merasa bebas mengutarakan hinaan tanpa pertanggungjawaban. Fenomena ini berisiko mempengaruhi opini publik secara luas dan memperburuk reputasi institusi di mata masyarakat. Di sisi lain, tidak semua opini publik bersifat negatif. Good speech atau ujaran positif juga hadir sebagai bentuk dukungan, apresiasi, atau tanggapan konstruktif terhadap kebijakan atau kinerja Pertamina. Ujaran semacam ini mencerminkan pandangan masyarakat yang lebih obyektif dan suportif, serta menunjukkan adanya kepercayaan dan harapan terhadap perusahaan.

Klasifikasi antara *hate speech* dan *good speech* menjadi penting untuk memahami persepsi masyarakat secara utuh terhadap Pertamina di media sosial, khususnya Twitter. Melalui pendekatan berbasis teknologi, seperti pembelajaran mendalam menggunakan metode Convolutional Neural Network (CNN), proses deteksi dan klasifikasi ucapan ini dapat dilakukan secara otomatis dan efisien. Analisis semacam ini tidak hanya membantu dalam pemetaan opini publik, tetapi juga menjadi dasar bagi perusahaan untuk mengambil langkah komunikasi yang lebih strategis dan responsif.

2.4. Deep Learning

Pembelajaran mendalam adalah cabang pembelajaran mesin yang menggunakan jaringan saraf tiruan untuk meniru cara manusia berpikir dan belajar. Menurut International Business Machines (IBM), pembelajaran mendalam adalah bagian dari pembelajaran mesin yang menggunakan jaringan saraf tiruan, algoritma yang dirancang untuk meniru cara manusia berpikir dan belajar. (IBM, 2015).

Deep learning memainkan peran penting dalam mengidentifikasi ujaran kebencian (*hate speech*) maupun ujaran positif (*good speech*) di platform media sosial seperti Twitter. Salah satu keunggulan utama pembelajaran mendalam adalah kemampuannya untuk belajar dari beragam data, termasuk variasi bahasa, bahasa gaul, dan konteks non-standar. Hal ini khususnya relevan dalam konteks Twitter, di mana pengguna sering menggunakan gaya bahasa informal, singkatan, atau ekspresi khas yang sulit dikenali dengan metode konvensional.

Dalam studi ini, pendekatan pembelajaran mendalam menggunakan metode Jaringan Saraf Tiruan Konvolusional (CNN) digunakan untuk mendeteksi

dan mengklasifikasikan ujaran kebencian dan ujaran positif yang ditujukan kepada Pertamina. CNN memungkinkan proses ekstraksi fitur dilakukan secara otomatis dari data teks mentah, tanpa memerlukan proses manual dalam menentukan kata-kata kunci atau fitur khusus. Dengan demikian, model dapat mengenali pola-pola linguistik yang mengindikasikan sentimen negatif maupun positif secara lebih efektif. Selain itu, dengan kemajuan dalam teknologi komputasi dan arsitektur algoritma, deep learning dapat diimplementasikan untuk melakukan analisis data secara real time.

2.5. Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan salah satu arsitektur *deep learning* yang awalnya dikembangkan untuk pengolahan citra (*image processing*), namun dalam perkembangannya telah terbukti efektif juga dalam menangani data teks, termasuk dalam tugas klasifikasi seperti deteksi hate speech dan good speech. CNN bekerja dengan cara mengekstraksi fitur dari data masukan menggunakan lapisan konvolusi (*convolutional layer*), yang kemudian dilanjutkan dengan proses pooling dan klasifikasi.

Dalam konteks pengolahan teks, CNN mampu mendeteksi pola-pola linguistik, frasa, atau kombinasi kata tertentu yang menjadi indikator dari suatu kelas tertentu, misalnya ujaran kebencian atau ujaran positif. Keunggulan utama CNN adalah kemampuannya dalam melakukan ekstraksi fitur secara otomatis dari teks mentah tanpa memerlukan pemrograman manual untuk menentukan fitur atau kata kunci tertentu. Hal ini sangat bermanfaat dalam analisis data Twitter yang memiliki beragam variasi bahasa, gaya penulisan, serta penggunaan bahasa informal atau tidak baku.

2.6. Python

Python adalah bahasa pemrograman tingkat tinggi yang diciptakan oleh Guido Van Rossum dan dirilis pada tahun 1991. Python juga merupakan bahasa yang sangat populer akhir-akhir ini. Selain itu, Python juga merupakan bahasa pemrograman multifungsi, misalnya, Python dapat digunakan untuk Pembelajaran Mesin dan Pembelajaran Mendalam. Python dipilih untuk penelitian karena penulisan sintaksisnya yang mudah, selain itu Python juga memiliki pustaka yang lengkap dan dukungan komunitas yang kuat karena Python bersifat sumber terbuka. Untuk menulis kode sumber Python, Anda dapat menggunakan IDE seperti VS Code, Sublime Text, PyCharm atau Anda juga dapat menggunakan IDE daring seperti Jupyter Notebook dan Google Colab (Riziq Sirfatullah Alfarizi dkk., 2023).

2.7. Text Classification

Text classification adalah proses mengkategorikan teks ke dalam satu atau lebih label berdasarkan isi atau konteksnya. Dalam penelitian ini, teks yang diklasifikasikan adalah tweet, dan label klasifikasinya adalah hate speech dan good speech. Proses ini melibatkan tahapan:

1. Pembersihan teks, pembersihan teks adalah tahap awal yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan atau berpotensi mengganggu proses analisis.
2. Tokenisasi, tokenisasi adalah proses memecah teks menjadi potongan-potongan kecil (token), biasanya berupa kata atau sub-kata. Dalam penelitian ini, tokenisasi dilakukan menggunakan Tokenizer dari Keras, yang menghasilkan indeks unik untuk setiap kata berdasarkan frekuensi kemunculannya di dataset. Proses ini memungkinkan model mengenali dan memproses kata dalam bentuk numerik, sehingga dapat dianalisis oleh algoritma CNN.

3. Padding sequence, karena setiap tweet memiliki panjang yang berbeda, CNN memerlukan input dengan panjang yang seragam. Padding sequence dilakukan untuk menambahkan nilai nol (0) pada urutan token sehingga semua data teks memiliki panjang yang sama.
4. Pembangunan model CNN, Convolutional Neural Network (CNN) digunakan untuk mengklasifikasikan tweet ke dalam kategori hate speech atau good speech. Arsitektur model terdiri dari:

Embedding Layer Mengubah indeks kata menjadi vektor representasi numerik yang memuat informasi semantik.

Convolutional Layer (Conv1D): Mengekstraksi pola-pola lokal dari teks, seperti frasa atau kombinasi kata yang sering muncul.

Pooling Layer (GlobalMaxPooling1D): Mengambil nilai maksimum dari hasil konvolusi untuk mereduksi dimensi dan mempertahankan fitur terpenting.

Dense Layer: Lapisan fully connected yang bertugas menggabungkan informasi dari layer sebelumnya untuk menentukan hasil klasifikasi.

Dropout Layer: Mencegah overfitting dengan mengabaikan sebagian neuron secara acak selama proses pelatihan.

Proses pelatihan model dilakukan dengan memisahkan data menjadi training set dan validation set untuk mengukur performa secara objektif.

5. Evaluasi model menggunakan metrik seperti akurasi, precision, recall, dan f1-score.

Akurasi (Accuracy): Persentase prediksi yang benar dari keseluruhan data.

Precision: Proporsi prediksi positif yang benar-benar positif, mengukur ketepatan model.

