

**KLASIFIKASI SENTIMEN NETIZEN MEDIA SOSIAL X TERHADAP  
KANDIDAT CAWAPRES PADA PILPRES 2024 MENGGUNAKAN  
INDOBERT**

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

**Disusun oleh:**

**Alfath Thoriq Kumara  
H06218009**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA  
2024**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Alfath Thoriq Kumara

NIM : H06218009

Program Studi : Sistem Informasi

Angkatan : 2018

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: "KLASIFIKASI SENTIMEN NETIZEN MEDIA SOSIAL X TERHADAP KANDIDAT CAWAPRES PADA PILPRES 2024 MENGGUNAKAN INDOBERT". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 20 Juni 2024  
Yang menyatakan,



Alfath Thoriq Kumara  
(H06218009)

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : Alfath Thoriq Kumara

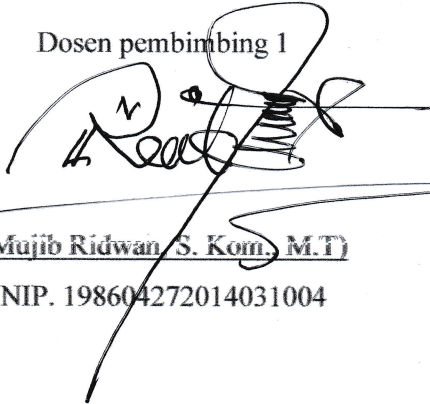
Nim : H06218009

Judul : Analisis Keberpihakan Netizen Media Sosial X Terhadap  
Kandidat Cawapres Pada Pilpres 2024 Menggunakan Indobert

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 20 Juni 2024

Dosen pembimbing 1



(Mujib Ridwan S. Kom, M.T)

NIP. 198604272014031004

Dosen pembimbing 2



(Dr. Eng. Anang Kunaefi, M. Kom)

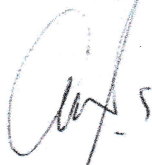
NIP. 197811132014031001

## PENGESAHAN TIM PENGUJI SKRIPSI

skripsi Alfath Thoriq Kumara telah dipertahankan  
di depan tim penguji skripsi  
di Surabaya, 21 Juni 2024

**Mengesahkan,  
Dewan Penguji**

Dosen penguji 1



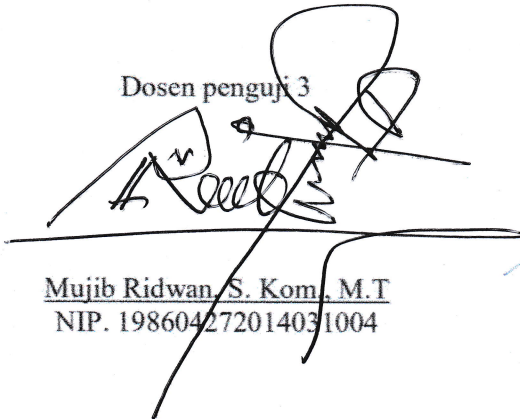
Khalid, M. Kom  
NIP. 197906092014031002

Dosen penguji 2



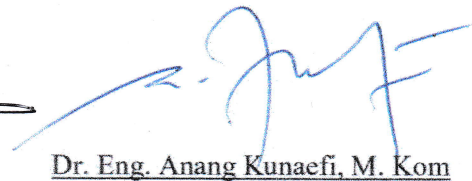
Subhan Nooriansyah, M. Kom  
NIP. 199012282020121010

Dosen penguji 3



Mujib Ridwan, S. Kom, M.T  
NIP. 198604272014031004

Dosen penguji 4



Dr. Eng. Anang Kunaefi, M. Kom  
NIP. 197811132014031001

**Mengetahui,**

Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya



Dr. A. Saepul Hamdani, M.Pd  
196507312000031002





UIN SUNAN AMPEL  
SURABAYA

KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA  
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Alfath Thoriq Kumara  
NIM : h062180009  
Fakultas/Jurusan : Saintek/Sistem Informasi  
E-mail address : h06218009@uinsby.ac.id

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi  Tesis  Desertasi  Lain-lain (.....)  
yang berjudul :

