JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 3, September 2024, Pp. 1431-1441

# IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK UNTUK ANALISIS SENTIMEN BACAPRES 2024 PADA KOLOM KOMENTAR YOUTUBE MATA NAJWA

**Dany Eka Saputra*[1)], Auliya Rahman Isnain[2)]**
1. Universitas Teknokrat Indonesia
2. Universitas Teknokrat Indonesia

**Article Info**
**Kata Kunci:** CNN; Demokrasi; Mata Najwa; Sentimen Analysis; Youtube

**Keywords:** *CNN; Democratic; Mata Najwa; Sentiment Analysis; Youtube*

**Article history:**
Received 8 June 2024
Revised 19 July 2024
Accepted 6 August 2024
Available online 1 September 2024

**DOI :**
https://doi.org/10.29100/jipi.v9i3.5420

* Corresponding author.
Corresponding Author
E-mail address:
dany_eka_saputra@teknokrat.ac.id

**ABSTRAK**
Indonesia sebagai salah satu negara berpenduduk padat dengan sistem demokrasi, Penelitian ini berfokus pada analisis sentimen terhadap calon presiden dan wakil presiden 2024 melalui komentar YouTube di "Mata Najwa." Memanfaatkan Convolutional Neural Network (CNN) pada 45.736 komentar, penelitian ini mencapai akurasi keseluruhan 91% yang mengesankan. Metode CNN, menggunakan fase arsitektur dan fine-tuning dengan pengoptimal Adam, secara efektif mengkategorikan sentimen ke dalam kelas positif, negatif, dan netral. Kemahiran model dalam menavigasi dinamika bahasa dan fluktuasi opini publik menunjukkan dampak positifnya pada tantangan analisis sentimen dalam konteks politik platform media sosial seperti YouTube. Penelitian ini menyoroti kemanjuran CNN dalam menangani seluk-beluk wacana politik dalam skala besar, menawarkan wawasan berharga tentang sentimen publik selama musim pemilihan.

**ABSTRACT**
Indonesia as one of the densely populated countries with a democratic system, this study focuses on sentiment analysis of 2024 Presidential and Vice Presidential candidates through YouTube comments in "Mata Najwa." Utilizing the Convolutional Neural Network (CNN) at 45,736 comments, the study achieved an impressive 91% overall accuracy. The CNN method, using architectural phases and fine-tuning with Adam's optimizer, effectively categorizes sentiments into positive, negative, and neutral classes. The model's proficiency in navigating language dynamics and fluctuations in public opinion shows its positive impact on sentiment analysis challenges in the political context of social media platforms such as YouTube. The research highlights CNN's efficacy in addressing the intricacies of political discourse at scale, offering valuable insights into public sentiment during election season.

## I. PENDAHULUAN

INDONESIA sebagai negara dengan populasi terbesar keempat mencapai angka 274.790.244, menerapkan sistem demokrasi. Pelaksanaan pemilihan umum di Indonesia dilakukan setiap lima tahun [1]. Pemilihan umum tersebut mencakup suara mayoritas untuk presiden, DPR, gubernur, bupati, hingga kepala desa. Pemilihan umum yang merupakan ciri khas negara demokratis, diadakan secara berkala. Tahun 2024 ialah momen puncak pesta demokrasi untuk seluruh masyarakat Indonesia, sebab masa periode kepengurusan presiden dan wakil presiden akan segera berakhir. Banyak masyarakat Indonesia ingin menyuarakan dukungan kepada calon presiden dan wakil presiden.

Youtube merupakan platform yang mendominasi dalam menyajikan laporan dalam bentuk video dengan jumlah pengguna paling besar. Para pengguna dimungkinkan berinteraksi melalui berbagai video, memberi respon berupa *like* ataupun *dislike,* serta meningkatkan jumlah penonton dengan *subscribe* pada saluran (*channel*) tertentu [2]. Youtube menyediakan sarana bagi pengguna untuk menyampaikan pendapat pada video melalui komentar. Kolom komentar di platform tersebut dapat digunakan sebagai sumber data untuk melakukan analisis media sosial. Data komentar yang dihasilkan oleh pengguna merupakan gambaran langsung dari pandangan, perasaan, dan opini mereka terhadap topik yang dibahas dalam video.

Pemanfaatan data komentar Youtube dapat memberikan wawasan tentang pandangan masyarakat terhadap berbagai topik, termasuk politik dan pemilihan umum. Dalam konteks ini, salah satu Chanel Youtube yang membahas tentang pemilihan umum yaitu Mata Najwa yang berjudul 3 Bacapres Bicara Gagasan, terkait dengan calon presiden dalam pemilihan umum 2024. Data komentar youtube tidak hanya berfungsi sebagai sumber informasi, tetapi juga bahan untuk memahami opini dan sikap masyarakat terhadap calon presiden dalam pemilihan umum 2024. Penerapan analisis sentimen terhadap data komentar

Youtube memberikan kesempatan bagi peneliti untuk memahami lebih dalam dinamika opini publik dalam skala yang luas. Melalui pendekatan ini, dapat mengidentifikasi tren dan pola sentimen yang muncul, serta memperoleh wawasan yang lebih akurat tentang sikap masyarakat terhadap berbagai isu politik.

Analisis sentimen ialah teknik yang dimanfaatkan menganalisis sentimen, *feedback,* opini, sekaligus emosi terhadap produk, layanan, ataupun situasi tertentu. Tujuan mendasar analisis sentimen ialah untuk mengkategorikan kata, frasa, hingga kalimat yang merupakan materi sentimen ataupun opini ke dalam kategori negatif, netral, ataupun positif [3]. Analisis sentimen muncul sebagai alat penting untuk memahami keragaman pandangan. Analisis sentimen, sering pula dinamakan *opinion mining,* yakni teknik pemrosesan bahasa alami guna menafsirkan sekaligus mengkategorikan emosi dalam data subjektif seperti email, sosial postingan media, dan hasil survei. Analisis sentimen bertujuan untuk menentukan alat otomatis untuk mengekstrak informasi subjektif teks yang ditulis dalam bahasa alami, seperti sentimen maupun opini, untuk mengembangkan pengetahuan terstruktur serta berharga baik sebagai sistem pendukung keputusan atau pengambil keputusan [4]. Teknik ini memungkinkan untuk mengklasifikasikan opini masyarakat dalam kategori positif, negatif, dan netral. Namun, di tengah kompleksitas dinamika sosial-politik, permasalahan muncul dalam merespons fluktuasi cepat opini publik, dinamika bahasa yang beragam, serta tantangan dalam memastikan akurasi analisis sentimen.

