

Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode *Support Vector Machine*

Nur Aisyah Wahyuni¹, Dinda Putri Ayu², Hafidz Irsyad³

^{1,2,3}Program Studi Informatika, Universitas Multi Data Palembang, Indonesia

Email : nuraisyahwah@mhs.mdp.ac.id, dindaputri@mhs.mdp.ac.id, hafiziryad@mdp.ac.id

Article Information

Article history

Received 20 June 2024

Revised 25 June 2024

Accepted 30 June 2024

Available 30 June 2024

Keywords

Analysis

Single Tuition Fee Increase (STF)

Sentiment

Support Vector Machine

YouTube

Corresponding Author:

Dinda Putri Ayu,

Universitas Multi Data Palembang,

Email :

dindaputri@mhs.mdp.ac.id

Abstract

Students and the general public usually have different responses to the increase in Single Tuition Fees (UKT) at universities. Protests, complaints, and support for this increase may be expressed through various social media platforms, such as YouTube. Using the Support Vector Machine (SVM) method, this study analyzes comments on the YouTube platform related to the increase in UKT. Comment data is divided into three categories: positive, negative, and neutral. The evaluation results show that the SVM model achieves an accuracy of 0.88; it also demonstrates good ability to recognize negative sentiment with a precision of 0.83, recall of 0.90, and f1-score of 0.86. For neutral sentiment, the model shows a precision of 0.86, recall of 0.75, and f1-score of 0.80. Nevertheless, the SVM model achieves perfect scores for precision, recall, and f1-score of 1.00 for positive sentiment. Although the SVM model has proven effective in analyzing sentiment towards the increase in UKT on YouTube, further improvements are needed to enhance accuracy in identifying neutral sentiment.

Keywords : *Analysis, Single Tuition Fee Increase (UKT), Sentiment, Support Vector Machine, YouTube.*

Abstrak

Mahasiswa dan masyarakat umum biasanya memiliki tanggapan yang berbeda terhadap kenaikan Uang Kuliah Tunggal (UKT) di perguruan tinggi. Ada kemungkinan protes, keluhan, dan dukungan terhadap kenaikan ini diungkapkan melalui berbagai platform media sosial, seperti YouTube. Dengan menggunakan metode Support Vector Machine (SVM), penelitian ini menganalisis komentar di platform YouTube yang berkaitan dengan kenaikan UKT. Data komentar dibagi menjadi tiga kategori: positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa model SVM mencapai akurasi sebesar 0.88; itu juga menunjukkan kemampuan yang baik untuk mengenali sentimen negatif dengan presisi sebesar 0.83, recall sebesar 0.90, dan f1-score sebesar 0.86; dan untuk sentimen netral, model menunjukkan presisi sebesar 0.86, recall sebesar 0.75, dan f1-score sebesar 0.80. Meskipun demikian, model SVM mencapai nilai sempurna untuk presisi, recall, dan f1-score sebesar 1.00. Meskipun model SVM terbukti efektif dalam menganalisis sentimen terhadap kenaikan UKT di YouTube, perbaikan lebih lanjut diperlukan untuk meningkatkan akurasi dalam mengidentifikasi sentimen netral.

Kata Kunci : *Analisis, Kenaikan Uang Kuliah Tunggal (UKT), sentimen, Support Vector Machine, Youtube.*

Copyright©2024 Nur Aisyah Wahyuni, Dinda Putri Ayu, Hafidz Irsyad

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Dalam beberapa tahun terakhir, penggunaan teknologi digital telah berkembang pesat di Indonesia, khususnya dalam platform media sosial dan video streaming seperti *YouTube*. Platform ini tidak hanya digunakan sebagai sarana hiburan, tetapi juga sebagai media untuk menyampaikan dan berbagi pendapat mengenai berbagai isu sosial, politik, dan ekonomi. Salah satu isu yang sering menjadi perbincangan adalah kenaikan Uang Kuliah Tunggal (UKT) di berbagai perguruan tinggi di Indonesia.

Kenaikan UKT seringkali memicu reaksi yang beragam dari berbagai kalangan, terutama dari mahasiswa dan orang tua. Reaksi ini biasanya diekspresikan melalui berbagai komentar dan video di *YouTube*. Mengingat besarnya volume data yang dihasilkan, diperlukan suatu metode yang efektif untuk menganalisis sentimen dari komentar-komentar tersebut guna memahami pola dan kecenderungan opini publik.

Analisis sentimen merupakan komputasional dari opini orang lain, pemberian nilai, serta emosi yang terdapat dalam entitas, event, dan atribut yang dimiliki (Nugraha, Harani, & Habibi, 2020). Ini sangat penting untuk memahami persepsi penonton terhadap masalah tersebut. *YouTube* adalah salah satu media sosial yang besar dan menjadi sumber data untuk menganalisis sentimen. Pengguna aktif dapat memberikan ulasan komentar secara spontan. Oleh karena itu, melakukan penelitian tentang komentar penonton tentang masalah kenaikan UKT dapat memberikan informasi yang sedang dicari dan dapat memberikan wawasan tentang penggunaan teknik yang tepat.