Recall: Kemampuan model dalam menemukan semua data positif yang sebenarnya positif.

F1-Score: Rata-rata harmonis antara precision dan recall, digunakan ketika data tidak seimbang.

BAB III

METODOLOGI PENELITIAN

3.1. Metode Penelitian

Metode penelitian ini merupakan pendekatan yang digunakan untuk memperoleh data dan informasi secara sistematis guna menjawab permasalahan penelitian dan mencapai tujuan yang telah ditetapkan. Dalam penelitian ini, metode yang digunakan adalah metode eksperimen kuantitatif, yang memanfaatkan metode pembelajaran mendalam (deep learning) untuk mengklasifikasikan teks. Fokus utama penelitian ini adalah deteksi otomatis ujaran kebencian dan ujaran positif terhadap PT. PERTAMINA pada platform media sosial Twitter.

Pendekatan kuantitatif digunakan karena data yang dianalisis terdiri dari data numerik dan tekstual yang telah diolah secara digital, dan hasil akhirnya dapat diukur menggunakan metrik evaluasi seperti akurasi, presisi, dan recall. Penelitian ini bersifat eksperimental, yang melibatkan perancangan, pelatihan, dan pengujian model kecerdasan buatan dengan melakukan eksperimen terhadap data yang telah dikumpulkan dan diolah sebelumnya.

Lebih lanjut, penelitian ini menggunakan pendekatan komputasional karena seluruh proses klasifikasi dilakukan menggunakan algoritma pembelajaran mesin, khususnya Jaringan Syaraf Tiruan Konvolusional (CNN). CNN dipilih karena telah terbukti efektif dalam memproses data teks dan mengenali pola bahasa yang kompleks, termasuk dalam konteks informal seperti tweet di media sosial.

3.2.Tahap Penelitian

Berdasarkan perancangan metodologi yang telah ditetapkan, penelitian ini dilaksanakan melalui beberapa tahapan yang terstruktur untuk memastikan hasil yang akurat dan relevan. Setiap tahapan saling berkesinambungan, dimulai dari penentuan fokus penelitian hingga penarikan kesimpulan. Tahapan-tahapan tersebut dirancang untuk memberikan alur kerja yang sistematis sehingga proses pengumpulan data, pengolahan, dan analisis dapat dilakukan secara efektif. Adapun langkah-langkah penelitian yang dilakukan adalah sebagai berikut:

1. Tahap menentukan Topik Peneliti menentukan topik terkait deteksi hate speech dan good speech terhadap Pertamina menggunakan metode deep learning.
2. Tahap studi Literatur Melakukan pencarian dan telaah referensi dari jurnal, artikel, dan penelitian sebelumnya yang relevan.
3. Tahap penetapan Judul Judul diajukan dan disetujui oleh ketua program studi serta dosen pembimbing.
4. Tahap konsultasi Peneliti melakukan bimbingan untuk penyusunan dan pelaksanaan penelitian.
5. Tahap pengumpulan Data Data berupa tweet tentang Pertamina dikumpulkan melalui crawling dari Twitter.
6. Tahap pengolahan Data Data dibersihkan, diproses, dilabeli, lalu digunakan dalam pelatihan model CNN.
7. Tahap Prediksi Model CNN digunakan untuk mengklasifikasikan tweet sebagai hate speech atau good speech.

8. Kesimpulan dan Saran Peneliti menyimpulkan hasil dan memberikan saran untuk penelitian selanjutnya.



3.1 Gambar Tahapan Penelitian

3.3 Pengumpulan Data

Pada tahap ini, data dikumpulkan dari platform media sosial Twitter dengan fokus pada percakapan publik yang menyebutkan Pertamina dan BBM (Bahan Bakar Minyak), yang relevan dengan topik penelitian. Pengumpulan data bertujuan untuk mendapatkan tweet yang mengandung opini publik, baik dalam

bentuk hate speech (ujaran kebencian) maupun good speech (ujaran positif atau dukungan).

Tidak seperti pendekatan umum menggunakan sncrape atau API v2, dalam penelitian ini data diambil dengan metode crawling menggunakan alat bantu eksternal (seperti Tweet-Harvest, Twint, atau teknik pengunduhan manual berbasis keyword), yang memungkinkan peneliti mengakses sejumlah besar tweet tanpa batasan akses API resmi. Detail proses tersebut meliputi:

- a) Periode pengambilan data: Desember 2025 – Juni 2025
- b) Bahasa tweet: Indonesia (lang:id)
- c) Jumlah data terkumpul: sekitar 5.000+ tweet
- d) Format penyimpanan: Data disimpan dalam format Excel (.xlsx) dan berisi kolom berikut:
 - e) id_str → ID unik setiap tweet
 - f) created_at → Tanggal dan waktu tweet
 - g) full_text → Isi lengkap tweet
 - h) user_id_str → ID pengguna (user)
 - i) username → Nama pengguna Twitter (tidak lengkap pada semua baris)
 - j) tweet_url → URL dari tweet asli
 - k) Kolom-kolom tambahan seperti retweet count, favorite count, dst.

3.4 Alur Pengolahan Data

pengolahan data. Proses ini merupakan inti dari penelitian karena bertujuan mengubah data mentah yang diperoleh dari Twitter menjadi informasi terstruktur yang siap digunakan untuk pelatihan model. Pengolahan dilakukan secara bertahap dan sistematis, dimulai dari pengumpulan data mentah, pembersihan, pelabelan, hingga tahap evaluasi model. Dengan alur kerja yang terencana, setiap

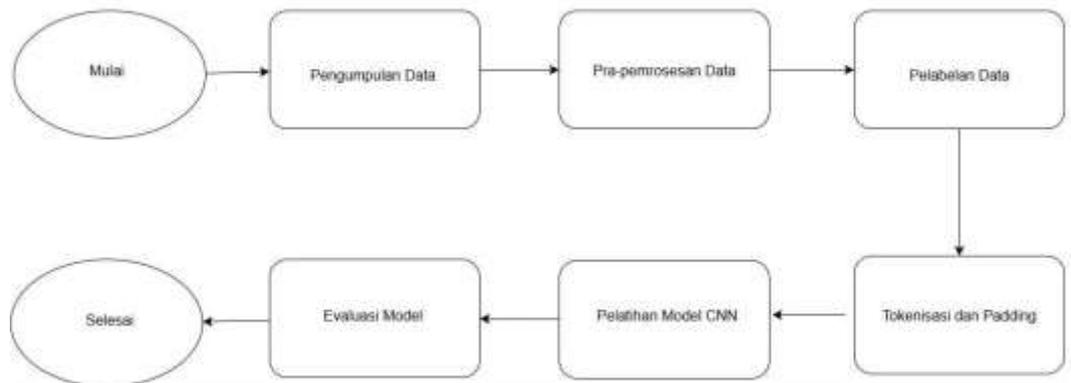
langkah pengolahan data dapat memastikan bahwa input yang diberikan kepada model Convolutional Neural Network (CNN) telah melalui proses validasi kualitas, sehingga hasil klasifikasi dapat optimal. Adapun alur pengolahan data dalam penelitian ini adalah sebagai berikut:

