



## Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn)

Adit Permana Putra<sup>1</sup>, Aurelia Farrah Syafira<sup>2</sup>

<sup>1,2</sup>program Studi Statistika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

### Abstrak

Received: 17 September 2023  
Revised: 24 September 2023  
Accepted: 30 September 2023

*Sentiment analysis is a method used to identify, derive, and evaluate sentiments, opinions, or emotions contained in text. In the context of political topics, sentiment analysis can help understand public views and attitudes towards issues, presidential elections, presidential candidates, parties, etc. This information is valuable for decision makers, including the government, in the image of political figures and political life in Indonesia. This information is invaluable for decision makers, including the government, in the image of political figures and political life in Indonesia-data retrieval, data labeling, data pre-processing, data extraction, CNN classification, and naive bayes This shows that the accuracy value of the label between the negative and the requested information gives a greater true negative value than other sentiments. The large percentage recall value found in Naive Bayes with negative sentiment proves that, the ability of the model the number of correct ones is 0.95. Meanwhile, the highest percentage value of f1-score is obtained in Naive Bayes with a value of 0.81. Naive Bayes obtained an accuracy of 0.69 while CNN obtained an accuracy of 0.68. The above results show that there is a fairly thin difference in the accuracy obtained. So it can be concluded that the two models above for twitter sentiment analysis on political topics are good enough to use the above methods.*

**Kata kunci:** *sentiment analysis, naive bayes, convolutional neural networks, politics*

(\*) Corresponding Author: [Aditpermana17@Student.Uns.Ac.Id](mailto:Aditpermana17@Student.Uns.Ac.Id)

**How to Cite:** Putra A P, & Syafira A F. (2023). Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn). <https://doi.org/10.5281/zenodo.8396579>.

## PENDAHULUAN

### Latar Belakang

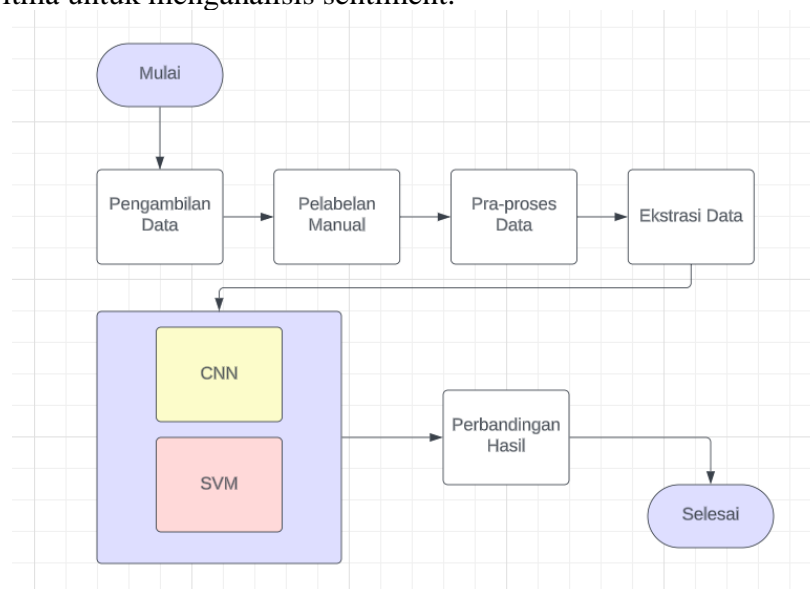
Media sosial menjadi sumber informasi yang penting dalam berbagai aspek kehidupan.. Twitter merupakan salah satu media sosial, *platform* ini termasuk populer, memberikan ruang bagi pengguna untuk berbagi pendapat, pemikiran, dan perasaan terkait topik tertentu, termasuk topik politik. Dengan jutaan cuitan yang dibagikan setiap hari, Twitter menjadi sumber data yang berharga untuk menganalisis sentimen publik.

Analisis sentimen adalah metode yang digunakan untuk mengidentifikasi, memperoleh, dan mengevaluasi sentimen, opini, atau emosi yang terkandung dalam teks. Dalam konteks topik politik, analisis sentimen dapat membantu memahami pandangan dan sikap publik terhadap isu-isu, pemilihan presiden, calon presiden, partai, dll. Informasi ini sangat berharga bagi pengambil keputusan, termasuk pemerintah, dalam citra tokoh politik dan kehidupan politik di Indonesia.