Salah satu teknik atau metode yang paling efektif untuk analisis sentimen ini adalah *Support Vector Machine* (SVM). SVM merupakan teknik *machine learning* yang banyak digunakan untuk mengklasifikasi data. Selain itu, multi-vector matching (SVM) adalah algoritma pembelajaran mesin yang sangat akurat untuk klasifikasi data. SVM memilih hyperplane terbaik antar kelas berdasarkan vektor dukungan terdekat hyperplane (Dang, Moreno-García, & de la Prieta, 2020). Keunggulan SVM adalah kemampuan untuk mengidentifikasi berbagai hyperplane untuk memaksimalkan jarak antar kelas (Hao, Shaohong, & Jinping, 2011). Dalam konteks analisis sentimen, SVM dapat membantu mengkategorikan komentar-komentar penonton menjadi sentimen positif, negatif ataupun netral dengan akurasi tinggi.

Penelitian ini bertujuan untuk menerapkan metode SVM dalam menganalisis sentimen terhadap kenaikan UKT berdasarkan data komentar dari *YouTube*. Dengan menggunakan SVM, diharapkan hasil yang lebih jelas tentang persepsi publik terhadap kenaikan UKT akan digunakan sebagai dasar pengambilan kebijakan.

2. Kajian Terdahulu

2.1 Analisis Sentimen

Analisis sentimen adalah proses untuk menentukan sikap ataupun opini dari sebuah teks terhadap subjek atau isu tertentu. Teknik ini menggunakan metode pemrosesan bahasa alami dan analisis teks untuk mengidentifikasi informasi subjektif dari berbagai sumber data. Dalam konteks media sosial seperti *Youtube*, analisis sentimen ini berfungsi untuk memahami bagaimana penonton bereaksi terhadap konten yang diunggah, termasuk video, komentar dan ulasan.

Penelitian analisis sentimen sudah dilakukan sebelumnya pada layanan perbankan oleh Sari dan Irhamah (2020). Penelitian tersebut mengklasifikasi data *Twitter* menjadi sentimen positif dan negatif menggunakan representasi kata *Term Frequency Inverse Document Frequency* (TF IDF) sebagai input algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) dengan SMOTE (Sari & Irhamah, 2020).

2.2 Youtube

Youtube merupakan media sosial berbagi video terbesar dengan jutaan pengguna aktif setiap harinya yang dapat memberikan reaksi, komentar ataupun ulasan dari perfilman hingga konten video. Komentar yang diberikan dapat beragam opini dari penonton. Dengan adanya youtube bisa menjadi sumber data, tetapi daya ini terkadang berjumlah banyak sehingga memerlukan teknik analisis yang efektif dan efisien (Yosef, Sihombing, & Hanggara, 2020).

2.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Algoritma ini mencari *hyperplane* yang memisahkan data ke dalam kelas-kelas dengan margin terbesar, bekerja efektif pada data berdimensi tinggi. SVM juga menggunakan teknik *kernel* untuk memungkinkan pemisahan *non-linear* jika data tidak dapat dipisahkan secara *linear*.

Penelitian tentang analisis sentimen dan klasifikasi terhadap berita hoax *Covid-19* dengan menggunakan algoritma *Support Vector Machine*. Berdasarkan hasil dari penelitian ini, algoritma SVM dengan *kernel linear* menunjukkan hasil prediksi yang baik pada skenario (80:20). Pasalnya, model tersebut mampu membedakan berita yang hoax dan bukan hoax tentang *Covid-19* dengan akurasi tertinggi yaitu 97,06%. Kemudian hasil akurasi *kernel RBF* paling rendah pada skenario (90:10) yaitu 90,46%, model ini tidak cocok untuk mengklasifikasikan berita hoax atau tidak hoax *Covid-19* (Ropikoh et al., 2021).

Dalam analisis sentimen, SVM memproses teks menjadi vektor fitur menggunakan TF-IDF, lalu mengklasifikasikan komentar ke dalam kategori sentimen seperti positif, negatif, atau netral. Kekuatan SVM terletak pada kemampuannya

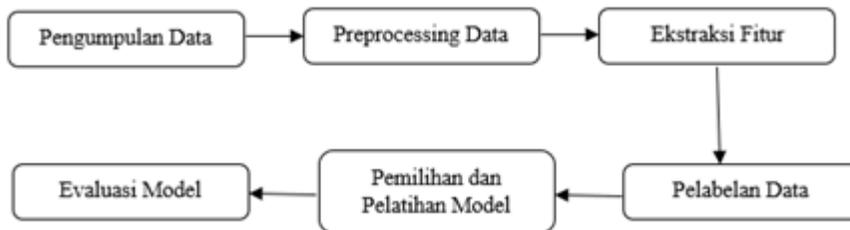
menangani data kompleks dan menghasilkan klasifikasi yang akurat, menjadikannya pilihan populer dalam berbagai aplikasi klasifikasi sentimen.