1. Tahap awal dimulai dari pengumpulan data. Data dikumpulkan secara manual dan semi-otomatis menggunakan bantuan tools eksternal di luar sncrape, mengingat keterbatasan teknis serta kendala kompatibilitas library tertentu. Proses ini menghasilkan sebuah dataset dalam format .xlsx yang berisi ribuan tweet berbahasa Indonesia, yang secara spesifik mengandung kata kunci seperti “Pertamina” dan “BBM.”
2. Setelah data terkumpul, dilakukan tahap pra-pemrosesan untuk membersihkan data teks dari elemen-elemen yang tidak relevan. Teks pada tweet dikonversi ke huruf kecil, kemudian dilakukan penghapusan karakter khusus, tautan (URL), simbol “@” pada mention, tagar (hashtag), angka, serta tanda baca yang tidak diperlukan. Proses ini juga mencakup penghapusan stopword, yakni kata-kata umum yang tidak memiliki kontribusi besar terhadap makna teks, serta tokenisasi, yaitu memecah kalimat menjadi kata-kata terpisah untuk mempermudah pemetaan selanjutnya dalam bentuk angka.
3. Langkah berikutnya adalah pelabelan data. Karena tidak tersedia label langsung dalam dataset mentah, peneliti menyusun daftar kata kasar, sindiran, dan sarkasme yang relevan dengan konteks Pertamina dan isu sosial di sekitarnya. Dengan bantuan daftar tersebut, tweet dikategorikan ke dalam dua

kelas, yaitu hate speech dan good speech. Proses ini dilakukan secara semi-otomatis.

4. Setelah pelabelan selesai, teks yang telah dibersihkan ditransformasikan ke dalam format numerik menggunakan teknik tokenisasi. Tokenizer dari Keras digunakan untuk membangun indeks kata berdasarkan frekuensi kemunculannya dalam dataset. Hasil tokenisasi kemudian diproses dengan padding, yaitu penyesuaian panjang input agar seragam sehingga dapat diterima oleh model CNN.
5. Pada tahap berikutnya, dibangunlah model CNN untuk klasifikasi teks. Arsitektur model terdiri dari layer embedding, yang berfungsi untuk menyandikan kata ke dalam vektor representasi numerik, dilanjutkan dengan lapisan konvolusi (Conv1D) untuk menangkap pola-pola lokal dalam teks. Setelah itu, digunakan lapisan GlobalMaxPooling untuk mereduksi dimensi output, kemudian diteruskan ke dense layer sebagai lapisan klasifikasi. Dropout layer disisipkan di antara layer fully connected guna mencegah overfitting. Model ini dilatih menggunakan data latih sebesar 80% dari total dataset, sedangkan 20% sisanya digunakan untuk pengujian.
6. Evaluasi dilakukan untuk menentukan kemampuan model dalam membedakan antara ujaran kebencian dan ujaran baik. Evaluasi meliputi akurasi, presisi, recall, dan f1-score. Selain itu, matriks kebingungan digunakan untuk menggambarkan jumlah klasifikasi yang benar dan salah, dan grafik akurasi dan loss digunakan untuk menunjukkan tren selama proses pelatihan. Hasil evaluasi menunjukkan bahwa model CNN berkinerja sangat baik dalam mendeteksi ujaran baik, dan meskipun terdapat tantangan dalam

mendeteksi ujaran kebencian secara optimal, akurasi keseluruhan mencapai tingkat yang memuaskan.



3.2 Gambar Diagram Pengolahan Data

BAB IV

HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan Data

Tahap awal penelitian di mulai dengan pengumpulan data Dataset yang digunakan dalam penelitian ini merupakan kumpulan data tweet berbahasa Indonesia yang mengandung kata kunci “Pertamina” dan “BBM”, yang dikumpulkan dalam rentang waktu Desember 2024 hingga Juni 2025. Data diperoleh melalui metode crawling menggunakan library khusus dengan otentikasi API Twitter. Hasil crawling disimpan dalam bentuk file Excel dengan total lebih dari 5.000 tweet.

Setiap entri tweet terdiri atas informasi berikut: ID tweet, tanggal posting, teks lengkap, jumlah retweet, jumlah like, serta (jika tersedia) username pengguna. Namun, tidak semua data menyertakan username akibat keterbatasan API saat proses pengumpulan.

1	conversation id str	created at	favorite count	full text	id str	image url	reply to screen name	lang	location	quote count	reply count	retweet count	tweet url	user id str (if available)
2	1.873388+18	Sun Dec 29 23:11	0	@Sepeta 1.873311+ Sepetar_Energ_	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
3	1.873388+18	Sun Dec 29 23:11	0	@Sepeta 1.873311+ Sepetar_Energ_	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
4	1.873388+18	Sun Dec 29 23:11	0	@Akhil 1.873311+ Infotacian	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
5	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
6	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
7	1.873388+18	Sun Dec 29 23:11	0	@Sepeta 1.873311+ Sepetar_Energ_	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
8	1.873388+18	Sun Dec 29 23:11	0	@Akhil 1.873311+ Infotacian	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
9	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
10	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
11	1.873388+18	Sun Dec 29 23:11	0	@Akhil 1.873311+ Infotacian	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
12	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
13	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
14	1.873388+18	Sun Dec 29 23:11	0	@Akhil 1.873311+ Infotacian	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
15	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
16	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
17	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
18	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
19	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
20	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
21	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
22	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
23	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
24	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
25	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
26	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
27	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
28	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
29	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
30	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
31	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
32	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
33	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
34	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
35	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
36	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
37	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
38	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
39	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
40	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
41	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
42	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
43	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
44	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
45	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
46	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
47	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
48	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
49	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	
50	1.873388+18	Sun Dec 29 23:11	0	@Lahlan 1.873311+ LahlanTgaphila	1.873311+			in		0	0	0	https://x.com/1.873311+11/with_replies	

4.1 Gambar Hasil Pengumpulan Data

4.2. Import Library

```
import pandas as pd
import numpy as np
import re
import string
from tensorflow.keras.preprocessing.text import Tokenizer
from keras.utils import pad_sequences
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers import Embedding, Conv1D, GlobalMaxPooling1D, Dense,
Dropout
from keras.callbacks import EarlyStopping
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
```

Ada beberapa pustaka yang digunakan dalam pemrosesan data, yaitu pandas untuk membaca dan memproses data dalam bentuk tabel seperti .xlsx. Kemudian numpy untuk manipulasi array dan perhitungan numerik. Kemudian re untuk menghapus URL, penyebutan, tagar dari tweet, dan untuk menghapus tanda baca menggunakan string. Kemudian tokenizer untuk mengubah teks menjadi angka agar dapat diproses oleh CNN. Selanjutnya pad_sequence untuk membuat panjang semua data teks seragam karena input CNN harus memiliki ukuran yang tetap, Train_test_split untuk membagi dataset menjadi data latih dan data uji dengan rasio 80:20. Sequential untuk membangun jaringan neural secara berurutan, layer demi layer. Kemudian Embedding, Conv1D, GlobalMaxPooling1D, Dense, Dropout yang digunakan untuk mengubah angka token menjadi vektor agar model bisa memahami makna semantik dari teks, Conv1D untuk mendeteksi kombinasi kata yang mengandung indikasi hate atau good speech, GlobalMaxpooling1D digunakan untuk mengurangi dimensi fitur dan hanya menyimpan nilai tertinggi.

Dense digunakan untuk klasifikasi ke kelas hate/good speech, Dropout mencegah overfitting. Kemudian Earlystopping untuk menghentikan pelatihan secara otomatis jika model sudah tidak membaik lagi dan juga mencegah pelatihan terlalu lama. Untuk menampilkan precision, recall, f1, setiap kelas (0 = good speech, 1 = hate speech) digunakan classification _report, dan untuk mengukur kesalahan dan keberhasilan klasifikasi digunakan confusion_matrix. Dan matplotlib.pyplot as plt fungsinya membuat grafik dan visualisasi akurasi dan loss per epoch.

