

ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE TERKAIT JUDI ONLINE MENGGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK(CNN)*

Abhinaya Raditya Putra, Aria Hendrawan

1,2) Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia

Article Info	ABSTRACT
<p>Article history:</p> <p>Received: 12 Desember 2025 Revised: 17 Desember 2025 Accepted: 23 Desember 2025</p>	<p>Abstrak</p> <p>Perkembangan teknologi informasi dan komunikasi telah mendorong meningkatnya penggunaan media sosial, termasuk YouTube, yang kini menjadi salah satu platform paling populer untuk berbagi informasi, hiburan, serta opini publik. Banyaknya komentar yang diunggah oleh pengguna mencerminkan berbagai ekspresi sentimen terhadap suatu konten, namun volume data yang besar membuat analisis manual menjadi tidak efisien. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen pada komentar YouTube dengan menggunakan pendekatan <i>deep learning</i> berbasis <i>Convolutional Neural Network (CNN)</i>. Proses penelitian diawali dengan pengumpulan data melalui metode <i>web scraping</i> untuk memperoleh komentar pengguna, kemudian dilakukan tahap <i>preprocessing</i> meliputi pembersihan teks, penghapusan tanda baca, tokenisasi, dan normalisasi. Setelah itu, data diberi label berdasarkan kategori sentimen positif, negatif, dan netral. Model CNN diterapkan untuk mengekstraksi fitur penting dari teks melalui lapisan konvolusi dan <i>pooling</i> sehingga mampu mengenali pola kata dan konteks kalimat secara efektif. Evaluasi kinerja dilakukan menggunakan metrik <i>accuracy</i>, <i>precision</i>, <i>recall</i>, dan <i>F1-score</i> untuk mengukur tingkat keberhasilan model dalam klasifikasi sentimen. Hasil penelitian menunjukkan bahwa CNN mampu memberikan hasil yang baik pada komentar netral yaitu 1.527 data yang dari total data diambil 1.648 data.</p> <p>Kata Kunci: Youtube, Deep Learning, Analisis Sentimen, CNN.</p> <p>Abstract</p> <p>The development of information and communication technology has driven the increasing use of social media, including YouTube, which has become one of the most popular platforms for sharing information, entertainment, and public opinions. The vast number of comments posted by users reflects various expressions of sentiment toward certain content; however, the large volume of data makes manual analysis inefficient. Therefore, this study aims to analyze sentiments in YouTube comments using a deep learning approach based on the Convolutional Neural Network (CNN) method. The research process begins with data collection through web scraping to obtain user comments, followed by preprocessing stages that include text cleaning, punctuation removal, tokenization, and normalization. The data are then labeled according to sentiment categories: positive, negative, and neutral. The CNN model is applied to extract important textual features through convolution and pooling layers, enabling it to effectively recognize word patterns and sentence contexts. Model performance is evaluated using accuracy, precision, recall, and F1-score. The research results show that CNN is able to provide good performance on neutral comments, with 1,527 data correctly classified out of a total of 1,648 data.</p> <p>Keywords: Youtube, Deep Learning, Sentiment Analysis, CNN.</p> <p>Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 (CC-BY-NC-SA).</p>



1. PENDAHULUAN

Digitalisasi mempermudah masyarakat mengakses berbagai kebutuhan hanya melalui smartphone, sehingga mendorong meningkatnya jumlah pelaku judi online dari berbagai kelompok usia tanpa membedakan jenis kelamin. Menurut data PPATK, pada Oktober 2023 nilai transaksi terkait judi online melonjak 52,69% dibandingkan tahun 2022, mencapai Rp104,42 triliun.[1] Meski pemerintah telah berupaya memblokir sejumlah situs judi online, platform-platform baru terus bermunculan dengan nama berbeda. Meningkatnya kasus kecanduan judi online turut memicu bertambahnya tindak kejahatan seperti penipuan dan pencurian, karena para pecandu kerap mencari cara untuk memenuhi kebutuhan berjudi mereka. Meski demikian, upaya penegakan hukum kerap menghadapi berbagai hambatan, termasuk sulitnya memblokir situs judi online yang beroperasi dari luar negeri dan minimnya koordinasi antar lembaga terkait. Pada pasal 27 ayat 2 dan Pasal 303 bis KUHP, yang khusus mengatur perjudian online, mengancam para pemain judi dengan pidana penjara paling lama 10 tahun dan/atau denda pidana paling banyak satu milyar rupiah.[2]

