ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION

SKRIPSI

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

disusun oleh

NOVANDA RENALDY SAPUTRA

21.11.3928

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika

disusun oleh

NOVANDA RENALDY SAPUTRA

21.11.3928

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION

yang disusun dan diajukan oleh

NOVANDA RENALDY SAPUTRA

21.1.3928

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 03 Juni 2025

Dosen Pembimbing,

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

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION

yang disusun dan diajukan oleh

NOVANDA RENALDY SAPUTRA

21.1.3928

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Juni 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Anggit Dwi Hartanto, S.Kom., M.Kom. **NIK. 190302163**	
Rumini, S.Kom., M.Kom. **NIK. 190302246**	
Ferian Fauzi Abdulloh, S.Kom., M.Kom. **NIK. 190302276**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Novanda Renaldy Saputra
NIM : 21.11.3928

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION

Dosen Pembimbing : Ferian Fauzi Abdulloh, S.Kom., 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, 18 Juni 2025

Yang Menyatakan,

Novanda Renaldy Saputra

HALAMAN PERSEMBAHAN

Alhamdulillah, segala puji dan syukur penulis panjatkan kepada Allah SWT atas rahmat, hidayah, dan kemudahan yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Karya ini merupakan bentuk rasa syukur, penghargaan, dan terima kasih yang mendalam kepada semua pihak yang telah memberikan dukungan, doa, dan semangat selama proses ini berlangsung.

Dengan penuh ketulusan, penulis mempersembahkan skripsi ini kepada:

1. Allah SWT, atas limpahan rahmat dan kasih sayang-Nya yang selalu menyertai setiap langkah, memberikan kekuatan dan keteguhan hati dalam menghadapi tantangan selama menyusun skripsi ini.
2. Ayah dan Ibu tercinta, Achmad Baedowi dan Sanny Puspita Wardhani, yang selalu menjadi sumber kekuatan, semangat, dan inspirasi. Terima kasih atas doa yang tiada henti, cinta yang tak terbatas, dan kepercayaan yang diberikan kepada penulis hingga mampu menyelesaikan studi ini. Segala pencapaian ini adalah berkat ridho dan pengorbanan kalian.
3. Dosen pembimbing saya, Bapak Ferian Fauzi Abdulloh, S.Kom., M.Kom., atas bimbingan, kesabaran, dan ilmu yang sangat berharga dalam proses penyusunan skripsi ini. Terima kasih telah membimbing dengan penuh perhatian dan ketulusan.
4. Teman-teman seperjuangan di Yogyakarta, terutama Febri, Ando, Adnan, Dika, Hasan, dan Fakhri, yang telah menjadi bagian penting dalam perjalanan ini. Terima kasih atas canda, semangat, serta kebersamaan yang membuat hari-hari penuh tekanan menjadi lebih ringan dan bermakna.
5. Teman-teman tercinta di Bumiayu, seperti Ii, Ndon, Aka, Ahan, dan Ndut. Terima kasih atas persahabatan yang tulus, dukungan yang hangat, serta semangat yang tak pernah putus meski jarak memisahkan. Kalian adalah rumah yang selalu menenangkan untuk kembali. Diri saya sendiri, atas keteguhan untuk terus melangkah meski dalam keadaan sulit, atas keberanian untuk menghadapi rasa ragu, dan atas kesediaan untuk terus belajar dan tumbuh. Terima kasih telah bertahan dan tidak menyerah.

Semoga skripsi ini menjadi langkah kecil yang berarti dalam perjalanan panjang kehidupan dan ilmu pengetahuan.

KATA PENGANTAR

Alhamdulillah, puji syukur ke hadirat Allah SWT atas rahmat dan karunia-Nya, yang telah memberikan kekuatan, kesehatan, serta kemudahan dalam menyelesaikan skripsi ini yang berjudul: "ANALISIS SENTIMEN TERHADAP KOMENTAR YOUTUBE MENGENAI UPAH MINIMUM INDONESIA MENGGUNAKAN PENDEKATAN ALGORITMA SVM DAN LOGISTIC REGRESSION". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Amikom Yogyakarta.

Penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, S.Kom., M.Kom., selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Bapak Ferian Fauzi Abdulloh, S.Kom., M.Kom., selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, serta ilmu yang sangat berharga selama penyusunan skripsi ini.
5. Bapak Achmad Baedowi dan Ibu Sanny Puspita Wardhani, orang tua penulis tercinta, atas segala doa, kasih sayang, dukungan moral maupun material yang tiada henti.

