

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
KEMENTERIAN PUPR PASCA ISU DITETAPKAN
UU TAPERA MENGGUNAKAN
ALGORITMA BERT**

SKRIPSI

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



disusun oleh

NDARU WINDRA SAYIDINAA

21.11.4089

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
KEMENTERIAN PUPR PASCA ISU DITETAPKAN
UU TAPERA MENGGUNAKAN
ALGORITMA BERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NDARU WINDRA SAYIDINAA

21.11.4089

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEMENTERIAN
PUPR PASCA ISU DITETAPKAN UU TAPERA MENGGUNAKAN
ALGORITMA BERT**

yang disusun dan diajukan oleh

NDARU WINDRA SAYIDINAA

21.11.4089

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Januari 2025

Dosen Pembimbing,


Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN MASYARAKAT TERHADAP KEMENTERIAN
PUPR PASCA ISU DITETAPKAN UU TAPERA MENGGUNAKAN
ALGORITMA BERT

yang disusun dan diajukan oleh

NDARU WINDRA SAYIDINAA

21.11.4089

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Januari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda-Tangan

Firman Ashurudin, S.Kom., M.Kom.
NIK. 190302315

Andriyan Dwi Putra, S.Kom., M.Kom.
NIK. 190302270

Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ndaru Windra Sayidinaa
NIM : 21.11.4089

Menyatakan bahwa Skripsi dengan judul berikut:

**ANALISIS SENTIMEN MASYARAKAT TERHADAP KEMENTERIAN
PUPR PASCA ISU DITETAPKAN UU TAPERA MENGGUNAKAN
ALGORITMA BERT**

Dosen Pembimbing : Ferian Fauzi Abdulloh, M.Kom

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, 20 Januari 2025

Yang Menyatakan,



Ndaru Windra Sayidinaa

HALAMAN PERSEMBAHAN

Alhamdulillah, segala puji dan syukur hanya kepada Allah SWT, yang telah melimpahkan rahmat, hidayah, dan kekuatan kepada penulis sehingga skripsi ini dapat diselesaikan dengan baik. Dalam rasa syukur ini, penulis mempersembahkan skripsi ini sebagai ungkapan terima kasih kepada semua pihak yang telah memberikan dukungan, doa, dan motivasi sepanjang proses penelitian dan penulisan skripsi ini, sehingga dapat terselesaikan dengan baik dan lancar.

Dengan penuh rasa hormat dan cinta, penulis khususnya mempersembahkan terima kasih kepada:

1. Dengan penuh rasa puji dan syukur terhadap Allah SWT atas semua rahmat, karunia dan ridho-Nya. Penulis diberikan kemampuan, kemudahan serta kelancaran untuk menyelesaikan skripsi ini dengan baik.
2. Ayahanda dari penulis, Bapak Widodo, atas semua perjuangan, dukungan, inspirasi, semangat, dan doa yang tiada henti yang beliau berikan kepada penulis. Terimakasih penulis sampaikan karena sebuah kepercayaan penuh yang telah diberikan kepada penulis untuk menyelesaikan masa studi. Semoga karya ini dapat menjadi ungkapan kecil dari rasa terima kasih atas segala cinta dan kebaikan Ayahanda selama ini.
3. Ibu dari penulis, Ibu Indra Wahyuning Tiyas, atas segala perjuangan, pengorbanan, doa, dan kasih sayang yang tiada putus telah menjadi penyemangat dan sumber kekuatan penulis dalam menyelesaikan skripsi ini. Kehangatan, perhatian, dan dukungan yang Ibu berikan di setiap langkah perjalanan ini adalah anugerah yang tak ternilai. Semoga karya ini menjadi ungkapan kecil dari rasa syukur dan terima kasih atas segala cinta dan dedikasi yang Ibu curahkan sepanjang hidup penulis.
4. Seluruh sahabat dan kerabat penulis, yang telah memberikan dukungan, semangat, dan doa selama masa perkuliahan dan proses penyelesaian skripsi ini. Kehadiran kalian menjadi sumber kebahagiaan dan motivasi yang luar biasa, baik melalui kata-kata penyemangat, bantuan, maupun kebersamaan

yang penuh makna. Semoga kebaikan dan perhatian yang kalian berikan selalu dibalas dengan keberkahan dan kebahagiaan. Skripsi ini penulis persembahkan sebagai bentuk penghargaan atas dukungan luar biasa yang telah kalian berikan.