Masalah ini tidak hanya berdampak secara individual, tetapi juga berpotensi menimbulkan kerugian bagi masyarakat luas. Berbagai kajian sebelumnya telah mencoba memahami fenomena judi online dari sudut pandang hukum, ekonomi, hingga psikologi. [3]. Dampak-dampak seperti meningkatnya tindakan kriminal, penurunan produktivitas, serta terganggunya ketertiban sosial menunjukkan bahwa judi online merupakan ancaman serius bagi stabilitas sosial dan masa depan generasi muda. Sudah banyak orang yang merasakan dampak buruk dari aksi merugikan ini salah satunya seperti penipuan, hacking, phishing dan masih banyak akses peretasan lainnya[4] Dua sisi argumen sama-sama memiliki dukungan kuat dari berbagai kalangan masyarakat, dan sulit menentukan bagaimana tindakan terbaik terkait pemblokiran situs judi online karena kedua sisi sama-sama setuju pada poin bahayanya judi online[5]

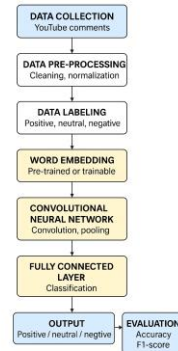
Analisis sentimen merupakan salah satu cabang text mining yang berfokus pada pengelompokan opini dalam teks ke dalam kelas tertentu, seperti positif, negatif, dan netral. Penerapannya pada komentar YouTube menjadi penting karena tingginya tingkat interaksi pengguna pada platform tersebut yang mampu merepresentasikan pandangan dan respons publik terhadap suatu topik secara luas. Dengan dukungan teknik machine

learning dan deep learning, analisis sentimen dapat menyajikan pemetaan opini masyarakat secara lebih objektif dan sistematis. Perkembangan teknologi deep learning, khususnya metode Convolutional Neural Network (CNN), semakin banyak dimanfaatkan dalam analisis sentimen karena kemampuannya mengekstraksi fitur secara otomatis serta menghasilkan performa akurasi yang lebih baik dibandingkan pendekatan konvensional, sehingga analisis sentimen komentar YouTube berbasis deep learning menjadi solusi efektif untuk memahami pola opini publik dan mendukung pengambilan keputusan berbasis data.

Salah satu metode analisis sentimen adalah menggunakan Deep Learning. Analisis sentimen adalah cabang penelitian text mining yang melakukan proses dalam klasifikasi pada dokumen teks[6]. Analisis sentimen pada media sosial saat ini menjadi dapat dilakukan karena semakin banyaknya masyarakat yang menggunakan media sosial sebagai media untuk menyuarakan pendapat pribadi mereka[7]. Deep learning merupakan salah satu cabang dari machine learning yang memiliki kelebihan dibandingkan dengan machine learning biasa yaitu dapat melakukan ekstraksi fitur secara otomatis. Metode CNN dipilih karena lebih cocok digunakan pada data berformat matriks seperti data citra dan data teks yang diubah dalam bentuk numerik. CNN dapat digunakan untuk mengolah data teks selain digunakan untuk memproses data citra[8]. CNN akan digunakan dengan memanfaatkan representasi vektor kata (word embedding) sebagai bentuk pemetaan teks komentar. Model CNN kemudian dilatih menggunakan kumpulan komentar yang telah diberi label sentimen, sehingga mampu mengelompokkan setiap komentar ke dalam kategori sentimen yang tepat. Metode CNN terbukti memiliki akurasi yang lebih tinggi[9].

2. METODE PENELITIAN

Penelitian ini dijalankan menggunakan google colab. Alur penelitian dengan metode CNN disajikan pada gambar 1.



Gambar 1. Alur penelitian dengan metode CNN

2.1. Scraping Data

pengumpulan data berupa komentar publik dari platform YouTube. Komentar diambil dari video yang relevan dengan topik penelitian dengan tema tertentu seperti “judi online”. Data diambil dengan menggunakan YouTube Data API v3 yang telah disediakan oleh google dan juga library APIs & Services digunakan pada python untuk mengakses data komentar dari video tertentu[10]. Data yang dikumpulkan berupa komentar pengguna yang relevan dengan topik penelitian, hasil dari proses scraping menghasilkan sebanyak 8.239 data komentar yang dijadikan sebagai dataset penelitian. Data yang diperoleh kemudian disimpan dalam format yang mudah diolah seperti .csv atau .json.