2.4 Implementasi Analisis Sentimen pada pengguna *Youtube*

Youtube sebagai platform berbagi video memungkinkan pengguna untuk memberikan umpan balik dalam bentuk komentar. Setelah itu, komentar dibersihkan dari tanda baca, angka, dan kata-kata tidak relevan. Kemudian, teks komentar diubah menjadi vektor numerik menggunakan TF-IDF. Model SVM dilatih dengan data yang sudah diberi label sentimen (positif, negatif, atau netral) dan dievaluasi menggunakan metrik seperti akurasi dan presisi. Setelah model dievaluasi, digunakan untuk memprediksi sentimen dari komentar baru. Hasil prediksi dianalisis untuk memahami reaksi penonton terhadap video, membantu pembuat konten dalam membuat keputusan strategis yang lebih baik.

3. Metodologi Penelitian

Bagian ini menjelaskan enam langkah yang digunakan untuk menganalisis komentar-komentar, yang terdiri dari pengumpulan data, *preprocessing*, ekstraksi fitur, pelabelan data, pemilihan dan pelatihan model svm, dan evaluasi model, yang dijelaskan lebih lanjut dalam sub bab berikutnya.



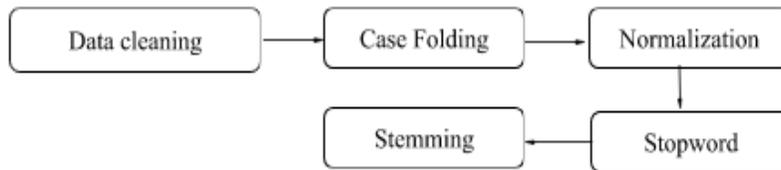
Gambar 1. Tahap Penelitian

3.1 Pengumpulan Data

Data diperoleh dari video *Youtube* membahas UKT naik. Komentar-komentar dikumpulkan menggunakan *Nethtytic* sebagai alat bantu pengumpulan data. Pengumpulan data dilakukan dengan memanfaatkan fitur dari *Nethtytic* untuk mengunduh data komentar dari *Youtube* secara otomatis. Jumlah data yang berhasil terkumpul mencapai 121 komentar. Untuk *id*, *author*, *description*, *guid*, *to*, *likecount*, *link*, *pubdate*, *replycount*, *title*, dan *authorChannelUrl* adalah semua data yang dikumpulkan. Dengan proses pengumpulan data, diharapkan dapat dilakukan analisis sentimen untuk memahami pendapat dan reaksi masyarakat terhadap kenaikan UKT di perguruan tinggi.

3.2 Preprocessing

Pada tahap ini *preprocessing* atau *text preprocessing* yakni proses memberikan serta membuat data agar lebih terstruktur dan bersih sebelum diolah nantinya (Fitri, Andreswari, & Hasibuan, 2019). Proses ini terdiri dari beberapa tahapan, seperti yang ditunjukkan pada gambar 2.



Gambar 2. Tahap *Preprocessing*

Tahapan pada *preprocessing* ada 5 proses yang dilalui **data cleaning**, **case folding** untuk mengubah karakter menjadi huruf kecil (*lower case*), **normalization** untuk proses mengubah teks menjadi bentuk normal, **stopword** untuk mengurangi teks yang kurang relevan, dan **stemming** sebagai pengurangan variasi kata yang sama hal ini dapat membantu dalam meningkatkan keefisienan analisis untuk peneliti.

3.3 Ekstraksi Fitur

Setelah *preprocessing*, fitur-fitur penting dari teks komentar diekstraksi menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk mengukur seberapa penting kata dalam sebuah dokumen dibandingkan dengan kumpulan dokumen. Nilai tinggi TF-IDF menunjukkan bahwa kata tersebut memiliki signifikansi tinggi dalam dokumen (Saputra, Subhi, & Winatama, 2019).

3.4 Pelabelan Data

Pada tahap ini, proses dilakukan secara manual menggunakan *Google Spreadsheet* dengan menambahkan fitur “sentimen” pada setiap data dan memberikan label sesuai dengan sentimennya, baik itu positif, negatif, atau netral. Langkah ini memiliki peran penting dalam pelatihan model nantinya karena membantu dalam mengklasifikasikan data berdasarkan sentimennya.