5. Diri sendiri Ndaru Windra Sayidinaa, atas keteguhan hati, kerja keras, dan kesabaran yang telah ditunjukkan dalam menghadapi setiap tantangan yang datang selama perjalanan ini. Meskipun rasa lelah dan keraguan sering dirasakan, langkah untuk terus maju tetap diambil, dan setiap bagian diselesaikan dengan usaha terbaik. Pengalaman ini dijadikan sebagai bukti bahwa batasan dapat dilampaui, dan pembelajaran untuk menjadi pribadi yang lebih kuat serta bijaksana terus dilakukan. Skripsi ini dianggap bukan hanya sebuah karya akademik, tetapi juga simbol dari perjuangan, tekad, dan pertumbuhan diri selama ini. Keyakinan pada mimpi dan kemampuan diharapkan tetap dijaga, karena perjalanan ini baru saja dimulai.



KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas rahmat, karunia, dan hidayah-Nya, sehingga skripsi berjudul "Analisis Sentimen Masyarakat terhadap Kementerian PUPR Pasca Isu UU Tapera Menggunakan Algoritma BERT" ini dapat diselesaikan. Skripsi ini disusun sebagai syarat kelulusan S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Proses penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak prof. Dr. M. Suyanto, M.M., selaku rektor Universitas Amikom Yogyakarta, atas kesempatan dan fasilitas yang diberikan selama penulisan skripsi ini.
2. Bapak Ferian Fauzi Abdulloh, M.Kom., sebagai dosen pembimbing, yang telah memberikan arahan, masukan, dan bimbingan yang menjadi pendorong utama dalam penyelesaian skripsi ini. Ucapan terima kasih yang mendalam ditujukan atas kesediaan dan dedikasi beliau.
3. Ayahanda, Bapak Widodo, dan Ibunda, Ibu Indra Wahyuning Tiyas Terima kasih atas doa, dukungan, dan kasih sayang yang tiada henti. Semua semangat dan pengorbanan yang diberikan menjadi kekuatan penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini bermanfaat dan memberikan kontribusi bagi pengembangan ilmu pengetahuan.

Yogyakarta, 20 Januari 2025



Ndaru Windra Sayidinaa

DAFTAR ISI

COVER HITAM PUTIH	i
COVER WARNA	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
DAFTAR LAMBANG DAN SINGKATAN	xvi
DAFTAR ISTILAH	xvii
INTISARI	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7

2.1	Studi Literatur	7
2.2	Dasar Teori.....	18
2.2.1	Analisis Sentimen	18
2.2.2	Kementerian PUPR.....	18
2.2.3	Tapera	19
2.2.4	<i>Bidirectional Encoder Representation from Transformers (BERT)</i> 19	
2.2.5	<i>IndoBERT</i>	21
2.2.6	<i>Smote</i>	22
2.2.7	<i>K-Fold Cross Validation</i>	22
2.2.8	<i>Confusion Matrix</i>	23
BAB III METODE PENELITIAN		25
3.1	Alur Penelitian	25
3.1.1	<i>Scraping Data</i>	27
3.1.2	<i>Labelling Data</i>	29
3.1.3	<i>Preprocessing Text</i>	29
3.1.4	<i>SMOTE</i>	30
3.1.5	<i>Wordcloud</i>	31
3.1.6	<i>Splitting Data</i>	31
3.1.7	<i>IndoBERT</i>	31
3.1.8	<i>K-Fold Cross validation</i>	32
3.1.9	<i>Evaluasi Confusion Matrix</i>	32
3.2	Alat dan Bahan.....	32
3.2.1	Alat Penelitian.....	32
3.2.2	Data Penelitian	33