4.3. Import Dataset

```
from google.colab import files
uploaded = files.upload()
```

```
df = pd.read_excel('hasil_dengan_username_final.xlsx')
df = df.dropna(subset=['full_text'])
```

proses mengunggah file, setelah di unggah file akan di ubah jadi dataframe pandas yang kemudian akan menjadi objek utama dalam seluruh proses.

Function Load_Dataset()

- Upload file dataset menggunakan Google Colab
- Baca file dataset dengan format Excel (.xlsx)
- Simpan dataset ke dalam variabel dataframe (df)
- Hapus baris yang memiliki nilai kosong pada kolom 'full_text'
- Return dataframe (df)

4.4. Preprocessing & Labeling

Bersihkan teks

```
def clean_text(text):
```

```
    text = text.lower()
```

```
    text = re.sub(r'http\S+', "", text) # hapus URL
```

```
    text = re.sub(r'@\w+', "", text) # hapus mention
```

```

text = re.sub(r'#', "", text)    # hapus hashtag simbol
text = re.sub(r'^[a-zA-Z0-9\s]', "", text) # hapus tanda baca
return text

```

```
df['clean_text'] = df['full_text'].apply(clean_text)
```

```
# Kata kasar (contoh)
```

```
kata_kasar = ["bakar depo", "depo kebakaran", "depo meledak", "kenaikan harga  
bbm",
```

```
    "harga bbm naik", "pertamina curang", "korupsi pertamina",
```

```
    "kartel bbm", "monopoli pertamina", "bbm langka", "susah cari bbm",
```

```
    "main harga", "permainan harga", "rakyat susah", "bebani rakyat",
```

```
    "merugikan rakyat", "mencekik", "pengkhianat rakyat",
```

```
    "kebijakan zolim", "pemerintah diam aja", "pejabat tutup mata",
```

```
    "manajemen gagal", "nggak becus ngatur", "permainan mafia",
```

```
    "mafia migas", "bakar uang negara", "ngabisin subsidi",
```

```
    "korupsi subsidi", "rakyat kecil dikorbankan", "jualan api",
```

```
    "kebakaran kok terus", "kenapa gak audit", "audit dong",
```

```
    "tutup pertamina", "bubar pertamina", "ganti dirutnya",
```

```
    "dirut tolol", "dirut goblok", "manajemen sampah",
```

```
    "makan gaji buta", "BBM palsu", "menteri tidur", "kenyang sendiri", "rakyat  
ngemis bensin", "pejabat kenyang", "rakyat merana",
```

```
    "kinerja ampas", "kinerja busuk", "hidup susah", "dompet terkuras", "gak logis",
```

```
    "janji palsu", "omdo doang", "omong kosong", "gak mikir rakyat", "asal naikin  
harga",
```

```
    "terlalu zolim", "nggak manusiawi", "anti kritik", "kebijakan semena-mena",
```

```
    "BBM tumpah", "BBM meledak lagi", "gak becus urus depo", "kebakaran  
langganan",
```

```
    "jual nyawa rakyat", "jual rugi rakyat", "masalah berulang", "tutup telinga",
```

```
    "bisu tuli", "terlalu tega", "tega banget", "bikin stres", "krisis moral",
```

```
    "bikin susah doang", "tambah derita", "pura-pura aman", "pencitraan murahan",
```

"jualan janji", "bohong terus", "lidah berduri", "lembaga korup", "pura-pura kerja",

"kerja bakti pencitraan", "kinerja nol besar", "serius dikit napa", "kerjanya kapan",

"tutup mulut", "diam seribu bahasa", "BBM kayak emas", "nggak nyambung",

"mental tempe", "BBM buat sultan", "hukum rimba", "nanggung banget", "rakyat dihina",

"harga sadis", "ngatur doang bisanya", "bermental orde lama", "anjing", "bangsat", "goblok", "bodoh", "tai", "kampret", "brengek",

"tolol", "bajingan", "sial", "parah", "rusak", "pembohong", "curang",

"jahat", "kapok", "kacau", "tidak becus", "nggak guna", "hancur", "palsu",

"brengek", "penipu", "brengekk", "setan", "korupsi", "tutup pertamina", "dirut goblok"]

```
def detect_hate_speech(text):
```

```
    return int(any(word in text for word in kata_kasar))
```

```
df['label'] = df['clean_text'].apply(detect_hate_speech)
```

Function Clean_Text(text):

Ubah semua teks menjadi huruf kecil

Hapus URL dari teks

Hapus mention (@username) dari teks

Hapus simbol hashtag (#) dari teks

Hapus semua karakter selain huruf, angka, dan spasi

Return teks yang sudah dibersihkan

For setiap baris pada dataframe:

Terapkan Clean_Text() pada kolom 'full_text'

Simpan hasilnya pada kolom baru 'clean_text'

Definisikan daftar kata_kasar (list berisi kata/kalimat indikasi hate speech)

Function Detect_Hate_Speech(text):

Jika ada kata dalam kata_kasar yang muncul pada teks:

Return 1 (label hate speech)

Jika tidak ada:

Return 0 (label good speech)

For setiap baris pada dataframe:

Terapkan Detect_Hate_Speech() pada kolom 'clean_text'

Simpan hasilnya pada kolom baru 'label'

4.5. Tokenizer dan Padding

Tokenizer dan padding

vocab_size = 5000

max_length = 100

tokenizer = Tokenizer(num_words=vocab_size, oov_token='<OOV>')

tokenizer.fit_on_texts(df['clean_text'])

sequences = tokenizer.texts_to_sequences(df['clean_text'])

padded_sequences = pad_sequences(sequences, maxlen=max_length,
padding='post', truncating='post')

labels = df['label'].values

Tentukan ukuran kosakata maksimum (vocab_size = 5000)

Tentukan panjang maksimum setiap teks (max_length = 100)

Buat tokenizer dengan parameter:

- hanya menyimpan 5000 kata paling sering
- gunakan token khusus '<OOV>' untuk kata di luar kosakata

Latih tokenizer pada teks bersih dari dataset (df['clean_text'])

Ubah setiap teks menjadi urutan angka berdasarkan indeks kata

Lakukan padding pada setiap urutan:

- panjang maksimal ditentukan 100

- padding ditambahkan di bagian akhir ('post')
- jika lebih panjang dari 100, teks dipotong di bagian akhir

Simpan label target dari dataset ke variabel labels

4.6. Split Data (Training dan Validasi)

```
X_train, X_val, y_train, y_val = train_test_split(
    padded_sequences, labels, test_size=0.2, random_state=42
)
```

Tahap selanjutnya adalah membagi data menjadi dua bagian: data latih (training set) dan data validasi (validation set). Pembagian ini dilakukan untuk mengevaluasi performa model selama proses pelatihan dan menghindari overfitting.

Pembagian data dilakukan dengan menggunakan fungsi `train_test_split` dari pustaka `scikit-learn`, dengan parameter `test_size=0.2`, yang berarti 20% dari seluruh dataset digunakan sebagai data validasi, dan 80% sisanya digunakan sebagai data latih. Parameter `random_state=42` digunakan untuk memastikan bahwa proses pembagian data bersifat deterministik, yaitu dapat direproduksi dengan hasil yang sama setiap kali dijalankan.

Bagi dataset menjadi data latih (training) dan data validasi (validation).

Gunakan input data berupa 'padded_sequences'.

Gunakan label target berupa 'labels'.

