

**ANALISIS KOMPARATIF MODEL INDOBERTWEET DAN SVM  
UNTUK SENTIMEN PUBLIK TERHADAP PEMBANGUNAN IKN  
DI PLATFORM X**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**FARREL SHAHIZIDAN ARDANTIANTO**

**21.11.4364**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**ANALISIS KOMPARATIF MODEL INDOBERTWEET DAN SVM  
UNTUK SENTIMEN PUBLIK TERHADAP PEMBANGUNAN IKN  
DI PLATFORM X**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**FARREL SHAHIZIDAN ARDANTIANTO**

**21.11.4364**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS KOMPARATIF MODEL INDOBERTWEET DAN SVM  
UNTUK SENTIMEN PUBLIK TERHADAP PEMBANGUNAN IKN  
DI PLATFORM X**

yang disusun dan diajukan oleh

**FARREL SHAHIZIDAN ARDANTIANTO**

**21.11.4364**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 22 Agustus 2025

Dosen Pembimbing,



Majid Rahardi, S.Kom., M.Eng

**NIK. 190302393**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**ANALISIS KOMPARATIF MODEL INDOBERTWEET DAN SVM**  
**UNTUK SENTIMEN PUBLIK TERHADAP PEMBANGUNAN IKN**  
**DI PLATFORM X**

yang disusun dan diajukan oleh

**FARREL SHAHIZIDAN ARDANTIANTO**

**21.11.4364**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 22 Agustus 2025

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Nuri Cahyono, M.Kom.**  
**NIK. 190302278**

**Dwi Nurani, S.Kom., M.Kom.**  
**NIK. 190302236**

**Majid Rahardi, S.Kom., M.Eng.**  
**NIK. 190302393**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 22 Agustus 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



**Prof. Dr. Kusriani, M.Kom.**  
**NIK. 190302106**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : FARREL SHAHIZIDAN ARDANTIANTO  
NIM : 21.11.4364

Menyatakan bahwa Skripsi dengan judul berikut:

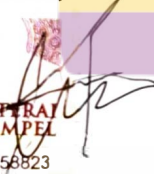
**Analisis Komparatif Model Indobertweet Dan Svm Untuk Sentimen Publik  
Terhadap Pembangunan IKN Di Platform X**

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Agustus 2025

Yang Menyatakan,



0AAMX412458823

FARREL SHAHIZIDAN ARDANTIANTO

## HALAMAN PERSEMBAHAN

Dengan rasa syukur dan hormat, karya ilmiah ini saya persembahkan kepada:

1. Kedua orang tua saya, Bapak Iwan Fajar dan Ibu Tri Wiyanti atas kasih sayang, doa, dan dukungan moral serta material yang tidak pernah henti dalam setiap langkah hidup saya.
2. Dosen pembimbing saya, Bapak Majid Rahardi, S.Kom., M.Eng, yang telah memberikan ilmu, bimbingan, dan arahan selama pengerjaan skripsi saya.
3. Teman-teman seperjuangan, yang telah menjadi sumber semangat dan kebersamaan dalam melewati setiap tantangan akademik.
4. Almamater tercinta, tempat saya tumbuh dan belajar, serta mengembangkan potensi diri di bidang ilmu pengetahuan dan teknologi.

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Analisis Komparatif Model IndoBERTweet dan SVM untuk Sentimen Publik terhadap Pembangunan IKN di Platform X*” dengan baik.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Informatika. Proses penyusunan skripsi ini tentunya tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Majid Rahardi sebagai dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan yang berharga selama proses penyusunan skripsi ini.
2. Bapak/Ibu dosen penguji yang telah meluangkan waktu, memberikan kritik, saran, dan masukan yang sangat membangun demi penyempurnaan skripsi ini.
3. Seluruh pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam penyusunan skripsi ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa yang akan datang.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang berkepentingan.