3.5 Pemilihan dan Pelatihan Model

Dalam proses pemilihan dan pelatihan model untuk analisis sentimen terhadap Kenaikan UKT di *YouTube* dipilih menggunakan metode *Support Vector Machine* (SVM) karena kemampuannya yang baik dalam menangani data berdimensi tinggi dan performa yang unggul dalam berbagai tugas klasifikasi. Proses ini dimulai dengan memisahkan dataset menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk membangun model, sementara data uji digunakan untuk mengukur kinerja model.

Selama pelatihan, model SVM dilatih menggunakan data latih yang telah diberi label sentimen (positif, negatif, atau netral).

Setelah model dilatih, dilakukan optimasi *hyperparameter*, yaitu menyesuaikan parameter-parameter penting dalam SVM seperti *kernel*, *C*, dan *gamma* untuk meningkatkan kinerja model. Pemilihan *hyperparameter* yang tepat sangat penting karena dapat mempengaruhi kemampuan model dalam mengklasifikasikan data dengan akurasi tinggi. Setelah optimasi, model diuji menggunakan data uji untuk mengevaluasi kinerjanya. Berbagai matrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* digunakan untuk mengukur seberapa baik model dalam mengklasifikasikan sentimen. Dengan demikian, proses pemilihan dan pelatihan model ini memastikan bahwa model SVM yang digunakan mampu memberikan hasil yang akurat dan andal dalam analisis sentimen komentar penonton terhadap Kenaikan UKT.

Selain itu, metode SVM, yang merupakan algoritma klasifikasi, menggunakan persamaan matematika untuk menemukan hyperplane optimal yang memisahkan data ke dalam berbagai kelas. Untuk menghitung hyperplane dalam klasifikasi Mesin Support Vector Linear, digunakannya rumus persamaan (1):

$$w \cdot x_i + b = 0 \quad (1)$$

Persamaan (1), digunakan untuk mencari *hyperplane* pada svm dengan memisahkan dua kelas dengan margin maksimum. w merupakan vektor bobot, x adalah vektor fitur, dan b merupakan bias (Qosim, 2021).

3.6 Evaluasi Model

Dalam evaluasi model Support Vector Machine (SVM) yang digunakan untuk analisis sentimen Kenaikan UKT di *YouTube*, beberapa metrik penting diterapkan untuk memastikan kinerja model yang optimal. Dibawah ini adalah penjelasan tentang akurasi, presisi, dan *recall*:

a. Akurasi.

Perbandingan antara data yang diklasifikasikan dengan benar dibandingkan dengan keseluruhan data disebut akurasi (Khaerullah, Suarna, & Nurdiawan, 2023). Rumus untuk menghitung akurasi adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

b. Presisi

Presisi adalah rasio antara prediksi positif yang benar dengan jumlah keseluruhan prediksi positif (Prasetya, Ali, Rohmat, & Nurdiawan, 2020). Rumus untuk menghitung presisi adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

c. *Recall*

Recall adalah rasio antara prediksi positif yang benar dengan total data yang benar-benar positif (Syafi'i, Nurdiawan, & Dwilestari, 2022). Rumus untuk menghitung *recall* adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

d. *F1-Score*

F1-Score adalah perbandingan presisi dan *recall* rata-rata yang dibobotkan. Persamaan untuk menghitungnya adalah sebagai berikut:

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

Untuk memastikan bahwa evaluasi model dilakukan secara menyeluruh, teknik *K-Fold Cross-Validation* diterapkan. Teknik ini melibatkan pembagian dataset menjadi K bagian (*folds*), melatih model pada K-1 bagian, dan menguji pada bagian yang tersisa. Proses ini diulang K kali dengan bagian pengujian yang berbeda setiap kali, dan hasilnya dirata-ratakan untuk mendapatkan estimasi performa yang lebih stabil.

Selain itu, Grid Search dan Cross-Validation digunakan untuk mengidentifikasi kombinasi parameter SVM terbaik, seperti parameter C dan tipe *kernel*. Ini memastikan bahwa model bekerja pada parameter terbaik yang memungkinkan kinerja maksimal dalam klasifikasi sentimen.

True positives (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN) adalah semua contoh klasifikasi model yang diberikan oleh *Confusion Matrix*. Ini membantu dalam mengidentifikasi pola kesalahan dan area untuk perbaikan.

4. Hasil dan Pembahasan

4.1 Preprocessing Data

Preprocessing data merupakan proses membersihkan dan mempersiapkan data awal untuk analisis. Data dibuat setelah data diproses, terutama fitur "description" seperti Tabel 1.

