

**SENTIMENT ANALYSIS OF ECONOMIC POLICY
COMMENTS ON YOUTUBE USING ENSEMBLE MACHINE
LEARNING**

SKRIPSI NON-REGULER JALUR SCIENTIST

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



Disusun oleh :

KETY NANDINI

22.11.4595

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**SENTIMENT ANALYSIS OF ECONOMIC POLICY COMMENTS
ON YOUTUBE USING ENSEMBLE MACHINE LEARNING**

SKRIPSI NON-REGULER JALUR SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



Disusun oleh :

KETY NANDINI

22.11.4595

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**SENTIMENT ANALYSIS OF ECONOMIC POLICY COMMENTS ON
YOUTUBE USING ENSEMBLE MACHINE LEARNING**

yang disusun dan diajukan oleh

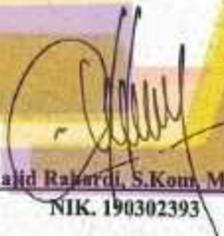
Kety Nandini

22.11.4595

telah disetujui oleh Dosen Pembimbing

pada tanggal 20 Oktober 2025

Dosen Pembimbing,



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

HALAMAN PENGESAHAN

JALUR NON-REGULER

SENTIMENT ANALYSIS OF ECONOMIC POLICY COMMENTS ON
YOUTUBE USING ENSEMBLE MACHINE LEARNING

yang disusun dan diajukan oleh

Kety Nandini
22.11.4595

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Oktober 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifyanto Hadinegoro, S.Kom., M.T.
NIK. 190302289

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248

Majid Rahardj, S.Kom., M.Eng.
NIK. 190302393

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Oktober 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriati, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Kety Nandini

NIM : 22.11.4595

Menyatakan bahwa Laporan dengan judul berikut:

Sentiment Analysis of Economic Policy Comments on YouTube Using Ensemble Machine Learning

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 kegiatan 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 ketidak-benaran 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,

Yang Menyatakan,



Kety Nandini

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas limpahan rahmat, kebijaksanaan, dan kekuatan yang telah mengiringi setiap langkah dalam proses penyusunan laporan ini. Laporan ini merupakan hasil dari proses penelitian sekaligus perjalanan pembelajaran dan tanggung jawab akademik. Penulis menyampaikan terima kasih yang tulus atas bimbingan, dukungan, dan kepercayaan kepada:

1. Orang tua dan adik tercinta, yang menjadi sumber kekuatan dalam diam, doa yang tidak pernah terputus, dan kasih yang tidak pernah berkurang nilainya oleh waktu.
2. Prof. Dr. Mohammad Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Prof. Dr. Kusriani, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Eli Pujastuti, M.Kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
5. Majid Rahardi, S.Kom, M.Eng selaku Dosen Pembimbing yang bukan hanya memberikan arahan ilmiah, tetapi juga menanamkan nilai ketelitian, kejujuran, dan keteguhan dalam berpikir.
6. Tim Dosen Penguji, atas pandangan kritis dan saran konstruktif yang memperkaya cara pandang penulis terhadap penelitian ini.

Penulis menyadari bahwa karya ini masih jauh dari sempurna, namun semoga laporan ini menjadi sarana pembelajaran sekaligus cerminan semangat untuk terus berkembang dalam bidang keilmuan.