Yogyakarta, 9 Juni 2025

Penulis



Farrel Shahizidan Ardantianto

## DAFTAR ISI

|                                           |      |
|-------------------------------------------|------|
| HALAMAN JUDUL .....                       | i    |
| HALAMAN PERSETUJUAN.....                  | ii   |
| HALAMAN PENGESAHAN .....                  | iii  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI ..... | iv   |
| HALAMAN PERSEMBAHAN .....                 | v    |
| KATA PENGANTAR .....                      | vi   |
| DAFTAR ISI.....                           | vii  |
| DAFTAR TABEL.....                         | x    |
| DAFTAR GAMBAR .....                       | xi   |
| DAFTAR LAMPIRAN.....                      | xii  |
| DAFTAR LAMBANG DAN SINGKATAN .....        | xiii |
| DAFTAR ISTILAH .....                      | xiv  |
| INTISARI .....                            | xv   |
| ABSTRACT.....                             | xvi  |
| BAB I PENDAHULUAN.....                    | 1    |
| 1.1 Latar Belakang.....                   | 1    |
| 1.2 Rumusan Masalah.....                  | 2    |
| 1.3 Batasan Masalah .....                 | 2    |
| 1.4 Tujuan Penelitian .....               | 3    |
| 1.5 Manfaat Penelitian .....              | 3    |
| 1.6 Sistematika Penulisan .....           | 4    |
| BAB II TINJAUAN PUSTAKA .....             | 5    |

|                                 |                                                         |    |
|---------------------------------|---------------------------------------------------------|----|
| 2.1                             | Studi Literatur .....                                   | 5  |
| 2.2                             | Dasar Teori.....                                        | 10 |
| 2.2.1                           | Analisis Sentimen .....                                 | 10 |
| 2.2.2                           | Media Sosial Platform X.....                            | 10 |
| 2.2.3                           | Akuisisi Data Tweet Harvest .....                       | 11 |
| 2.2.4                           | Pra-pemrosesan Data .....                               | 12 |
| 2.2.5                           | Pemindahan Ibu Kota Negara (IKN) .....                  | 13 |
| 2.2.6                           | Machine Learning untuk Analisis Teks .....              | 14 |
| 2.2.7                           | Support Vector Machine (SVM).....                       | 14 |
| 2.2.8                           | Term Frequency Inverse Document Frequency (TF-IDF)..... | 16 |
| 2.2.9                           | Hyperparameter Tuning untuk SVM .....                   | 17 |
| 2.2.10                          | Augmentasi Teks Synonym Replacement .....               | 18 |
| 2.2.11                          | Natural Language Processing (NLP) .....                 | 19 |
| 2.2.12                          | IndoBERTweet .....                                      | 19 |
| 2.2.13                          | Fine Tuning .....                                       | 20 |
| 2.2.14                          | Hyperparameter Optuna .....                             | 21 |
| 2.2.15                          | Metrik Evaluasi Model .....                             | 21 |
| 2.2.16                          | Confusion Matrix .....                                  | 23 |
| BAB III METODE PENELITIAN ..... |                                                         | 24 |
| 3.1                             | Objek Penelitian.....                                   | 24 |
| 3.2                             | Alur Penelitian .....                                   | 24 |
| 3.2.1                           | Persiapan Data .....                                    | 26 |
| 3.2.2                           | Pra-pemrosesan Data .....                               | 26 |
| 3.2.3                           | Pembagian dan Augmentasi Data .....                     | 27 |
| 3.2.4                           | Pembuatan dan Pelatihan Model.....                      | 28 |
| 3.2.5                           | Evaluasi Model .....                                    | 28 |
| 3.3                             | Alat dan Bahan.....                                     | 29 |
| 3.3.1                           | Bahan Penelitian .....                                  | 29 |
| 3.3.2                           | Alat dan Instrumen Penelitian.....                      | 30 |