Tabel 1. Data hasil *preprocessing* data

No	Sebelum	Sesudah
1.	HIDUP ORANG KAYA!!!! 🎉	hidup orang kaya
2.	Kebabasan kampus, ... Kenapa para petinggi d negara ini diam... Ada apa...?	kebabasan kampus kenapa para petinggi negara ini diam

3.	TIDAK ADA CARA LAIN SELAIN MEMBUAT ANARKIS	tidak ada acara lain buat anarkis
4.	@@darno234 kalau menurut pandangan saya sih jokowi nya udah bener ...cuma mentri mentri nya yang gak bener dableg kabeh....	pandangan jokowi bener cuma mentri bener dableg kabeh
5.	Pengen kuliah di kampus2 populer, tapi ga mau bayar mahal. Hukum pasar berlaku bro. Kampus murah masih banyak, kualitasnya juga bersaing.	pengen kuliah kampus populer ga mau bayar mahal hukum pasar berlaku bro kampus murah masih banyak kualitas bersaing

Pada Preprocessing data seperti Tabel 1, Bertujuan untuk meningkatkan kualitas data asli agar saat menganalisis mendapatkan hasil yang akurat. Selain itu, proses ini juga melibatkan penghapusan data yang tidak relevan seperti menghilangkan stiker, tanda baca, dan angka pada komentar masyarakat di *YouTube*. Dengan menggunakan tahapan ini data akan menjadi lebih jelas, bersih dan terstruktur.

4.2 Pelabelan Data

Pada Tabel 2, data yang sudah dilakukan *processing* dilakukan pelabelan data sesuai komentar yang diberikan seperti komentar yang memiliki makna positif, negatif dan netral.

Tabel 2. Data yang sudah dilabeli

	author	description	pubdate	sentimen
0	@MELALIS4	butuh pendidikan gratis bukan makanan gratis	6/5/2024 20:12	negatif
1	@amni.z0r0	ukt mahal/naik bukan ulah dosen. jangan dinara.	6/5/2024 3:45	negatif
2	@hrwstudio4104	mas mentri.. mas mentri.. segera mundur	6/1/2024 9:47	negatif
3	@chowmariana8197	ukt tahun naik, bagaimana tahun depan, terus t...	6/1/2024 0:25	negatif
4	@chowmariana8197	waduhhhh... masa sih putra putri bangsa yg mau.	6/1/2024 0:15	negatif
...
113	@severinusmaha4020	undang undang diobrak abrik uang negara habis	5/25/2024 1:02	negatif
114	@siehongsiehon g7098	tadi nongkrong warung kopi, mabar game mobile	5/25/2024 0:26	negatif
117	@UNKHOWN-mn9ro	bila mafia campuri urusan pendidikan.lalu apa	5/24/2024 23:41	negatif
118	@therotras5881	negara berbisnis rakyat, kementerian 34-38 mas...	5/24/2024 23:21	negatif
120	@azwarannas6297	pengen kuliah kampus2 populer, ga mau bayar ma.	5/24/2024 22:47	positif

Adapun fungsi dari pelabelan data pada sentimen agar dapat membantu penelitian dalam memahami persepsi komentar masyarakat di *youtube* terhadap kenaikan UKT di tahun 2024.

4.3 Ekstraksi Fitur

Proses mengubah data teks mentah menjadi representasi numerik yang dapat digunakan dalam model pembelajaran mesin dikenal sebagai ekstraksi fitur. Proses ini sangat penting dalam menganalisis sentimen, terutama menganalisis topik mengenai kenaikan Uang Kuliah Tunggal (UKT) di *Youtube*. Penelitian ini mengumpulkan dan memproses komentar pengguna di *Youtube* untuk mengekstrak informasi yang relevan, kemudian digunakan untuk mengidentifikasi pola sentimen seperti positif, negatif atau netral. Hasil analisis ini memberikan wawasan tentang reaksi publik terhadap kebijakan kenaikan UKT, membantu pemerintah maupun institusi pendidikan memahami persepsi serta kekhawatiran orang tua dan mahasiswa. Dengan menerapkan ekstraksi fitur yang tepat, penelitian ini mengubah data tidak terstruktur menjadi informasi yang mendukung analisis yang akurat dan informatif terkait informasi sentimen masyarakat terhadap isu pendidikan. Dapat dilihat juga pada Gambar 3 untuk hasil dari ekstraksi fitur yang telah dilakukan.

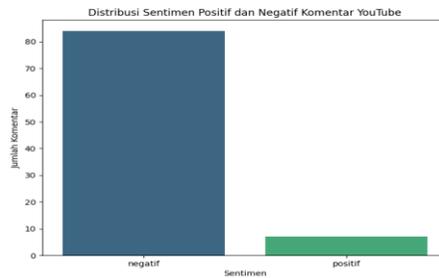
```

Dimensi matriks fitur: (121, 961)
Beberapa fitur (kata unik) yang diekstrak: ['abis' 'abrik' 'action' 'adakah' 'adilan' 'aetinya' 'agama' 'aj' 'aja'
'ajaib']
Nilai TF-IDF dari beberapa dokumen dan fitur:
  abis  abrik  action  adakah adilan  aetinya  agama  aj  aja  ajaib
0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
4  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
    
```