Penelitian ini bertujuan untuk mengembangkan dan membandingkan kinerja dua metode klasifikasi umum dalam analisis sentimen, yaitu Metode Naive Bayes dan CNN. Metode Naive Bayes adalah metode statistik yang menggunakan teorema Bayes dan mengasumsikan independensi kondisional antara fitur-fitur dalam data. CNN merupakan salah satu algoritma. Oleh karena itu, penelitian ini bertujuan untuk memberikan wawasan berharga bagi pengambil keputusan di bidang politik. Dalam konteks magang atau praktik kerja, penelitian ini diinisiasi untuk mengatasi tantangan yang dihadapi perusahaan dalam mengumpulkan dan menganalisis data Twitter terkait topik politik.

## METODE PENELITIAN

Sebelum memasuki tahap hasil dan pembahasan berikut adalah tahapan algoritma untuk menganalisis sentiment.



Gambar 2.1 Algoritma Penelitian

### 2.1 Pengambilan Data

Langkah pertama yang dilakukan dalam analisis sentimen adalah Pengumpulan data pada saat ini banyak metode yang bisa dilakukan. Terdapat crawling data yang menggunakan pemrograman python untuk mengambil dan metode scraping web. Data ini diambil pada saat magang dimulai yaitu per-tanggal 14 Maret 2023.

### 1.2 Labelling data

Tahap selanjutnya memberikan label atau kategori sentimen pada data yang akan dianalisis. Misalnya, data dari Twitter diamati dan diberikan label positif, negatif, atau netral. Dalam pengerjaan tahap ini ada beberapa daftar *rules* dan polaritasnya untuk mempermudah dalam pemberian label. Contohnya seperti “jika ada data yang sekilas menyatakan ajakan kebersamaan atau dukungan, tetapi disertai dengan ekspektasi yang berlawanan dan di mana si penulis pun terbukti secara logika tidak akan melakukan ajakan tersebut, kita akan padankan data tersebut dengan polaritas negatif dengan konteks menyindir”. Selain itu, ada juga panduan konteks kalimat tanya, *adjectives*, *how to differ context*, dan lain-lain. Pemberian label dalam penelitian ini juga dilakukan secara manual dan membutuhkan waktu satu bulan untuk memberi label pada 10520 data tweet.

**Tabel 1.** Hasil labelling manual

@yusuf_dumdum Prabowo aja gak gubris apa lagi kita2 gak nek.	Negatif
@PreciosaKanti Jgn dukung partai pelanggar konstitusi. Itu pasti partai anti demokrasi ! <a href="https://t.co/aPQktCMc5X">https://t.co/aPQktCMc5X</a>	Negatif
Gestur Memberi Hormat Prabowo ke Jokowi Dipuji Warganet - SINDOnews Nasional <a href="https://t.co/ENZjdNrptI">https://t.co/ENZjdNrptI</a> #PrabowoPeduli #BangkitBersama	Positif
Kalau dilihat silsilah para capres yg mau maju 2024. Ada garis turun Majapahit dan Mataram. Tergantung wahyu ratunya mau ikut siapa. Wahyu Ratu dikejar mlayu. Ditinggal melu.	Negatif

### 1.3 Pra-Proses Data

Setelah data diberi label, yang perlu dilakukan adalah *preprocessing* pada data tersebut untuk membersihkannya dan mengoptimalkan kualitasnya. Beberapa langkah yang umum dilakukan dalam *pre-processing* data adalah sebagai berikut:

- Menghapus tanda baca dan karakter khusus yang tidak relevan.
- Mengubah semua teks menjadi huruf kecil agar tidak ada perbedaan antara huruf besar dan kecil.
- Menghapus kata-kata yang tidak berguna atau stop words seperti "dan", "di", "yang", dan sebagainya.
- Mengatasi masalah dengan kata-kata yang salah eja atau singkatan.
- Mengatasi kata yang ada imbuhan

### 1.4 Ekstraksi Fitur

Selanjutnya, yang perlu dilakukan adalah mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin. Beberapa teknik ekstraksi fitur yang umum digunakan dalam analisis sentimen adalah:

- *Bag of Words* (BoW): Menghitung frekuensi kemunculan kata-kata dalam setiap teks.
- *TF-IDF* (*Term Frequency-Inverse Document Frequency*): Menghitung bobot kata berdasarkan frekuensi kemunculannya dalam teks dan keseluruhan dokumen.
- *Word Embeddings*: Menggunakan model bahasa seperti Word2Vec atau GloVe untuk menghasilkan vektor representasi kata-kata.