Yogyakarta, 10 Oktober 2025



Kety Nandini

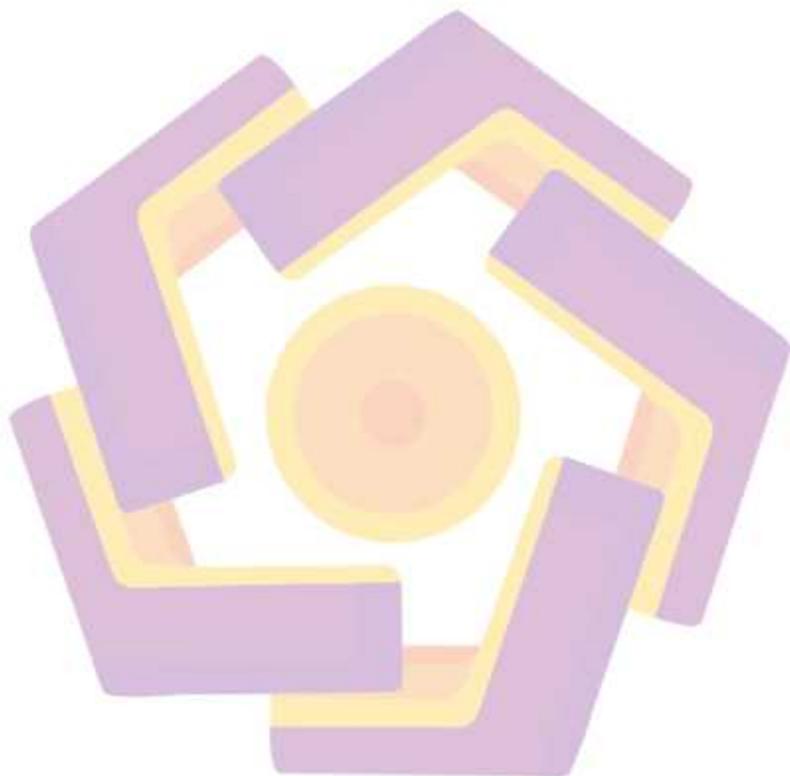
DAFTAR ISI

| | |
|--|------|
| Halaman Persetujuan..... | ii |
| Halaman Pengesahan | iii |
| Halaman Pernyataan Keaslian Karya..... | iv |
| Kata Pengantar | v |
| Daftar Isi..... | vi |
| Daftar Tabel..... | viii |
| Daftar Gambar..... | ix |
| Daftar Lampiran..... | x |
| Daftar Lambang dan Singkatan..... | xi |
| Daftar Istilah..... | xii |
| Intisari | xiv |
| <i>Abstract</i> | xv |
| Bab I Pendahuluan | 16 |
| 1.1. Gambaran Umum..... | 16 |
| 1.2. Rumusan Masalah | 16 |
| 1.3. Batasan Masalah..... | 17 |
| 1.4. Tujuan..... | 17 |
| Bab II Tinjauan Pustaka | 18 |
| 2.1. Studi Literatur..... | 18 |
| 2.2. Landasan Teori | 19 |
| BAB III Metode Penelitian | 24 |
| 3.1. Metode Penelitian..... | 24 |
| 3.2. Objek Penelitian | 25 |
| 3.3. Alur Penelitian..... | 26 |
| 3.3.1. Sumber Data | 28 |
| 3.3.2. Teknik Preprocessing Data | 28 |
| 3.3.3. Metode Pelabelan Data | 29 |
| 3.3.4. Teknik Augmentasi Data..... | 30 |
| 3.3.5. Feature Engineering..... | 30 |
| 3.3.6. Pembagian Dataset..... | 31 |
| 3.3.7. Algoritma Machine Learning..... | 31 |

| | |
|---|-----------|
| 3.3.8. Ensemble Learning | 33 |
| 3.3.9. Implementasi Teknik..... | 34 |
| 3.3.10. Evaluasi Model..... | 34 |
| BAB IV Pembahasan | 36 |
| 4.1. Kinerja Model Individual | 36 |
| 4.2. Analisis Per-Kelas Sentimen | 39 |
| 4.3. Dampak Data Augmentation | 41 |
| 4.4. Efektivitas Ensemble Learning..... | 42 |
| 4.5. Analisis Kesalahan dan Limitation..... | 44 |
| BAB V Kesimpulan | 46 |
| 5.1. Kesimpulan..... | 46 |
| 5.2. Saran..... | 47 |
| Referensi | 49 |
| Lampiran | 55 |
| Curriculum Vitae..... | 55 |
| Lampiran dan Bukti Pendukung..... | 56 |
| a. Letter of Acceptance (LOA)..... | 56 |
| b. Lembar Review | 57 |
| c. Bukti Terbit/Terindex | 60 |
| d. Bukti Pembayaran | 60 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 3.2. 1 Distribusi Sentimen..... | 25 |
| Tabel 3.2. 2 Contoh Dataset Hasil Crawling..... | 26 |
| Tabel 3.5. 1 Dataset Hasil Preprocessing..... | 29 |
| Tabel 4. 1 Performa Kinerja Model..... | 36 |
| Tabel 4.3 1 Distribusi Data dengan Tiga Metode Augmentasi..... | 42 |