2.2 pre processing data

Preprocessing merupakan tahapan proses untuk membersihkan data dari kata-kata atau komentar youtube yang tidak di perlukan serta kata-kata yang tidak memiliki makna[11] Proses ini mencakup beberapa langkah penting, antara lain: *case folding* (mengubah seluruh huruf menjadi huruf kecil), *cleaning* Pada proses data cleansing dilakukan pembersihan kata dengan dilakukannya penghilangan karakter-karakter selain huruf alfabet, angka tanda baca seperti delimiter titik (.), koma (,), dan tanda baca lainnya[12]

2.3. labeling data

Tahap pelabelan data merupakan langkah penting dalam membangun dataset yang siap digunakan untuk pelatihan model. Pada tahap ini klasterisasi K-Means digunakan untuk mengelompokkan sentimen ke dalam kelas negatif, positif, dan netral[13]. Misalnya, komentar yang mengandung pujian, ekspresi senang, atau apresiasi dikategorikan sebagai positif; komentar bernada datar, informatif, atau tidak

mengekspresikan emosi tertentu dikategorikan sebagai netral; sedangkan komentar yang mengandung kritik, ketidaksenangan, atau emosi negatif masuk dalam kategori negatif. Apabila terdapat perbedaan pendapat antarannotator, maka dilakukan diskusi atau validasi ulang untuk mencapai kesepakatan akhir.

2.4 word embedding

Komentar yang telah diberi label kemudian diubah menjadi bentuk numerik agar dapat diproses oleh algoritma CNN. Alih-alih hanya melihat kata-kata yang berdekatan, GloVe mempelajari seberapa sering setiap pasangan kata muncul bersama dalam konteks global, lalu memetakan setiap kata ke dalam vektor berdimensi tetap[14]. Dalam penelitian ini digunakan dua jenis embedding, yaitu pre-trained embedding seperti *fastText Bahasa Indonesia* atau *Word2Vec* yang telah dilatih pada korpus besar, serta trainable embedding yang disesuaikan langsung dengan data komentar YouTube. Model GloVe yang digunakan pada penelitian ini memiliki dimensi 300 (GloVe 300D), artinya setiap kata akan diwakili oleh vektor dengan 300 elemen[15]

2.5 CNN

Convolutional Neural Network adalah pengembangan dari Multi Layer Perceptron (MLP) yang digunakan untuk mengolah data dua dimensi. Karena kedalaman jaringan yang tinggi dan banyak diterapkan untuk analisis sentiment sehingga CNN termasuk dalam dari Deep Neural Network[16]. CNN akan melakukan pemrosesan lapisan konvolusional dengan penggabungan beberapa lapisan dalam prosesnya[17]. Convolutional Neural Network (CNN) terdiri dari lapis masukan (input layer), lapis keluaran (output layer) dan sejumlah lapisan tersembunyi (hidden Layers)[18]

3. HASIL DAN PEMBAHASAN

3.1 pengambilan data

Pengambilan data ini menggunakan google colab dengan youtube api v3. Berikut adalah contoh hasil pengambilan data:

Gambar 2 hasil pengambilan data

Langkah selanjutnya yaitu pre processing data, disini ada tahap clean text yaitu pembersihan data dari symbol atau lambang yang tidak perlu. Tahap cleant text ada pada gambar 3 dibawah ini.

Gambar 3 hasil dari clean text

Pada bagian ini hasil yang sudah diambil dan sudah diolah atau sudah dibersihkan dari simbol dan emote selanjutnya akan diberi label positif,netral,negative. Hasil dari labeling data akan ditampilkan pada gambar 4.