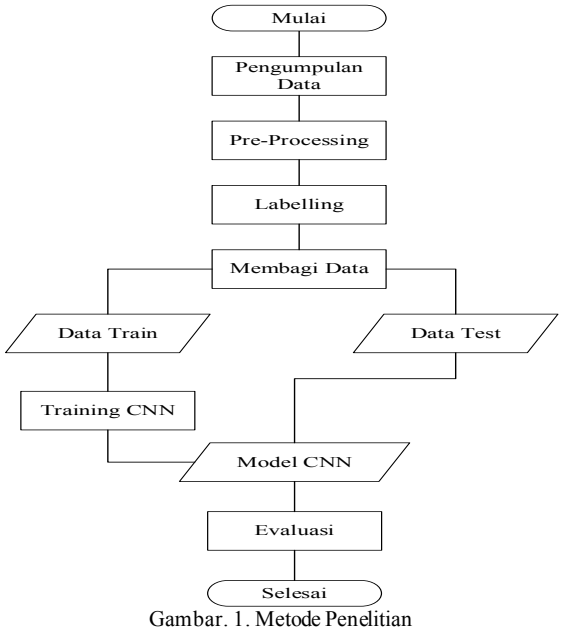
Meskipun analisis sentimen merupakan alat yang kuat untuk memahami pandangan dan reaksi masyarakat, terdapat sejumlah tantangan dan batasan yang perlu diperhatikan dalam melakukan penelitian ini. Salah satu tantangan utama adalah fluktuasi cepat opini publik. Opini dan sikap masyarakat dapat berubah dengan cepat sebagai respons terhadap peristiwa atau informasi baru yang muncul. Oleh karena itu, analisis sentimen harus dilakukan secara teratur dan dengan cepat untuk mengikuti perubahan dalam opini publik. Dinamika bahasa yang beragam juga menjadi faktor penting dalam analisis sentimen. Bahasa sering kali memiliki nuansa, makna ganda, dan konteks yang kompleks, yang dapat memengaruhi interpretasi sentimen secara signifikan. Oleh karena itu, diperlukan pendekatan yang cermat dalam memahami dan mengklasifikasikan teks, terutama dalam konteks analisis sentimen yang melibatkan data komentar YouTube yang mungkin penuh dengan beragam gaya bahasa dan ekspresi. Selain itu, ada batasan teknis dan metodologis yang perlu diperhatikan dalam analisis sentimen, seperti akurasi klasifikasi sentimen dan interpretasi konteks. Meskipun algoritma dan teknik analisis sentimen terus berkembang, masih ada kemungkinan terjadinya kesalahan dalam klasifikasi sentimen, terutama dalam kasus teks yang kompleks atau ambigu. Penting untuk memperhatikan konteks dan makna di balik teks, karena hal ini dapat memengaruhi interpretasi sentimen secara keseluruhan.

Salah satu metode analisis sentimen adalah menggunakan *Deep Learning*. *Deep learning* merupakan salah satu cabang dari *machine learning* yang memiliki kelebihan dibandingkan dengan *machine learning* biasa yaitu dapat melakukan ekstraksi fitur secara otomatis dari kata mentah secara detail [5]. *Deep Learning* memiliki berbagai macam metode salah satunya yaitu *Convolutional Neural Network* (CNN). Secara umum, CNN dapat diartikan pula sebagai suatu algoritma yang sering dimanfaatkan untuk mengolah data berupa gambar maupun teks [6]. Metode CNN juga memiliki kelebihan dibandingkan dengan metode Deep Learning lainnya, CNN memiliki lebih sedikit parameter, dan mudah untuk dilakukan proses training [7]. Metode CNN dipilih karena lebih cocok digunakan pada data berformat matriks seperti data citra dan data teks yang diubah dalam bentuk *numerik*. CNN akan diterapkan dengan menggunakan representasi vektor kata (*word embeddings*) untuk mewakili teks komentar. Jaringan CNN akan dilatih menggunakan data komentar yang diberi label sentimen untuk mengklasifikasikan setiap komentar ke dalam kategori sentimen yang sesuai. Metode CNN terbukti memiliki akurasi yang lebih tinggi dibandingkan dengan metode lain. Dalam penelitian yang dilakukan oleh (Simbolon et al., 2021) CNN memiliki akurasi 86% sedangkan SVM memiliki akurasi 83% [8]. Bahkan menurut penelitian yang dilakukan oleh Irawan & Rochmah (2022), yang berjudul "penerapan algoritma *CNN* untuk mengetahui sentimen Masyarakat terhadap kebijakan vaksin *covid-19*" menghasilkan akurasi terbaik 98,66%, rata *precision* 98,33%, *recall* 98,33%, serta *fi-score* 98,66% [9]. Oleh karena itu penelitian ini menggunakan metode CNN yang diterapkan pada data komentar Youtube Mata Najwa. penelitian ini bertujuan untuk mengklasifikasikan sentimen terhadap calon presiden 2024 ke dalam tiga kategori: positif, negatif, dan netral, dan melakukan prediksi dengan mencari nilai klasifikasi yang efisien dan akurat berdasarkan respons komentar.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 3, September 2024, Pp. 1431-1441

## II. METODE PENELITIAN

Metode penelitian merujuk pada serangkaian kegiatan yang dilaksanakan dengan metode yang teratur guna mencapai tujuan penelitian. Metode penelitian terlihat pada Gambar 1.




