

---

## Analisis Sentimen Pergantian Kementerian Keuangan Indonesia dengan Menggunakan Algoritma Support Vector Machine

Haikal Nurul Barki<sup>1</sup>, Haidar Aslam<sup>2</sup>,

Adhi Prasetyo Wibowo<sup>3</sup>, Faqih Al Araf<sup>4</sup>, Abdul Hamid Musawir<sup>5</sup>

<sup>1</sup>Ilmu Komputer, Universitas Bina Sarana Informatika, [haikalbarki0@gmail.com](mailto:haikalbarki0@gmail.com)

<sup>2</sup>Ilmu Komputer, Universitas Bina Sarana Informatika, [aslamhaidar98@gmail.com](mailto:aslamhaidar98@gmail.com)

<sup>3</sup>Ilmu Komputer, Universitas Bina Sarana Informatika, [adhiprasetyow07@gmail.com](mailto:adhiprasetyow07@gmail.com)

<sup>4</sup>Ilmu Komputer, Universitas Bina Sarana Informatika, [faqihalaraf51@gmail.com](mailto:faqihalaraf51@gmail.com)

<sup>5</sup>Ilmu Komputer, Universitas Bina Sarana Informatika, [hamid03abd@gmail.com](mailto:hamid03abd@gmail.com)

---

### Keywords :

Sentiment Analysis,  
Classification,  
Multiclass SVM,  
Confusion Matrix,  
Ministry of Finance,

### ABSTRACT

The leadership transition within the Ministry of Finance of the Republic of Indonesia is a strategic event that has triggered various public responses. These reactions reflect the level of public trust in the government's fiscal and economic policy directions. This study aims to analyze public sentiment regarding the leadership change using the *Support Vector Machine* (SVM) algorithm as the main classification method. The data were collected from public comments on YouTube videos discussing the Minister of Finance transition. The research stages included data collection, text preprocessing, sentiment labeling, feature extraction using the *Term Frequency-Inverse Document Frequency* (TF-IDF) method, and sentiment classification using SVM. Model evaluation was carried out through accuracy, precision, and recall metrics. The results show that the SVM algorithm achieved optimal classification performance with an accuracy of approximately 97%. The majority of the public expressed positive sentiment toward the leadership transition, while a smaller portion conveyed negative responses related to uncertainty about future economic policy directions.

---

### Kata Kunci :

Analisis Sentimen,  
Klasifikasi,  
Multiclass SVM,  
Confusion Matrix,  
Kementerian Keuangan,

### ABSTRAK

Pergantian kepemimpinan di Kementerian Keuangan Republik Indonesia merupakan peristiwa strategis yang berpotensi memengaruhi arah kebijakan fiskal dan stabilitas ekonomi nasional serta memicu beragam respons dari publik. Permasalahan dalam penelitian ini adalah belum terpetakannya secara objektif kecenderungan sentimen masyarakat terhadap pergantian kepemimpinan tersebut, mengingat respons publik yang tersebar di media daring bersifat tidak terstruktur dan sulit dianalisis secara manual. Penelitian ini punya tujuan untuk menganalisis sentimen masyarakat soal pergantian tersebut. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi utama. Data penelitian ini diambil dari komentar publik di YouTube yang membahas isu pergantian Menteri Keuangan. Tahapan penelitian mencakup pengumpulan data. Lalu ada pembersihan teks melalui textbook preprocessing. Selanjutnya pelabelan sentimen. Kemudian ekstraksi fitur pakai *Term Frequency - Inverse Document Frequency* (TF-IDF). Akhirnya klasifikasi dengan SVM. Evaluasi model penelitian ini dilakukan lewat metrik akurasi, presisi, dan recall. Hasilnya menunjukkan SVM punya performa klasifikasi yang optimal. Akurasinya mencapai (97%). Mayoritas masyarakat tampilkan sentimen positif terhadap pergantian kepemimpinan ini. Sementara sebagian kecil beri respons negatif. Respons negatif itu terkait ketidakpastian arah kebijakan ekonomi.

---

### Korespondensi Penulis:

Haikal Nurul Barki,  
Universitas Bina Sarana Informatika,  
Jl. Margonda Raya No.8, Kota Depok  
Telepon : +6285892644370  
Email: [haikalbarki0@gmail.com](mailto:haikalbarki0@gmail.com)

Submitted: 13-11-2025; Accepted: 20-12-2025;  
Published: 26-01-2026

*Copyright* (c) 2026 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

---

## 1. PENDAHULUAN

Pergantian kepemimpinan di Kementerian Keuangan Republik Indonesia merupakan peristiwa strategis yang berpotensi memengaruhi arah kebijakan fiskal, persepsi publik, serta stabilitas ekonomi nasional. Dalam era digital, respons masyarakat terhadap peristiwa kebijakan publik semakin banyak terekam di platform daring, salah satunya

YouTube. Komentar pengguna pada platform tersebut menjadi sumber data yang merepresentasikan opini dan sentimen masyarakat secara luas, namun bersifat tidak terstruktur dan sulit dianalisis secara manual [1].

Pengukuran sentimen masyarakat diperlukan untuk memperoleh gambaran yang objektif dan terukur mengenai kecenderungan persepsi publik terhadap pergantian kepemimpinan tersebut. Tanpa adanya analisis sentimen yang sistematis, penilaian terhadap respons publik berpotensi bersifat subjektif dan tidak representatif. Oleh karena itu, penelitian ini menerapkan algoritma *Support Vector Machine (SVM)* untuk melakukan klasifikasi sentimen masyarakat berdasarkan komentar publik di *YouTube*. *Support Vector Machine* dipilih karena kemampuannya dalam menangani klasifikasi teks berdimensi tinggi secara efektif dan akurat, sehingga diharapkan mampu memberikan hasil analisis sentimen yang reliabel.