Gambar 3. Hasil dari Ekstraksi fitur

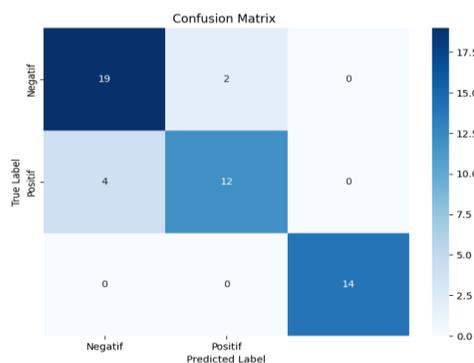
4.4 Klasifikasi dan Evaluasi

Proses ini menggunakan 90 data yang telah diproses sebelumnya, terdiri dari 87 data dengan sentimen negatif dan 7 data dengan sentimen positif. Data-data ini kemudian digunakan sebagai bahan latihan untuk mengembangkan model. Selain itu, pada prose penelitian Klasifikasi dan evaluasi pada data hasil oversampling dengan menggunakan SMOTE dalam analisis sentimen bertujuan untuk meningkatkan kinerja model dalam mengidentifikasi sentimen dari teks yang tidak seimbang secara kuantitas antara kelas positif, negatif, dan netral. dapat dilihat pada Gambar 4. untuk data hasil *oversampling* menggunakan SMOTE.



Gambar 4. Data hasil *oversampling* menggunakan SMOTE

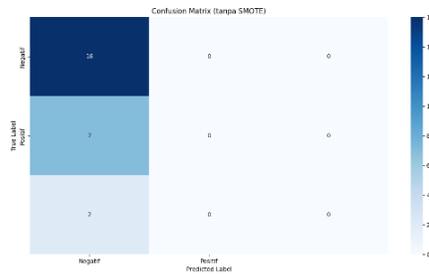
Pengembangan model SVM dimulai dengan melakukan *oversampling* pada data untuk mengatasi ketidakseimbangan kelas, sehingga dapat membantu meningkatkan kinerja model, terutama pada kelas minoritas yaitu memiliki jumlah sampel yang lebih sedikit dari kelas seperti sentimen positif. Selain itu, dengan menggunakan SMOTE juga dapat melihat perbedaan dalam performa model setelah *oversampling* dapat dilihat di Gambar 4.



Gambar 5. *Confusion Matrix* menggunakan SMOTE

Pada Gambar 5, Menjelaskan mengenai nilai *confusion matrix* dari sentimen positif dan sentimen negatif. Pada gambar *confusion matrix* tersebut terdapat nilai sebesar 19 pada *predicted label* dan *true label* serta berwarna biru gelap yang menandakan bahwa nilai 19 pada sentimen analisis merupakan sentimen negatif dan telah berhasil diprediksi sebagai negatif oleh model.

Selanjutnya, pada Gambar 6 memperlihatkan hasil *confusion matrix* tanpa SMOTE dari sentimen positif dan sentimen negatif. Pada gambar tersebut terdapat nilai sebesar 16 untuk *true label* dan *predicted label* dengan warna biru gelap yang menandakan bahwa hasil SVM tanpa menggunakan SMOTE.



Gambar 6. SVM tanpa menggunakan SMOTE

Selain *Confusion matrix* dapat dilihat pada Tabel 3, yang menunjukkan nilai *Precision*, *Recall*, *F1-Score* dan *Accuracy* algoritma SVM menggunakan SMOTE dalam sentimen masyarakat terhadap kenaikan biaya UKT 2024:

Tabel 3. Tabel *Precision*, *Recall*, *F1-Score* dan *Accuracy*

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negatif</i>	0.83	0.90	0.86	21
<i>Netral</i>	0.00	0.00	0.00	16
<i>Positif</i>	1.00	1.00	1.00	14
<i>Accuracy</i>	-	-	0.88	51
<i>Macro Avg</i>	0.89	0.88	0.89	51
<i>Weighted Avg</i>	0.41	0.88	0.80	51

Pada Tabel 3. Menjelaskan bahwa hasil dari klasifikasi dan evaluasi menunjukkan kinerja yang solid. Model memiliki tingkat *precision* yang tinggi untuk kelas positif mencapai 1.00 yang menunjukkan bahwa mayoritas prediksi positif yang dibuat oleh model benar. Namun, *precision* untuk kelas negatif sedikit lebih rendah yaitu 0.83, menunjukkan bahwa ada beberapa kesalahan dalam memprediksi kelas negatif. Selain itu, model juga memiliki tingkat *recall* yang tinggi untuk kelas positif mencapai 1.00 yang menunjukkan bahwa model dapat mendeteksi sebagian besar *instance* yang sebenarnya positif. Namun, hasil *recall* untuk kelas negatif menunjukkan sedikit lebih rendah mencapai 0.90 yang berarti ada beberapa *instance* negatif yang tidak terdeteksi oleh model. *F1-score* menunjukkan hasil yang baik dan merupakan rata-rata harmonis dari *precision* dan *recall*, terutama untuk kelas positif mencapai 1.00 sehingga sedikit lebih rendah untuk kelas negatif mencapai 0.86. Sedangkan, hasil Akurasi keseluruhan model adalah 0.88 yang menunjukkan bahwa model memberikan prediksi yang benar untuk sebagian besar *instance*. Meskipun demikian, terdapat ruang untuk meningkatkan kinerja model terutama dalam memprediksi kelas negatif untuk mencapai keseimbangan yang lebih baik antara *precision* dan *recall* untuk kedua kelas. Selain itu, dapat dilihat pada Tabel