Gambar. 1. Metode Penelitian

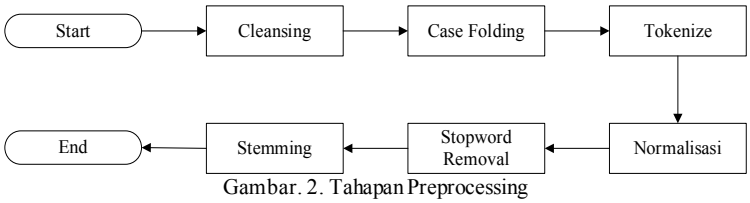
### A. *Pengumpulan Data*

Pengumpulan Data dalam penelitian ini dilakukan dengan Crawling Data. Crawling adalah istilah dalam teknologi informasi yang merujuk pada proses pengumpulan data dari internet secara otomatis dengan menggunakan perangkat lunak khusus yang disebut web crawler atau spider [10]

Data dikumpulkan dengan memanfaatkan *Google Colaboratory* untuk melakukan *crawling* data komentar youtube melalui chanel mata najwa dengan judul "3 Bacapres Bicara Gagasan". Dibutuhkan Youtube Data API V3 dan *link URL chanel* youtube mata najwa untuk melakukan *crawling*. Youtube API adalah sebuah layanan yang disediakan oleh youtube untuk para pengembang, API yang disediakan oleh youtube bersifat public API yang berarti youtube memberikan akses secara terprogram terhadap data dan layanan yang ada di website tersebut. Youtube API disediakan oleh youtube untuk para pengembang untuk membuat application yang dapat berinteraksi dengan resource video – video yang ada di Youtube [11].

### B. *Preprocessing*

Pemrosesan data atau *preprocessing teks* memiliki peran krusial dalam *Text Mining* dan *Natural Language Processing* (NLP). *Preprocessing* merujuk pada tugas penting yang melibatkan transformasi data mentah menjadi bentuk yang dapat digunakan. Dalam konteks ini, langkah-langkah *preprocessing* melibatkan beberapa tahap yang disesuaikan dengan kebutuhan penelitian, seperti *case folding*, *tokenizing*, *stop word removal*, *slang word*, *normalization,* dan *stemming* [12]. Tahapan *preprocessing* data ditunjukkan pada Gambar 2.




Gambar. 2. Tahapan Preprocessing

### C. *Labelling*

Pelabelan dilakukan dengan membagi data menjadi tiga kategori, yaitu memberikan label 1 pada data netral, label 2 pada data positif, dan label 0 pada data negatif. Model arsitektur transformer memungkinkan label diberikan secara otomatis. Transformator memanfaatkan beberapa mekanisme perhatian untuk membentuk arsitektur yang secara signifikan mengungguli tugas terjemahan mesin yang canggih tanpa menggunakan lapisan berulang [13].

### D. *Data Training dan Data Testing*

Setelah melalui proses *preprocessing* dan *labeling* serta memastikan keseimbangan dataset, dataset siap untuk diolah lebih lanjut. Tahap selanjutnya melibatkan pembagian dataset ke dalam 2 bagian, yaitu *data testing* serta *data training*. *Data testing* berfungsi menguji kinerja model, sementara *data training* berfungsi melatih sekaligus mengembangkan model.

### E. *Modelling CNN*

Pada pengembangan model analisis sentimen yang mampu mengklasifikasikan sentimen berkategori positif, netral, ataupun negatif, dibutuhkan penerapan algoritma klasifikasi. Adapun penelitian ini menggunakan algoritma CNN satu dimensi dengan empat lapisan esensial, termasuk *embedding, conv1D, fully connected, serta pooling.*

### F. *Evaluasi*

Evaluasi dilakukan dengan melakukan pengujian menggunakan data yang sudah dibagi menjadi *data testing* maupun *data training*. Selanjutnya dijalankan pengujian tingkat performa dengan *confusion matrix* dan menggunakan *Word2Vec*. Performa diukur dengan metrix seperti *accuracy*, *precicion*, *recall*, dan *f1-score*.

*Word2Vec* adalah sebuah metode atau algoritma *natural language processing* yang digunakaan untuk menciptakan representasi vektor kata (word embeddings) dari teks [14]. Tujuan dari *Word2vec* adalah untuk menghubungkan setiap kata dalam teks ke dalam ruang vektor, di mana kata-kata dengan makna yang mirip atau sering kali muncul bersama-sama akan memiliki representasi vektor yang berdekatan satu sama lain [15].

*Confusion matrix* ialah metode yang dimanfaatkan dalam pengukuran kinerja metode klasifikasi. *Confusion matrix* berisi data perbandingan output klasifikasi sistem dengan tujuan mengukur tingkat akurasi. Metode ini umumnya digunakan dalam penghitungan akurasi didalam proses *data mining* [16]. Ada 4 istilah terkait output klasifikasi dalam *confusion matrix* yaitu:

*Precision* ialah metrik kepastian yang mengukur persentase tuple yang benar-benar diberi label positif.

$$Precision = \frac{TP}{FP+TP} \quad (1)$$

*Recall* ialah metrik kelengkapan yang mengukur keberhasilan sistem didalam menemukan informasi berulang.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1-Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan.

$$F1 = 2X\frac{recall\ x\ precision}{recall + precision} \quad (3)$$

Accuracy merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data.