BAB IV HASIL DAN PEMBAHASAN	34
4.1 <i>Scrapping Data</i>	34
4.2 <i>Labelling Dataset</i>	34
4.3 <i>Preprocessing Text</i>	36
4.3.1 <i>Cleaning Data</i>	36
4.3.2 <i>Case Folding</i>	38
4.3.3 <i>Word Normalization</i>	39
4.3.4 <i>Tokenizing</i>	40
4.3.5 <i>Stopword Removal</i>	41
4.3.6 <i>Stemming</i>	42
4.4 <i>Smote</i>	43
4.5 <i>WordCloud</i>	45
4.6 <i>Split Data</i>	45
4.7 <i>Implementasi Bert dan K-Fold Cross Validation</i>	46
4.8 <i>Evaluasi Confusion Matrix</i>	55
BAB V PENUTUP	61
5.1 Kesimpulan	61
5.2 Saran	62
REFERENSI	63
LAMPIRAN.....	67

DAFTAR TABEL

Tabel 2.1 - Keaslian Penelitian	10
Tabel 2.2 - Confusion Matrix.....	23
Tabel 3.1 - Jumlah data <i>scraping</i>	27
Tabel 4.1 - <i>Sample Dataset Hasil Scraping</i>	34
Tabel 4.2 - <i>Sample Dataset Hasil Labelling</i>	35
Tabel 4.3 - Hasil <i>Cleaning Text</i>	38
Tabel 4.4 - Hasil <i>Case Folding</i>	38
Tabel 4.5 - Hasil <i>Word Normalization</i>	40
Tabel 4.6 - Hasil <i>Tokenizing</i>	41
Tabel 4.7 - Hasil <i>Stopword Removal</i>	42
Tabel 4.8 - Hasil <i>Stemming</i>	43
Tabel 4.10 - Tokenisasi <i>Bert</i>	46
Tabel 4.11 - <i>Hyperparamater</i>	48
Tabel 4.12 - Tabel Keseluruhan <i>Classification Report</i>	60