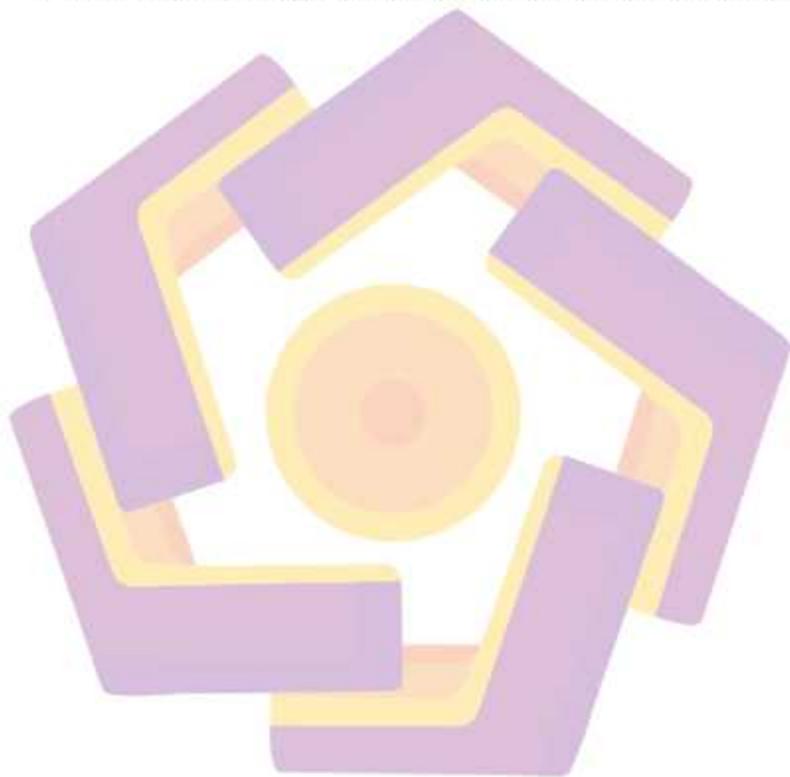
DAFTAR GAMBAR

| | |
|---|----|
| Gambar 4.1 1 Classification Report Baseline Model..... | 38 |
| Gambar 4.1 2 Baseline Model Confusion Matrix | 39 |
| Gambar 4.2. 1 Classification Report Model Ensemble..... | 40 |
| Gambar 4.2. 2 Confusion Matrix Model Ensemble | 40 |
| Gambar 4.3 1 Distribusi Data Augmentasi | 41 |
| Gambar 4.4 1 Confidence Score Model Ensemble | 43 |
| Gambar 4.4 2 Hasil Uji McNemar | 44 |



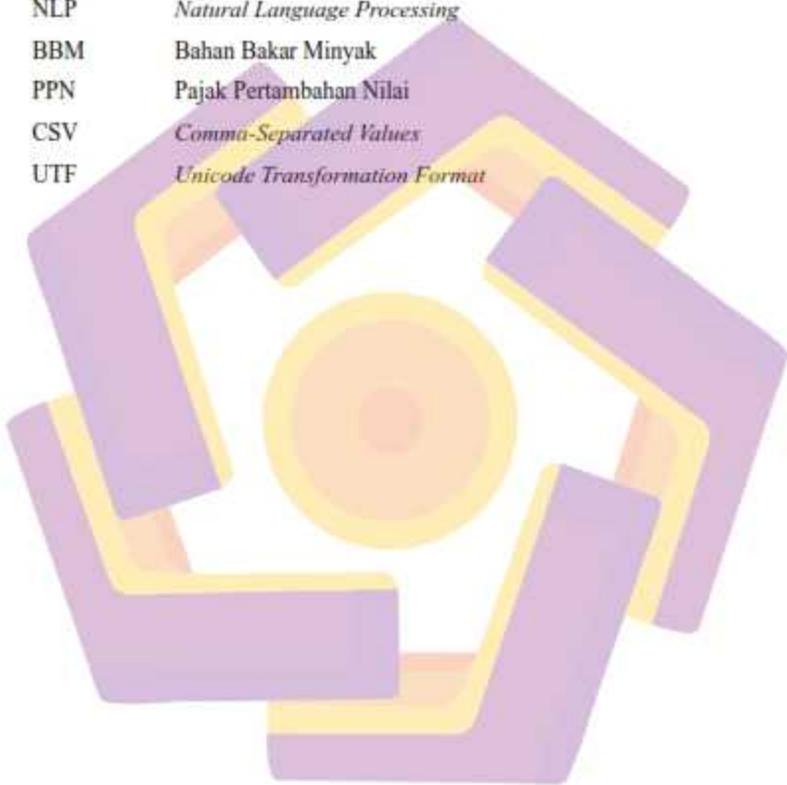
DAFTAR LAMPIRAN

| | |
|---|----|
| Lampiran 1. Curriculum Vitae | 55 |
| Lampiran 2. Letter of Acceptance (LOA)..... | 56 |
| Lampiran 3. Lembar Review..... | 57 |
| Lampiran 4. Bukti Terindex | 60 |
| Lampiran 5. Bukti Terbit..... | 60 |
| Lampiran 6. Bukti Pembayaran..... | 61 |