KLASIFIKASI SENTIMEN NETIZEN MEDIA SOSIAL X TERHADAP KANDIDAT

CAWAPRES PADA PILPRES 2024 MENGGUNAKAN INDOBERT

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 9 September 2024

Penulis

(Alfath Thoriq Kumara)  
nama terang dan tanda tangan

## ABSTRAK

### KLASIFIKASI SENTIMEN NETIZEN MEDIA SOSIAL X TERHADAP KANDIDAT CAWAPRES PADA PILPRES 2024 MENGGUNAKAN INDOBERT

By:  
Alfath Thoriq K

Penelitian ini bertujuan untuk mengklasifikasi sentimen masyarakat pada masa kampanye pemilu 2024 melalui *tweet* berbahasa Indonesia pada media sosial X (Twitter). Penelitian ini diharapkan dapat menjadi telaah untuk masyarakat atau peneliti lain mengenai masa pemilu 2024. Penelitian ini juga diharapkan menjadikan masyarakat lebih mawas terhadap informasi yang beredar di media sosial. Data yang digunakan berjumlah  $\pm 4500$  *tweet* yang diambil pada waktu yang telah ditentukan. Analisis penelitian ini berbasis deep learning, terkhusus menggunakan metode IndoBERT dari *huggingface* yang berbasis pada BERT dengan arsitektur Transformer. Hasil penelitian menunjukkan bahwa model IndoBERT menghasilkan performa yang baik hingga mencapai akurasi 93,67%. Hasil analisa sentimen positif dan negatif, keduanya menunjukkan bahwa *tweet* berisi kata-kata dengan konotasi positif seperti “pilih”, “menang”, dan “jadi” dan berisi kata yang merujuk pada subjek seperti “pak”, “mas”, “gus”, dan “prof”.

**Kata kunci:** sentimen, Indobert, *tweet*, pemilu, cawapres



## ABSTRACT

### SENTIMENTS CLASSIFICATION OF SOCIAL MEDIA NETIZENS X TOWARDS THE VICE PRESIDENT CANDIDATE IN THE 2024 PRESIDENTIAL ELECTION USING INDOBERT

By:  
**Alfath Thoriq K**

*This study aims to classify public sentiment during the 2024 election campaign through Indonesian-language tweets on social media X (Twitter). This research is expected to be a study for the community or other researchers regarding the 2024 election period. This research is also expected to make people more aware of the information circulating on social media. The data used amounted to  $\pm$  4500 tweets taken at a predetermined time. This research analysis is based on deep learning, specifically using the IndoBERT method from huggingface based on BERT with Transformer architecture. The results show that the IndoBERT model produces good performance to achieve 93.67% accuracy. The results of positive and negative sentiment analysis both show that the tweets contain words with positive connotations such as "pilih", "menang", and "jadi" and contain words that refer to subjects such as "pak", "mas", "gus", and "prof".*

**Keywords:** *sentiment, IndoBERT, tweet, election, vice president candidate*

## DAFTAR ISI

PERNYATAAN KEASLIAN.....	i
LEMBAR PERSETUJUAN PEMBIMBING.....	ii
PENGESAHAN TIM PENGUJI SKRIPSI.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	viii
DAFTAR TABEL.....	ix
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB 2 TINJAUAN PUSTAKA.....	4
2.1 Penelitian Terdahulu.....	4
2.2 Teori Dasar yang Digunakan.....	5
2.2.1 Text Mining.....	5
2.2.2 Analisis Sentimen.....	6
2.2.3 BERT.....	10
2.2.4 IndoBERT.....	14
2.2.5 Confusion Matrix.....	15
2.2.6 Pseudo Labelling.....	19
2.2.7 Python.....	20
2.2.8 K-Fold Cross Validation.....	21



2.3 Integrasi Keilmuan .....	21
<b>BAB 3 METODOLOGI PENELITIAN.....</b>	<b>23</b>
3.1 Jenis dan Sumber Data .....	23
3.2 Metodologi Penelitian .....	23
3.2.1 Tahapan Awal Penelitian .....	24
3.2.2 Data Processing .....	25
3.2.3 Klasifikasi Teks.....	28
3.2.4 Analisis Hasil .....	29
<b>BAB 4 HASIL DAN PEMBAHASAN.....</b>	<b>30</b>
4.1 Scraping Data .....	30
4.2 Data Processing.....	34
4.2.1 Penghapusan Duplikasi Data.....	34
4.2.2 Preprocessing .....	35
4.2.3 Pelabelan .....	50
4.2.4 Validasi Label .....	52
4.3 Klasifikasi Teks.....	55
4.3.1 Modelling .....	55
4.4 Analisis Hasil .....	56
4.4.1 Confusion Matrix .....	56
<b>BAB 5 PENUTUP.....</b>	<b>68</b>
5.1 Kesimpulan .....	68
5.2 Saran.....	68
<b>DAFTAR PUSTAKA .....</b>	<b>70</b>