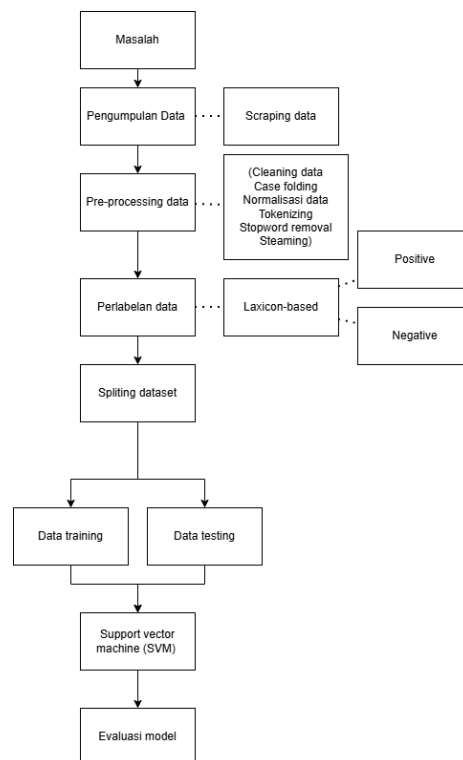
*Analisis sentimen (sentiment analysis)* adalah teknik pemrosesan teks yang digunakan untuk menentukan polaritas opini (*positif/negatif*) dalam *korpus teks*. Untuk komentar singkat berbahasa Indonesia—yang kerap mengandung bahasa tidak baku, singkatan, dan *slang*—*pra-pemrosesan* khusus (*normalisasi slang, stopword removal, stemming Bahasa Indonesia*) serta ekstraksi fitur seperti *TF-IDF* terbukti krusial untuk meningkatkan kualitas representasi fitur sebelum pelatihan model klasifikasi [2], [3].

Di antara metode *supervised learning* untuk klasifikasi teks, *Support Vector Machine (SVM)* tetap menjadi pilihan sebagai *baseline* karena kemampuannya menangani data berdimensi tinggi dan memberikan generalisasi yang baik pada masalah klasifikasi biner. Studi-studi terkini yang mengaplikasikan *SVM* pada komentar *YouTube*, ulasan aplikasi, dan data media Indonesia melaporkan performa kompetitif *SVM* dipadukan dengan *pipeline pra-pemrosesan* dan *tuning hyperparameter* (mis. *GridSearchCV*) [4], [5]. Oleh karena itu, *SVM* dipilih sebagai metode utama dalam penelitian ini.

Kontribusi utama penelitian ini adalah: (1) penyusunan *dataset* komentar *YouTube* yang terfokus pada isu pergantian Kementerian Keuangan (dengan prosedur etis dan anonimisasi), (2) adaptasi *pipeline pra-pemrosesan* yang mencakup *normalisasi slang Bahasa Indonesia* dan penggunaan *stemmer* lokal, dan (3) evaluasi empiris *SVM* (dengan *GridSearchCV* dan *cross-validation*) untuk mengukur kecenderungan sentimen publik terhadap pergantian kepemimpinan. Hasil diharapkan memberikan wawasan empiris terkait persepsi masyarakat serta rekomendasi teknis untuk pemantauan opini publik berbasis teks di konteks kebijakan.

## 2. METODE PENELITIAN

*Sistem klasifikasi* yang dikembangkan pada penelitian ini dilaksanakan melalui beberapa tahapan prosedural agar proses penelitian berjalan secara sistematis, sebagaimana ditunjukkan pada *Gambar 1*.



Gambar 1. Diagram Alir Proses Analisis Sentimen

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *analisis sentimen* berbasis *machine learning*, khususnya algoritma *Support Vector Machine (SVM)*. Proses penelitian mencakup tahapan pengumpulan data

komentar publik dari platform *YouTube*, pembersihan data, ekstraksi fitur teks menggunakan *TF-IDF* (*Term Frequency–Inverse Document Frequency*), pelatihan model *SVM*, serta evaluasi performa model. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap pergantian Kementerian Keuangan Indonesia menjadi kategori *positif* dan *negatif*.

## 2.1 Pengambilan Data

Proses pengambilan data merupakan tahap awal yang krusial dalam penelitian *analisis sentimen*, karena kualitas dan representativitas data sangat memengaruhi hasil klasifikasi yang diperoleh. Pada penelitian ini, data dikumpulkan dari platform *YouTube*, khususnya pada kolom komentar dari beberapa video yang membahas mengenai pergantian kepemimpinan di Kementerian Keuangan Republik Indonesia. Pemilihan *YouTube* sebagai sumber data didasarkan pada tingginya tingkat interaksi publik dan keterbukaan pendapat masyarakat terhadap isu-isu pemerintahan di platform tersebut.

Pengumpulan data dilakukan menggunakan *YouTube Data API v3*, yang memungkinkan peneliti untuk mengekstraksi komentar secara otomatis dari sejumlah video yang relevan berdasarkan kata kunci seperti “Pergantian Menteri Keuangan Indonesia”, “Sri Mulyani digantikan”, dan “Kementerian Keuangan 2025”. Data yang diperoleh kemudian disimpan dalam format *CSV* (*Comma-Separated Values*) agar mudah diproses pada tahap *pra-pemrosesan teks*. Sebagai batasan, penelitian ini mengambil sekitar 7.000 komentar dari 25 video teratas yang paling relevan dan memiliki tingkat keterlibatan tinggi (jumlah *views*, *likes*, dan *comments* yang signifikan).