$$Accuracy = \frac{TP+TN}{TP + TN+FP+FN} \quad (4)$$

## III. HASIL PEMBAHASAN

### A. *Pengumpulan Data*

Pengumpulan data atau proses crawling data komentar Youtube menggunakan Youtube Data API v3 dan proses pengambilan data dibantu dengan Bahasa pemrograman python. Youtube Data API v3 adalah antarmuka pemrograman aplikasi (API) yang disediakan oleh Youtube untuk mengakses dan mengelola data dari platform Youtube. Youtube Data API v3 itu sendiri memiliki suatu *api_key, video_id* untuk digunakan mengakses data komentar Youtube. Dalam penelitian ini api_key didapatkan melalui https://console.cloud.google.com/ dan video_id didapatkan melalui chanel youtube mata najwa dengan judul 3 Bacapres Bicara Gagasan yang dapat dilihat pada Gambar 3.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 3, September 2024, Pp. 1431-1441

```
# isikan dengan api key Anda
api_key = 'AIzaSyBIk_Ff7W0BsM01x95H2bkYgDEb8oZSZHE'

# Enter video id
video_id = "C2aZPjVdqyA" #isikan dengan kode / ID video

# Call function
comments = video_comments(video_id)

comments
```

Gambar. 3. Youtube Data API v3

Selain membutuhkan Youtube Data API v3 untuk dapat melakukan pengambilan data komentar Youtube, peneliti menggunakan tools pendukung untuk menganalisis sentimen seperti Google Colaboratory. Google colaboratory lebih umum disebut sebagai "google colab" atau hanya sekadar "colab" adalah proyek penelitian untuk prototipe model pembelajaran mesin pada opsi perangkat keras yang kuat seperti GPU dan TPU. ini menyediakan lingkungan jupyter notebook tanpa server untuk pengembangan interaktif. Google colab bebas digunakan seperti produk G Suite lainnya [17]. Google Colab adalah coding environment bahasa pemrograman Python dengan format "notebook" (mirip dengan Jupyter notebook), atau dengan kata lain Google seakan meminjami komputer secara gratis untuk membuat program oleh Google.

Dengan adanya Google Colaboratory peneliti dapat melakukan proses crawling data menggunakan bahasa pemrograman Python seperti Gambar 4.

```
[ ] import pandas as pd
    from googleapiclient.discovery import build
```

```
[ ] def video_comments(video_id):
        replies = []
        youtube = build('youtube', 'v3', developerKey=api_key)
        video_response = youtube.commentThreads().list(part='snippet,replies', videoId=video_id).execute()
        while video_response:
            for item in video_response['items']:
                published = item['snippet']['topLevelComment']['snippet']['publishedAt']
                user = item['snippet']['topLevelComment']['snippet']['authorDisplayName']
                comment = item['snippet']['topLevelComment']['snippet']['textDisplay']
                likeCount = item['snippet']['topLevelComment']['snippet']['likeCount']
                replies.append([published, user, comment, likeCount])
                replycount = item['snippet']['totalReplyCount']
                if replycount>0:
                    for reply in item['replies']['comments']:
                        published = reply['snippet']['publishedAt']
                        user = reply['snippet']['authorDisplayName']
                        repl = reply['snippet']['textDisplay']
                        likeCount = reply['snippet']['likeCount']
                        replies.append([published, user, repl, likeCount])
            if 'nextPageToken' in video_response:
                video_response = youtube.commentThreads().list(
                        part = 'snippet,replies',
                        pageToken = video_response['nextPageToken'],
                        videoId = video_id
                    ).execute()
            else:
                break
        return replies
```

Gambar. 4. Source Code Crawling Data

Pada Gambar 4 proses pendeklarasian library pandas yang berguna untuk pengolahan data dan library googleapiclient.discovery merupakan bagian dari paket google-api-python-client yang menyediakan alat untuk berinteraksi dengan barbagai API Google, termasuk Youtube Data API v3.

Pada saat melakukan crawling, peneliti mengambil kriteria komentar dengan mengambil komentar utama dari setiap benang komentar (*comment thread*) dan juga mengambil balasan untuk setiap komentar utama yang memiliki balasan. Dengan demikian, peneliti memastikan bahwa tidak hanya komentar langsung terhadap video yang direpresentasikan, tetapi juga balasan yang diberikan oleh pengguna atas komentar tertentu. Selain itu, peneliti juga mengambil informasi penting dari setiap komentar, seperti waktu publikasi, nama pengguna, teks komentar, dan jumlah suka yang diterima. Dengan demikian, data yang diambil dari crawling ini dapat memberikan representasi yang memadai dari komentar yang terkait dengan video YouTube yang diteliti. Pada Gambar 5 menunjukkan hasil Crawling Data.

| | Unnamed: 0 | publishedAt | authorDisplayName | text | likeCount |
|---|---|---|---|---|---|
| 0 | 0 | 2024-02-19T05:44:33Z | @annaskubosze | Closing paling keren ya <br>Pak menhan | 0.0 |
| 1 | 1 | 2024-02-19T05:03:42Z | @annaskubosze | Terimakasih <br><br> Saya nonton langsung skip... | 0.0 |
| 2 | 2 | 2024-02-18T01:34:51Z | @juliaroma9852 | Setelah pemilihan baru sempat lihat. Acara ini... | 0.0 |
| 3 | 3 | 2024-02-16T12:25:38Z | @muhamadrizki8881 | Mungkin maksud prabowo, dia ga mau melihat cer... | 0.0 |
| 4 | 4 | 2024-02-16T06:39:32Z | @elngmrggsti | <a href="https://www.youtube.com/watch?v=C2aZP... | 1.0 |
| ... | ... | ... | ... | ... | ... |
| 45731 | 44880 | 2023-09-19T15:29:25Z | @Demoncy45 | Pilih dubur apa pilih extremist,, | 0.0 |
| 45732 | 44881 | 2023-09-19T15:31:43Z | @DwiAndikaYT18 | SIAPA DISINI YANG CAPRESNYA EMOSIAN, PENCULIK ... | 0.0 |
| 45733 | 44882 | 2023-09-19T15:31:46Z | @tabyathherms5946 | Banteng still the winner. FACT!!! | 1.0 |
| 45734 | 44884 | 2023-09-19T15:24:04Z | @animestel | Aldi Taher for The prosperity of Indonesia <a ... | 2.0 |
| 45735 | 44885 | 2023-09-19T15:33:47Z | @DwiAndikaYT18 | ANIES - ALDI TAHER COCOK | 0.0 |

Gambar. 5. Hasil Crawling Data

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 3, September 2024, Pp. 1431-1441

Hasil Crawling data yang diambil dalam rentang waktu September hingga Februari 2024 diperoleh data sebanyak 45.736 komentar dan data disimpan dalam bentuk file csv.