## DAFTAR GAMBAR

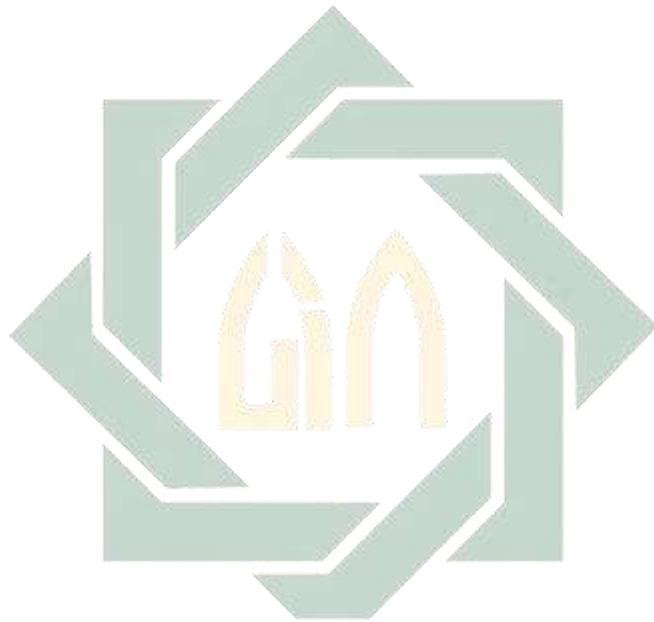
Gambar 2. 1 Arsitektur Transformer.....	11
Gambar 2. 2 Representasi Input BERT .....	12
Gambar 2. 3 Ilustrasi Alur Proses Pre-training dan Fine-tuning.....	13
Gambar 3. 1 Metodologi Penelitian .....	24
Gambar 4. 1 Irisan Data Tiap Subjek.....	35
Gambar 4. 2 Confusion Matrix Model 1 pada Dataset 1 .....	57
Gambar 4. 3 Confusion Matrix Model 2 pada Dataset 1 .....	58
Gambar 4. 4 Confusion Matrix Model 3 pada Dataset 1 .....	58
Gambar 4. 5 Confusion Matrix Model 4 pada Dataset 1 .....	59
Gambar 4. 6 Confusion Matrix Model 1 pada Dataset 2 .....	60
Gambar 4. 7 Confusion Matrix Model 2 pada Dataset 2 .....	61
Gambar 4. 8 Confusion Matrix Model 3 pada Dataset 2 .....	62
Gambar 4. 9 Confusion Matrix Model 4 pada Dataset 2 .....	62
Gambar 4. 10 Confusion Matrix Model 1 pada Dataset 3 .....	64
Gambar 4. 11 Confusion Matrix Model 2 pada Dataset 3 .....	64
Gambar 4. 12 Confusion Matrix Model 3 pada Dataset 3 .....	65
Gambar 4. 13 Confusion Matrix Model 4 pada Dataset 3 .....	66

UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	4
Tabel 2. 2 Tabel Confusion Matrix .....	15
Tabel 4. 1 sampel hasil scraping subjek 1 .....	31
Tabel 4. 2 sampel hasil scraping subjek 2.....	32
Tabel 4. 3 sampel hasil scraping subjek 3.....	33
Tabel 4. 4 contoh hasil proses cleansing pada dataset 1 .....	36
Tabel 4. 5 contoh hasil proses cleansing pada dataset 2 .....	37
Tabel 4. 6 contoh hasil proses cleansing pada dataset 3 .....	38
Tabel 4. 7 contoh hasil proses case folding pada dataset 1 .....	39
Tabel 4. 8 contoh hasil proses case folding pada dataset 2.....	40
Tabel 4. 9 contoh hasil proses case folding pada dataset 3.....	40
Tabel 4. 10 contoh hasil proses normalization pada dataset 1 .....	41
Tabel 4. 11 contoh hasil proses normalization pada dataset 2 .....	42
Tabel 4. 12 contoh hasil proses normalization pada dataset 3 .....	43
Tabel 4. 13 contoh hasil filtering pada dataset 1.....	44
Tabel 4. 14 contoh hasil proses filtering pada dataset 2 .....	44
Tabel 4. 15 contoh hasil proses filtering pada dataset 3 .....	45
Tabel 4. 16 contoh hasil proses stemming pada dataset 1.....	46
Tabel 4. 17 contoh hasil proses stemming pada dataset 2.....	46
Tabel 4. 18 contoh hasil proses stemming pada dataset 3.....	47
Tabel 4. 19 hasil tokenisasi metode IndoBERT pada dataset 1 .....	47
Tabel 4. 20 hasil tokenisasi metode IndoBERT pada dataset 2 .....	48
Tabel 4. 21 hasil tokenisasi metode IndoBERT pada dataset 3 .....	49
Tabel 4. 22 hasil pelabelan menggunakan algoritma dan pakar pada dataset 1....	51
Tabel 4. 23 hasil pelabelan menggunakan algoritma dan pakar pada dataset 2....	51
Tabel 4. 24 hasil pelabelan menggunakan algoritma dan pakar pada dataset 3....	52
Tabel 4. 25 hasil validasi antara label pakar dan label algoritma pada dataset 1..	53
Tabel 4. 26 hasil validasi antara label pakar dan label algoritma pada dataset 2..	53
Tabel 4. 27 hasil validasi antara label pakar dan label algoritma pada dataset 3..	54
Tabel 4. 28 Nilai hyperparameter yang digunakan pada tahap fine-tuning .....	56
Tabel 4. 29 Hasil Confusion Matrix Dataset 1.....	57