## 2.2 Pelabelan Sentimen

*Pelebelan sentimen* dilakukan secara terstruktur untuk menghasilkan *data latih* (*ground truth*) yang andal bagi model *SVM*. Skema pelabelan yang digunakan bersifat biner: *Positif* (dukungan/penilaian positif terhadap pergantian) dan *Negatif* (kritik/penilaian negatif terhadap pergantian). Pemilihan skema biner dipertimbangkan agar lebih stabil pada komentar pendek dan informal seperti komentar *YouTube* [1]. Pada penelitian ini, pelabelan data dilakukan dengan memberikan *label* pada setiap *comments* pengguna *YouTube* secara otomatis oleh *lexicon* berbahasa Indonesia untuk menentukan *comments* termasuk ke dalam sentimen *positif* dan *negatif*.

## 2.3 Pembentukan Data Training dan Data Testing

Proses pembagian data pelatihan dan pengujian dilakukan menggunakan *Google Colab* dengan memanfaatkan *train\_test\_split* dan *random\_split*. Kedua fungsi tersebut digunakan untuk membagi *dataset* menjadi dua bagian sesuai proporsi yang telah ditentukan serta memastikan proses pengacakan dilakukan secara konsisten agar hasil tetap replikasi saat dijalankan. Dalam penelitian ini, digunakan 739 data komentar yang kemudian dibagi menjadi 80% data pelatihan (591 komentar) dan 20% data pengujian (148 komentar). Seluruh hasil pemisahan data disimpan dalam format *Comma-Separated Values* (*CSV*) untuk memudahkan proses pemanggilan dan pengolahan data pada tahap selanjutnya.

## 2.4 Preprocessing

Pada penelitian ini, tahapan *pra-pemrosesan* mencakup: (1) *case folding* yang mengonversi seluruh teks menjadi huruf kecil; (2) *cleansing* untuk menghapus *URL*, *mention*, *hashtag*, angka, serta karakter khusus; (3) *tokenization* untuk memecah teks menjadi unit kata; (4) *stop-word removal* menggunakan kamus Bahasa Indonesia yang telah diperluas dengan daftar *slang word*; dan (5) *stemming* menggunakan metode Bahasa Indonesia, misalnya pustaka *Sastrawi*, untuk mengonversi kata ke bentuk dasar. Metode serupa ditunjukkan efektif dalam penelitian “*Text-Preprocessing Model YouTube Comments in Indonesian*” yang menyatakan bahwa skema *pra-pemrosesan* termasuk *normalisasi slang*, *stopword removal*, dan *stemming* memberikan peningkatan akurasi dalam klasifikasi *sentimen* komentar *YouTube* [6]. Dengan demikian, tahap *preprocessing* menjadi fondasi penting sebelum data dimasukkan ke dalam algoritma *SVM* untuk klasifikasi pergantian Kementerian Keuangan Indonesia pada studi ini. Data yang sudah diproses dengan baik akan menghasilkan model yang mampu mengidentifikasi komentar *positif* maupun *negatif* dengan lebih akurat.

## 2.5 Pembobotan Term Frequency–Inverse Document Frequency (TF-IDF)

Setelah teks dibersihkan dan dinormalisasi, fitur teks dikonversi ke representasi numerik menggunakan *TF-IDF* (*Term Frequency–Inverse Document Frequency*). *TF-IDF* menurunkan bobot kata yang umum muncul di seluruh *korpus* dan meningkatkan bobot kata yang lebih informatif untuk dokumen tertentu. Praktik yang dianjurkan meliputi: *unigram* ± *bigram* (untuk menangkap frasa penting), pengaturan *min\_df* untuk menghapus kata yang sangat jarang, dan *max\_features* untuk mengontrol dimensi *vector*; pendekatan ini umum dan efektif pada *dataset* komentar *YouTube* berbahasa Indonesia [4], [6]. Studi komparatif pada *dataset* Indonesia menunjukkan kombinasi *TF-IDF* + *SVM* sering memberikan hasil kompetitif dibanding metode sederhana *bag-of-words* [4].

$$w_{td} = t_{fd} \times idf_t \quad (1)$$

$$idf_t = \log \left( \frac{n}{df_t} \right) \quad (2)$$

Disini,  $wd_t$  = bobot dari kata  $t$  dalam setiap dokumen  $d$ ,  $tf_{td}$  = banyaknya kemunculan kata  $t$  yang dicari dalam dokumen  $d$ ,  $idf_t$  = bobot *invers df*,  $t$  = kata  $t$  dari kata kunci,  $d$  = dokumen  $d$ ,  $D$  = banyaknya semua dokumen,  $df_t$  = banyaknya dokumen yang mengandung kata  $t$ .

### 2.6 Support Vector Machine (SVM)

*Support Vector Machine (SVM)* adalah algoritma *supervised learning* untuk klasifikasi yang beroperasi dengan menemukan *hyperplane* pemisah optimal antara kelas dalam ruang fitur berdimensi tinggi. *SVM* efektif untuk teks karena representasi teks (mis. *TF-IDF*) biasanya menghasilkan *vektor* berdimensi besar; penggunaan *regularisasi* dan teknik *kernel* memungkinkan *SVM* memberikan generalisasi yang baik serta ketahanan terhadap *overfitting* jika di-*tuning* dengan benar [2], [7]. Pada praktik *analisis sentimen*, dua varian yang umum dipakai adalah *LinearSVC* (efisien untuk data *sparse/TF-IDF* besar) dan *SVC* dengan *kernel non-linear (RBF / polynomial)* bila pemisahan linier tidak memadai; pemilihan dan *tuning kernel* (parameter  $C$ ,  $\gamma$ ) biasanya dilakukan melalui *GridSearchCV* atau *RandomizedSearch* dengan *stratified k-fold CV* [7], [8].