|                                                                  |    |
|------------------------------------------------------------------|----|
| BAB IV HASIL DAN PEMBAHASAN .....                                | 34 |
| 4.1 Akusisi Data.....                                            | 34 |
| 4.2 Preparation Data.....                                        | 35 |
| 4.2.1 Penghapusan Data Duplikat (Drop Duplicate Data).....       | 35 |
| 4.2.2 Pelabelan Data (Data Labeling) .....                       | 36 |
| 4.3 Pra-pemrosesan Data .....                                    | 37 |
| 4.3.1 Pipeline Pra-pemrosesan untuk SVM .....                    | 37 |
| 4.3.2 Pipeline Pra-pemrosesan untuk IndoBERTweet.....            | 40 |
| 4.3.3 Pembagian Data Latih dan Data Uji (Splitting Dataset)..... | 42 |
| 4.3.4 Augmentasi Data Synonym Replacement .....                  | 44 |
| 4.3.5 Penyimpanan Dataset Hasil Pra-pemrosesan.....              | 46 |
| 4.4 Pemodelan dengan Machine Learning dan Deep Learning .....    | 46 |
| 4.4.1 Implementasi Model Support Vector Machine (SVM).....       | 46 |
| 4.4.2 Implementasi Model IndoBERTweet .....                      | 53 |
| 4.5 Evaluasi Model Machine Learning dan Deep Learning .....      | 58 |
| 4.5.1 Hasil Evaluasi Model SVM .....                             | 59 |
| 4.5.2 Hasil Evaluasi Model IndoBERTweet.....                     | 60 |
| 4.5.3 Perbandingan Kinerja Model .....                           | 61 |
| 4.5.4 Kesimpulan Model Optimal.....                              | 61 |
| BAB V PENUTUP .....                                              | 63 |
| 5.1 Kesimpulan .....                                             | 63 |
| 5.2 Saran .....                                                  | 63 |
| REFERENSI .....                                                  | 65 |
| LAMPIRAN.....                                                    | 70 |

## DAFTAR TABEL

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Tabel 2.1 Keaslian Penelitian                                             | 7  |
| Tabel 2.2 Confusion matrix biner                                          | 23 |
| Tabel 3.1 Contoh variable dataset <i>full_text</i>                        | 29 |
| Tabel 3.2 Perangkat Keras                                                 | 30 |
| Tabel 3.3 Perangkat Lunak                                                 | 31 |
| Tabel 4.1 Hasil labeling menggunakan model pre-trained                    | 37 |
| Tabel 4.2 Hasil labeling menggunakan model pre-trained                    | 38 |
| Tabel 4.3 Perbandingan teks sebelum dan sesudah preprocessing SVM         | 40 |
| Tabel 4.4 Perbandingan teks sebelum dan sesudah preprocessing IndoBERTwee | 40 |
| Tabel 4.5 Perbandingan F1-Score antara Model SVM dan IndoBERTweet         | 61 |

