



Analisis Sentimen Opini Publik Terhadap Program Bantuan Subsidi Upah (BSU) di TikTok dengan Pendekatan *Support Vector Machine* (SVM)

Tiara Septya Madiana¹, Noor Latifah², Fajar Nugraha³

^{1,2,3} Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus

Article Info:

Dikirim: 28 Mei 2025

Direvisi: 1 Juni 2025

Diterima: 15 Juni 2025

Tersedia Online: 30 Juni 2025

Penulis Korespondensi:

Tiara Septya Madiana

Program Studi Sistem Informasi,

Fakultas Teknik,

Universitas Muria Kudus

Email: 202153047@std.umk.ac.id

Abstrak: Program Bantuan Subsidi Upah (BSU) merupakan inisiatif pemerintah Indonesia untuk mendukung pekerja berpenghasilan rendah sebagai bentuk perlindungan sosial, khususnya di tengah pemulihan pasca-pandemi. Namun, pelaksanaannya menuai berbagai respons dari masyarakat, terutama yang disampaikan melalui media sosial seperti TikTok. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi sentimen opini masyarakat terhadap BSU 2025 menggunakan algoritma *Support Vector Machine* (SVM), guna memperoleh pemetaan persepsi publik secara otomatis dan berbasis data. Data komentar diperoleh melalui proses web scraping dari platform TikTok, kemudian diproses melalui tahapan CRISP-DM yang mencakup business understanding, eksplorasi data, preprocessing (cleaning, tokenizing, stemming, dan lainnya), pelabelan awal menggunakan pendekatan lexicon-based, transformasi fitur dengan metode TF-IDF, serta pemodelan dan evaluasi. Model SVM yang dibangun dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model memiliki performa tinggi dengan akurasi 91% pada data latih dan 88% pada data uji. Sistem ini berhasil mengklasifikasikan komentar ke dalam tiga kelas sentimen: positif, netral, dan negatif. Dengan hasil tersebut, sistem ini dapat menjadi alat pendukung kebijakan yang membantu pemerintah dalam memahami dan merespons opini masyarakat secara lebih cepat, akurat, dan berbasis bukti.

Kata kunci: Analisis Sentimen, TikTok, Bantuan Subsidi Upah (BSU), Opini Publik, *Support Vector Machine* (SVM), Data Mining

Abstract: The Wage Subsidy Assistance (BSU) program is an Indonesian government initiative to support low-income workers as a form of social protection, particularly during the post-pandemic recovery. However, its implementation has drawn mixed responses from the public, particularly those expressed through social media platforms like TikTok. This study aims to develop a public opinion sentiment classification system for the 2025 BSU using the *Support Vector Machine* (SVM) algorithm to automatically and data-drivenly map public perception. Comment data was obtained through web scraping from the TikTok platform, then processed through the CRISP-DM stages, which include business understanding, data exploration, preprocessing (cleaning, tokenizing, stemming, and others), initial labeling using a lexicon-based approach, feature transformation using the TF-IDF method, and modeling and evaluation. The developed SVM model was evaluated using accuracy, precision, recall, and F1-score metrics. The evaluation results showed high performance with 91% accuracy on training data and 88% on test data. The system successfully classified comments into three sentiment classes: positive, neutral, and negative. With these results, this system can become a policy support tool that helps the government understand and respond to public opinion more quickly, accurately, and evidence-based.

Keywords: Sentiment Analysis, TikTok, Wage Subsidy Assistance (BSU), Public Opinion, *Support Vector Machine* (SVM), Data Mining

1. PENDAHULUAN

Pemerintah Indonesia kembali meluncurkan Program Bantuan Subsidi Upah (BSU) sebagai upaya mendukung daya beli masyarakat dan mempercepat pemulihan ekonomi pasca-pandemi. Program ini menasar pekerja dengan penghasilan di bawah Rp3,5 juta, termasuk guru honorer dan pekerja sektor informal. Bantuan sebesar Rp600 ribu diberikan dalam dua tahap selama bulan Juni dan Juli. Secara umum, kebijakan ini disambut baik karena menunjukkan kepedulian pemerintah terhadap kelompok ekonomi rentan.

Namun, pelaksanaan BSU tidak lepas dari berbagai tantangan. Masyarakat mengeluhkan beberapa permasalahan seperti keterlambatan penyaluran, ketidaktepatan sasaran, hingga kurangnya transparansi dalam proses pendataan. Beragam reaksi ini memperlihatkan bahwa meskipun kebijakan memiliki niat baik, implementasinya tetap menjadi sorotan publik yang membutuhkan perhatian serius agar manfaat bantuan benar-benar tepat sasaran.

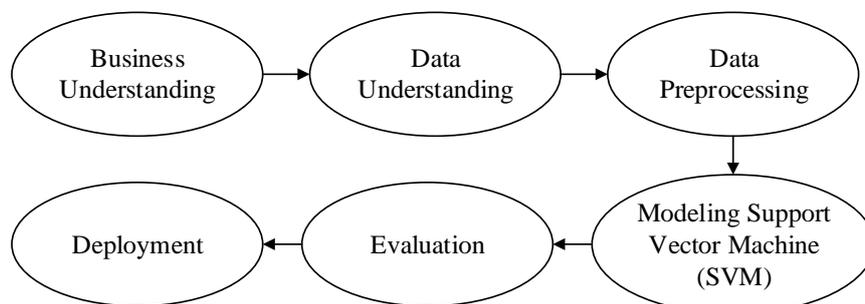
Di tengah dinamika tersebut, media sosial, khususnya TikTok, menjadi ruang ekspresi yang ramai digunakan masyarakat untuk menyampaikan pendapat mereka terkait BSU. Sebagai *platform* yang awalnya fokus pada hiburan berbasis video pendek, TikTok kini berkembang menjadi media diskusi publik yang dinamis. Pengguna dari berbagai latar belakang memanfaatkan fitur komentar, video unggahan, hingga siaran langsung untuk merespons kebijakan ini, menjadikan TikTok sebagai sumber data publik yang sangat kaya.