Referensi praktis menunjukkan *SVM* tetap menjadi *baseline* yang kuat pada banyak studi *analisis sentimen* komentar *YouTube* dan *review* berbahasa Indonesia, terutama bila dikombinasikan dengan *pra-pemrosesan* yang teliti serta pembobotan fitur yang efektif (*TF-IDF/unigram-bigram*) [4], [9].

### 2.7 Evaluasi

Pengujian dan evaluasi dilakukan untuk mengetahui kinerja model *klasifikasi* yang telah dibuat sehingga dapat ditentukan pengembangan model selanjutnya. Pada penelitian ini, metode evaluasi yang digunakan adalah *confusion matrix* seperti *Tabel 1*, di mana cara kerjanya yaitu melakukan perbandingan hasil *klasifikasi* dari sistem dengan hasil *klasifikasi* yang sebenarnya. Setiap baris pada *Tabel 1* merepresentasikan kelas yang sebenarnya dan setiap kolom merepresentasikan kelas dari hasil *prediksi* atau *klasifikasi*, yang mana *TP* adalah jumlah data *positif* yang terklasifikasi dengan benar, *TN* adalah jumlah data *negatif* yang terklasifikasi dengan benar, *TNt* adalah jumlah data *netral* yang terklasifikasi dengan benar, *FP* adalah jumlah data *positif* namun terklasifikasi salah oleh sistem, *FN* adalah jumlah data *negatif* namun terklasifikasi salah oleh sistem, dan *FNt* adalah jumlah data *netral* namun terklasifikasi salah oleh sistem.

Tabel 1. *Confusion Matrix*

Aktual	Klasifikasi	
	Positif	Negatif
Positif	True Positif (TP)	False Negatif (FN)
Negatif	False Positif (FP)	True Negatif (TN)

Berdasarkan proses *confusion matrix*, maka akan didapatkan nilai dari *precision*, *recall*, *f1-score* untuk ketiga *sentimen* dan keseluruhan *accuracy* yang dihasilkan [10]. *Precision* merupakan rasio *prediksi* atau *klasifikasi* benar *positif* dibandingkan dengan keseluruhan hasil yang diprediksi *positif* [10] yang ditulis dalam *persamaan (3)*.

$$Precision = \frac{TP}{TF} \times 100\% \quad (3)$$

*Recall* merupakan rasio *prediksi* atau *klasifikasi* benar *positif* dibandingkan dengan keseluruhan data yang benar *positif* [10] dalam bentuk *persamaan (4)*.

$$Recall = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

*F1-score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan [10] dalam *persamaan (5)*.

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + precision} \quad (5)$$

*Accuracy* merupakan rasio *prediksi* benar (*positif* dan *negatif*) dengan keseluruhan data [11] dalam *persamaan (6)*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

### 2.8 Visualisasi

Pada penelitian ini, kedua *sentimen* hasil *klasifikasi* yaitu *positif* dan *negatif* divisualisasikan dalam bentuk *word cloud*, *bar chart*, dan *pie chart* dengan menggunakan *library Python* yaitu *Matplotlib* dan *WordCloud* dengan

tujuan agar mudah dibaca dan dipahami oleh masyarakat sehingga informasi yang disampaikan dapat tersampaikan dengan baik.

### 3. HASIL DAN ANALISIS

Berdasarkan rangkaian eksperimen yang telah dilakukan, bagian ini menyajikan evaluasi performa model klasifikasi sentimen komentar *YouTube* terkait Pergantian Kementerian Keuangan Indonesia beserta pembahasannya. Analisis dimulai dari hasil pelabelan berbasis *lexicon* dan kualitas *corpus* dua kelas yang terbentuk, dilanjutkan dengan vektorisasi *TF-IDF*, pelatihan *LinearSVC* dengan penalaan *hyperparameter*, hingga pengujian pada *hold-out test set*. Interpretasi metrik (*accuracy, precision, recall, F1-score*, dan *confusion matrix*) disertai visualisasi (*label distribution, confusion matrix, most influential features*, dan *word cloud*) untuk mengidentifikasi kekuatan serta keterbatasan model dalam mengklasifikasikan sentimen secara akurat.

#### 3.1 Pengumpulan Data

Proses pengambilan data merupakan tahap awal yang krusial dalam penelitian analisis sentimen, karena kualitas dan representativitas data sangat memengaruhi hasil klasifikasi yang diperoleh. Pada penelitian ini, data dikumpulkan dari platform *YouTube*, khususnya pada kolom komentar dari beberapa video yang membahas mengenai pergantian kepemimpinan di Kementerian Keuangan Republik Indonesia. Pemilihan *YouTube* sebagai sumber data didasarkan pada tingginya tingkat interaksi dan keterbukaan pendapat masyarakat terhadap isu-isu pemerintahan di platform tersebut.