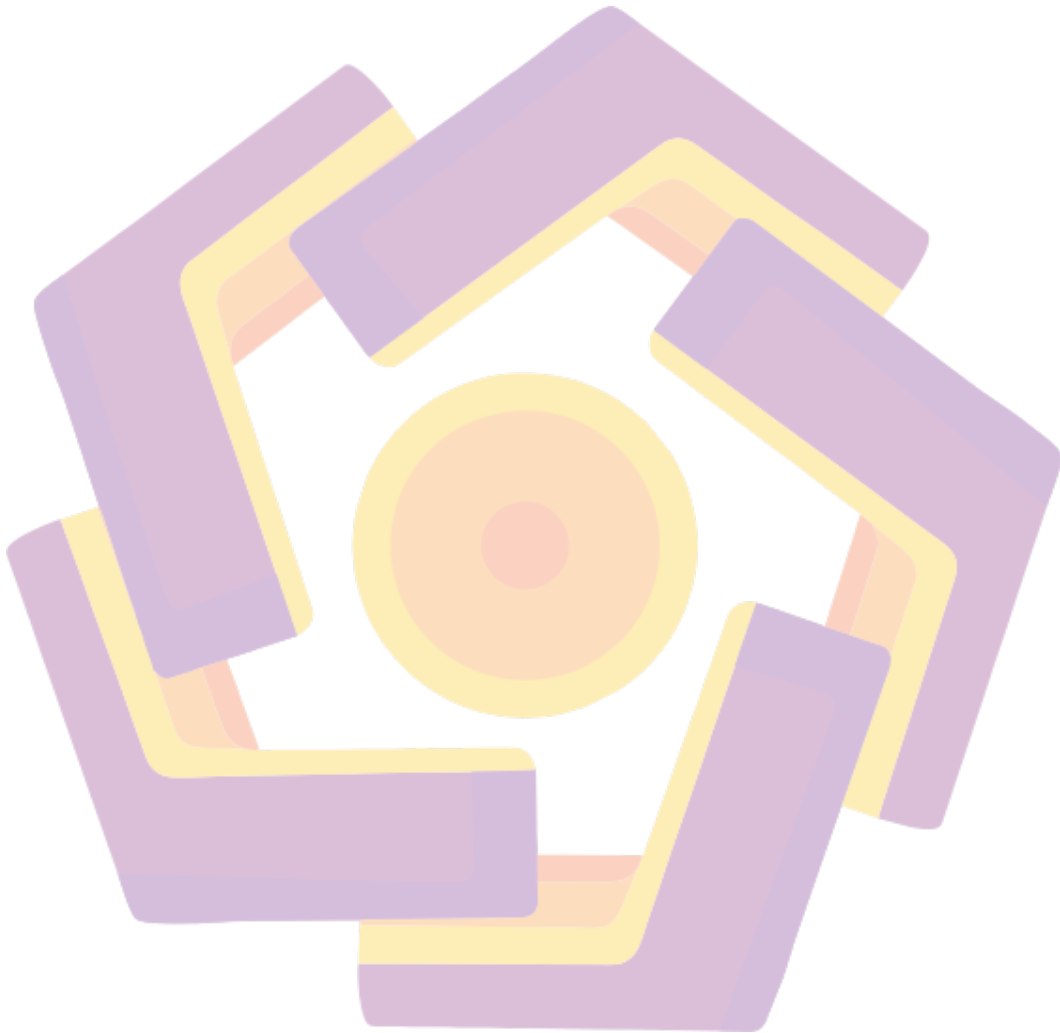
## DAFTAR GAMBAR

|                                                                                         |        |
|-----------------------------------------------------------------------------------------|--------|
| Gambar 2.1 Laporan penggunaan platform X 2025                                           | 11     |
| Gambar 2.2 Laporan hasil survei pemindahan Ibu Kota Negara (IKN) tahun 2020 sampai 2022 | 14     |
| Gambar 2.3 Proses klasifikasi SVM                                                       | 16     |
| Gambar 2.4 Grafik untuk perbandingan Akurasi                                            | 18     |
| <br>Gambar 3.1 Alur Penelitian                                                          | <br>25 |
| <br>Gambar 4.1 Proses esekusi Tweet Harvest                                             | <br>35 |
| Gambar 4.2 Jumlah data awal                                                             | 36     |
| Gambar 4.3 Hasil proses drop duplicate                                                  | 36     |
| Gambar 4.4 Sebelum dan sesudah preprocessing model SVM                                  | 40     |
| Gambar 4.5 Sebelum dan sesudah preprocessing model IndoBERTweet                         | 42     |
| Gambar 4.6 Pemisahan Data model SVM 80/20                                               | 43     |
| Gambar 4.7 Pemisahan Data model IndoBERTweet 70/10/20                                   | 44     |
| Gambar 4.8 Data Imbalance                                                               | 44     |
| Gambar 4.9 Distribusi data sesudah augmentasi                                           | 45     |
| Gambar 4.10 Data model SVM 80/20                                                        | 47     |
| Gambar 4.11 Proses menemukan kombinasi parameter terbaik                                | 48     |
| Gambar 4.12 Proses re-tuning model                                                      | 50     |
| Gambar 4.13 Hasil evaluasi tahap awal                                                   | 51     |
| Gambar 4.14 Hasil evaluasi re-tuning model                                              | 51     |
| Gambar 4.15 Hasil Confusion Matrix tahap awal                                           | 52     |
| Gambar 4.16 Hasil Confusion Matrix tahap re-tuning                                      | 53     |
| Gambar 4.17 Pembagian data model IndoBERTweet 70/10/20                                  | 54     |
| Gambar 4.18 Pencarian hyperparameter terbaik                                            | 55     |
| Gambar 4.19 Hasil pelatihan hyperparameter terbaik                                      | 56     |
| Gambar 4.20 Evaluasi hasil hyperparameter                                               | 57     |
| Gambar 4.21 Hasil Confusion Matrix                                                      | 58     |
| Gambar 4.22 Hasil evaluasi final pada model SVM                                         | 59     |
| Gambar 4.23 Hasil Confusion Matrix final pada model SVM                                 | 59     |
| Gambar 4.24 Hasil evaluasi final pada model IndoBERTweet                                | 60     |
| Gambar 4.25 Hasil Confusion Matrix final pada model IndoBERTweet                        | 60     |