Pada penelitian ini, ekstraksi fitur yang digunakan adalah *TF-IDF Vectorizer*, untuk syntax lengkapnya adalah seperti berikut:

### 1.5 Klasifikasi

Klasifikasi adalah proses pengelompokan objek atau data ke dalam kelas atau kategori yang telah ditentukan sebelumnya berdasarkan fitur-fitur atau atribut-atribut yang dimiliki oleh objek tersebut. Tujuan utama dari klasifikasi adalah untuk mengidentifikasi pola atau karakteristik yang membedakan antara kelas-kelas yang berbeda, sehingga dapat memprediksi kelas atau kategori baru untuk objek yang belum diketahui kelasnya. Dalam konteks pengolahan bahasa alami, klasifikasi sering digunakan untuk mengklasifikasikan teks ke dalam kategori-kategori tertentu berdasarkan fitur-fitur linguistik yang ada. Misalnya, dalam analisis sentimen, teks dapat diklasifikasikan ke dalam kategori positif, negatif, atau netral berdasarkan evaluasi sentimen yang terkandung dalam teks tersebut. Dalam penerjemahan mesin, teks dapat diklasifikasikan ke dalam kategori bahasa sumber dan bahasa target. Klasifikasi juga digunakan dalam banyak tugas NLP lainnya seperti identifikasi entitas nama, klasifikasi topik, dan analisis sentimen.

### 1.6 Naïve bayes

Naive Bayes adalah metode klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi naif (*naive assumption*) yaitu asumsi bahwa setiap fitur dalam data adalah independen secara kondisional terhadap kelasnya. Metode ini sangat populer dalam analisis teks dan pengenalan pola, terutama dalam aplikasi yang melibatkan klasifikasi teks berdasarkan konten atau sentimen. Secara umum, Naive Bayes bekerja dengan menghitung probabilitas kelas dari suatu objek berdasarkan kemunculan kombinasi nilai-nilai fitur yang dimiliki oleh objek tersebut. Dengan menggunakan teorema Bayes, probabilitas kelas dapat diestimasi dengan memanfaatkan probabilitas kemunculan fitur-fitur tersebut dalam kelas yang bersangkutan. Kelebihan dari metode Naive Bayes adalah kemampuannya dalam mengatasi dimensi fitur yang tinggi, serta kecepatan dan efisiensinya dalam melakukan klasifikasi. Meskipun memiliki asumsi yang sederhana, Naive Bayes sering memberikan performa yang baik dalam banyak kasus, terutama dalam aplikasi analisis sentimen, klasifikasi teks, dan filtrasi spam.

### 1.7 CNN

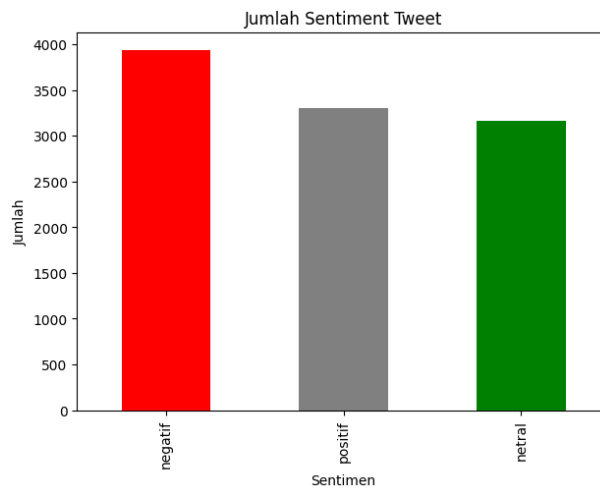
*Convolutional Neural Networks* atau disingkat dengan CNN adalah salah satu algoritma *deep learning* yang cara kerjanya melakukan training data pada kumpulan data yang besar. Metode ini sudah tidak asing di ranah *deep learning* karena memiliki performa yang lebih baik dalam menganalisis sentiment (Qudsi et al., 2021). Tidak hanya itu metode ini juga salah satu metode yang efektif dalam klasifikasi karena memiliki lapisan-lapisan yang memiliki kemampuan membaca data dengan jumlah yang banyak. Pada metode ini dapat tidak hanya untuk teks tetapi, bisa dengan mengidentifikasi wajah, gambar, tanda-tanda lalu lintas, dan masih banyak lagi.

## HASIL DAN PEMBAHASAN

### A. Pegujian Metode

Pengujian model ini dilakukan dengan pembagian data testing dan datah latih 30:70 berlaku tidak hanya untuk *Convolutional Neural Network* tetapi, *Naïve Bayes*. Pada hasil ini didapatkan hasil *precision*, *recall*, *f1-score*, dan *accuracy*. Pada *Convolutional Neural Network* memakai layer *GlobalMax Polling Layer*.