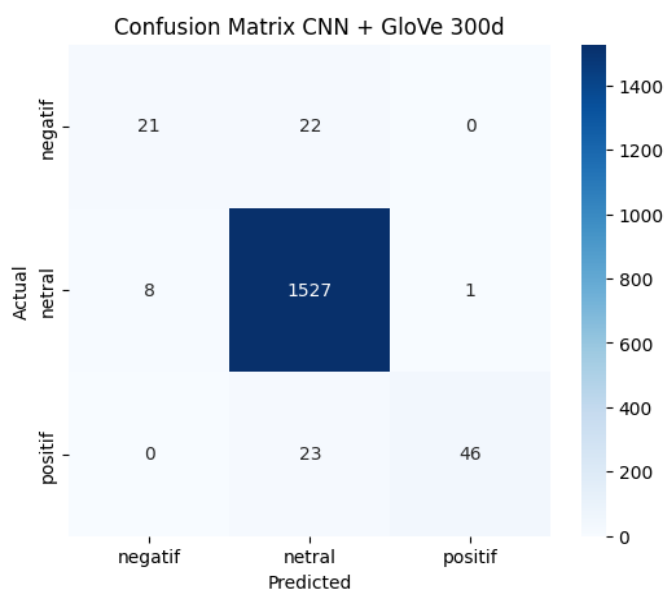
Gambar 4 hasil dari labeling data

setelah labeling data proses selanjutnya adalah menghitung f1 score dan juga akan ada visualisasi confusion matrix dari CNN dan word embedding glove 300d. Penggabungan CNN dengan word embedding GloVe berdimensi 300 memberikan kemampuan yang lebih baik bagi model dalam memahami makna teks. GloVe (Global Vectors for Word Representation) adalah teknik embedding yang merepresentasikan kata dalam vector 300 dimensi untuk menangkap hubungan semantik dalam teks[19]. CNN kemudian

memproses representasi tersebut untuk menangkap pola penting, seperti frasa yang menandakan sentimen tertentu. Proses yang terjadi pada convolutional layer adalah melakukan dot product dari vector dengan bobot atau kernel vector setiap kata pada kalimat untuk mendapatkan fitur yang lain[20]. Kombinasi ini membuat analisis menjadi lebih akurat dan efisien karena CNN dapat mengekstraksi fitur utama secara otomatis, sementara GloVe memberikan fondasi representasi kata yang stabil. Alhasil, model mampu menjalankan tugas analisis sentimen dan klasifikasi teks dengan performa yang lebih optimal. Hasil akan ditampilkan pada tabel dan gambar di bawah.

Tabel 1 hasil model CNN dengan glove 300d

	precision	recall	f1-score	support
negatif	0.72	0.49	0.58	43
netral	0.97	0.99	0.98	1536
positif	0.98	0.67	0.79	69



Gambar 5 hasil dari confusion matrix dengan glove 300d

Hasil confusion matrix menunjukkan bahwa model CNN dengan GloVe 300d bekerja paling baik pada kelas netral, dengan 1.527 komentar berhasil diklasifikasikan dengan benar. Pada kelas negatif, model mengidentifikasi 21 komentar secara tepat, namun 22 komentar salah diprediksi sebagai netral. Untuk kelas positif, terdapat 46 prediksi benar,

sementara 23 komentar positif keliru diklasifikasikan sebagai netral. CNN sering mengalami kendala dalam mengenali kelas minoritas karena distribusi data yang tidak seimbang, di mana kelas mayoritas mendominasi proses pelatihan. Situasi ini membuat fungsi loss dan pembaruan bobot lebih terfokus pada kelas dengan jumlah data lebih besar, sehingga pengaruh kelas minoritas menjadi terbatas. Akibatnya, model tidak dapat mempelajari fitur yang representatif dan beragam untuk kelas minoritas secara optimal. Selain itu, dominasi kelas mayoritas meningkatkan kecenderungan bias dan overfitting, sehingga CNN lebih sering mengklasifikasikan data dari kelas minoritas ke dalam kelas mayoritas, terutama ketika karakteristik antar kelas memiliki kemiripan.

4. SIMPULAN

model CNN yang menggunakan word embedding GloVe 300d menunjukkan performa yang solid dalam melakukan klasifikasi sentimen. Model berhasil mengenali komentar netral dengan tingkat akurasi yang sangat tinggi, yaitu sebanyak 1.527 data, serta mampu mengklasifikasikan komentar negatif dan positif secara konsisten, masing-masing sebanyak 21 dan 46 data. Temuan ini mengindikasikan bahwa kombinasi CNN dengan GloVe 300d mampu menangkap pola-pola penting dalam teks, memahami konteks linguistik, dan memberikan hasil prediksi sentimen yang reliabel. Dengan demikian, metode ini layak digunakan sebagai pendekatan yang efektif untuk analisis sentimen pada komentar YouTube maupun jenis data teks lainnya.