## DAFTAR LAMPIRAN

Lampiran 1 link *google drive* program

70

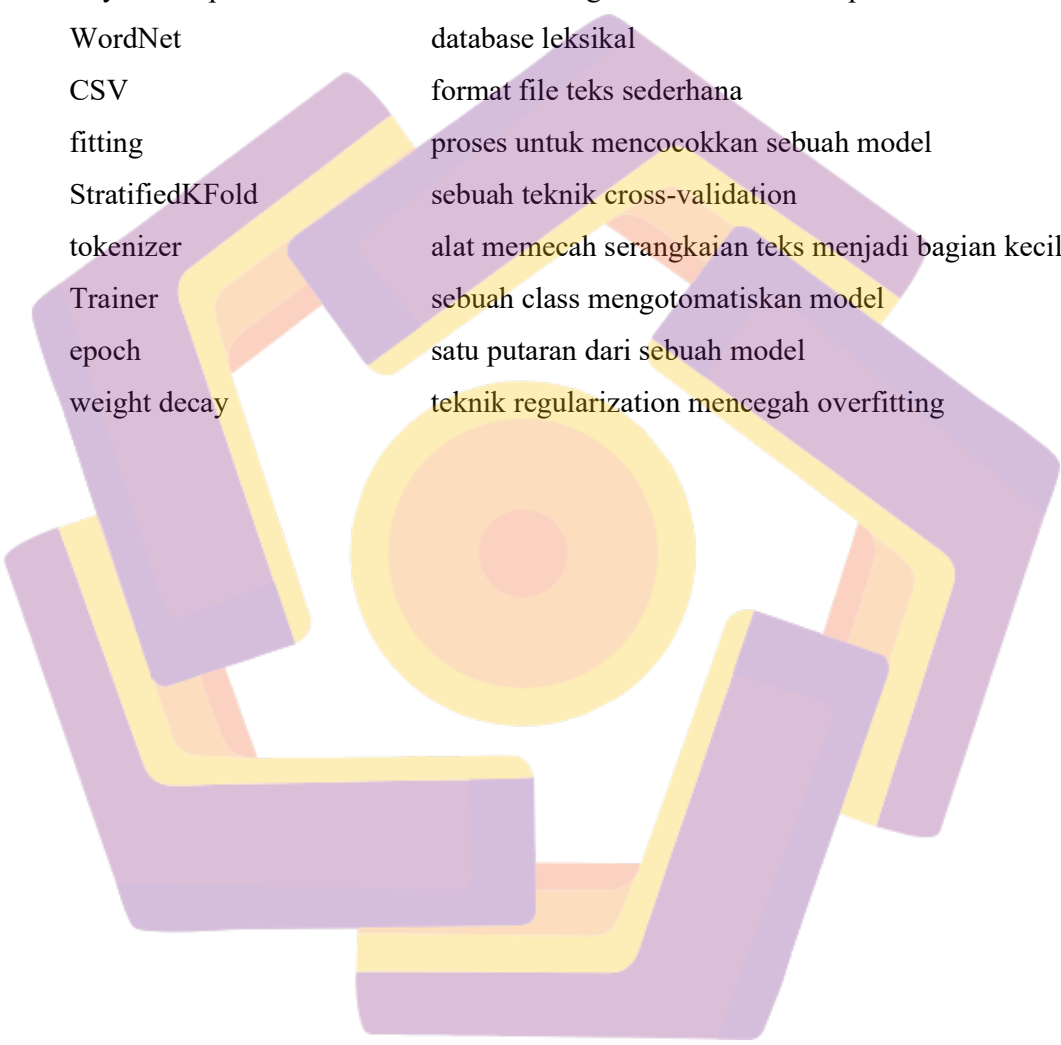


## DAFTAR LAMBANG DAN SINGKATAN



|        |                                           |
|--------|-------------------------------------------|
| SVM    | Support Vector Machines                   |
| ML     | Machine Learning                          |
| NLP    | Natural Language Processing               |
| IKN    | Ibu Kota Negara                           |
| API    | Application Programming Interface         |
| CSV    | Comma Separated Values                    |
| TF-IDF | Term Frequency-Inverse Document Frequency |
| TF     | Term Frequency                            |
| IDF    | Inverse Document Frequency                |
| LLM    | Large Language Model                      |
| NLTK   | Natural Language Toolkit                  |