## B. *Preprocessing*

*Preprocessing* ialah Langkah awal *text mining* yang mencakup persiapan input data tekstual supaya bisa diproses ke dalam langkah selanjutnya [18]. *Processing* ialah langkah yang melibatkan sejumlah teknik, termasuk *cleaning data*, *case folding*, *tokenize data*, *stopword removal data*, *normalization data*, dan *stemming data*. Hasil dari teknik-teknik tersebut akan diuraikan secara rinci pada bagian ini.

### 1) *Cleansing Data*

*Cleansing Data* yakni proses pembersihan data teks yang tidak relevan maupun konsisten. Fungsi *cleansing* yakni menghapus karakter yang dinilai tidak penting, mencakup *hashtag, username,* angka, *url, emoticon,* serta tanda baca [19]. Hasil tahap *cleansing data* bisa teramati melalui Tabel I.

TABEL I
CLEANSING DATA

| Komentar (sebelum | Cleansing Data |
|---|---|
| Pecahh pak prabowo😍😍 | Pecah pak prabowo |
| Prabowo gas poll mbk Nana pasti 1 putaran. | Prabowo gas pol mbk Nana pasti putaran |
| Closing paling keren ya <br>Pak menhan | Closing paling keren ya Pak menhan |
| Ngomong sich nyaman, tindakan nyatanya mana.. | Ngomong sich nyaman tindakan nyatanya mana |

### 2) *Case Folding*

Case folding merupakan proses mengubah huruf kapital menjadi huruf kecil, bertujuan memudahkan perbandingan ataupun komparasi teks [20]. Didalam langkah ini, terjadi modifikasi huruf kapital ke dalam huruf kecil. Jika tidak dilakukan case folding, dua string akan dianggap berbeda. Tetapi, setelah melalui case folding, keduanya dianggap identik sebab telah dikonversi menjadi huruf besar ataupun kecil. Rincian proses case folding bisa diamati melalui Tabel II.

TABEL II
CASE FOLDING

| Cleansing Data | Case Folding |
|---|---|
| Pecah pak prabowo | pecah pak prabowo |
| Prabowo gas pol mbk Nana pasti putaran | prabowo gas pol mbk nana pasti putaran |
| Closing paling keren ya Pak menhan | closing paling keren ya pak menhan |
| Ngomong sich nyaman tindakan nyatanya mana | ngomong sich nyaman tindakan nyatanya mana |

### 3) *Tokenize*

*Tekonize* ialah proses split kalimat tertentu kedalam beberapa potongan [21]. Tujuan dari *tokenize* ialah untuk menyederhanakan proses pengolahan maupun analisis suatu teks dengan menekan kompleksitas sekaligus memaksimalkan efektivitas pemrosesan data. Selama proses ini, dilakukan pula penghapusan angka, tanda baca, maupun karakter lainnya yang dinilai tidak berpengaruh bagi pemrosesan teks. Penggunaan *tokenize* dapat diamati pada Tabel III.

TABEL III
TOKENIZE

| Case Folding | Tokenize |
|---|---|
| pecah pak prabowo | [pecah, pak, prabowo] |
| prabowo gas pol mbk nana pasti putaran | [prabowo, gas, pol, mbk, nana, pasti, putaran] |
| closing paling keren ya pak menhan | [closing, paling, keren, ya, pak, menhan] |
| ngomong sich nyaman tindakan nyatanya mana | [ngomong, sich, nyaman, tindakan, nyatanya, mana] |

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 3, September 2024, Pp. 1431-1441

*4) Normalization*

Normalisasi bertujuan mengubah data yang tidak sesuai dengan ejaan Bahasa Indonesia menjadi kata yang baku. Langkah ini melibatkan pembuatan kamus data yang mencakup koreksi terhadap kesalahan umum dalam ejaan, serta penambahan entri untuk kesalahan yang terdapat dalam dataset. Proses normalisasi dapat dilihat Tabel IV.

TABEL IV
NORMALIZATION

| Tokenize | Normalization |
|---|---|
| [pecah, pak, prabowo] | [pecah, bapak, prabowo] |
| [prabowo, gas, pol, mbk, nana, pasti, putaran] | [prabowo, gas, pol, mbak, nana, pasti, putaran] |
| [closing, paling, keren, ya, pak, menhan] | [penutupan, paling, keren, iya, bapak, menteri pertahanan] |
| [ngomong, sich, nyaman, tindakan, nyatanya, mana] | [mengomong, sih, nyaman, tindakan, nyatanya, mana] |

*5) Stopwords Removal*

*Stopwords Removal* bermanfaat dalam mengurangi kata umum yang tidak memberikan kontribusi signifikan dan dianggap tidak memiliki makna [22]. Tujuan dari *stopwords removal* ialah mempercepat pemrosesan teks melalui reduksi kuantitas kata. Stopwords melibatkan jenis kaya seperti kata depan, kata ganti, dan kata penghubung, seperti contoh ‘dan’,‘atau’,’di’.’ke’. Hasil *stopwords* terlihat pada Tabel V.