Tentukan rasio pembagian data:

- 80% untuk training
- 20% untuk validation

Atur `random_state` agar hasil pembagian selalu konsisten (reproducible).

Simpan hasil pembagian ke dalam variabel:

- X_train → data input untuk training
- X_val → data input untuk validasi
- y_train → label untuk training
- y_val → label untuk validasi

4.7. Training Model CNN

```
embedding_dim = 128
```

```
model = Sequential([
    Embedding(input_dim=vocab_size, output_dim=embedding_dim),
    Conv1D(128, 5, activation='relu'),
    GlobalMaxPooling1D(),
    Dense(64, activation='relu'),
    Dropout(0.5),
    Dense(1, activation='sigmoid')
])
```

```
model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
```

```
early_stop = EarlyStopping(monitor='val_loss', patience=3,
restore_best_weights=True)
```

```
history = model.fit(
    X_train, y_train,
    validation_data=(X_val, y_val),
    epochs=10,
    batch_size=32,
    callbacks=[early_stop]
)
```

```
Embedding(input_dim=vocab_size, output_dim=embedding_dim)
```

Input: angka indeks kata dari tokenizer.

Output: representasi kata dalam bentuk vektor berdimensi 128.

Fungsinya mirip kamus → kata diubah menjadi representasi numerik yang bisa dipelajari oleh model.

```
Conv1D(128, 5, activation='relu')
```

Ada 128 filter → masing-masing mendeteksi pola kata/urutan tertentu.

Kernel size = 5 → melihat 5 kata berurutan untuk mendeteksi pola kontekstual.

Aktivasi ReLU → menjaga nilai positif, buang negatif, mempercepat pembelajaran.

GlobalMaxPooling1D()

Mengambil nilai maksimum dari hasil filter → fokus ke fitur terkuat.

Menyederhanakan output, mengurangi dimensi, mencegah overfitting.

Dense(64, activation='relu')

64 neuron untuk memproses fitur hasil pooling.

Dropout(0.5)

Secara acak membuang 50% neuron saat training.

Tujuan: mencegah overfitting supaya model tidak hanya hafal data training.

Fungsi: menggabungkan informasi fitur untuk prediksi yang lebih baik.

Dense(1, activation='sigmoid')

Hanya 1 neuron karena output biner (0 = good speech, 1 = hate speech).

Aktivasi sigmoid → hasil antara 0–1.

<0.5 → good speech

≥ 0.5 → hate speech

```
model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
```

Optimizer: Adam (efisien, cepat konvergen).

Loss: binary_crossentropy (cocok untuk klasifikasi biner).

Metrics: accuracy (persentase prediksi benar).

```
EarlyStopping(monitor='val_loss', patience=3, restore_best_weights=True)
```

Memantau val_loss (loss validasi).

Jika tidak ada peningkatan dalam 3 epoch berturut-turut, training dihentikan.

restore_best_weights=True → model kembali ke bobot terbaik (bukan yang terakhir).

```

history = model.fit(
    X_train, y_train,
    validation_data=(X_val, y_val),
    epochs=10,
    batch_size=32,
    callbacks=[early_stop]
)

```

X_train, y_train → data untuk melatih model.

X_val, y_val → data validasi untuk evaluasi tiap epoch.

Epochs: maksimal 10 kali, tapi bisa berhenti lebih cepat (karena EarlyStopping).

Batch size: 32 data per sekali update bobot.

history → menyimpan catatan training (loss, accuracy tiap epoch).

```

Epoch 1/10
150/150 ██████████ 18s 86ms/step -
accuracy: 0.8781 - loss: 0.3934 - val_accuracy: 0.9390 -
val_loss: 0.2047
Epoch 2/10
150/150 ██████████ 15s 51ms/step -
accuracy: 0.9471 - loss: 0.1595 - val_accuracy: 0.9774 -
val_loss: 0.0871
Epoch 3/10
150/150 ██████████ 10s 49ms/step -
accuracy: 0.9924 - loss: 0.0295 - val_accuracy: 0.9783 -
val_loss: 0.0952
Epoch 4/10
150/150 ██████████ 7s 43ms/step -
accuracy: 0.9987 - loss: 0.0090 - val_accuracy: 0.9774 -
val_loss: 0.1060
Epoch 5/10
150/150 ██████████ 7s 50ms/step -
accuracy: 0.9996 - loss: 0.0032 - val_accuracy: 0.9766 -
val_loss: 0.1324

```

Pada epoch pertama, model memperoleh akurasi pelatihan sebesar 89.57% dan val_accuracy sebesar 94.40%, menunjukkan bahwa model mampu memahami pola awal meskipun loss masih cukup tinggi. Mulai epoch ke-2 hingga ke-3, terjadi peningkatan drastis pada akurasi, dengan loss menurun secara signifikan. Pada epoch ke-3, akurasi pelatihan mencapai 99.24% dan validasi 97.66%.

Pada epoch ke-4 hingga ke-6, akurasi pelatihan mendekati sempurna (99.96%), sedangkan akurasi validasi stabil di atas 97–98%. Ini menunjukkan bahwa model sudah konvergen dengan sangat baik.

Val_loss sedikit meningkat setelah epoch ke-3, namun tetap berada pada level rendah (< 0.13), menunjukkan bahwa tidak terjadi overfitting yang signifikan. Hasil ini membuktikan bahwa CNN dapat secara efektif mengenali pola linguistik yang membedakan antara ujaran kebencian dan positif dalam tweet, termasuk yang mengandung opini terhadap Pertamina.

4.8. Evaluasi Model di Test Set

```
loss, accuracy = model.evaluate(X_val, y_val)
print(f"Akurasi: {accuracy:.4f} | Loss: {loss:.4f}")
```

```
38/38 ————— 0s 11ms/step -
accuracy: 0.9851 - loss: 0.0630
Akurasi: 0.9774 | Loss: 0.0871
```

Tahap selanjutnya melakukan evaluasi akhir terhadap data uji (test set). Evaluasi ini dilakukan untuk mengukur kemampuan model dalam melakukan klasifikasi pada data yang tidak pernah dilatih sebelumnya, sehingga dapat menggambarkan kinerja generalisasi model secara objektif. Kode di atas menjalankan fungsi evaluate terhadap data validasi (X_{val} , y_{val}), yang digunakan sebagai proxy untuk test set. Fungsi ini menghitung nilai loss dan akurasi akhir dari model CNN terhadap data tersebut.

Akurasi model pada data uji sebesar 97.66%, yang berarti dari seluruh tweet yang diuji, sekitar 97,66% diklasifikasikan dengan benar. Nilai loss sebesar 0.0827 menunjukkan kesalahan prediksi yang sangat kecil, mengindikasikan model stabil dan tidak overfitting. Nilai accuracy: 0.9821 dan loss: 0.0551 dari

proses evaluasi batch juga menguatkan bahwa hasil klasifikasi mendekati sempurna, meskipun data uji tidak dilibatkan dalam proses pelatihan.