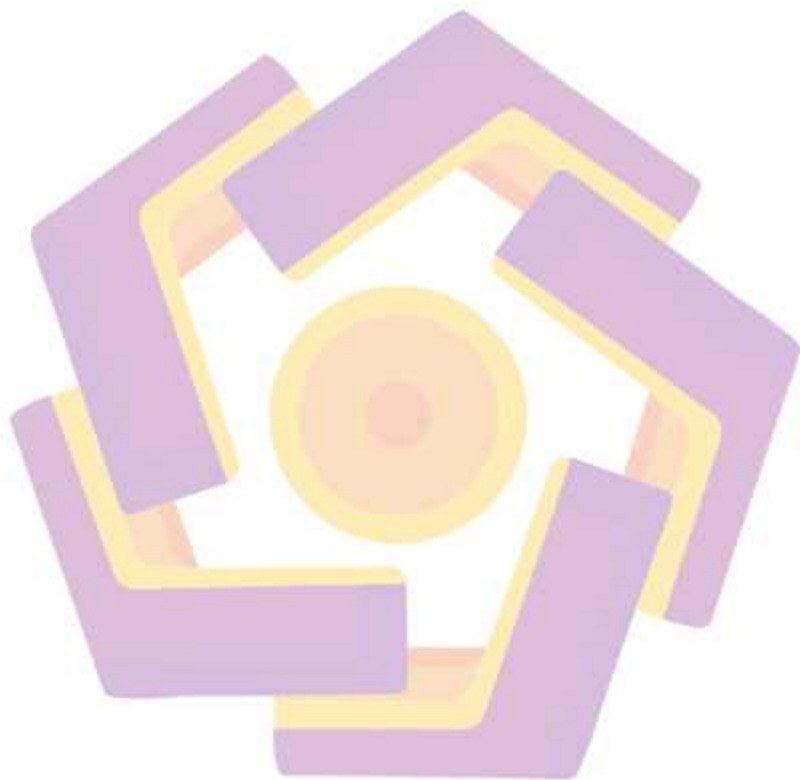
DAFTAR GAMBAR

Gambar 2.1 - Arsitektur <i>Transformers</i>	20
Gambar 2.2 - Reprerentasi <i>Input Bert</i>	20
Gambar 2.3 - Ilustrasi <i>Pre-training</i> dan <i>Fine-tuning</i>	21
Gambar 2.4 - Proses <i>K-Fold Cross Validation</i>	23
Gambar 3.1 - <i>Flowchart</i> Metode Penelitian	26
Gambar 3.2 - Grafik Persebaran Jumlah Dataset	29
Gambar 4.1 - Kode untuk menghitung jumlah nilai <i>null</i>	36
Gambar 4.2 - <i>Output</i> dari nilai <i>null</i>	36
Gambar 4.3 - Kode untuk menghapus nilai <i>null</i>	37
Gambar 4.4 - Kode untuk menghitung kalimat duplikat	37
Gambar 4.5 - Kode untuk menghapus kalimat duplikat	37
Gambar 4.6 - Kode <i>Cleaning Text</i>	37
Gambar 4.7 - Kode <i>Case Folding</i>	38
Gambar 4.8 - <i>Library Word Normalization</i>	39
Gambar 4.9 - Kode <i>Word Normalization</i>	39
Gambar 4.10 - Kode <i>Tokenizing</i>	40
Gambar 4.11 - Kode <i>Stopword Removal</i>	41
Gambar 4.12 - Kode <i>Stemming</i>	42
Gambar 4.13 - Distribusi Data Sebelum <i>Smote</i>	44
Gambar 4.14 - Distribusi Data Setelah <i>Smote</i>	44
Gambar 4.15 - Visualisasi Kata Negatif	45
Gambar 4.16 - Visualisasi Kata Positif	45
Gambar 4.17 - Ilustrasi <i>Fine-tuning</i> klasifikasi <i>IndoBERT</i>	48
Gambar 4.18 - <i>Training loss</i> dan <i>validation loss</i> skema 1 data 6000	49
Gambar 4.19 - <i>Training acc</i> dan <i>validation acc</i> skema 1 data 6000	49
Gambar 4.20 - <i>Training loss</i> dan <i>validation loss</i> skema 1 data 10000	50
Gambar 4.21 - <i>Training acc</i> dan <i>validation acc</i> skema 1 data 10000	50
Gambar 4.22 - <i>Training loss</i> dan <i>validation loss</i> skema 1 data 14000	50
Gambar 4.23 - <i>Training acc</i> dan <i>validation acc</i> skema 1 data 14000	51

Gambar 4.24 - <i>Training loss</i> dan <i>validation loss</i> skema 2 data 6000	52
Gambar 4.25 - <i>Training acc</i> dan <i>validation acc</i> skema 2 data 6000	52
Gambar 4.26 - <i>Training loss</i> dan <i>validation loss</i> skema 2 data 10000	53
Gambar 4.27 - <i>Training acc</i> dan <i>validation acc</i> skema 2 data 10000	53
Gambar 4.28 - <i>Training loss</i> dan <i>validation loss</i> skema 2 data 14000	54
Gambar 4.29 - <i>Training acc</i> dan <i>validation acc</i> skema 2 data 14000	54
Gambar 4.30 - <i>Confusion matrix</i> skema 1 data 6000.....	55
Gambar 4.31 - <i>Confusion matrix</i> skema 1 data 10000.....	55
Gambar 4.32 - <i>Confusion matrix</i> skema 1 data 14000.....	56
Gambar 4.33 - <i>Classification report</i> skema 1 data 6000	56
Gambar 4.34 - <i>Classification report</i> skema 1 data 10000	57
Gambar 4.35 - <i>Classification report</i> skema 1 data 14000	57
Gambar 4.36 - <i>Confusion matrix</i> skema 2 data 6000.....	58
Gambar 4.37 - <i>Confusion matrix</i> skema 2 data 10000.....	58
Gambar 4.38 - <i>Confusion matrix</i> skema 2 data 14000.....	59
Gambar 4.39 - <i>Classification report</i> skema 2 data 6000	59
Gambar 4.40 - <i>Classification report</i> skema 2 data 10000	59
Gambar 4.41 - <i>Classification report</i> skema 2 data 14000	60