Untuk mengelola data dalam jumlah besar dan dalam bentuk teks tidak terstruktur, diperlukan pendekatan teknologi yang mampu menangani kompleksitas bahasa alami [1]. Dalam konteks ini, metode *Support Vector Machine* (SVM) menjadi salah satu algoritma yang efektif untuk melakukan analisis sentimen [2]. SVM memiliki keunggulan dalam mengklasifikasikan data teks ke dalam kategori sentimen positif, negatif, atau netral dengan akurasi yang tinggi [3]. Dengan memanfaatkan metode ini, opini masyarakat terkait BSU dapat diklasifikasikan secara otomatis, sehingga membantu pemerintah dan pemangku kepentingan dalam memahami arah sentimen publik serta menjadikannya sebagai bahan evaluasi kebijakan.

Melalui pendekatan berbasis data ini, diharapkan proses pengambilan keputusan dapat menjadi lebih responsif, partisipatif, dan berbasis bukti. Analisis sentimen terhadap komentar masyarakat di TikTok bukan hanya berperan sebagai alat bantu akademik, tetapi juga dapat digunakan secara praktis untuk menyempurnakan kebijakan sosial ekonomi agar lebih tepat sasaran. Oleh karena itu, penting dilakukan penelitian yang fokus pada bagaimana persepsi publik terhadap BSU 2025 terbentuk dan tersebar melalui media sosial, serta bagaimana teknologi *machine learning* dapat dimanfaatkan untuk membaca dinamika opini tersebut secara objektif dan efisien.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, pendekatan yang digunakan adalah *machine learning* dengan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*) sebagai pedoman utama dalam proses pengembangan model klasifikasi sentimen opini masyarakat terhadap Program Bantuan Subsidi Upah (BSU) 2025 [4]. Proses analisis dilakukan melalui tahapan CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu suatu metodologi standar dalam proses data mining yang terdiri dari enam tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [5]. Pendekatan ini dipilih karena bersifat sistematis dan terstruktur, sehingga mampu menyesuaikan dengan kebutuhan pengolahan data teks tidak terstruktur dari *platform* TikTok. Proses klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM), yang bertujuan untuk mengklasifikasikan opini masyarakat ke dalam kategori sentimen positif, negatif, dan netral secara efektif dengan akurasi yang tinggi. Alur pelaksanaan proses data mining mengikuti tahapan CRISP-DM, seperti yang akan ditunjukkan pada Gambar 1 berikut ini.



Gambar 1. Diagram Alir Proses CRISP-DM

2.1 Business Understanding

Tahapan *business understanding* bertujuan untuk memahami konteks dan kebutuhan utama dari penelitian [6]. Dalam hal ini, fokus utama adalah mengetahui dan memetakan persepsi masyarakat terhadap Program Bantuan Subsidi Upah (BSU) 2025 yang diluncurkan pemerintah sebagai bentuk perlindungan sosial bagi pekerja berpenghasilan rendah. Mengingat besarnya jumlah opini masyarakat yang disampaikan melalui media

sosial TikTok, dibutuhkan pendekatan analitis berbasis teknologi untuk mengklasifikasikan opini publik secara otomatis ke dalam sentimen positif, negatif, atau netral. Informasi ini diharapkan dapat menjadi bahan evaluasi dan penyusunan kebijakan berbasis data.

2.2 Data Understanding

Tahap ini mencakup pengumpulan, pemahaman dan eksplorasi awal data yang akan digunakan dalam penelitian [6]. Data utama penelitian ini sendiri berasal dari komentar pengguna TikTok pada video-video yang membahas topik BSU 2025. Pengumpulan dilakukan melalui proses *web scraping* atau API, dengan menggunakan kata kunci seperti “BSU 2025”, “Bantuan Subsidi Upah”, dan istilah serupa. Selain komentar teks, metadata video seperti deskripsi dan jumlah interaksi juga dapat dikumpulkan. Untuk memperkaya konteks, data sekunder berupa artikel berita dan dokumen kebijakan turut dikaji. Tujuan dari tahap ini adalah memahami struktur, kualitas, serta potensi pola yang muncul dari data opini publik yang dikumpulkan.

2.3 Data Preprocessing

Pada tahap data preparation ini merupakan langkah dimana data mentah akan diseleksi dan diproses lebih lanjut [7]. Setiap kata dalam ulasan akan dipecah menjadi bagian-bagian yang lebih kecil untuk membuat maknanya lebih jelas dan terperinci. Proses ini meliputi beberapa tahapan sebagaimana berikut [8]:

1. *Cleaning* adalah teknik penghapusan karakter yang tidak relevan seperti emoji, simbol, angka, URL, dan tanda baca agar teks menjadi bersih dan mudah diolah.
2. *Case Folding* adalah teknik yang dapat mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menghindari duplikasi kata akibat perbedaan kapitalisasi.
3. *Tokenizing* adalah proses pemecahan kalimat atau teks menjadi unit kata-kata (token) individual untuk memudahkan analisis kata per kata.
4. *Normalization* adalah pengkonversian kata-kata tidak baku atau kata gaul menjadi bentuk baku sesuai dengan kamus normalisasi, agar makna kata lebih konsisten.
5. *Stemming* adalah proses mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan seperti awalan atau akhiran (semisal: bermain menjadi main).
6. *Stopword Removal & Filtering* adalah teknik yang dapat menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis (seperti: dan, yang, di) serta menyaring token yang tidak relevan.
7. *Join Filtering Tokens* adalah proses penggabungan kembali token-token hasil filter menjadi satu kalimat atau *string* yang telah dibersihkan dan siap digunakan untuk analisis selanjutnya.
8. *Lexicon-Based Scoring* adalah teknik pemberian label sentimen awal (positif, negatif, atau netral) berdasarkan jumlah kata positif dan negatif yang ditemukan pada komentar menggunakan kamus kata (*lexicon*) yang telah disiapkan.
9. *TF-IDF (Term Frequency-Inverse Document Frequency)* adalah salah satu teknik utama dalam *feature extraction* dari teks, yang mengubah setiap dokumen/komentar menjadi representasi angka (fitur) yang bisa diproses oleh model.

2.4 Modeling Support Vector Machine

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang digunakan dalam klasifikasi data berdasarkan pola dari hasil pelatihan sebelumnya [9]. Data yang digunakan berbentuk numerik, dengan pembobotan menggunakan TF-IDF untuk membentuk pola klasifikasi. Model SVM menghasilkan *hyperplane*, yaitu garis pemisah yang memisahkan dua kelas data, misalnya antara ulasan pengguna TikTok yang positif (label 1), negative (label 2) dan netral (0). Rumus dasar untuk *hyperplane*:

$$f(x) : w \cdot x + b = 0 \quad (1)$$

Keterangan:

w = Parameter *hyperplane* yang mewakili vektor normal atau garis tegak lurus antara *hyperplane* dan titik *support vector*

x = Data input untuk SVM (misalnya, x_1 = indeks kata, x_2 = bobot kata)

b = Bias, parameter tambahan untuk mengatur posisi *hyperplane*

f = Fungsi *hyperplane*

2.5 Evaluation

Confusion Matrix merupakan metode evaluasi model yang menilai kinerjanya berdasarkan akurasi, presisi, *recall*, dan *F1-Score*. Presisi menunjukkan ketepatan prediksi pada data relevan, sementara *recall* mengukur kemampuan model dalam mendeteksi data yang benar [10].

Tabel 1. Confusion Matrix

	True	False
True (Positif)	TP (True Positif) Correct result	FP (False Positif) Unexpected result
False (Negatif)	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

Berikut adalah rumus dalam menghitung *precision*, *recall*, *accuracy*, dan *F1-Score*

$$1. \text{ Precision} : \frac{TP}{TP + FP} \quad (2)$$

$$2. \text{ Recall} : \frac{TP}{TP + FN} \quad (3)$$

$$3. \text{ Accuracy} : \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$4. \text{ F1-Score} : \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Recall} + \text{Precision}} \quad (5)$$

Dalam *confusion matrix*, *True Positive* (TP) adalah data positif yang diprediksi dengan benar, sedangkan *False Positive* (FP) merupakan data negatif yang salah diklasifikasikan sebagai positif. *False Negative* (FN) menunjukkan data positif yang salah terdeteksi sebagai negatif, sementara *True Negative* (TN) adalah data negatif yang berhasil diklasifikasikan dengan benar. Model dikatakan berkinerja baik jika *precision*, *recall*, *F1-Score*, dan *accuracy* bernilai tinggi, karena mencerminkan ketepatan prediksinya.

2.6 Deployment

Tahap *deployment* merupakan bagian akhir dari alur kerja CRISP-DM yang bertujuan untuk menyajikan hasil analisis dalam bentuk yang bisa digunakan secara langsung oleh pengguna atau pembuat kebijakan [11]. Dalam penelitian ini, proses *deployment* dilakukan dengan membangun aplikasi web interaktif berbasis *Streamlit* yang memungkinkan klasifikasi sentimen terhadap komentar TikTok terkait program BSU dilakukan secara real-time.

3. HASIL DAN PEMBAHASAN

Pada pembahasan hasil ini, dilakukan proses analisis sentimen terhadap komentar masyarakat mengenai Program Bantuan Subsidi Upah (BSU) 2025 yang diperoleh dari platform TikTok. Proses analisis dilakukan melalui tahapan CRISP-DM dengan menggunakan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi. Hasil yang diperoleh mencakup distribusi sentimen (positif, negatif, dan netral), performa model, serta interpretasi terhadap kecenderungan opini publik. Pembahasan dalam bab ini juga akan mengaitkan temuan data dengan konteks sosial dan kebijakan terkait BSU.

3.1. Business Understanding

Pada tahap ini, fokus utama adalah memahami tujuan dan konteks bisnis atau penelitian. Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan opini publik mengenai program Bantuan Subsidi Upah (BSU) dengan menggunakan data yang diambil dari platform TikTok. Program BSU adalah bantuan pemerintah yang penting, sehingga mengetahui bagaimana masyarakat meresponsnya sangat krusial untuk evaluasi dan perbaikan kebijakan, karena volume data komentar dan opini dari TikTok sangat besar dan tidak memungkinkan untuk dianalisis secara manual, maka diperlukan penerapan algoritma machine learning. Algoritma ini akan membantu

melakukan klasifikasi opini secara otomatis, misalnya membedakan komentar yang bersifat positif, negatif, atau netral. Dengan demikian, hasil analisis dapat diperoleh lebih cepat, efisien, dan akurat.