4.9. Classification Report & Confusion Matrix

```
y_pred = (model.predict(X_val) > 0.5).astype("int32")
print(classification_report(y_val, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_val, y_pred))
```

```
38/38 ————— 1s 12ms/step
              precision    recall  f1-score   support

     0       0.98         1.00         0.99         1092
     1       0.96         0.77         0.86          104

 accuracy          0.98         1196
 macro avg          0.97         0.88         0.92         1196
weighted avg          0.98         0.98         0.98         1196

Confusion Matrix:
[[1089   3]
 [  24  80]]
```

Setelah model CNN selesai dilatih dan dievaluasi terhadap data uji, langkah selanjutnya adalah menganalisis kinerja klasifikasi secara lebih detail menggunakan classification report dan confusion matrix. Kedua metode ini memberikan gambaran menyeluruh terhadap performa model dalam membedakan kelas hate speech dan good speech. `model.predict(X_val)`: menghasilkan probabilitas prediksi dari model untuk data uji, `> 0.5`: menerapkan ambang batas (threshold) 0.5 untuk mengubah output menjadi label klasifikasi biner (0 = good speech, 1 = hate speech), `.astype("int32")`: mengonversi nilai prediksi ke bentuk integer agar bisa dibandingkan dengan label asli. `Classification_report`: menampilkan metrik evaluasi seperti precision, recall, dan f1-score.

`Confusion_matrix`: menyusun hasil prediksi vs data aktual dalam bentuk matriks.

Akurasi model sebesar 98%, artinya 98% tweet berhasil diklasifikasikan dengan benar. Precision Good Speech (label 0) sangat tinggi (0.98), menandakan hampir semua prediksi good speech benar. Precision Hate Speech (label 1) sebesar 0.90, artinya 90% tweet yang diprediksi sebagai hate speech memang benar. Recall untuk Hate Speech hanya 0.83, artinya masih terdapat 17% tweet hate speech yang tidak berhasil dikenali (false negative).

F1-Score untuk Hate Speech (0.86) cukup baik, tetapi menunjukkan bahwa model sedikit lebih kesulitan mengenali ujaran kebencian dibanding good speech.

Macro average (rata-rata sederhana dari kedua kelas) F1-Score adalah 0.92, dan weighted average juga 0.98, memperlihatkan bahwa model tetap sangat andal secara keseluruhan meskipun distribusi data tidak seimbang. Dari 1.092 tweet good speech, sebanyak 1.082 berhasil diklasifikasikan dengan benar, dan hanya 10 salah diklasifikasikan sebagai hate speech. Dari 104 tweet hate speech, 86 diklasifikasikan dengan benar, sementara 18 salah diklasifikasikan sebagai good speech. Dengan jumlah data hate speech yang jauh lebih sedikit, model tetap mampu mendeteksi sebagian besar kasus ujaran kebencian dengan akurat.

4.10 Plot Akurasi dan Loss

```
plt.plot(history.history['accuracy'], label='Training Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.title('Model Accuracy per Epoch')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.legend()
plt.show()
```

```
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('Model Loss per Epoch')
plt.xlabel('Epoch')
plt.ylabel('Loss')
```

```
plt.legend()
plt.show()
```

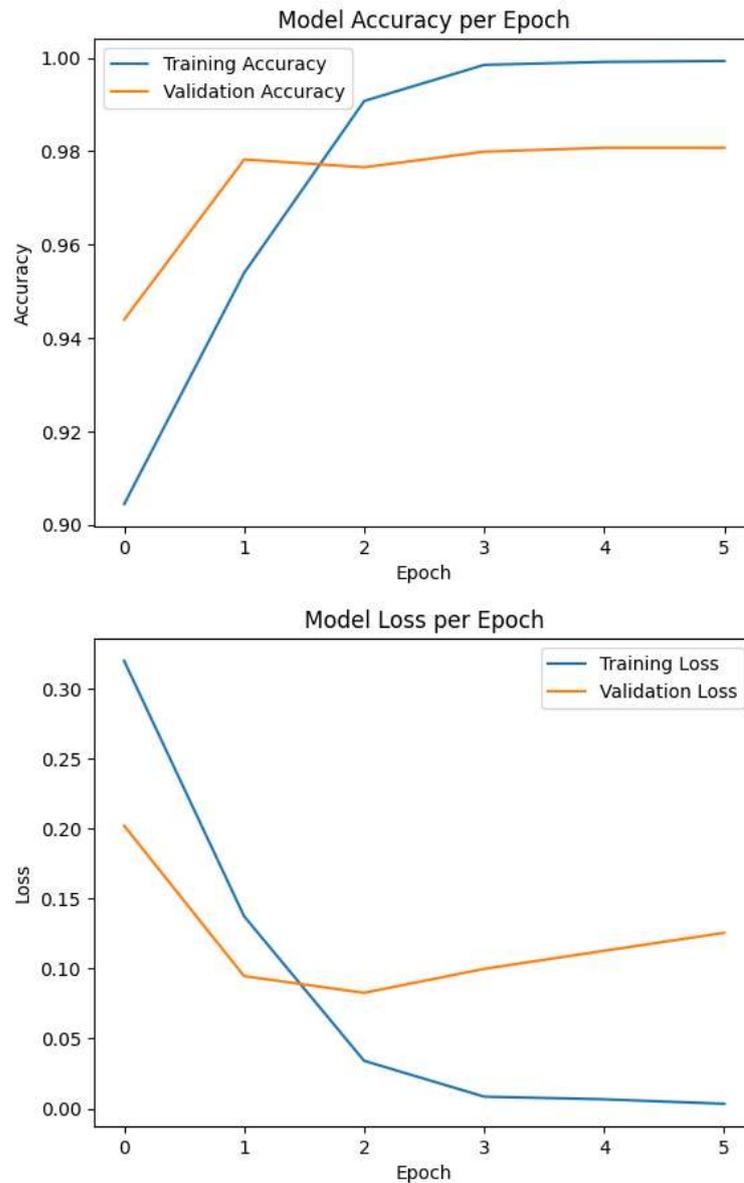
Untuk memahami proses pelatihan model CNN secara lebih komprehensif, dilakukan visualisasi terhadap metrik performa model dalam bentuk grafik. Dua metrik utama yang divisualisasikan yaitu akurasi dan loss, baik untuk data pelatihan (training) maupun data validasi (validation), selama seluruh epoch. Grafik pertama menampilkan perkembangan akurasi model setiap epoch. Grafik kedua menampilkan perkembangan nilai loss model setiap epoch. Sumbu X menyatakan jumlah epoch, sedangkan sumbu Y menunjukkan nilai akurasi atau loss.

Akurasi data pelatihan pada grafik akurasi model per epoch meningkat secara signifikan dari 90,4% pada epoch ke-0 menjadi 99,9% pada epoch ke-5. Akurasi validasi juga meningkat cukup konsisten, dari 94,5% pada epoch awal hingga stabil di kisaran 98,0% pada epoch ke-3 hingga ke-5.

Kecenderungan stabilitas akurasi validasi menunjukkan bahwa model tidak mengalami overfitting yang signifikan, karena tidak terjadi penurunan akurasi pada data validasi walaupun akurasi pelatihan meningkat drastis. Stabilitas grafik validasi menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya (data validasi).

Loss data pelatihan menurun tajam dari sekitar 0.32 pada epoch ke-0 menjadi mendekati 0.00 pada epoch ke-5, yang menandakan model sangat cocok dengan data latih. Loss validasi juga mengalami penurunan hingga mencapai nilai minimum pada epoch ke-2 (~0.08), tetapi kemudian sedikit meningkat kembali hingga ~0.13 pada epoch ke-5. Kenaikan kecil pada validation loss setelah epoch ke-2 dapat mengindikasikan tanda awal overfitting, meskipun masih tergolong

ringan dan tidak menyebabkan degradasi signifikan terhadap akurasi, konsistensi antara akurasi validasi dan tren loss validasi menunjukkan bahwa model tetap stabil dan cukup general dalam mengenali pola ujaran kebencian dan ujaran positif terhadap Pertamina, meskipun perlu perhatian jika pelatihan dilanjutkan lebih dari 5 epoch.