REFERENCES

- [1] eptu Haudli Bakhtiar and Azizah Nur Adilah, 'Fenomena Judi Online: Faktor, Dampak, Pertanggungjawaban Hukum', *INNOVATIVE: Journal Of Social Science Research*, 2024.
- [2] Annisa Laras, Najwa Salvabillah, Cindy Caroline, Jusini Delas H, Farra Dinda, and Mic Finanto, 'Analisis Dampak Judi Online di Indonesia', *Concept: Journal of Social Humanities and Education*, vol. 3, no. 2, pp. 320–331, Jun. 2024, doi: 10.55606/concept.v3i2.1304.
- [3] 2Eti Kusmiati, 3Dwi Nurhayati 1,2,3 Garut University, Garut, Indonesia 1Agung Hermawan, '90094-203339-1-SM', *Indonesian Journal of Digital Business*, 2025.
- [4] F. Nur and F. Firmasnyah, 'Analisis Sentimen dan Jaringan Sosial pada Komen Video "Youtube Tangan Kotor Influencer" dalam Judi Online Milik Ferri Irwandi', 2024.
- [5] Robert Antonius, A. R. Zulkarnain, and H. Irsyad, 'Pendekatan TF-IDF, SMOTE, dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online', *Buletin Ilmiah Informatika Teknologi*, vol. 2, no. 3, pp. 115–122, Jun. 2024, doi: 10.58369/biit.v2i3.65.
- [6] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiarsari, 'Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE', *AITI: Jurnal Teknologi Informasi*, vol. 18, no. Agustus, pp. 173–184, 2021.
- [7] I. N. Husada and H. Toba, 'Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia', *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6, no. 2, Aug. 2020, doi: 10.28932/jutisi.v6i2.2743.
- [8] K. Muludi, M. Naufal Humam, D. A. Shofiana, A. Syarif, J. I. Komputer, and U. Lampung, 'Perbandingan Kinerja CNN dan Naïve Bayes pada Analisis Sentimen Performa Manchester United di Twitter', 2023.

-
- [9] D. E. Saputra and A. R. Isnain, 'IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK UNTUK ANALISIS SENTIMEN BACAPRES 2024 PADA KOLOM KOMENTAR YOUTUBE MATA NAJWA', *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1431–1441, Aug. 2024, doi: 10.29100/jipi.v9i3.5420.
 - [10] R. Danyana Umrona, S. Naja Anwar, and R. Soelistijadi, 'ANALISIS SENTIMEN KOMENTAR YOUTUBE TERKAIT KASUS PAGAR LAUT MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)', *Jurnal Informatika Teknologi dan Sains*.
 - [11] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, 'Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS', *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 8, no. 3, 2023.
 - [12] G. A. Suwito, I. Cholissodin, and P. P. Adikara, 'Analisis Sentimen Citayam Fashion Week pada Komentar YouTube dengan Metode Convolutional Neural Network', 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
 - [13] K. K. Safra and E. Zuliarso, 'ANALISIS SENTIMEN TERHADAP PELAKSANAAN PILKADA 2024 PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE DECISION TREE', *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 1, pp. 117–126, 2025.
 - [14] H. Asrawi and M. Fataya Bilqisthi, 'Implementasi Fitur Ekstraksi Glove Pada Teks Klasifikasi Menggunakan Algoritma Long-Short Term Memory', vol. 07, 2025, doi: 10.54209/jatilima.v7i03.1694.
 - [15] I. Bagus Jatiarso and Y. Azhar, 'Klasifikasi Emosi Pada Tweet Pengguna Platform X Menggunakan Metode LSTM-GloVe Berbasis SMOTE', *REPOSITOR*, vol. 7, no. 3, pp. 401–406, 2025.
 - [16] M. Resa Arif Yudianto and H. Al Fatta, 'ANALISIS PENGARUH TINGKAT AKURASI KLASIFIKASI CITRA WAYANG DENGAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK'.
 - [17] N. K. Negoro, E. Utami, and A. Yaqin, 'KLASIFIKASI DETEKSI PENGGUNAAN MASKER MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK', *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 2, pp. 664–674, May 2023, doi: 10.29100/jipi.v8i2.3748.
 - [18] M. Rafly Alwanda, R. Putra, K. Ramadhan, and D. Alamsyah, 'Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle', 2020.
 - [19] R. Kusuma Wardana, N. Cahyono, U. Anggoro Saputro, A. Hadi Negoro, and N. Aini, 'IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI MENGGUNAKAN KOMPARASI N-GRAM DAN GLOVE PADA SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI', 2025.
 - [20] P. S. T. I. H. B. T. I. I. S. dan T. T. S. dan E. I. S. T. I. I. S. dan T. T. S. Muhammad Arief Rahman, 'Aspect Based Sentimen Analysis Opini Pub', *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*.