Pengumpulan data dilakukan menggunakan *YouTube Data API v3*, yang memungkinkan peneliti untuk mengekstraksi komentar secara otomatis dari sejumlah video yang relevan berdasarkan kata kunci seperti “Pergantian Menteri Keuangan Indonesia”, “Sri Mulyani digantikan”, dan “Kementerian Keuangan 2025”. Data yang diperoleh kemudian disimpan dalam format *CSV (Comma-Separated Values)* agar mudah diproses pada tahap *pre-processing* teks. Sebagai data, penelitian ini mengambil sekitar 7.000 komentar dari 25 video teratas yang paling relevan dan memiliki tingkat keterlibatan tinggi (jumlah *views, likes*, dan *comments* yang signifikan).

video_id	video_title	comment_id	author	text	publishedat	likecount	is_reply	parent_id	
0	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgwoEAx9rb-mdIPPLE4AaABAg	@toprizal6019	Semogah Indonesia sejah tera	2025-10-23T23:16:18Z	0	False	NaN
1	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	Ugy6zhnMe5TFJVTdw4AaABAg	@srihandayani5599-q9x	Ibu pajak diganti	2025-09-24T07:57:22Z	0	False	NaN
2	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgwJIS1afZL0jEB2uFh4AaABAg	@NDOKORTECH	Di lokal lawan loe DPR, konglomerat dan mafia2...	2025-09-20T03:49:55Z	0	False	NaN
3	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	Ugy97nKUXyIXDfMPTx4AaABAg	@TuripnoJayaraya	Semangat pa pur, the best ♥	2025-09-14T12:48:21Z	0	False	NaN
4	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgzcgwcbEH9A9pNrcj94AaABAg	@ericakbar7283	Pajak ke rakyat jangan kancang" boss.Vltu dar...	2025-09-11T11:30:33Z	0	False	NaN
5	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgwuwNXat1IXNazRcpF4AaABAg	@mansyurtono1052	Bu Sri Mulyani pergi bawah uang haram di masa ...	2025-09-11T09:44:30Z	0	False	NaN
6	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgyZcyfCCShFINKex8x4AaABAg	@AgusTina-vj2qk	Semoga ppkenalkan gaji Asn danPensiun cepat di...	2025-09-11T06:50:25Z	0	False	NaN
7	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgyjsT7_aGyF1h1x2F4AaABAg	@ImronHapiz-18h	Pensiunan menteri keuangan adalah beban negara	2025-09-10T16:20:06Z	0	False	NaN
8	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	UgzPcp7g3sc3VCV1VF4AaABAg	@paksuparno4700	Aku kok ragu hal kapasitas Purbaya. Baru omong...	2025-09-10T07:57:10Z	0	False	NaN
9	PSw89vFqQVA	Profil Purbaya Yudhi Sadewa, Menteri Keuangan ...	Ugyxkx0z9MsSrxJac4AaABAg	@Alchemiste	Ini pelajaran buat probowol Kalo mau milih men...	2025-09-10T03:34:56Z	0	False	NaN

Gambar 2. Hasil Pengumpulan Data

#### 3.2 Pelebelan sentiment dengan lexicon based

Komentar yang diperoleh dari *YouTube* diberi label awal menggunakan kamus sentimen positif dan negatif yang telah diperluas untuk konteks Pergantian Kementerian Keuangan (mis. *baik, rakyat, negara, parah, ganti, mundur, korupsi*). Kamus di-*stem* agar selaras dengan hasil *pre-processing*, sedangkan *scoring* mempertimbangkan *negator (tidak, bukan, nggak), intensifier/downtoner*, dan pemisahan klausa kontras (*tapi/namun*). Hasil awal menghasilkan dua label. Sesuai tujuan klasifikasi biner, seluruh entri netral dihapus sehingga terbentuk korpus dua kelas. Distribusi akhir setelah pelabelan ditunjukkan pada *Tabel 2* dan *output* ditampilkan pada *Gambar 3* (a dan b).

Tabel 2. Distribusi label setelah pelebelan dua kelas

Label	Jumlah	Proporsi
Positif	542	73,7%
Negatif	197	26,3%
Total	739	100%

=== Contoh Komentar POSITIF ===

	text	sentiment
12	Semoga Mentri baru bisa memberikan yang terbai...	positif
22	Jangan puas jabatan pak \nKalau belum bisa mem...	positif
23	Dia banyak hutang mangkaya pajak naik ,,udah b...	positif
43	Sri muliany SDH terlalu pinter,maka mulut nya ...	positif
55	Menko Perekonomian diganti juga dong...jangan ...	positif

(a)

=== Contoh Komentar NEGATIF ===

	text	sentiment
63	Parah ini...\nBlunder lagi, blunder lagi 😞	negatif
161	Parahhh.. Lebih parah	negatif
162	Parahhh.. Lebih parah	negatif
276	Mentri baru ini lebih parah guys hahahaha indo...	negatif
280	Cuma mau pesan kpd para pejabat baru..kalau se...	negatif

(b)

Gambar 3. Output Hasil Pelebelan Dua Kelas

### 3.3 Pembentukan Data Training dan Data Testing

Data komentar yang digunakan pada proses klasifikasi sebanyak 739 data dan dilakukan pembagian data menjadi dua bagian. *Gambar 4* menunjukkan proses pembagian data menjadi dua bagian yaitu data *training* sebesar 80% (591 data komentar) dan data *testing* 20% (148 data komentar) dengan memanfaatkan *library train\_test\_split* serta menggunakan *random\_state* untuk mengacak dan menjaga data agar tetap sama ketika *system* dijalankan kembali. Nilai *random\_state* yang digunakan adalah nilai tertinggi dari rentang nilai *random\_state*, yaitu 42.

```
# Cell 11: Split dataset
x = df['clean_text']
y = df['sentiment']

X_train, X_test, y_train, y_test = train_test_split(
    x, y,
    test_size=0.2,
    random_state=42,
    stratify=y # menjaga proporsi label
)

print("Data training:", len(X_train))
print("Data testing :", len(X_test))

Data training: 591
Data testing : 148
```

Gambar 4. Pembentukan Data Training dan Data Testing

### 3.4 Preprocessing

Tahap persiapan data atau *preprocessing* bertujuan untuk memastikan data yang akan digunakan sudah siap dan menghilangkan *noise* yang dapat mengganggu proses klasifikasi [2], [9].