TABEL V
STOPWORDS REMOVAL

| Normalization | Stopwords Removal |
|---|---|
| [pecah, bapak, prabowo] | [pecah, prabowo] |
| [prabowo, gas, pol, mbak, nana, pasti, putaran] | [prabowo, mbak, nana, putaran] |
| [penutupan, paling, keren, iya, bapak, menteri pertahanan] | [penutupan, keren, iya, menteri pertahanan] |
| [mengomong, sih, nyaman, tindakan, nyatanya, mana] | [mengomong, nyaman, tindakan] |

*6) Stemming*

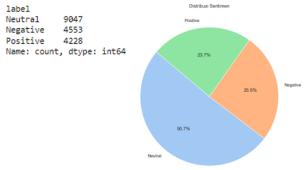
Dengan bantuan pustaka *python sastrawi*, *StemmerFactory* adalah modul yang digunakan dalam *stemming* untuk mengganti kata dengan imbuhan menjadi kata dasar**.** Tabel VI menunjukkan hasil *stemming*.

TABEL VI
STEMMING

| Stopwords Removal | Stemming |
|---|---|
| [pecah, prabowo] | prabowo |
| [prabowo, mbak, nana, putaran] | prabowo mbak nana putar |
| [penutupan, keren, iya, menteri pertahanan] | tutup keren iya menteri tahan |
| [mengomong, nyaman, tindakan] | omong nyaman tindak |

## C. Labeling Data

Setelah tahap *pre-processing*, selanjutnya data yang dikumpulkan akan diproses labelling dengan cara membagi ke dalam tiga kategori labelling data yaitu *positive*, *negative* dan *neutral*. Perolehan data yang didapat dari hasil *preprocessing data* sebanyak 17.828 data komentar.

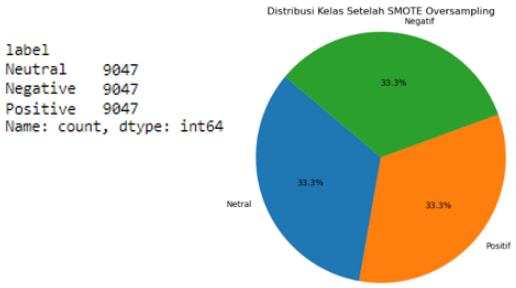



Gambar. 6. Diagram Labelling Data

Selanjutnya setelah melalui tahap pelabelan data dan pengelompokan menjadi tiga kategori label *positive*, *negative* dan *neutral*, diperoleh dataset dengan jumlah keseluruhan 17828 data. Dataset tersebut terdiri dari 9047 data dengan label *neutral*, 4553 data dengan label *negative*, dan 4228 data dengan label *positive*. Distribusi data yang telah diberi label dapat dilihat melalui diagram pada Gambar 6.

### *D. Random Oversampling*

Dalam penelitian ini penggunaan teknik random oversampling bertujuan untuk menyeimbangkan distribusi kelas dengan meningkatkan jumlah sampel dari kelas minoritas. Ini dapat membantu meningkatkan representasi kelas minoritas dan menghindari dominasi kelas mayoritas dalam model. Peningkatan jumlah sampel dari kelas minoritas dapat menyebabkan overfitting pada data pelatihan dan kemungkinan penurunan kinerja model pada data uji. SMOTE dipilih sebagai teknik oversampling karena mampu mengatasi kekurangan dari random oversampling dengan menciptakan sampel sintetis baru dari kelas minoritas. Ini membantu meningkatkan representasi kelas minoritas tanpa mengulang-ulang sampel yang ada. *SMOTE* adalah salah satu turunan dari oversampling. SMOTE pertama kali diperkenalkan oleh Nithes V. Chawla [23]. Pendekatan ini bekerja dengan membuat replikasi dari data minoritas. Replikasi tersebut dikenal dengan data sintetis (*syntetic data*). Metode SMOTE bekerja dengan mencari k-nearest neighbors (yaitu ketetanggaan terdekat data sebanyak k) untuk setiap data di kelas minoritas, setelah itu dibuat data sintetis sebanyak persentase duplikasi yang diinginkan antara data minor dan k-nearest neighbors yang dipilih secara acak. Hasil dari langkah *random oversampling* terlihat pada Gambar 7.




### *E. Data Pelatihan dan Data Pengujian*

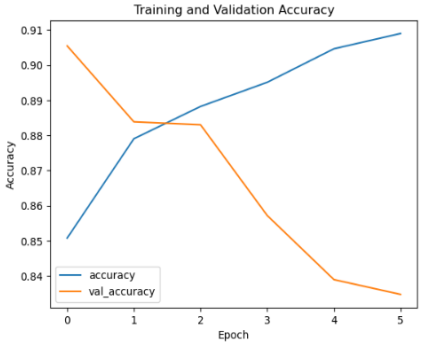
Dalam penelitian ini, data awal dibagi menjadi dua yaitu: data pelatihan *('train_df')* dan data pengujian *('test_df')*. Dilakukan menggunakan fungsi *library Scikit-learn* yang dikenal sebagai *"train_test_split"* untuk membagi data uji menjadi 20% dari total data, dan *"random_state"* digunakan untuk memastikan hasil pemisahan yang konsisten. Kode program dapat dilihat pada Tabel VII.

TABEL VII
KODE PROGRAM PEMBAGIAN DATA

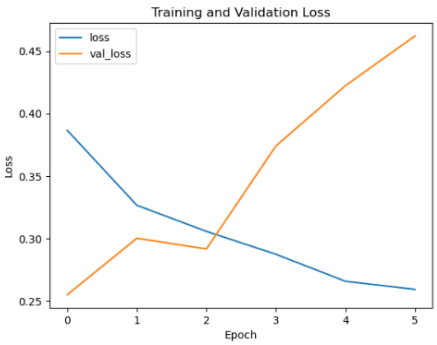
```
# Split data into train and test sets
train_df, test_df = train_test_split(df, test_size=0.2, random_state=42)
```

### *F. Klasifikasi Model CNN*

Pada tahap ini, model klasifikasi akan diuji pada data teks yang telah melewati proses *preprocessing* dan *balancing data*. Pengujian menggunakan arsitektur CNN yang dirancang khusus untuk tugas analisis sentimen pada teks. Proses klasifikasi dimulai dengan mengonversi teks menjadi representasi *vektor* dan melakukan *fine-tuning* menggunakan *optimizer* Adam.