3.2. Data Understanding

Tahap ini berfokus pada proses pengumpulan dan pemahaman data yang akan digunakan dalam penelitian. Data dikumpulkan dari platform TikTok dengan cara menggunakan teknik *web scraping* atau melalui API (*Application Programming Interface*), yang memungkinkan pengambilan data secara otomatis dan sistematis. Pengambilan data difokuskan pada konten yang berhubungan dengan program BSU dengan menggunakan kata kunci pencarian seperti “BSU”, “Bantuan Subsidi Upah”, “BSU 2025”, dan variasi lain yang relevan. Data yang diambil meliputi teks komentar pengguna serta deskripsi video yang memuat opini publik terkait BSU. Selain data sekunder dari TikTok, juga dilakukan studi dokumentasi dengan mengumpulkan data pendukung, seperti artikel berita, laporan resmi, atau dokumen kebijakan yang relevan. Data pendukung ini berguna untuk memberikan konteks dan memperkaya pemahaman terhadap opini publik serta validasi hasil analisis.

3.3. Data Preprocessing

Tahapan *data preparation* bertujuan untuk membersihkan dan mempersiapkan data teks agar dapat dianalisis menggunakan metode *machine learning*. Dalam penelitian ini, proses *preprocessing* dilakukan secara berurutan melalui langkah-langkah berikut:

1. Cleaning

Proses ini menghapus karakter khusus, URL, *mention*, emotikon, dan simbol-simbol tidak penting menggunakan *regex*. Juga dilakukan penghapusan angka, tanda baca, spasi berlebih, serta karakter tunggal. Hasil dari tahap ini adalah teks yang bersih dan lebih siap untuk diolah.

Fungsi yang diterapkan meliputi:

- `remove_tweet_special()`
- `remove_number()`
- `remove_punctuation()`
- `remove_whitespace_LT()`
- `remove_whitespace_multiple()`
- `remove_single_char()`

2. Case Folding

Langkah awal adalah *case folding*, yaitu mengubah seluruh teks menjadi huruf kecil (*lowercase*) agar konsistensi data terjaga dan perbedaan huruf kapital tidak mempengaruhi analisis.

Potongan Kode Output:

```
# ----- Case Folding -----  
# Using Series.str.lower() function on Pandas  
tweet['Text Case Folding'] = tweet['text'].str.lower()
```

Case Folding Result :

```
0          ujung ujungnya juga di korup elahh  
1                                     percaya?  
2          yang gaji nya 3,6  
3   knp gabikin larangan perusahaan ngegaji dibawa...  
4   mending kemarin, subsidi listrik 50% lebih jel...  
Name: Text Case Folding, dtype: object
```

3. Tokenizing

Kalimat-kalimat komentar kemudian dipecah menjadi token atau unit kata menggunakan fungsi *wordpunct_tokenize* dari pustaka NLTK (*Natural Language Toolkit*). Tahapan ini menghasilkan daftar kata dari setiap komentar.

Potongan Kode Output:

```
# ----- Tokenizing -----  
tweet['Text Tokenizing'] = tweet['Text Case  
Folding'].apply(word_tokenize_wrapper)
```

Tokenizing Result :

```
0          [ujung, ujungnya, juga, di, korup, elahh]  
1                                     [percaya]  
2          [yang, gaji, nya]
```

```
3 [knp, gabikin, larangan, perusahaan, ngegaji, ...
4 [mending, kemarin, subsidi, listrik, lebih, je...
Name: Text Tokenizing, dtype: object
```

4. Normalization

Tahapan ini bertujuan untuk mengubah kata-kata tidak baku atau slang seperti “gk”, “nggak”, “udh”, menjadi bentuk baku sesuai Kamus Bahasa Indonesia. Kamus normalisasi (Kamus_Baku.xlsx) digunakan sebagai acuan konversi.

```
Potongan Kode Output:
# ----- Normalization -----
tweet['Text Normalization'] = tweet['Text Tokenizing'].apply(normalized_term)
```

	Text
	Normalization
0	[ujung, ujungnya, juga, di, korupsi, elahh]
1	[percaya]
2	[yang, gaji, nya]
3	[kenapa, gabikin, larangan, perusahaan, mengga...]
4	[mending, kemarin, subsidi, listrik, lebih, je...]

5. Stemming

Setiap kata yang telah dinormalisasi kemudian dikembalikan ke bentuk dasarnya menggunakan algoritma *stemming* dari Sastrawi. Proses ini menghapus imbuhan (awalan, akhiran, dan sisipan), sehingga berbagai variasi kata disederhanakan ke akar katanya. Misalnya, “membantu”, “bantuan”, dan “dibantu” akan dikembalikan menjadi “bantu”.

```
Potongan Kode Output:
# ----- Stemming -----
tweet['Text Stemming'] = tweet['Text
Normalization'].swifter.apply(get_stemmed_term)
```

```
Output stemming akan dipotong hingga 5000 baris terakhir
menjadi : jadi
meskipun : meski
dua : dua
ke : ke
tangan : tangan
kita : kita
pajakjatah : pajakjatah
lain : lain
dilarang : larang
atau : atau
...
```