DAFTAR LAMBANG DAN SINGKATAN

| | |
|--------|--|
| EDA | <i>Easy Data Augmentation</i> |
| SVM | <i>Support Vector Machines</i> |
| API | <i>Application Programming Interface</i> |
| TF-IDF | <i>Term Frequency-Inverse Document Frequency</i> |
| NLP | <i>Natural Language Processing</i> |
| BBM | Bahan Bakar Minyak |
| PPN | Pajak Pertambahan Nilai |
| CSV | <i>Comma-Separated Values</i> |
| UTF | <i>Unicode Transformation Format</i> |

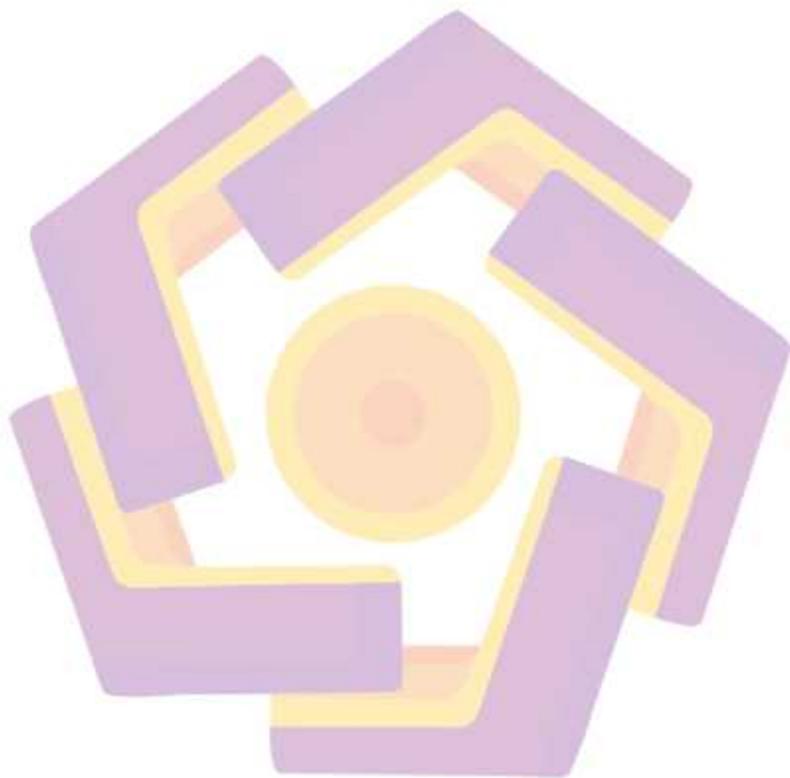


DAFTAR ISTILAH



| | |
|----------------------------|--|
| <i>Soft Voting</i> | teknik untuk menggabungkan prediksi dari beberapa model klasifikasi |
| <i>Machine Learning</i> | cabang kecerdasan buatan yang membuat sistem belajar dari data |
| <i>Ensemble Learning</i> | metode <i>machine learning</i> yang menggabungkan beberapa model |
| <i>Accuracy</i> | metrik evaluasi yang mengukur seberapa benar prediksi model secara keseluruhan |
| <i>Precision</i> | metrik evaluasi yang mengukur seberapa baik model memprediksi benar pada kelas positif |
| <i>Recall</i> | metrik evaluasi untuk menggambarkan seberapa baik model dalam mengidentifikasi kelas positif dengan benar |
| <i>F1-Score</i> | metrik evaluasi yang menggabungkan precision dan recall |
| <i>Stemming</i> | teknik prapemrosesan teks dalam Pemrosesan Bahasa Alami |
| <i>Stopword</i> | kata yang diabaikan dalam pemrosesan |
| <i>Gradient Boosting</i> | algoritma yang membangun model secara bertahap dengan memperbaiki kesalahan prediksi model sebelumnya |
| <i>Logistic Regression</i> | algoritma klasifikasi yang menggunakan fungsi logistik untuk memprediksi probabilitas suatu data masuk ke dalam kelas |
| <i>Naïve Bayes</i> | algoritma klasifikasi berbasis teorema Bayes untuk menghitung probabilitas suatu data |
| <i>Random Forest</i> | metode pembelajaran mesin ensemble yang membangun banyak pohon keputusan untuk mendapatkan prediksi yang lebih akurat dan stabil |

| | |
|------------------------------|---|
| <i>Word Embedding</i> | representasi kata dalam bentuk vector numerik berdimensi rendah |
| <i>Hyperparameter Tuning</i> | proses mencari kombinasi terbaik untuk memaksimalkan performa dan akurasi model |