Semoga karya ini dapat memberikan kontribusi dalam dunia ilmu pengetahuan, khususnya dalam bidang analisis sentimen dan penerapan machine learning, serta bermanfaat bagi masyarakat dan dunia pendidikan. Yogyakarta, Juni 2025 Penulis

Yogyakarta, 18 Juni 2025

Novanda Renaldy Saputra

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 xiii
DAFTAR LAMBANG DAN SINGKATAN xiv
DAFTAR ISTILAH xv
INTISARI xvi
ABSTRACT xvii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 5
1.6 Sistematika Penulisan 6
BAB II TINJAUAN PUSTAKA 9
2.1 Studi Literatur 9
2.2 Dasar Teori 18
2.2.1 Analisis Sentimen 18
2.2.2 Youtube 18
2.2.3 Text Mining 18

2.2.4 Support Vector Machine 19
2.2.5 Logistic Regression 21
2.2.6 K-fold Cross Validation 22
2.2.7 Split Data 23
2.2.8 Confusion Matrix 23
BAB III METODE PENELITIAN 25
3.1 Dataset 25
3.2 Alur Penelitian 27
3.2.1 Scraping Data 29
3.2.2 Labelling Data 29
3.2.3 Analisis Data Eksplorasi 30
3.2.4 Prepocessing Data 31
3.2.5 Wordcloud 35
3.2.6 Proses Pra-Modeling Data 35
3.2.7 Implementasi Pemodelan 38
3.2.8 Evaluasi Model 41
3.3 Alat dan Bahan 41
3.3.1 Alat Penelitian 41
3.3.2 Bahan Penelitian 42
BAB IV HASIL DAN PEMBAHASAN 43
4.1 Pengambilan Data 43
4.2 Labeling 45
4.3 Preprocessing Data 47
4.3.1 Text Cleaning 47
4.3.2 Case Folding 48
4.3.3 Normalitazion 49
4.3.4 Tokenezing 52
4.3.5 StopWords Removal 53
4.3.6 Stemming 54
4.3.7 Visualisasi Wordcloud 54
4.4 Pre-Modeling 56

4.4.1 Ekstrasi Fitur Menggunakan TF-IDF 56
4.4.2 Penerapan TF-IDF dengan K-Fold Cross Validation 57
4.4.3 Penerapan TF-IDF dengan Split Data 58
4.5 Modeling SVM dan Logistic Regression dengan K-Fold CV 59
4.5.1 Support Vector Machine (SVM) dengan Kernel RBF 60
4.5.2 Support Vector Machine (SVM) dengan Kernel Sigmoid 63
4.5.3 Support Vector Machine (SVM) dengan Kernel Polynomial 66
4.5.4 Perbandingan Hasil dari SVM dengan Berbagai Kernel 68
4.5.5 Logistic Regression pada K-Fold Cross Validation 69
4.6 Modeling SVM dan Logistic Regression dengan Split Data 72
4.6.1 Support Vector Machine (SVM) dengan Kernel RBF 72
4.6.2 Support Vector Machine (SVM) dengan Kernel Sigmoid 74
4.6.3 Support Vector Machine (SVM) dengan Kernel Polynomial 76
4.6.4 Perbandingan Hasil SVM pada Split Data 80:20 78
4.6.5 Logitic Regression pada Data Train 80% dan Data Test 20% 79
4.7 Perbandingan Hasil Model Klasifikasi 81
4.7.1 Evaluasi Hasil Model dengan K-Fold Cross Validation 82
4.7.2 Evaluasi Hasil Model dengan Split Data 80:20 85
BAB V PENUTUP 88
5.1 Kesimpulan 88
5.2 Saran 89
REFERENSI 90
LAMPIRAN 96

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian 13
Tabel 2. 2 Confusion Matrix 1 24
Tabel 3. 1 Dataset Dari Beberapa Channel Youtube 25
Tabel 3. 2 Labeling Data 30
Tabel 3. 3 Perhitungan Skor SVM Menggunakan TF-IDF 39
Tabel 3. 4 Contoh Dataset 42
Tabel 4. 1 Data Hasil Scraping 44
Tabel 4. 2 Hasil Proses Pelabelan 45
Tabel 4. 3 Jumlah Data yang Digunakan 46
Tabel 4. 4 Sebelum Dilakukan Text Cleaning 48
Tabel 4. 5 Sesudah Dilakukan Text Cleaning 48
Tabel 4. 6 Sesudah Dilakukan Case Folding2 48
Tabel 4. 7 Kata yang Perlu di Normalisasi 50
Tabel 4. 8 Hasil Normalisasi 51
Tabel 4. 9 Hasil Tokenizing 52
Tabel 4. 10 Hasil Stopwords Removal 53
Tabel 4. 11 Hasil Stemmin 54
Tabel 4. 12 Rata-Rata Confusion Matrix Model pada 5-Fold 83
Tabel 4. 13 Perbandingan Akurasi 86