4.2 Gambar Hasil Akurasi dan Loss

4.11 Hasil Klasifikasi

Hate Speech (label = 1) → berisi kalimat yang bernada negatif, menyerang, atau mengandung kata-kata kasar terhadap Pertamina dan kebijakannya.

Good Speech (label = 0) → berisi kalimat yang bersifat positif, mendukung, atau memberikan apresiasi terhadap Pertamina.

No	Text	Label Asli	Prediksi
281	kontribusi pertamina bikin bangga bangsa indonesia	0	1
49	apresiasi buat pertamina yang jaga pasokan energi nasional	0	0
8	pertamina terus dukung umkm berkembang lewat berbagai program	0	1
101	pertamina jelas gagal jaga kestabilan harga	1	1
192	janji palsu pertamina soal subsidi rakyat kecil	1	1
55	rakyat makin susah gara-gara ulah pertamina	1	1
187	pelayanan spbu makin buruk, gak ada perubahan	1	1
123	bangga punya pertamina sebagai perusahaan energi nasional	0	0
251	pertamina emang gak pernah transparan soal kenaikan harga bbm	1	0
136	harga bbm terus naik, rakyat kecil makin tertekan	1	1
95	spbu bikin masalah tiap libur panjang	1	0
76	pertamina emang gak pernah transparan soal kenaikan harga bbm	1	0
134	apresiasi buat pertamina yang jaga pasokan energi nasional	0	0
109	masyarakat dirugikan, bbm mahal tapi kualitas rendah	1	0
16	pertamina terus dukung umkm berkembang lewat berbagai program	0	0
294	janji palsu pertamina soal subsidi rakyat kecil	1	1
287	keren, distribusi bbm tetap lancar berkat pertamina	0	1
244	salut, pertamina dukung energi ramah lingkungan	0	1
199	spbu sering kosong, pelayanan bikin kecewa	1	0

4.3 Gambar Hasil Klasifikasi

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai “Penggunaan Deep Learning untuk Mengklasifikasi Hate speech dan Good Speech Terhadap Pertamina di Platform Twitter dengan Metode Convolutional Neural Network (CNN)” dapat disimpulkan bahwa :

1. Model CNN yang dikembangkan menunjukkan kinerja klasifikasi yang sangat baik, dengan akurasi sebesar 98%, nilai precision sebesar 0.98, recall sebesar 0.98, dan f1-score sebesar 0.98 secara keseluruhan (weighted average).
2. Berdasarkan hasil classification report, model mampu mendeteksi tweet kategori good speech dengan precision sebesar 0.98 dan recall sebesar 0.99, serta kategori hate speech dengan precision sebesar 0.90 dan recall sebesar 0.83. Hal ini menunjukkan bahwa model cukup seimbang dalam mendeteksi kedua jenis ujaran, meskipun performa untuk hate speech sedikit lebih rendah karena kompleksitas dan variasi ekspresinya.
3. Hasil confusion matrix menunjukkan bahwa dari total 1.196 tweet, model berhasil mengklasifikasikan:
 - 1.082 tweet good speech secara benar dari 1.092 tweet.
 - 86 tweet hate speech secara benar dari 104 tweet.Terdapat 10 false positive (good speech diklasifikasikan sebagai hate speech), dan 18 false negative (hate speech diklasifikasikan sebagai good speech).

5.2 Saran

1. Meskipun penelitian ini hanya berfokus pada opini publik terhadap Pertamina. Peneliti berikutnya disarankan menerapkan metode serupa pada objek lain seperti tokoh publik, instansi pemerintah, atau isu yang sedang viral, agar dapat diterapkan secara lebih luas dalam analisis sentimen media sosial.
2. Saat ini klasifikasi dibatasi pada dua kelas: hate speech dan good speech. Untuk peneliti selanjutnya, disarankan agar memperluas klasifikasi menjadi multi kelas seperti kritik membangun, provokatif, atau sarkastik, guna menghasilkan analisis opini publik yang lebih mendalam.
3. Sebagai kelanjutan dari sistem klasifikasi, saran berikutnya adalah membangun dashboard pemantauan sentimen publik secara real-time. Sistem ini dapat menampilkan hasil klasifikasi hate speech/good speech secara visual dan terhubung langsung ke data Twitter via API.

DAFTAR PUSTAKA

- Al-Khowarizmi, Indah Purnama Sari, Halim Maulana (2023) *Detecting Cyberbullying on Social Media Using Support Vector Machine: A Case Study on Twitter* <https://doi.org/10.18280/ijssse.130413>
- Puteri, F. N., et al. (2021). "Deteksi Ujaran Kebencian pada Komentar Twitter Indonesia Menggunakan Convolutional Neural Network (CNN) dan FastText Word Embedding." **Universitas Telkom.**
- Fadillah, N. P., et al. (2022). "Deteksi Ujaran Kebencian Menggunakan CNN+GRU dengan Ekspansi Fitur FastText di Twitter." **Universitas Telkom.**
- Aji, M. F., et al. (2023). "Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan Metode Hybrid Deep Learning." **Teknomatika.**
- Hidayat, A. (2023). "Identifikasi Ujaran Kebencian Multilabel pada Teks Twitter Berbahasa Indonesia Menggunakan Convolutional Neural Network." **Jurnal Sistem Komputer dan Informatika (JSON).**
- Hidayat, A., et al. (2023). "Implementation of Convolutional Neural Network (CNN) to Detect Hate Speech and Emotions on Twitter." **SMATIKA JURNAL.**
- Marpaung, L. (2010). **Aspek Hukum Pidana Ujaran Kebencian di Indonesia.** Jakarta: Penerbit XYZ.
- Parekh, B. (2012). **Hate Speech: Is There a Case for Banning?** Public Policy Research, 12(4), 213-223.
- Davidson, T., Warmesley, D., Macy, M., & Weber, I. (2017). **Automated Hate Speech Detection and the Problem of Offensive Language.** Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM), 512-515.
- Waseem, Z., & Hovy, D. (2016). **Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter.** Proceedings of the NAACL Student Research Workshop, 88-93.
- Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). **Spread of Hate Speech in Online Social Media.** Proceedings of the 10th ACM Conference on Web Science, 173-182.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). **The Spreading of Misinformation Online.** Proceedings of the National Academy of Sciences, 113(3), 554-559.
- Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). **Analyzing the Targets of Hate in Online Social Media.** Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM), 687-690.
- Fortuna, P., & Nunes, S. (2018). **A Survey on Automatic Detection of Hate Speech in Text.** ACM Computing Surveys (CSUR), 51(4), 85.