INTISARI

Analisis sentimen terhadap kebijakan publik semakin penting di era digital karena media sosial telah menjadi ruang utama diskusi masyarakat. Tantangan utama dalam penelitian ini adalah sulitnya menganalisis ribuan komentar YouTube secara manual untuk memahami persepsi publik terhadap kebijakan ekonomi, ditambah dengan kompleksitas bahasa Indonesia informal yang banyak menggunakan singkatan, *slang*, dan variasi penulisan. Pendekatan berbasis satu algoritma juga terbukti kurang mampu menangkap keragaman data teks, sehingga akurasi prediksi rendah dan hasil analisis kurang reliabel untuk mendukung pengambilan keputusan berbasis data empiris. Penelitian ini mengusulkan pendekatan *ensemble learning* dengan memanfaatkan 1.029 komentar YouTube terkait kebijakan ekonomi Tom Lembong, yang kemudian diperluas menjadi 2.169 komentar melalui teknik augmentasi data (*synonym replacement*, *random insertion*, *random deletion*). Tahap *preprocessing* meliputi normalisasi, penghapusan *stopword*, dan *stemming* yang disesuaikan dengan karakteristik bahasa Indonesia. Ekstraksi fitur dilakukan menggunakan *TF-IDF vectorization* (unigram, bigram, trigram) serta *CountVectorizer*, dengan pembagian data latih dan uji sebesar 80:20. Selanjutnya, beberapa algoritma *machine learning* (SVM, Logistic Regression, Random Forest, Gradient Boosting, dan Naïve Bayes) dievaluasi dengan *hyperparameter tuning*, kemudian dikembangkan model *ensemble* berbasis *soft voting* untuk meningkatkan performa klasifikasi. Hasil penelitian menunjukkan bahwa SVM dengan *TF-IDF bigram* memperoleh kinerja terbaik dengan akurasi 96,08% dan F1-score 96,03%. Model *ensemble* mencapai akurasi 95,16% dengan stabilitas prediksi yang lebih konsisten. Analisis per kelas sentimen juga menunjukkan hasil seimbang, dengan F1-score 0,97 untuk sentimen negatif, 0,95 untuk positif, dan 0,90 untuk netral. Penelitian ini berkontribusi pada pengembangan teknik NLP untuk bahasa Indonesia, khususnya dalam pengolahan teks informal media sosial, serta dapat dimanfaatkan oleh pembuat kebijakan, peneliti, maupun praktisi data science. Ke depan, penelitian dapat diarahkan pada pengembangan sistem analisis sentimen *real-time* dan integrasi dengan berbagai platform media sosial lainnya.

Kata kunci: analisis sentimen, ensemble learning, machine learning, YouTube, kebijakan ekonomi.

ABSTRACT

Sentiment analysis of public policy is increasingly important in the digital age because social media has become the main space for public discussion. The main challenge in this research is the difficulty of manually analyzing thousands of YouTube comments to understand public perceptions of economic policy, compounded by the complexity of informal Indonesian language, which makes extensive use of abbreviations, slang, and spelling variations. A single algorithm-based approach has also proven incapable of capturing the diversity of text data, resulting in low prediction accuracy and unreliable analysis results to support empirical data-based decision making. This study proposes an ensemble learning approach using 1,029 YouTube comments related to Tom Lembong's economic policies, which were then expanded to 2,169 comments through data augmentation techniques (synonym replacement, random insertion, random deletion). The preprocessing stage included normalization, stopword removal, and stemming tailored to the characteristics of the Indonesian language. Feature extraction was performed using TF-IDF vectorization (unigram, bigram, trigram) and CountVectorizer, with a training and testing data split of 80:20. Next, several machine learning algorithms (SVM, Logistic Regression, Random Forest, Gradient Boosting, and Naïve Bayes) were evaluated with hyperparameter tuning, then an ensemble model based on soft voting was developed to improve classification performance. The results showed that SVM with TF-IDF bigram achieved the best performance with an accuracy of 96.08% and an F1-score of 96.03%. The ensemble model achieved an accuracy of 95.16% with more consistent prediction stability. Analysis per sentiment class also showed balanced results, with an F1-score of 0.97 for negative sentiment, 0.95 for positive, and 0.90 for neutral. This research contributes to the development of NLP techniques for the Indonesian language, particularly in the processing of informal social media text, and can be utilized by policymakers, researchers, and data science practitioners. In the future, research can be directed towards the development of real-time sentiment analysis systems and integration with various other social media platforms.

Keyword: *sentiment analysis, ensemble learning, machine learning, YouTube, economic policy.*