#### 3.4.1 Data Cleaning

Data dibersihkan dari simbol atau bagian-bagian yang tidak diperlukan yang umumnya terkandung di dalam sebuah *dataset* untuk mengurangi *noise* sehingga mempermudah proses klasifikasi. *Tabel 3* menunjukkan hasil proses *data cleaning* di mana terlihat pada hasil *data cleaning* adanya pembersihan *username* maupun simbol-simbol seperti hilangnya *username*, *hashtag*, tanda baca, angka, dan *emoticon*.

Tabel 3. Data Cleaning pada Data Komentar

Komentar	Hasil Data Cleaning
Parah ini...\nBlunder lagi, blunder lagi 😞	Parah ini...\nBlunder lagi, blunder lagi

#### 3.4.2 Case Folding

Proses ini bertujuan menghindari kendala *font* yang tidak beraturan antara *uppercase* dan *lowercase* dengan cara mengubah semua kata pada data komentar menjadi huruf kecil (*lowercase*). Hasil data setelah melalui proses *case folding* ditunjukkan pada *Tabel 4*, di mana pada kolom hasil *case folding* semua kata pada data komentar hasil *data cleaning* telah berubah menjadi huruf kecil.

Tabel 4. Case Folding pada Data Komentar

Hasil Data Cleaning	Hasil Case Folding
Parah ini...\nBlunder lagi, blunder lagi	parah ini...\nblunder lagi, blunder lagi

### 3.4.3 Tokenizing

Proses ini menguraikan sebuah kalimat berdasarkan tiap kata yang menyusun kalimat, seperti ditunjukkan pada Tabel 5. Pada kolom hasil *tokenizing* terlihat bahwa kalimat hasil *case folding* telah diurai menjadi per kata yang ditunjukkan dengan adanya penggunaan tanda baca koma sebagai pemisah antar kata.

Tabel 5. Tokenizing pada Data Komentar

Hasil Case Folding	Hasil Tokenizing
parah ini...\nblunder lagi, blunder lagi	parah ini,blunder lagi, blunder lagi

### 3.4.4 Stemming

Kata-kata pada data komentar *YouTube* yang memiliki imbuhan dihilangkan sehingga hanya menyisakan kata dasar, seperti ditunjukkan pada Tabel 6. Contoh yang dapat dilihat yaitu kata hasil *tokenizing* ‘umatnya’ dengan imbuhan ‘-nya’ berubah menjadi kata ‘umat’ setelah melalui proses *stemming*.

Tabel 6. Stemming pada Data Komentar

Hasil Tokenizing	Hasil Stemming
parah ini,blunder lagi, blunder lagi	parah ini,blunder lagi, blunder lagi

### 3.4.5 Stopword Removal

Menghilangkan kata yang tidak penting dan tidak diperlukan dalam proses klasifikasi sentimen dilakukan dengan memanfaatkan *library NLTK*, dan hasil proses *stopword removal* ditunjukkan pada Tabel 7. Pada Tabel 7 terlihat bahwa terdapat kata-kata pada hasil *stemming* seperti ‘jadi’, ‘naik’, dan ‘hingga’ yang hilang setelah melalui proses *stopword removal*.

Tabel 7. Stopword Removal pada Data Komentar

Hasil Stemming	Hasil Stopword Removal
parah ini,blunder lagi, blunder lagi	Parah blunder blunder

## 3.5 Pembobotan TF-IDF

Gambar 5 menunjukkan proses pembobotan terhadap setiap kata atau *term* pada *dataset* dengan tujuan untuk mengubah kata-kata menjadi sebuah *vektor* yang merepresentasikan nilai dari banyaknya kata pada dokumen, di mana data yang digunakan pada proses pembobotan ini adalah data yang sudah melalui tahap *preprocessing*.

```
# Cell 12: TF-IDF vectorizer
vectorizer = TfidfVectorizer(
    max_features=5000, # batasi jumlah fitur agar tidak terlalu besar
    ngram_range=(1,2), # unigram + bigram untuk menangkap konteks
    sublinear_tf=True
)

x_train_tfidf = vectorizer.fit_transform(X_train)
x_test_tfidf = vectorizer.transform(X_test)

print("Bentuk x_train_tfidf:", x_train_tfidf.shape)
```

Gambar 5. Proses Pembobotan TF-IDF

## 3.6 Klasifikasi Support Vector Machine (SVM)

*Support Vector Machine (SVM)* merupakan algoritma *machine learning* yang termasuk dalam kategori *supervised learning* dan secara luas digunakan dalam berbagai permasalahan klasifikasi, termasuk analisis sentimen berbasis teks. *SVM* bekerja dengan cara mencari *hyperplane* optimal yang dapat memisahkan data dari dua atau lebih kelas dengan *margin* maksimum, sehingga tingkat kesalahan klasifikasi menjadi minimal[7]. Dalam konteks analisis

sentimen, setiap komentar atau teks direpresentasikan dalam bentuk *vektor fitur*, umumnya hasil dari pembobotan *TF-IDF*, yang kemudian menjadi masukan bagi model *SVM* untuk menentukan kelas sentimen seperti positif, negatif, atau netral [9].