6. Stopword Removal & Filtering

Kata-kata yang tidak memiliki makna penting dalam analisis sentimen, seperti kata hubung, kata bantu, atau kata sapaan, dihapus berdasarkan daftar *stopword* dari NLTK dan file eksternal (stopwordbahasa.txt). Filtering ini dilakukan untuk mengurangi *noise* dalam data.

```
Potongan Kode Output:
# ----- Stopword Removal & Filtering -----
tweet['Text Filtering'] = tweet['Text Stemming'].apply(stopwords_removal)
```

```
Text Filtering \
0 [ujung, ujung, korupsi, elahh]
1 [percaya]
2 [gaji]
3 [gabikin, larang, usaha, gaji, umr]
4 [mending, kemarin, subsidi, listrik, jelsebe...]
```

7. Join Filtering Tokens

Token yang tersisa kemudian digabung kembali menjadi kalimat utuh, yang disimpan dalam kolom Text Filtering Join. Format ini diperlukan untuk keperluan tampilan, analisis lanjutan, dan model klasifikasi teks.

Potongan Kode Output:

```
# ----- Join Filtering Tokens -----
tweet['Text Filtering Join'] = tweet['Text Filtering'].apply(lambda x: '
'.join(x))

                                Text Filtering Join
0                                ujung ujung korupsi elahh
1                                percaya
2                                gaji
3                                gabikin larang usaha gaji umr
4 mending kemarin subsidi listrik jelasbenarnya...
```

8. Labeling Awal (*Lexicon-Based*)

Sebelum dilakukan pelatihan model, data dilabeli terlebih dahulu menggunakan pendekatan lexicon-based scoring berdasarkan daftar kata positif dan negatif yang terdapat dalam dua file Excel (kamus_positive.xlsx dan kamus_negative.xlsx). Komentar yang memiliki skor > 0 diklasifikasikan sebagai positif, skor < 0 sebagai negatif, dan skor = 0 sebagai netral.

9. Term Frequency – Inverse Document Frequency (TF-IDF) Vectorization

Tahap selanjutnya adalah mengubah data teks tersebut menjadi bentuk numerik agar dapat dikenali oleh algoritma *machine learning*. Transformasi ini dilakukan menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). TF-IDF menghitung bobot pentingnya suatu kata dalam sebuah dokumen relatif terhadap kumpulan seluruh dokumen. Semakin sering sebuah kata muncul di satu dokumen tetapi jarang muncul di dokumen lain, maka semakin tinggi bobotnya. Dengan demikian, kata-kata umum akan diberi bobot rendah, sementara kata-kata yang lebih unik akan memiliki bobot tinggi.

Potongan Kode Output:

```
# ----- TF-IDF Vectorization -----
tfidf_vectorizer = TfidfVectorizer(
    max_features=2500,
    min_df=7,
    max_df=0.8
)

X = tfidf_vectorizer.fit_transform(tweet['Text Filtering Join'])
```

3.4. Modeling Support Vector Machine

Setelah data komentar TikTok yang dikumpulkan melalui proses scraping berhasil diproses dan direpresentasikan dalam bentuk TF-IDF *vector*, langkah selanjutnya adalah membangun model klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM). Algoritma ini dipilih karena keandalannya dalam menangani permasalahan klasifikasi teks, terutama ketika data memiliki dimensi tinggi seperti pada TF-IDF.

SVM bekerja dengan mencari *hyperplane* terbaik yang dapat memisahkan kelas-kelas data (dalam hal ini: sentimen positif, netral, dan negatif) secara optimal. Model berusaha memaksimalkan margin antara titik-titik data dari kelas yang berbeda, sehingga prediksi untuk data baru menjadi lebih akurat. Keluaran dari model ini adalah label sentimen, yang digunakan untuk menganalisis persepsi publik terhadap program Bantuan Subsidi Upah (BSU) berdasarkan komentar di TikTok. Hasil klasifikasi ini kemudian dievaluasi pada tahap berikutnya menggunakan metrik evaluasi performa model.

3.5. Evaluation

Evaluasi merupakan tahap penting dalam pengujian model klasifikasi, bertujuan untuk mengetahui sejauh mana kinerja model dalam memprediksi data baru secara akurat dan konsisten. Dalam penelitian ini, evaluasi dilakukan terhadap model *Support Vector Machine* (SVM) yang telah dilatih sebelumnya dengan data komentar TikTok yang telah melalui proses *preprocessing* dan transformasi TF-IDF.

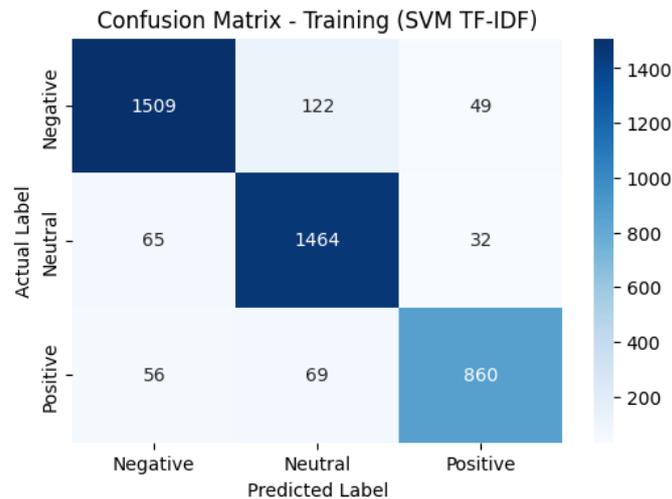
Untuk uji dari hasil evaluasi ada dua bagian yaitu data latih dan data uji, sebagaimana berikut:

1. Hasil Evaluasi Data Latih

Akurasi model terhadap data latih mencapai nilai tinggi, yang menunjukkan model mampu belajar pola dengan baik dari data yang diberikan. *Confusion matrix* divisualisasikan menggunakan *heatmap*, menunjukkan proporsi prediksi yang tepat untuk setiap kelas sentimen.

a. *Confusion Matrix – Training*

Confusion matrix digunakan untuk memvisualisasikan performa model dalam mengenali setiap kelas pada data pelatihan. Matriks ini menunjukkan distribusi prediksi yang dilakukan model terhadap label sebenarnya, serta membantu dalam mengukur akurasi klasifikasi per kelas. Berikut merupakan hasil dari *Confusion Matrix Training* yang ditunjukkan pada Gambar 2 dibawah:



Gambar 2. *Confusion Matrix Training*

Pada Gambar 2, terlihat bahwa model SVM dengan representasi fitur TF-IDF mampu mengklasifikasikan sebagian besar data pelatihan dengan akurat:

- 1) Dari 1680 komentar Negatif, sebanyak 1509 berhasil diprediksi dengan benar.
- 2) Dari 1561 komentar Netral, model mampu mengklasifikasikan 1464 secara tepat.
- 3) Dari 985 komentar Positif, sebanyak 860 berhasil dikenali dan diklasifikasikan dengan benar.

Hasil ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik terhadap data pelatihan, dengan tingkat kesalahan klasifikasi yang relatif rendah. *Confusion matrix* ini memperkuat bahwa model cukup efektif dalam mengenali dan membedakan pola sentimen dari setiap kelas secara seimbang.

b. *Classification Report – Training*

Untuk menilai kinerja model dalam tahap pelatihan, dilakukan evaluasi menggunakan metrik klasifikasi yang mencakup *precision*, *recall*, dan *f1-score* untuk masing-masing kelas sentimen: Negatif, Netral, dan Positif. Evaluasi ini juga menyertakan nilai akurasi keseluruhan, serta rata-rata makro (*macro average*) dan rata-rata tertimbang (*weighted average*) sebagai indikator umum performa model. Berikut merupakan hasil dari *Classification Report Training* yang ditunjukkan pada Tabel 2 dibawah:

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negatif</i>	0.93	0.90	0.91	1680
<i>Netral</i>	0.88	0.94	0.91	1561
<i>Positif</i>	0.91	0.87	0.89	985
<i>Accuracy</i>	-	-	0.91	4226
<i>Macro Avg</i>	0.91	0.90	0.91	4226
<i>Weighted Avg</i>	0.91	0.91	0.91	4226

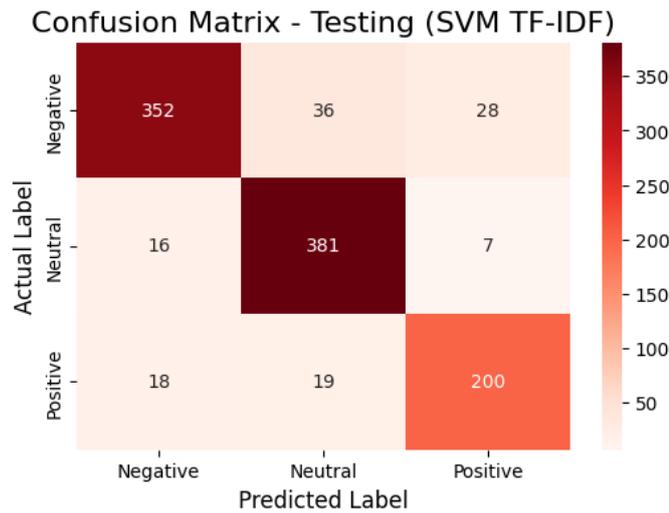
Berdasarkan tabel di atas, dapat disimpulkan bahwa model SVM memiliki performa klasifikasi yang stabil dan efektif pada data pelatihan. Kelas Netral memperoleh nilai *recall* tertinggi sebesar 0.94, yang menunjukkan bahwa sebagian besar data Netral berhasil dikenali. Di sisi lain, kelas Negatif mencatat nilai *precision* tertinggi sebesar 0.93, yang berarti prediksi untuk kelas ini sangat akurat. Kelas Positif juga menunjukkan hasil yang baik, dengan nilai *F1-score* sebesar 0.89. Secara keseluruhan, model mencapai akurasi sebesar 91%, menunjukkan bahwa proses pembelajaran dari data pelatihan berjalan dengan sangat baik tanpa adanya tanda-tanda *underfitting*.

2. Hasil Evaluasi Data Uji

Akurasi model terhadap data uji juga tergolong baik, menandakan bahwa model memiliki kemampuan generalisasi yang cukup kuat. *Confusion matrix* data uji ditampilkan dalam heatmap dengan 3 label *sentiment* yaitu Positif, Netral, dan Negatif.

a. *Confusion Matrix – Testing*

Confusion matrix pada tahap pengujian menggambarkan bagaimana model yang telah dilatih menerapkan pemahamannya terhadap data baru yang belum pernah dilihat sebelumnya. Hal ini penting untuk mengevaluasi konsistensi dan kemampuan generalisasi model. Berikut merupakan hasil dari *Confusion Matrix Testing* yang ditunjukkan pada Gambar 3 dibawah:



Gambar 3. *Confusion Matrix Testing*

Pada gambar 3, dapat dilihat performa model terhadap data pengujian sebagai berikut:

- 1) Dari 416 komentar Negatif, sebanyak 352 diklasifikasikan dengan benar.
- 2) Dari 404 komentar Netral, 381 berhasil diprediksi secara akurat.
- 3) Dari 237 komentar Positif, sebanyak 200 berhasil dikenali dengan benar oleh model.