4, yang menunjukkan nilai *Precision*, *Recall*, *F1-Score* dan *Accuracy* algoritma SVM tanpa menggunakan SMOTE dalam sentimen masyarakat terhadap kenaikan biaya UKT 2024:

Tabel 4. Tabel *Precision*, *Recall*, *F1-Score* dan *Accuracy* tanpa SMOTE

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
<i>Negatif</i>	0.64	1.00	0.78	16
<i>Netral</i>	0.00	0.00	0.00	7
<i>Positif</i>	0.00	0.00	0.00	2
<i>Accuracy</i>	-	-	0.64	25
<i>Macro Avg</i>	0.21	0.33	0.26	25
<i>Weighted Avg</i>	0.41	0.64	0.50	25

Tabel 4. menunjukkan berbagai hasil model support vector machine (SVM) yang digunakan untuk menganalisis sentimen komentar terkait kenaikan Uang Kuliah Seragam (UKT). Model ini menunjukkan kinerja yang baik dalam mendeteksi komentar negatif dengan presisi 0,64, recall 1,00, dan skor F1 0,78. Namun untuk kelas netral dan positif, nilai presisi, recall, dan skor F1 semuanya 0,00 yang menunjukkan adanya kesalahan total dalam klasifikasi. Akurasi keseluruhan model adalah 0,64. Presisi rata-rata makro (0,21, recall 0,33, skor F1 0,26) dan rata-rata tertimbang (presisi 0,41, recall 0,64, skor F1 0,50) menunjukkan kinerja yang tidak konsisten dalam mengklasifikasikan kelas netral dan positif. Untuk meningkatkan performa, diperlukan penggunaan teknik penyeimbangan data atau penyesuaian parameter model lebih lanjut.

5. Kesimpulan

Kesimpulan penelitian ini menunjukkan bahwa metode support vector machine (SVM) dengan menggunakan oversampling SMOTE berhasil meningkatkan akurasi klasifikasi sentimen terkait kenaikan Uang Kuliah Tunggal (UKT) tahun 2024 di YouTube. Berdasarkan evaluasi, model ini memiliki kinerja yang baik dengan akurasi 0,88. Model ini dapat memprediksi emosi positif, negatif, dan netral dengan baik. Untuk emosi negatif, model memiliki presisi 0,83, recall 0,90, dan skor F1 0,86, yang menunjukkan kemampuan yang baik dalam mendeteksi emosi negatif. Namun, ada juga beberapa contoh netral yang salah diprediksi sebagai negatif. Namun, untuk emosi positif, model mencapai presisi, perolehan, dan skor F1 sebesar 1,00, yang menunjukkan kemampuan yang baik dalam mengidentifikasi emosi positif. Secara keseluruhan, penggunaan SMOTE efektif dalam mengatasi ketidakseimbangan kelas dan meningkatkan akurasi prediksi emosi, serta memberikan wawasan berharga mengenai persepsi masyarakat terhadap kebijakan pendidikan ini. Model SVM ini dapat digunakan untuk menganalisis sentimen terkait kenaikan UKT di platform YouTube, dengan anotasi untuk lebih meningkatkan pendeteksian sentimen netral.

6. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penelitian ini.

7. Pernyataan Penulis

Kami selaku penulis, menekankan bahwa penelitian ini dilakukan secara cermat dan menyeluruh untuk memahami pandangan masyarakat terhadap kenaikan Uang Kuliah Tunggal (UKT) melalui platform YouTube. Data yang kami analisis melalui proses pra-pemrosesan yang menyeluruh untuk memastikan keakuratannya. Pada penelitian ini, dengan menggunakan teknik support vector machine (SVM) dan teknik oversampling SMOTE, kami mampu mencapai hasil yang memuaskan dalam klasifikasi emosi positif, negatif, dan netral. Kami berharap hasil penelitian ini dapat memberikan kontribusi yang berarti terhadap kebijakan pendidikan dan pengambilan keputusan di lembaga pendidikan. Kesimpulan kami didasarkan pada analisis data yang obyektif dan metodologi yang telah terbukti serta memberikan dasar untuk pengembangan lebih lanjut teknik analisis sentimen yang meningkatkan akurasi dan keandalan prediksi di masa depan.