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



Gambar 3.1 Jumlah Sentiment Tweet

Dari gambar diatas dapat disimpulkan bahwa komentar atau opini masyarakat tentang politik negatif masih yang paling tinggi dibanding dengan netral dan positif.

## B. Perbandingan Hasil

Pada perbandingan dapat dihasilkan seperti dibawah dan akan dilihat dalam bentuk tabel dibawah.

Tabel 2. Hasil Klasifikasi

	<i>Naive Bayes</i>			<i>Convolutional Neural Networks</i>		
	<i>pre</i> <i>cision</i>	<i>r</i> <i>ecall</i>	<i>f1</i> - <i>score</i>	<i>pre</i> <i>cision</i>	<i>r</i> <i>ecall</i>	<i>f1</i> <i>-score</i>
neg atif	0, 71	0 ,95	0,81	0,7 2	0 ,85	0, 78
net ral	0, 71	0 ,43	0,53	0,5 8	0 ,56	0, 61
pos itif	0, 66	0 ,64	0,65	0,7 5	0 ,50	0, 60
<i>acc</i> <i>uracy</i>			0.69			0, 68

Pada tabel 2 dapat dilihat bahwa nilai dari precision terbesar ada pada *Convolutional Neural Networks* dengan nilai 0.72 yang berlabel negatif. Hal ini menunjukkan bahwa nilai ketepatan pada label antara negative dengan informasi yang diminta memberikan nilai true negative yang lebih besar dari sentiment yang lain. Nilai persentase *recall* yang besar terdapat pada *Naive Bayes* dengan sentiment negative membuktikan bahwa, kemampuan model jumlah yang benar sebesar 0.95. Sedangkan, nilai persentase *f1-score* paling tinggi didapatkan pada

*Naive Bayes* dengan nilai 0.81. Hasil *f1-score* didapatkan dari penggabungan nilai *recall* dan *precission*.

## **KESIMPULAN DAN SARAN**

Pada *Naive Bayes* didapatkan akurasi sebesar 0.69 sedangkan CNN didapatkan akurasi sebesar 0.68. Hasil diatas menunjukkan bahwa adanya selisih yang cukup tipis untuk akurasi yang didapatkan. Sehingga dapat disimpulkan bahwa kedua model diatas terhadap analisis sentimen twitter topik politik cukup bagus untuk menggunakan metode diatas.

Saran dalam analisis ini perlu diperhatikan terkait pada saat pelabelan dilakukan. Pada saat pelabelan manual perlu dilakukan konsentrasi yang tinggi agar bisa mendapatkan akurasi yang lebih bagus lagi. Lalu, penggunaan filterisasi atau penghilangan kata yang tidak perlu ini sangat berpengaruh kepada hasil dari analisis, sehingga ketika analisis tersebut masih terlihat kurang untuk akurasi perlunya penambahan filterisasi atau penghilangan kata.

## **DAFTAR PUSTAKA**

- Apriani, R., & Gustian, D. (2019). Analisis Sentimen dengan *Naive Bayes* terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1), 54-62.
- Bird, S., Klein, E., & Loper, E. (2020). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media.
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma *Naive Bayes* untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *TEKNO KOMPAK Journal*, 15(1), 131-145. P-ISSN: 1412-9663, E-ISSN: 2656-3525.
- Pak, A., & Parvez, M. T. (2020). Sentiment Analysis of Twitter Data Using *Naive Bayes Classifier*. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.
- Qudsi, D. H., Lubis, J. H., Syaliman, K. U., & Najwa, N. F. (2021). Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen Di Perguruan Tinggi Menggunakan Sentiment Analysis In The Student's Reviews Of College Department Performance Using. *Jurnal Teknologi Informasi Dan Ilmu Komputer(JTIK)*, 8(5), 1067–1076. <https://doi.org/10.25126/jtiik.202184842>
- Pradana, M.G. (2020). PENGGUNAAN FITUR WORDCLOUD DAN DOCUMENT TERM MATRIX DALAM TEXT MINING. *Jurnal Ilmiah Informatika (JIF)*, 8(1), 38-43. P-ISSN : 2337-8379, E-ISSN: 2615-1049.
- Irawan, F.A., Rochmah, D.A. (2022). Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19. *JURNAL INFORMATIKA*, 9(2), 148-158, P-ISSN: 2355-6579, E-ISSN: 2528-2247.