Gambar 8. Grafik Statistik *Accuracy*




Gambar 9. Grafik Statistik *Loss*

Pada Gambar 8 diatas terlihat bahwa grafik hasil *accuracy* mendapatkan nilai *accuracy training* sebesar 0.90, sedangkan pada hasil *val_accuracy* mendapatkan nilai *accuracy validation* sebesar 0.91. Pada Gambar 9 terlihat bahwa grafik *loss* mendapatkan nilai *loss function* pada *training set* sebesar 0.2593, sedangkan pada hasil *val_loss* mendapatkan nilai *loss validation* sebesar 0.4621.

## *G.* Evaluasi Model

Pada tahap ini, dilakukan evaluasi model menggunakan arsitektur *Convolutional Neural Network* (CNN) dengan konfigurasi *ConvID, GlobalMaxPooling1D, dan Dropout*. Model ini menerapkan *optimizer* Adam dengan tingkat pelatihan sebesar 0,001, ukuran batch 32, dan melibatkan proses pelatihan selama 5 *epoch*. Evaluasi dilakukan terhadap hasil pelatihan model terhadap data validasi.

TABEL VIII
EVALUASI MODEL CNN

| Test Accuracy | Loss | Accuracy |
|---|---|---|
| 0,9090 | 0,2593 | 98.46% |

Setelah mengevaluasi model menggunakan data testing, diperoleh hasil loss sebesar 0.2593 dan akurasi sebesar 0.9090. Hasil evaluasi tersebut mencerminkan kinerja model pada dataset uji. Pada Tabel VIII menampilkan detail evaluasi model.

TABEL IX
KLASIFIKASI MODEL CNN

| | *Precission* | *Recall* | *F1-Score* | *support* |
|---|---|---|---|---|
| *Positive* | 0.94 | 0.83 | 0.88 | 926 |
| *Negative* | 0.89 | 0.94 | 0.91 | 1800 |
| *Neutral* | 0.91 | 0.92 | 0.92 | 839 |
| *Accuracy* | | | 0.91 | 3565 |
| *Macro avg* | 0.91 | 0.90 | 0.90 | 3565 |
| *Weighted avg* | 0.91 | 0.91 | 0.90 | 3565 |

Dapat dilihat pada Tabel IX Hasil evaluasi menunjukkan bahwa model bekerja dengan baik memiliki akurasi keseluruhan sebesar 91%, dan nilai *precision*, *recall*, dan *f1-score* yang sangat tinggi untuk setiap kelas. Hasil ini menunjukkan seberapa baik model dapat mengklasifikasikan data pada analisis sentimen.

Selanjutnya hasil evaluasi model *Convolutional Neural Network* (CNN) dilakukan prediksi label untuk mengetahui sejauh mana model tersebut dalam memprediksi teks baik *positive*, *negative* dan *neutral*. Tabel X menunjukkan hasil prediksi.

TABEL X
PREDIKSI MODEL CNN

| | |
|---|---|
| *Text:* | Pembangunan saat ini semakin bagus dan keren |
| *Predicted Label* | Positive |
| *Prediction Probabilities:* | |
| *Negative* | 0.0001 |
| *Neutral* | 0.0000 |
| *Positive* | 0.9999 |
| *Text:* | berantas koruptor hukum mati miskin suap penjara hukum potong tahan hukum indonesia baik |
| *Predicted Label* | *Negative* |
| *Prediction Probabilities:* | |
| *Negative* | 1.0000 |
| *Neutral* | 0.0000 |
| *Positive* | 0.0000 |
| *Text:* | indonesia butuh orang pintar indonesia butuh pimpin jujur berani |
| *Predicted Label* | *Neutral* |
| *Prediction Probabilities:* | |
| *Negative* | 0.1124 |
| *Neutral* | 0.6254 |
| *Positive* | 0.2622 |

Hasil akhir dari penelitian ini terdapat pada tabel prediksi model. Prediksi model untuk mengklasifikasikan sentimen. Terdapat penelitian terdahulu tentang analisis sentimen komentar youtube mengenai Vaksin Covid-19

menggunakan Support Vector Machine yang dilakukan oleh (Wijayanto & Defara, n.d.) menunjukkan akurasi sebesar 86% dengan klasifikasi dua kelas prediksi yaitu positif dan negatif [24].

Berdasarkan pengujian yang telah dilakukan, peneliti melakukan perbandingan pengujian yang dilakukan oleh Ardhi Wijayanto dan Afrima Dhia Defara yang menggunakan *Support Vector Machine* dan pengujian peneliti menggunakan metode *Convolutional Neural Network*. Pengujian yang dihasilkan oleh peneliti menghasilkan akurasi rata-rata mencapai 91%. Prediksi model yang dihasilkan peneliti mengklasifikasikan tiga kelas prediksi yaitu positif, negatif, dan netral. Menunjukkan bahwa algoritma *Convolutional Neural Network* telah mencapai akurasi yang baik. Probabilitas prediksi yang tinggi pada label yang benar menunjukkan bahwa model memiliki keyakinan yang kuat dalam klasifikasinya. Interpretasi probabilitas prediksi dapat memahami sejauh mana model yakin dengan klasifikasi yang diberikan pada setiap teks. Dengan akurasi yang baik model *Convolutional Neural Network* juga bisa dipakai berbagai jenis platform dan jenis data teks yang berbeda.

## IV. KESIMPULAN

Hasil penelitian mengindikasikan bahwa penggunaan algoritma *Convolutional Neural Network* untuk analisis sentimen bacapres 2024 pada kolom komentar youtube mata najwa menghasilkan *accuracy* keseluruhan sebesar 91%. Model CNN mampu mengklasifikasikan dan memprediksi sentimen komentar ke dalam tiga kategori: *positive*, *negative,* dan *neutral* masing-masing dengan tingkat akurasi yang memuaskan. Performa model dapat diukur melalui metrik seperti *precision, recall,* dan *f1-score* untuk masing-masing kelas yang menunjukkan kemampuan model dalam mengenali sentimen dengan baik. Hal ini membuktikan bahwa metode ini sangat efektif digunakan dalam penelitian yang terkait analisis sentimen. Saran dari penulis untuk penelitian selanjutnya, melakukan pengembangan model lebih lanjut, memperhatikan pemrosesan data lebih seperti *preprocessing* dan *balancing data* untuk meningkatkan kualitas input model.