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Hyperplane dan Margin Maksimal (Sumber: neliti.com) ...20
Gambar 2. 2 Konsep Dasar Logistic Regression 22
Gambar 2. 3 Ilustrasi Cross Validation 5 Fold (Sumber: mdpi.com) 23
Gambar 3. 1 Alur Penelitian 28
Gambar 3. 2 Alur Normalisasi 33
Gambar 3. 3 Alur Tokenizing 33
Gambar 3. 4 Alur Stopword Removal 34
Gambar 3. 5 Alur Stemming 35
Gambar 3. 6 Ilustrasi Cross Validation 5 Fold (Sumber: mdpi.com) 36
Gambar 4. 1 Diagram Hasil Labeling 47
Gambar 4. 2 Code Normalilasi 50
Gambar 4. 3 Wordcould Positif 55
Gambar 4. 4 Wordcould Negatif 56
Gambar 4. 5 Memvisualkan hasil TF-IDF 57
Gambar 4. 6 Hasil Pembagian Data Uji dan Latih Menggunakan K-Fold 58
Gambar 4. 7 Split Data Train & Test 59
Gambar 4. 8 Akurasi Tuning Hyperparameter SVM RBF dengan K-Fold 61
Gambar 4. 9 *Confusion Matrix Kernel RBF dengan K-Fold* 63
Gambar 4. 10 Akurasi Tuning SVM Sigmoid dengan K-Fold 64
Gambar 4. 11 Confusion Matrix Kernel Sigmoid dengan K-Fold 65
Gambar 4. 12 Akurasi Tuning SVM Polynomial dengan K-Fold 66
Gambar 4. 13 Confusion Matrix Kernel Polynomial dengan K-Fold 68
Gambar 4. 14 Evaluasi Kinerja Kernel SVM dengan K-Fold 69
Gambar 4. 15 Akurasi Hasil Tuning Logistic Regression dengan K-Fold 70
Gambar 4. 16 Confusion Matrix Model Logistic Regression dengan K-Fold 71
Gambar 4. 17 Akurasi Tuning SVM Sigmoid dengan Split Data 73
Gambar 4. 18 Confusion Matrix SVM Kernel RBF Dengan Split Data 80:20 74
Gambar 4.19 Akurasi Tuning SVM Kernel Sigmoid pada Split Data 75
Gambar 4.20 Confusion Matrix SVM Sigmoid Dengan Split Data 80:20 76

Gambar 4.21 Akurasi Tuning SVM Kernel Polynomial pada Split Data 77
Gambar 4. 22 Confusion Matrix SVM Polynomial Split 80:20 78
Gambar 4.23 Evaluasi Kinerja SVM dengan Split Data.....79
Gambar 4.24 Akurasi Tuning Logistic Regression pada K-Fold 80
Gambar 4.25 Confusion Matrix Logistic Regression Dengan Split Data 80:20....81
Gambar 4. 26 Perbandingan Akurasi Model dengan K-Fold Cross Validation 82
Gambar 4. 27 Perbandingan Akurasi Model dengan K-Fold Cross Validation 84
Gambar 4. 28 Perbandingan Akurasi Train dan Test pada Split Data 80:20.....85

DAFTAR LAMPIRAN

Lampiran 1 Kode Scraping Data.....96
Lampiran 2 Kode Pre-processing Data.....97
Lampiran 3 Kode Pre-Modeling Data.....99
Lampiran 4 Kode Pemodelan Data dengan K-Fold Cross Validation.....101
Lampiran 5 Kode Pemodelan Data dengan Split Data 80:20.....102

DAFTAR LAMBANG DAN SINGKATAN

N	Jumlah dokumen dalam korpus (TF-IDF).
f(t, d)	Frekuensi kemunculan suatu term t dalam dokumen d.
log	Logaritma untuk perhitungan.
SVM	Support Vector Machine.
RBF	Radial Basis Function.
TF-IDF	Term Frequency-Inverse Document Frequency.
API	Application Programming Interface.
CSV	Comma-Separated Values.
EDA	Exploratory Data Analysis.
TP	True Positive.
TN	True Negative.
FP	False Positive.
FN	False Negative.