Pada penelitian ini, algoritma *SVM* digunakan untuk mengklasifikasikan komentar masyarakat mengenai pergantian Kementerian Keuangan Indonesia yang diambil dari platform *YouTube*. Model dilatih dengan memanfaatkan data yang telah melalui proses *preprocessing* dan pembobotan fitur. Teknik ini dipilih karena *SVM* memiliki keunggulan dalam mengolah data berdimensi tinggi, seperti teks yang telah dikonversi ke ruang *vektor*, serta memiliki kemampuan generalisasi yang baik pada data berukuran kecil hingga menengah [12].

Proses klasifikasi dilakukan dengan menggunakan *kernel function* untuk memproyeksikan data ke ruang berdimensi lebih tinggi sehingga dapat dipisahkan secara linier. Beberapa jenis *kernel* yang umum digunakan adalah *linear*, *polynomial*, dan *radial basis function (RBF)*. Pada penelitian ini, fungsi *kernel linear* digunakan karena hasil studi sebelumnya menunjukkan bahwa *kernel linear* sering memberikan hasil yang optimal dalam klasifikasi teks Bahasa Indonesia. Parameter utama yang diatur dalam model ini adalah *C (cost)* yang mengontrol tingkat penalti terhadap kesalahan klasifikasi, dan *gamma* untuk menentukan pengaruh jarak antar titik data terhadap batas keputusan model.

Sebelum dilakukan pelatihan, *dataset* dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%, dengan proses *randomization* agar distribusi data tetap acak dan representatif. Model kemudian dilatih menggunakan *LinearSVC* dari pustaka *scikit-learn* di *Google Colab*, yang telah banyak digunakan dalam penelitian analisis sentimen karena efisiensi komputasinya. Setelah proses pelatihan, model diuji menggunakan data uji untuk memperoleh tingkat *akurasi*, *presisi*, *recall*, dan *F1-score* sebagai metrik evaluasi utama [4].

### 3.7 Evaluasi

Pengujian atau evaluasi hasil dilakukan untuk mengecek kinerja *multiclass SVM* dalam melakukan proses analisis sentimen dengan menggunakan *confusion matrix* yang mengelompokkan hasil klasifikasi ke dalam dua sentimen, yaitu sentimen positif dan negatif, seperti ditunjukkan pada *Tabel 8*.

Tabel 8. Confusion Matrix

Aktual	Klasifikasi	
	Positif	Negatif
Positif	108	1
Negatif	4	35

Berdasarkan hasil klasifikasi sentimen pada *Tabel 8*, setiap nilai *TP*, *FP*, *TN*, dan *FN* digunakan untuk memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* pada sentimen positif dan negatif menggunakan *persamaan (3-6)*. Hasil pengujian ditunjukkan pada *Gambar 6*.

```

Classification Report:
              precision    recall  f1-score   support

   positif      0.97      0.90      0.93        39
   negatif      0.96      0.99      0.98       109

 accuracy              0.97        148
 macro avg      0.97      0.94      0.96        148
 weighted avg   0.97      0.97      0.97        148
    
```

Gambar 6. Hasil Pengujian *Confusion Matrix*

Hasil evaluasi atau uji hasil analisis sentimen pada *Gambar 5* menggunakan perhitungan *confusion matrix* untuk 236 data *testing* menunjukkan bahwa *akurasi* keseluruhan yang dicapai pada proses klasifikasi adalah 97%, dengan *precision* 97% untuk kelas positif dan 96% untuk kelas negatif. Nilai *recall* yang diperoleh adalah 90% untuk kelas positif dan 99% untuk kelas negatif. Dalam hal ini, nilai *F1-score* adalah 93% untuk kelas positif dan 98% untuk kelas negatif.

### 3.8 Visualisasi

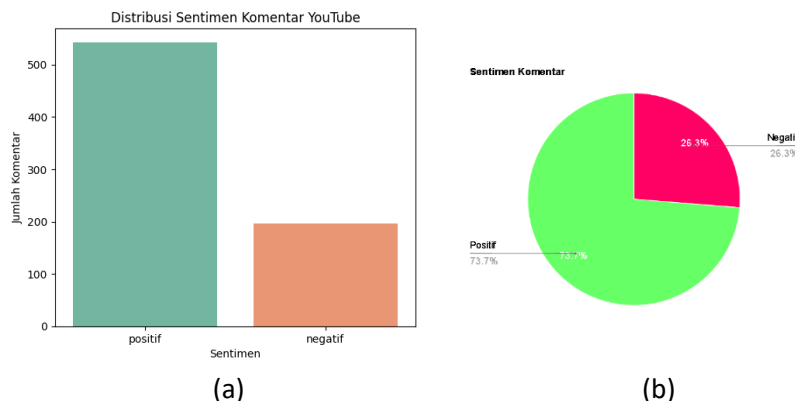
*Gambar 7 (a dan b)* menunjukkan hasil analisis sentimen yang divisualisasikan dalam bentuk *wordcloud* dan *chart* untuk memudahkan pembacaan informasi. *Wordcloud* menampilkan kedua sentimen hasil klasifikasi, yaitu positif dan negatif. Kata-kata seperti *baik*, *bagus*, dan *rakyat* mendominasi komentar dengan sentimen positif. Pada

sentimen negatif, kata-kata seperti *jelek*, *aneh*, *parah*, dan *korup* sering kali menjadi topik utama. Selain itu, kata-kata yang bersifat kontra terhadap kebijakan seperti *aneh*, *ganti*, dan *korupsi* juga muncul.