Penulis menyatakan bahwa tidak ada konflik kepentingan terkait publikasi artikel ini. Penulis menyatakan bahwa data dan makalah bebas dari plagiarisme serta penulis bertanggung jawab secara penuh atas keaslian artikel.

Bibliografi

- A. N. Yusuf, E. Supriyati, and T. Listyorini. 2020. "Analisis Sentimen Mengenai Layanan Provider Indihome Berdasarkan Pendapat Pelanggan Melalui Media Sosial Twitter dengan Metode Naïve Bayes Classifier," vol.4, pp. 75–78
- A. Ropikoh, R. Abdulkhakim, U. Enri, and N. Sulistiyowati, "Penerapan Algoritme Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Abdusyukur, Fatwa. (2023). Penerapan Algoritma *Support Vector Machine* (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter. *Jurnal Ilmiah Komputer dan Informatika*, 12(1), 1-10. <https://ojs.unikom.ac.id/index.php/komputa/article/view/9418/3696>
- Ardiansyah, M. Suharto, T. Farid, Ahmad, S. (2022). "Upaya Penanganan Uang Kuliah Tunggal Bermasalah Bagi Mahasiswa Yang Tidak Mampu Pada Perguruan Tinggi", *Jurnal Ilmiah Ilmu Pendidikan*.

- E. Yosef, S. Sihombing, and B. T. Hanggara, “Pemanfaatan API Youtube Dalam Pengembangan Aplikasi Portal Video Penangkaran Kenari Untuk Peternak Kenari Berbasis Android,” Vol. 4, No. 7, pp. 2067–2074, 2020.
- F. A. Nugraha, N. H. Harani, and R. Habibi, *Analisis Sentimen Pembatasan Sosial Menggunakan Deep Learning*, 1st ed. Bandung: Kreatif, 2020.
- Mualfah, D., Ramadhoni., Gunawan, R., Suratno, D. M.(2023). Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM. *Jurnal FASILKOM*,13(1) ,72-80. <https://ejournal.umri.ac.id/index.php/JIK/article/view/4920/2592>
- N. C. Dang, M. N. Moreno-García, and F. de la Prieta, “*Sentiment analysis based on deep learning: A comparative study*,” *Electronics* (Switzerland), vol. 9, no. 3, Mar. 2020, doi:10.3390/electronics9030483
- O. H. Rahman, Abdillah Gunawan & Komarudin Agus, “Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine”, *Jurnal Rekayasa Sistem dan Teknologi Informasi (RESTI)*, vol. 5, no. 1, pp. 17 – 23, 2021
- P. Y. Saputra, D. H. Subhi, and F. Z. A. 2019. Winatama, “Implementasi Sentimen Analisis Channel Video Pelayanan Pemerintah Di YouTube Menggunakan Algoritma Naïve Bayes,” *J. Inform. Polinema*, vol. 5, no. 4, pp.209–213, doi:10.33795/jip.v5i4.259
- Qosim, Ahmad. 2021, “Perbandingan Metode Klasifikasi Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC) Untuk Menentukan Kualitas Udara”, *Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim*.
- R. R. Khaerullah, N. Suarna, and O. Nurdiawan, “Analisa Pengelompokan Dataset Komputer Menggunakan Algoritma X-Means,” *J. Inform.dan Teknol. Inf.*, vol. 1, no. 2, pp. 125–132, 2023,doi: 10.56854/jt.v1i2.135.
- Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naive Bayes Classifier (NBC), dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2), D177–D184.I.
- Syafi'i, O. Nurdiawan, and G. Dwilestari, “Penerapan Machine Learning Untuk Menentukan Kelayakan Kredit Menggunakan Metode Support Vektor Machine,” *J. Sist. Inf. dan Manaj.*, vol. 10, no. 2, pp. 1–6, 2022
- T. Prasetya, I. Ali, C. L. Rohmat, and O.Nurdiawan, “Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K- Nearest Neighbor,” *INFORMATICS Educ. Prof.J. Informatics*, vol. 5, no. 1, p. 93, 2020, doi:10.51211/itbi.v5i1.1431
- V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “*Sentiment Analysis of Social Media Twitter with Case of Anti-LGBT Campaign In Indonesia Using NaïveBayes, Decision Tree, and*

Random Forest Algorithm,”*Procedia Comput. Sci.*, Vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.

Z. Hao, L. Shaohong, and S. Jinping, “Unitmodel of binary SVM with DS output and its application in multi-class SVM,” *Proc. – 2011 4th Int. Symp. Comput. Intell. Des. Isc.* 2011, vol.1, no. 2, pp. 101–104, 2011, doi:10.1109/ISCID.2011.34