DAFTAR LAMPIRAN

Lampiran 1 - Kode <i>Scrapping</i> Data	67
Lampiran 2 - Kode Proses <i>Fine-Tuning Bert</i>	69



DAFTAR LAMBANG DAN SINGKATAN



TAPERA	Tabungan Perumahan Rakyat
PUPR	Pekerjaan Umum dan Perumahan Rakyat
PP	Peraturan Pemerintah
UU	Undang - Undang
<i>BERT</i>	<i>Bidirectional Encoder From Transformers</i>
<i>NLP</i>	<i>Natural Language Processing</i>
<i>ML</i>	<i>Machine Learning</i>
<i>MLM</i>	<i>Masked Language Model</i>
<i>RNN</i>	<i>Recurrent Neural Network</i>
<i>CNN</i>	<i>Convolutional Neural Network</i>
<i>NSP</i>	<i>Next Sentence Prediction</i>
<i>AdamW</i>	<i>Adaptive Moment Estimation Weight Decay</i>
<i>CLS</i>	<i>Classification Token</i>
<i>PAD</i>	<i>Padding Token</i>
<i>SEP</i>	<i>Separator Token</i>
<i>DL</i>	<i>Deep Learning</i>
<i>AI</i>	<i>Artificial Intelligence</i>
<i>SMOTE</i>	<i>Synthetic Minority Oversampling Technique</i>
<i>API v3</i>	<i>Application Programming Interface Version 3</i>
<i>GPU</i>	<i>Graphics Processing Unit</i>
<i>CPU</i>	<i>Central Processing Unit</i>
<i>CSV</i>	<i>Comma-Separated Values</i>
<i>RAM</i>	<i>Random Access Memory</i>
<i>SSD</i>	<i>Solid-State Drive</i>