Gambar 7. WordCloud Sentimen

Gambar 8 (a) merupakan visualisasi hasil klasifikasi dalam bentuk *bar chart* yang menunjukkan jumlah komentar untuk masing-masing sentimen, yaitu sentimen positif sebanyak 542 komentar dan sentimen negatif sebanyak 197 komentar. Hal ini menunjukkan bahwa mayoritas masyarakat menunjukkan pandangan positif terhadap pergantian Kementerian Keuangan Indonesia. Gambar 8 (b) menampilkan visualisasi dalam bentuk *pie chart* yang memberikan sudut pandang berbeda dengan menampilkan proporsi sentimen dalam bentuk persentase, yaitu sentimen positif sebesar 73,7% dan sentimen negatif sebesar 26,3%. Kedua sentimen tersebut divisualisasikan secara menyatu dalam *chart* untuk memperlihatkan perbandingan secara proporsional.



Gambar 8. Chart Hasil Sentimen

#### 4. KESIMPULAN

Analisis sentimen yang dilakukan menunjukkan bahwa *Multiclass SVM* dengan *kernel linear* mampu melakukan klasifikasi terhadap data komentar *YouTube* yang mengandung sentimen terkait pergantian Kementerian Keuangan Indonesia dengan tingkat akurasi sebesar 97%. Hasil analisis menunjukkan bahwa sentimen positif mendominasi dengan jumlah 542 komentar (73,7%), sedangkan sentimen negatif berjumlah 197 komentar (26,3%). Temuan ini mengindikasikan bahwa mayoritas masyarakat menunjukkan respon positif terhadap pergantian Kementerian Keuangan Indonesia.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan, antara lain penggunaan satu sumber data, yaitu komentar publik di *YouTube*, sehingga hasil penelitian belum sepenuhnya merepresentasikan opini masyarakat secara menyeluruh. Selain itu, penelitian ini hanya menerapkan satu algoritma klasifikasi tanpa perbandingan dengan metode lain, serta membatasi klasifikasi sentimen pada dua kelas, yaitu positif dan negatif. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas sumber data dari berbagai platform media sosial, melakukan perbandingan performa dengan algoritma klasifikasi lain, serta mengembangkan skema klasifikasi sentimen yang lebih kompleks agar diperoleh hasil analisis yang lebih komprehensif dan mendalam.

#### UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada seluruh pihak yang telah memberikan bantuan, dukungan, serta kontribusi dalam pelaksanaan penelitian dan penulisan artikel ini sehingga dapat diselesaikan dengan baik dan tepat waktu. Penghargaan juga diberikan kepada pihak-pihak yang telah membantu dalam proses publikasi artikel ini.

**REFERENSI**

- [1] K. Nandini dan M. Rahardi, "Sentiment Analysis of Economic Policy Comments on YouTube Using Ensemble Machine Learning," 2025. [Daring]. Tersedia pada: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [2] A. Nur, A. Saputra, R. E. Saputro, dan S. Saputra, "Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments," *Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 2, 2025, doi: 10.33395/sinkron.v9i2.14513.
- [3] F. Rifaldy, Y. Sibaroni, dan S. S. Prasetyowati, "Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, hlm. 863–874, Mar 2025, doi: 10.29100/jipi.v10i2.6055.
- [4] S. Raihan Putri dan L. Rosnita, "Sentiment Analysis of Youtube and Gotube Reviews on Google Play Using the Support Vector Machine (SVM) Method in Indonesia," 2025. [Daring]. Tersedia pada: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] M. Ma'rufudin dan A. Yudhistira, "Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, hlm. 845–857, Mar 2025, doi: 10.52436/1.jpti.717.
- [6] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, S. Khomsah, dan A. S. Aribowo, "Terakreditasi SINTA Peringkat 2 Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *masa berlaku mulai*, vol. 1, no. 3, hlm. 648–654, 2017.
- [7] D. I. Sumantiawan, J. E. Suseno, dan W. A. Syafei, "Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction," *J. Sistem Info. Bisnis*, vol. 13, no. 1, hlm. 1–9, Jun 2023, doi: 10.21456/vol13iss1pp1-9.
- [8] N. Fajriyah, N. T. Lapatta, D. W. Nugraha, dan R. Laila, "IMPLEMENTASI SVM DAN SMOTE PADA ANALISIS SENTIMEN MEDIA SOSIAL X TERHADAP PELANTIKAN AGUS HARIMURTI YUDHOYONO," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, hlm. 1359–1370, Mar 2025, doi: 10.29100/jipi.v10i2.6246.
- [9] S. Chairani Siregar, R. T. Adek, dan Z. Fitri, "Sentiment Analysis of Comments on Youtube Channel Beauty Vlogger in Indonesian Language Using Support Vector Machine Method," vol. 2, hlm. 1–6, 2024, doi: 10.29103/icomden.v2.xxxx.
- [10] A. Syarif, A. I. Akaf, R. Prabowo, dan K. Muludi, "Analisis Sentimen Opini Masyarakat Terhadap Pelayanan BPJS Kesehatan Provinsi Lampung Berbasis Twitter," *Jurnal Pepadun*, vol. 3, no. 3, hlm. 380–388, Des 2022, doi: 10.23960/pepadun.v3i3.136.
- [11] P. D. Wulandari, U. Enri, dan A. Primajaya, "Analisis Sentimen Terhadap Kenaikan Iuran Bpjs Kesehatan Pada Twitter Menggunakan Naïve Bayes Classifier," *Jurnal Ilmu Komputer dan Teknologi*, vol. 1, no. 2, hlm. 18–22, Nov 2020, doi: 10.35960/ikomti.v1i2.575.
- [12] A. W. Nugroho dan N. Norhikmah, "Sentiment Analysis using the Support Vector Machine Algorithm on Covid\_19," *SISTEMASI*, vol. 13, no. 4, hlm. 1758, Jul 2024, doi: 10.32520/stmsi.v13i4.3778.