## DAFTAR ISTILAH



|                       |                                                    |
|-----------------------|----------------------------------------------------|
| C                     | parameter kontinu                                  |
| gamma                 | menentukan pengaruh data pelatihan                 |
| Bayesian Optimization | sebuah strategi menemukan nilai optimal            |
| WordNet               | database leksikal                                  |
| CSV                   | format file teks sederhana                         |
| fitting               | proses untuk mencocokkan sebuah model              |
| StratifiedKFold       | sebuah teknik cross-validation                     |
| tokenizer             | alat memecah serangkaian teks menjadi bagian kecil |
| Trainer               | sebuah class mengotomatiskan model                 |
| epoch                 | satu putaran dari sebuah model                     |
| weight decay          | teknik regularization mencegah overfitting         |

## INTISARI

Pemindahan Ibu Kota Negara (IKN) telah memicu diskursus publik yang luas di platform media sosial seperti X, namun pemahaman akurat terhadap sentimen masyarakat terhambat oleh kompleksitas bahasa informal dan ketidakseimbangan data. Penelitian ini melakukan analisis komparatif untuk menentukan model klasifikasi sentimen yang paling optimal antara *Support Vector Machine* (SVM) yang berbasis fitur *TF-IDF* dan *IndoBERTweet*, sebuah model *Transformer* yang telah di-*fine-tuning*. Metodologi penelitian mencakup akuisisi data melalui platform X, pra-pemrosesan data yang disesuaikan untuk setiap model, augmentasi data latih menggunakan *synonym replacement* untuk mengatasi ketidakseimbangan kelas, serta optimasi *hyperparameter* sistematis menggunakan *BayesSearchCV* untuk model *SVM* dan *Optuna* untuk model *IndoBERTweet*. Hasil evaluasi pada data uji yang tidak seimbang menunjukkan bahwa *IndoBERTweet* secara signifikan mengungguli *SVM*, dengan capaian akurasi 89,17% dan *F1-score* makro sebesar 0,77 dibandingkan model *SVM* dengan akurasi 83,59% dan *F1-score* makro 0,62. Keunggulan utama *IndoBERTweet* terletak pada kemampuannya yang superior dalam mengklasifikasikan sentimen minoritas (positif dan negatif), membuktikan efektivitas arsitektur berbasis konteks dalam menangani opini publik media sosial khususnya pada platform X. Temuan ini memberikan rekomendasi model yang lebih andal bagi pemerintah dan analisis kebijakan untuk memantau opini publik secara akurat, mendukung pengambilan keputusan berbasis data yang lebih baik.

**Kata kunci:** Analisis Sentimen, Ibu Kota Negara (IKN), *IndoBERTweet*, *Support Vector Machine* (SVM), Platform X.

## ***ABSTRACT***

The relocation of the Ibu Kota Negara (IKN) has sparked widespread public discourse on social media platforms such as X. However, accurately capturing public sentiment is hindered by the complexity of informal language and data imbalance. This study conducts a comparative analysis to determine the most optimal sentiment classification model between the *TF-IDF* based *Support Vector Machine* (SVM) and *IndoBERTweet*, a fine-tuned *Transformer* model. The research methodology includes data acquisition from platform X, data preprocessing tailored to each model, training data *augmentation* using synonym replacement to address class imbalance, and systematic hyperparameter optimization using *BayesSearchCV* for the *SVM* model and *Optuna* for *IndoBERTweet*. Evaluation results on imbalanced test data show that *IndoBERTweet* significantly outperforms *SVM*, achieving an *accuracy* of 89.17% and a *macro F1-score* of 0.77, compared to *SVM's accuracy* of 83.59% and *macro F1-score* of 0.62. The main advantage of *IndoBERTweet* lies in its superior ability to classify minority sentiments (positive and negative), demonstrating the effectiveness of context-based architectures in handling public opinion on social media, particularly on platform X. These findings offer a more reliable model recommendation for governments and policy analysts to monitor public sentiment accurately, supporting better data-driven decision-making.

**Keyword:** Sentiment Analysis, Ibu Kota Negara (IKN), IndoBERTweet, Support Vector Machine (SVM), Platform X.