## DAFTAR PUSTAKA

[1] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook," Eksplora Informatika, vol. 9, no. 1, pp. 60–69, Sep. 2019, doi: 10.30864/eksplora.v9i1.254.
[2] K. A. B. Permana, M. Sudarma, and W. G. Ariastina, "Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan STRUCT-SVM," Majalah Ilmiah Teknologi Elektro, vol. 18, no. 1, p. 113, May 2019, doi: 10.24843/mite.2019.v18i01.p17
[3] G. Sanjaya and K. M. Lhaksmana, "Analisis Sentimen Komentar YouTube tentang Terpilihnya Menteri Kabinet Indonesia Maju Menggunakan Lexicon Based," vol. 7, no. 3, pp. 9698–9710, 2020
[4] Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2017). Sentiment analysis in social networks. Morgan Kaufmann
[5] Cholissodin, I., Sutrisno., Soebroto, A. A., Hasanah, U., & Febiola, Y. I., 2019.Ai, Machine Learning& Deep Learning(Teori & Implementasi).[e-book]Tersediadi:https://www.researchgate.net/profile/Imam-Cholissodin/publication/348003841_Buku_Ajar_AI_Machine_Learning_Deep_Learning/links/5fee9968299bf14088610ab0/Buku-Ajar-AI-Machine-Learning-Deep-Learning.pdf
[6] *Hasan Badjrie, S., Pratiwi, O.N. and Anggana, H.D. (2021) Analisis Sentimen Review Customer Terhadap Produk Indihome Dan First Media Menggunakan Algoritma Convolutional Neural Network Review Analysis Sentiment Customer Product Indihome And First Media Using Convolutional Neural Network Algorithm.*
[7] Ouyang, et al., 2015. Sentiment Analysis Using Convolutional NeuralNetwork. IEEE International Conference on Computer and Information Technology [online] Tersedia di https://ieeexplore.ieee.org/document/7363395
[8] Simbolon et al., 2021. "ANALISIS SENTIMEN APLIKASI E-LEARNING SELAMA PANDEMI COVID-19 DENGAN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN CONVOLUTIONAL NEURAL NETWORK". SEMINASTIKA 2021. DOI: 10.47002/seminas-tika.v3i1.236.
[9] Irawan, F. A., & Rochmah, D. A. (2022), "Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19," JURNAL INFORMATIKA, Vol. 9 No. 2 Oktober 2022, Halaman 148~158 ISSN: 2355-6579 | E-ISSN: 2528-2247.
[10] A. Savirani and N. Kurnia, "BIG DATA UNTUK ILMU SOSIAL ANTARA METODE RISET DAN REALITAS SOSIAL," in Social Science / General, Social Science / Sociology / General, UGM PRESS, 2021, pp. 11–26. [Online]. Available: https://www.google.co.id/books/edition/BIG_DATA_UNTUK_ILMU_SOSIAL/yHxJEAAAQBAJ?hl=en&gbpv=1&dq=analisi+sentimen&pg=PA213&printsec=frontcover.
[11] Hadjon, R. P. (2014). Implementasi Metode Rest Request pada Youtube Web Services untuk Represntasi Informasi Berbasis Timeline. Jurnal Informat-ika.
[12] N. Petty Wahyuningtyas, D. Eka Ratnawati, and N. Yudi Setiawan, "Root Cause Analysis (RCA) berbasis Sentimen menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Pengunjung Kolam Renang Brawijaya)," 2023. [Daring]. Tersedia pada: http://j-ptiik.ub.ac.id.
[13] Vaswani, A. et al. (2017) 'Attention Is All You Need'. Available at: http://arxiv.org/abs/1706.03762.
[14] Hidayatul Qudsi, D. et al. (2019) 'Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Convolutional Neural Network', Jurnal teknologi Informasi dan Ilmu Komputer (JTIIK), 8(5), pp. 1067–1076. Available at: https://doi.org/10.25126/jtiik.202184842.
[15] Mikolov, T., Yih, W.-T. and Zweig, G. (2013) Linguistic Regularities in Continuous Space Word Representations. Association for Computational Linguistics. Available at: http://research.microsoft.com/en-.
[16] Rahman, M.F. et al. (2017) „Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)", Jurnal Informatika, 11(1), p. 36. Available at: https://doi.org/10.26555/jifo.v11i1.a5452.
[17] Bisong, E. (2019). Google Colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4470-8_7
[18] S. Arafah and Fathoni, "Sentiment Analysis Pada Masyarakat Terhadap LRT Kota Palembang Menggunakan Metode Improved K-Nearest Neighbor", Jurnal Media Informatika Budidarma, vol. 6, no. 3, pp. 1554-1561, July 2022.
[19] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," Eksplora Inform., vol. 9, no. 1, pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
[20] I. Mawanta, T. S. Gunawan and Wanayumini, "Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF", Jurnal Media Informatika Budidarma, vol. 5, no. 2, pp. 726-738, April 2021.

[21] M. I. Fikri, T. S. Sabrila and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter", SMATIKA, vol. 10, no. 2, pp. 71-76, 2020.
[22] A. Firdaus, "Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19", Jurnal Riset Statistika (JRS), vol. 2, no. 2, pp. 85-92, Dec. 2022.
[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, & W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
[24] Wijayanto & Defara, n.d., "Analisis Sentimen Komentar Youtube Mengenai Vaksin Covid-19 Menggunakan Support Vector Machine", Jurnal Pilar Teknologi, Vol 7, no 1, pp. 25-31, 2022.