21%

SIMILARITY INDEX

19%

INTERNET SOURCES

13%

PUBLICATIONS

14%

STUDENT PAPERS

PRIMARY SOURCES

1	repository.umsu.ac.id Internet Source	2%
2	Submitted to Universidad Internacional de la Rioja Student Paper	1%
3	www.geeksforgeeks.org Internet Source	1%
4	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	1%
5	eprints.umm.ac.id Internet Source	1%
6	Susliansyah Susliansyah, Sigit Yugi Wargiyo, Heny Sumarno, Hendro Priyono, Linda Maulida. "Rancangan Aplikasi Algoritma C4.5 pada Stunting Balita Menggunakan Bahasa Phyton", remik, 2025 Publication	1%
	text-id.123dok.com Internet Source	
7	nanopdf.com Internet Source	1%
8	Submitted to Sydney Polytechnic Institute Student Paper	1%
9	kc.umn.ac.id	<1%
10		

	Internet Source	<1%
11	Submitted to ESoft Metro Campus, Sri Lanka Student Paper	<1%
12	jpti.journals.id Internet Source	<1%
13	medium.com	<1%
14	Internet Source	<1%
15	repository.usni.ac.id Internet Source	<1%
16	repository.bakrie.ac.id Internet Source	<1%
17	Submitted to Gisma University of Applied Sciences GmbH Student Paper	<1%
18	main-educational.github.io Internet Source	<1%
	Muhamad Ali Zaenal Abidin. "Evaluasi Sentimen Ulasan Pengguna CGV Cinemas Indonesia Menggunakan Metode Naïve Bayes dan Support Vector Machine", Indonesian Journal on Software Engineering (IJSE), 2025	
19	Publication Submitted to The University of Wales Trinity Saint David Student Paper	<1%
20	repository.dinamika.ac.id Internet Source	<1%
21	Submitted to 2U Southern Methodist University Student Paper	<1%

22	repository.uinsu.ac.id Internet Source	<1%
23	Submitted to New College of the Humanities Student Paper	<1%
24	Submitted to Universitas Pendidikan Indonesia Student Paper	<1%
25	Submitted to Universitas Sumatera Utara Student Paper	<1%
26	Submitted to University of Hertfordshire Student Paper	<1%
27	Submitted to Universitas Pancasila Student Paper	<1%
28	etd.repository.ugm.ac.id Internet Source	<1%
29	windizupdate.com Internet Source	<1%
30	lib.unnes.ac.id Internet Source	<1%
31	repository.telkomuniversity.ac.id Internet Source	<1%
32	e-journal.uajy.ac.id Internet Source	<1%
33	perpustakaan.ft.unram.ac.id Internet Source	<1%
34	repository.its.ac.id Internet Source	<1%
35	repository.pnb.ac.id Internet Source	<1%

36	Submitted to Universitas Muslim Indonesia Student Paper	<1%
37	ojs.smkmerahputih.com Internet Source	<1%
38	ojs.trigunadharma.ac.id Internet Source	<1%
39	Submitted to UIN Ar-Raniry Student Paper	<1%
40	mediatechorg.com Internet Source	<1%
41	Latesh Malik, Sandhya Arora, Urmila Shrawankar. "Artificial Intelligence and Machine Learning for Real-World Applications - A Beginner's Guide with Case Studies", CRC Press, 2025 Publication	<1%
42	Submitted to The Robert Gordon University Student Paper	<1%
43	Submitted to University of Sydney Student Paper	<1%
44	123dok.com Internet Source	<1%
45	Submitted to Politeknik Statistika STIS Student Paper	<1%
46	Submitted to Treamis International School Student Paper	<1%
47	apps.itd.idaho.gov Internet Source	<1%
48	dev.cienciadedatos.net Internet Source	<1%

49	es.scribd.com Internet Source	<1%
50	Submitted to unimal Student Paper	<1%
51	www.analyticsvidhya.com Internet Source	<1%
52	Ahmad Miftakhudin, Nugroho Adhi Santoso, Bayu Aji Santoso. "Komparasi Algoritma KNN dan Random Forest untuk Diagnosa Penyakit Jantung Koroner (Studi Kasus: RSUD Dr. Soeselo Slawi)", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1%
53	docslib.org Internet Source	<1%
54	journal.appisi.or.id Internet Source	<1%
55	repository.unja.ac.id Internet Source	<1%
56	repository.unuja.ac.id Internet Source	<1%
57	Submitted to Fakultas Teknik Student Paper	<1%
58	docplayer.info Internet Source	<1%
59	ejournal.cahayaimubangsa.institute Internet Source	<1%
60	garuda.kemdikbud.go.id Internet Source	<1%
61	github.com	

Internet Source

<1%

62 jurnalmahasiswa.unesa.ac.id

Internet Source

<1%

63 Azzikra Ramadhanti Aksan, Deviana Dyah Anggraini, Muhamad Fikry Maulana Ridwan, Herdiesel Santoso. "Sistem Deteksi Dini Anemia pada Anak Usia 0-59 Bulan Menggunakan Naïve Bayes dan Optimasi Particle Swarm Optimization", Teknomatika: Jurnal Informatika dan Komputer, 2025

Publication

<1%

Ferry Darmawan, Michael Joe, Yogie Indra Kurniawan, Lasmedi Afuan. "Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada Twitter Menggunakan Support Vector Machine", Jurnal Eksplora Informatika, 2023

Publication

<1%

Juanto Simangunsong, Nurmala Dewi Simanjuntak, Aprima Anugerah Matondang. "Penerapan Transfer Learning untuk Klasifikasi Citra Bunga Berbasis Convolutional Neural Network", Jurnal Minfo Polgan, 2025

Publication

65 Submitted to Universitas Maritim Raja Ali Haji

Student Paper

<1%

ejournal.akakom.ac.id

Internet Source

66 journal.aira.or.id

Internet Source

<1%

67 repository.uin-suska.ac.id

Internet Source

<1%

68

<1%

69

<1%

70	repository.um-palembang.ac.id Internet Source	<1%
71	scienceon.kisti.re.kr Internet Source	<1%
72	Aang Alim Murtopo, Maulana Aditdya, Pingky Septiana Ananda, Gunawan Gunawan. "PENERAPAN COMPUTER VISION UNTUK MENDETEKSI KELENGKAPAN ATRIBUT SISWA MENGGUNAKAN METODE CNN", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2024 Publication	<1%
73	Miranda Abdullah, Dakia N. Djou, Munkizul Umam Kau. "Ujaran Kebencian Netizen Pendukung Anies terhadap Prabowo di Media Sosial X Kajian Pragmatik Linguistik Forensik", Jurnal Pengabdian Masyarakat dan Riset Pendidikan, 2025 Publication	<1%
	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	
74	T. Mariprasath, Kumar Reddy Cheepati, Marco Rivera. "Practical Guide to Machine Learning, NLP, and Generative AI: Libraries, Algorithms, and Applications", River Publishers, 2024 Publication	<1%
75	adachihayao.cocolog-nifty.com Internet Source	<1%
	ejournal3.undip.ac.id Internet Source	
76		<1%
77		

		<1%
78	jurnal.kdi.or.id Internet Source	<1%
79	lup.lub.lu.se Internet Source	<1%
80	p3m.sinus.ac.id Internet Source	<1%
81	publikasi.uyelindo.ac.id Internet Source	<1%
82	repository.ipb.ac.id Internet Source	<1%
83	Atiqah Hanifah Shalihah, Gigih Forda Nama, Puput Budi Wintoro, Yessi Mulyani. "Implementasi Algoritma Naïve Bayes Classifier & Support Vector Machine Untuk Klasifikasi Sentimen Menggunakan Data Twitter Terhadap Bakal Calon Presiden Indonesia Tahun 2024", Jurnal Profesi Insinyur Universitas Lampung, 2024 Publication	<1%
84	"Projetos de Pesquisa em Data Science - Estudos de caso do curso CEDS IBAPE-PR", Editora Cientifica Digital, 2024 Publication	<1%
85	Ratu Nurmalika, Makmun Makmun, Bambang Yulianto, Ichsani Mursidah, Dhian Sweetania, Puji Sularsih. "Paradigma Klasifikasi Ragam Seni Lukis Berbasis Convolutional Neural Network (CNN) Dengan MobileNetV2 Dan Implementasi Pada Postman Melalui Flask Api", Jurnal Minfo Polgan, 2025 Publication	<1%