Meskipun jumlah data pengujian lebih sedikit dibandingkan data pelatihan, model tetap menunjukkan performa yang konsisten. Hal ini menunjukkan bahwa model tidak mengalami *overfitting* dan mampu mempertahankan akurasi yang baik dalam mendeteksi sentimen pada data yang belum pernah digunakan sebelumnya.

b. *Classification Report – Testing*

Evaluasi model juga dilakukan pada data pengujian untuk melihat seberapa baik model dapat melakukan generalisasi terhadap data baru. Metode evaluasi yang digunakan tetap sama, yaitu *precision*, *recall*, *f1-score*, akurasi keseluruhan, serta *macro* dan *weighted average*. Berikut merupakan hasil dari *Classification Report Testing* yang ditunjukkan pada tabel 3 dibawah:

Tabel 3. *Classification Report Testing*

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.91	0.85	0.88	416
Netral	0.87	0.94	0.91	404
Positif	0.85	0.84	0.85	237
Accuracy	-	-	0.88	1057
Macro Avg	0.88	0.88	0.88	1057
Weighted Avg	0.88	0.88	0.88	1057

Hasil evaluasi pada data pengujian menunjukkan bahwa model mempertahankan performa yang baik dan seimbang, dengan akurasi keseluruhan sebesar 88%. Nilai ini hanya sedikit lebih rendah dibandingkan dengan data pelatihan, yang menandakan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik. Kelas Netral kembali menunjukkan *recall* tertinggi (0.94), sementara kelas Negatif mencatat *precision* sebesar 0.91, mirip dengan hasil pada data pelatihan. Semua metrik untuk ketiga kelas berada di atas 0.84, memperlihatkan bahwa model cukup handal dalam mengenali berbagai sentimen dalam data uji.

3.6. Deployment

Bagian ini menyajikan berbagai tampilan antarmuka dari sistem analisis sentimen terhadap opini masyarakat mengenai Program Bantuan Subsidi Upah (BSU) 2025. Sistem ini dikembangkan dengan algoritma *Support Vector Machine* (SVM) dan ditujukan untuk mengklasifikasikan komentar masyarakat dari *platform* TikTok ke dalam tiga kategori yaitu positif, netral, dan negatif. Antarmuka dibangun menggunakan *framework* *Streamlit* yang memungkinkan pengguna mengakses sistem melalui web secara interaktif, ringan, dan intuitif. Terdapat tiga tab menu utama dalam sistem, yaitu:

1. Tampilan Halaman Tab Beranda

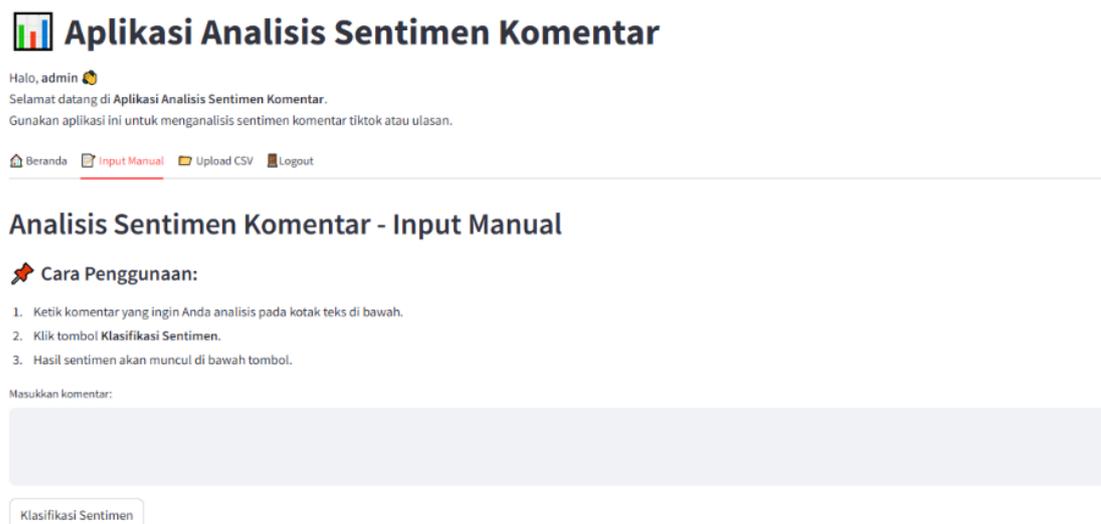
Halaman Beranda berfungsi sebagai titik awal penggunaan sistem. Tampilan ini memberikan penjelasan singkat mengenai latar belakang sistem, tujuan analisis, serta metode yang digunakan, yaitu pendekatan pemrosesan teks berbasis TF-IDF dan algoritma klasifikasi SVM. Informasi ini membantu pengguna memahami konteks dan cara kerja sistem sebelum menggunakan fitur lainnya. Selain itu, halaman ini juga dilengkapi ilustrasi atau ikon penunjang untuk meningkatkan daya tarik visual. Tampilan halaman ini dapat dilihat pada gambar 4 berikut:



Gambar 4. Tampilan Halaman *Tab* Beranda

2. Tampilan Halaman *Tab* Input Manual

Halaman Input Manual dirancang bagi pengguna yang ingin menguji komentar secara individu. Pengguna dapat mengetik komentar secara langsung pada kolom input teks, lalu menekan tombol "Klasifikasi Sentimen". Sistem akan melakukan pra-pemrosesan teks secara otomatis, menerapkan model SVM, dan menampilkan hasil klasifikasi dalam bentuk label warna (hijau untuk positif, kuning untuk netral, dan merah untuk negatif). Fitur ini sangat berguna untuk pengujian cepat atau demo secara langsung. Ilustrasi tampilan halaman ditunjukkan pada gambar 6 di bawah ini:



Gambar 5. Tampilan Halaman *Tab* Input Manual

3. Tampilan Halaman *Tab* Input CSV

Halaman *Upload* CSV ditujukan untuk pengguna yang ingin menganalisis komentar dalam jumlah besar secara sekaligus. Pengguna cukup mengunggah file berformat CSV yang berisi komentar-komentar masyarakat, dengan struktur kolom teks yang sesuai. Setelah file berhasil diunggah, sistem akan memproses seluruh komentar dan menampilkan hasil klasifikasi dalam bentuk tabel dan grafik distribusi sentimen.

Pengguna juga dapat mengunduh hasil analisis untuk kebutuhan dokumentasi atau pelaporan lebih lanjut. Tampilan halaman ini disajikan pada gambar 6 berikut:



Gambar 6. Tampilan Halaman Tab Input CSV

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa pengembangan sistem klasifikasi sentimen opini masyarakat terhadap Program Bantuan Subsidi Upah (BSU) 2025 menggunakan algoritma *Support Vector Machine (SVM)* mampu memberikan solusi analitis yang efektif dan akurat dalam membaca persepsi publik secara otomatis. Sistem ini dirancang mengikuti tahapan CRISP-DM secara sistematis, dimulai dari pemahaman konteks kebijakan, eksplorasi data komentar dari TikTok, *preprocessing* teks secara bertahap, pelabelan awal menggunakan pendekatan leksikon, transformasi data melalui TF-IDF, pemodelan dengan SVM, evaluasi performa model, hingga implementasi dalam bentuk aplikasi web interaktif berbasis *Streamlit*. Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang baik dengan akurasi sebesar 91% pada data latih dan 88% pada data uji, serta nilai *precision*, *recall*, dan *f1-score* yang seimbang di setiap kelas sentimen. Sistem ini tidak hanya membantu dalam mengidentifikasi kecenderungan opini masyarakat, tetapi juga berpotensi menjadi alat pendukung kebijakan berbasis data yang lebih transparan, partisipatif, dan responsif terhadap dinamika sosial yang berkembang di media digital.

DAFTAR PUSTAKA

- [1] Y. B. Widodo, F. Widyahastuti, M. Narji, and S. Sibuea, "Artificial Intelligence for Unstructured Data Processing," *J. Teknol. Inform. dan Komput.*, vol. 11, no. 1, pp. 202–213, 2025.
- [2] M. Iqbal, M. Afdal, and R. Novita, "Implementasi Algoritma Support Vector Machine Untuk Analisa Sentimen Data Ulasan Aplikasi Pinjaman Online di Google Play Store: Implementation of Support Vector Machine Algorithm for Sentiment Analysis of Online Loan Application Review Data on Google Play," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1244–1252, 2024.
- [3] M. A. A. Maldini and S. Andryana, "Analisis Sentimen Ulasan Pengguna Aplikasi Perbankan Menggunakan Algoritma Support Vector Machine Dan Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4098–4105, 2025.
- [4] M. Jamil, R. Rosihan, and M. S. Hanafi, "Analisis Sentimen Calon Kepala Daerah Maluku Utara dengan Metode CRISP-DM," *bit-Tech*, vol. 7, no. 3, pp. 995–1003, 2025.
- [5] I. Amelia, F. M. Sarimole, and others, "Analisis Sentimen Tanggapan Pengguna Media Sosial X Terhadap Program Beasiswa KIP-Kuliah dengan Menggunakan Algoritma Support Vector Machine (SVM)," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 2994–3003, 2024.
- [6] Z. Purwanti and others, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *J. Indones. Manaj. Inform. Dan Komun.*, vol. 5, no. 3, pp. 3065–3079, 2024.
- [7] F. F. Wati, S. Suleman, and A. E. Widodo, "Analisis Sentimen Ulasan Pengguna Aplikasi Deepseek Menggunakan Algoritma Random Forest Dan Naive Bayes," *CONTEN Comput. Netw. Technol.*, vol. 5, no. 1, pp. 8–15, 2025.
- [8] M. T. Nugraha, N. N. Sulistiyowati, and U. U. Enri, "Analisis Sentimen Ulasan Aplikasi Satu Sehat Pada

- Google Play Store Menggunakan Naive Bayes Classifier,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 7, no. 5, pp. 3593–3601, 2023.
- [9] S. Suryani and M. Mustakim, “Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine,” *Bull. Informatics Data Sci.,* vol. 1, no. 2, pp. 81–88, 2022.
- [10] P. Anggraini and W. Winarsih, “Komparasi Naive Bayes, Support Vector Machine, Dan Random Forest Dalam Analisis Sentimen Aplikasi Shopee Di Google Play Store,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 9, no. 3, pp. 4451–4457, 2025.
- [11] M. R. Amalsyah, D. Kurniawan, A. Rifai, and P. Sari, “Sentiment Analysis of Fintech Application User Reviews using the CRISP-DM Framework for Product Development Prioritization,” *SISTEMASI,* vol. 14, no. 2, pp. 813–825, 2025.