DAFTAR ISTILAH

<i>Self-Attention</i>	Memahami hubungan antar-kata dalam teks.
<i>Encoder</i>	Mengolah <i>input</i> menjadi representasi tersembunyi.
<i>Decoder</i>	Menghasilkan <i>output</i> dari representasi tersembunyi.
<i>Embedding</i>	Vektor numerik untuk merepresentasikan teks.
<i>Embedding Layer</i>	Lapisan untuk membuat vektor <i>embedding</i> .
<i>Fine-Tuning</i>	Menyesuaikan model pra-latih untuk tugas spesifik.
<i>Pre-Trained</i>	Model dilatih sebelumnya pada dataset besar.
<i>Overfitting</i>	Model terlalu cocok dengan data latih, buruk di data baru.
<i>Confusion Matrix</i>	Matriks untuk mengevaluasi performa klasifikasi.
<i>Precision</i>	Proporsi prediksi benar dari kelas tertentu.
<i>Recall</i>	Proporsi sampel relevan yang ditemukan.
<i>F1-Score</i>	Rata-rata harmonis <i>precision</i> dan <i>recall</i> .
<i>Accuracy</i>	Proporsi prediksi benar dari total data.
<i>Wordcloud</i>	Visualisasi kata berdasarkan frekuensinya.
<i>Scraping</i>	Ekstraksi data dari web.
<i>Preprocessing</i>	Persiapan data sebelum analisis.
<i>Case Folding</i>	Mengubah teks ke huruf kecil.
<i>Word Normalization</i>	Mengubah kata ke bentuk baku.
<i>Tokenizing</i>	Memecah teks jadi kata/token.
<i>Stopword Removal</i>	Menghapus kata-kata umum yang tidak penting.
<i>Stemming</i>	Mengubah kata ke bentuk dasar.
<i>AdamW</i>	Optimasi dengan regularisasi bobot.
<i>Hyperparameter</i>	Parameter yang diatur sebelum pelatihan.
<i>Epoch</i>	Satu siklus pelatihan penuh.
<i>Batch Size</i>	Jumlah data per iterasi pelatihan.
<i>Epsilon</i>	Nilai kecil untuk mencegah pembagian nol.
<i>Optimizer</i>	Algoritma pembaruan bobot.
<i>Learning Rate</i>	Kecepatan pembaruan parameter model.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Kementerian PUPR pasca ditetapkannya Peraturan Pemerintah Nomor 21 Tahun 2024 tentang Tabungan Perumahan Rakyat (Tapera). Masalah yang diangkat adalah adanya pro dan kontra di masyarakat terkait kebijakan iuran Tapera sebesar 3%, yang terdiri dari 2,5% dari gaji pekerja swasta dan 0,5% dari pemberi kerja. Kebijakan ini memunculkan pandangan yang beragam terhadap kinerja Kementerian PUPR dan implementasi UU Tapera. Dampaknya, opini yang tidak jelas atau tidak terukur dapat memengaruhi persepsi masyarakat dan pengambilan keputusan oleh pemangku kebijakan. Untuk menyelesaikan masalah ini, penelitian menggunakan metode analisis sentimen dengan menerapkan algoritma BERT, memanfaatkan model *pre-trained* dari Hugging Face, yaitu *indobenchmark/indobert-base-pl*, dan dataset IndoNLU. Data diperoleh dari *scraping* komentar YouTube, menghasilkan 15.984 baris data yang dilabeli secara manual menjadi sentimen positif dan negatif. Penelitian ini juga mencakup *preprocessing* teks, *balancing* data menggunakan teknik SMOTE, serta validasi model dengan *k-fold cross validation*. Hasil penelitian menunjukkan bahwa kombinasi teknik SMOTE dan *k-fold cross validation* mampu meningkatkan performa model, dengan akurasi mencapai 99%. Distribusi data akhir memberikan hasil analisis yang representatif terhadap opini masyarakat. Penelitian ini diharapkan dapat memberikan wawasan komprehensif bagi pemangku kebijakan dalam mengevaluasi kinerja Kementerian PUPR serta menjadi dasar untuk merumuskan regulasi yang lebih efektif. Penelitian lebih lanjut dapat dilakukan dengan memperluas data atau menggunakan teknik lain untuk meningkatkan akurasi analisis sentimen.

Kata kunci: analisis sentimen, bert, kementerian pupr, tapera, transformers.

ABSTRACT

This study aims to analyze public sentiment towards the Ministry of PUPR after the enactment of Government Regulation Number 21 of 2024 concerning Public Housing Savings (Tapera). The problem raised is that there are pros and cons in the community regarding the Tapera contribution policy of 3%, consisting of 2.5% of private workers' salaries and 0.5% from employers. This policy has led to mixed views on the performance of the PUPR Ministry and the implementation of the Tapera Law. As a result, unclear or unmeasured opinions can affect public perception and decision-making by policy makers. To solve this problem, the research used sentiment analysis methods by applying the BERT algorithm, utilizing pre-trained models from Hugging Face, namely indobenchmark/indo-bert-base-p1, and the IndoNLU dataset. Data was obtained from scraping YouTube comments, resulting in 15,984 rows of data that were manually labeled into positive and negative sentiments. This research also includes text preprocessing, data balancing using the SMOTE technique, and model validation with k-fold cross validation. The results show that the combination of SMOTE and k-fold cross validation techniques can improve model performance, with accuracy reaching 99%. The final data distribution provides analysis results that are representative of public opinion. This research is expected to provide comprehensive insights for policy makers in evaluating the performance of the Ministry of PUPR and become the basis for formulating more effective regulations. Further research can be conducted by expanding the data or using other techniques to improve the accuracy of sentiment analysis.

Keyword: bert, ministry of pupr, sentiment analysis, tapera, transformers.