DAFTAR ISTILAH

Scraping	Teknik mengumpulkan data dari web.
Preprocessing	Pembersihan dan pemrosesan teks sebelum analisis.
SVM	Algoritma klasifikasi berbasis hyperplane.
Logistic Regression	Algoritma klasifikasi biner yang menghitung probabilitas.
K-Fold	Teknik validasi model dengan pembagian beberapa kali.
Split Data	Membagi dataset untuk pelatihan dan pengujian.
Data Train	Data yang digunakan untuk melatih model.
Data Test	Data yang digunakan untuk menguji model.
TF-IDF	Penimbangan bobot kata berdasarkan frekuensi dalam dokumen.
Wordcloud	Visualisasi kata yang sering muncul.
Confusion Matrix	Tabel evaluasi model klasifikasi (TP, FP, TN, FN).
Akurasi	Persentase prediksi benar oleh model.
Precision	Proporsi prediksi positif yang benar.
Recall	Proporsi data positif yang terdeteksi dengan benar.
F1-Score	Rata-rata harmonis dari precision dan recall.
Tokenizing	Memecah teks menjadi kata-kata.
Stopword Removal	Menghapus kata umum yang tidak bermakna.
Stemming	Mengubah kata ke bentuk dasar.
Normalisasi	Mengubah kata tidak baku ke bentuk standar.
Hyperplane	Garis pemisah dalam SVM untuk klasifikasi.
Labeling	Proses memberikan label atau kategori pada data.

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan Upah Minimum Indonesia dengan menggunakan algoritma Support Vector Machine (SVM) dan Logistic Regression pada komentar YouTube. Kebijakan upah minimum merupakan isu strategis yang sering menjadi bahan diskusi publik karena dampaknya terhadap kesejahteraan pekerja dan keberlangsungan usaha. Perkembangan teknologi informasi, khususnya media sosial seperti YouTube, telah memudahkan masyarakat dalam menyampaikan opini secara terbuka. Komentar-komentar yang muncul mencerminkan beragam sentimen publik, sehingga diperlukan metode klasifikasi yang akurat untuk memahami persepsi masyarakat secara menyeluruh.

Metode penelitian meliputi pengumpulan data komentar dari YouTube menggunakan API, preprocessing teks berupa text cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming, serta pembagian data menggunakan K-Fold Cross Validation (5-Fold) dan Split Data 80:20. Pemodelan dilakukan menggunakan algoritma SVM dengan kernel Polynomial, RBF, dan Sigmoid, serta algoritma Logistic Regression. Evaluasi model dilakukan berdasarkan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model SVM dengan kernel RBF memiliki akurasi terbaik, yaitu 80,47% pada K-Fold, sedangkan Logistic Regression unggul tipis pada Split Data dengan akurasi 78,41%.

Hasil penelitian ini menunjukkan bahwa SVM RBF merupakan pilihan unggul untuk klasifikasi sentimen komentar kebijakan ekonomi di YouTube. Penelitian ini dapat dimanfaatkan oleh pengambil kebijakan, peneliti, dan pengembang sistem analisis opini publik berbasis media sosial.

Kata kunci: analisis sentimen, SVM, Logistic Regression, YouTube, upah minimum.

ABSTRACT

This study aims to analyze public sentiment towards Indonesia's minimum wage policy using Support Vector Machine (SVM) and Logistic Regression algorithms on YouTube comments. The minimum wage policy is a strategic issue frequently discussed in public discourse due to its impact on both worker welfare and business sustainability. The rapid development of information technology, especially through social media platforms like YouTube, has made it easier for the public to express opinions openly. The variety of public comments reflects diverse sentiments, thus requiring accurate classification methods to better understand public perception.

The research methods include data collection using the YouTube API, text preprocessing such as text cleaning, case folding, normalization, tokenization, stopword removal, and stemming, as well as data splitting using K-Fold Cross Validation (5-Fold) and an 80:20 Split. The modeling process involves the application of SVM with Polynomial, RBF, and Sigmoid kernels, along with the Logistic Regression algorithm. Model performance is evaluated based on accuracy, precision, recall, and F1-score. The results show that the SVM model with the RBF kernel achieved the best performance with an accuracy of 80.47% using K-Fold validation, while Logistic Regression slightly outperformed on the 80:20 data split with an accuracy of 78.41%.

These findings indicate that SVM with the RBF kernel is a superior choice for sentiment classification of public comments on economic policies via YouTube. This research can benefit policymakers, researchers, and developers of social media-based opinion analysis systems.

Keywords: *sentiment analysis, SVM, Logistic Regression, YouTube, minimum wage.*