Tabel 4. 30 perbandingan predicted positive dan predicted negative dataset 1 ....	59
Tabel 4. 31 Hasil Confusion Matrix Dataset 2.....	60
Tabel 4. 32 perbandingan predicted positive dan predicted negative dataset 2 ...	63
Tabel 4. 33 Hasil Confusion Matrix Dataset 3.....	63
Tabel 4. 34 perbandingan predicted positive dan predicted negative dataset 3 ...	66



UIN SUNAN AMPEL  
S U R A B A Y A



## DAFTAR PUSTAKA

- Abdillah, W. F., Premana, A., & Bhakti, R. M. H. (n.d.). Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur. *2021/11/30*, 3(02).  
<https://doi.org/10.46772/intech.v3i02.556>
- Aldayel, A., & Magdy, W. (2019). Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW). <https://doi.org/10.1145/3359307>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *jurnal komputasi*, 8(2).  
<https://doi.org/10.23960/komputasi.v8i2.2615>
- Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes*, 11(1), 1–34. <https://doi.org/10.1080/01638538809544689>
- Darwish, K., Magdy, W., Rahimi, A., Baldwin, T., & Abokhodair, N. (2018). Predicting Online Islamophobic Behavior after #ParisAttacks. *The Journal of Web Science*, 4(3), 34–52. <https://doi.org/10.1561/106.00000013>
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the Twelfth International Conference on World Wide Web - WWW '03*, 519.  
<https://doi.org/10.1145/775152.775226>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge university press.

*Hootsuite (We are Social): Indonesian Digital Report 2023*. (2023).

<https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2023/>

*Kementerian Komunikasi dan Informatika*. (2022).

[https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan\\_media](https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media)

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP* (arXiv:2011.00677). arXiv. <https://doi.org/10.48550/arXiv.2011.00677>

Küçük, D., & Can, F. (2021). Stance Detection: Concepts, Approaches, Resources, and Outstanding Issues. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2673–2676. <https://doi.org/10.1145/3404835.3462815>

Lahoti, P., Garimella, K., & Gionis, A. (2018). Joint Non-Negative Matrix Factorization for Learning Ideological Leaning on Twitter. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 351–359. <https://doi.org/10.1145/3159652.3159669>

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, 2(2010), 627–666.

Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4), Art. 4. <https://doi.org/10.3390/technologies9040081>

Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Teknologi & Informasi ITSsmart*, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.

Reveilhac, M., & Schneider, G. (2023). Replicable semi-supervised approaches to state-of-the-art stance detection of tweets. *Information Processing & Management*, 60(2), 103199. <https://doi.org/10.1016/j.ipm.2022.103199>

Triawati, C., Bijaksana, M. A., Indrawati, N., & Saputro, W. A. (2009). Pemodelan Berbasis Konsep Untuk Kategorisasi Artikel Berita Berbahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (n.d.). *Attention is All you Need*.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding* (arXiv:2009.05387). arXiv. <https://doi.org/10.48550/arXiv.2009.05387>

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation* (arXiv:1609.08144). arXiv. <https://doi.org/10.48550/arXiv.1609.08144>

Yuill, S., & Harpin, H. (2006). *Python*.