

Analisis Sentimen Pada Media Sosial Menggunakan Metode *Support Vector Machine*

Muhamad Hadi Arfian*, Hendrian Sofu Maruhawa, Aulia Aisyah Putri, Dinah Ratulugina, Ellsza Ridzky Khoirunnisa, Verell Hermawan, Ahmad Julaibib Syahr

Teknik Informatika, Universitas Esa Unggul, Jakarta
 *muhamad.arfian@esaunggul.ac.id

Abstrak — Pada era kemajuan teknologi digital saat ini, media sosial telah menjadi platform utama bagi individu untuk berbagi opini dan pengalaman mereka, yang dikenal sebagai sentimen. Sentimen ini memberikan wawasan berharga tentang berbagai topik. Penelitian ini fokus pada analisis sentimen terhadap penerima Kartu Indonesia Pintar-Kuliah (KIP-K) di X. Program KIP-K bertujuan untuk meningkatkan akses pendidikan tinggi bagi masyarakat dari berbagai latar belakang ekonomi, dan perhatian masyarakat terhadap penerima program ini semakin meningkat. Dengan menggunakan algoritma Support Vector Machine (SVM), yang terbukti memiliki akurasi tinggi dalam analisis sentimen dibandingkan metode lain, penelitian ini menganalisis tanggapan publik untuk memahami persepsi mereka terhadap penerima KIP-K. Hasil menunjukkan bahwa mayoritas sentimen di X adalah negatif, mengindikasikan ketidaksetujuan terhadap penerima program ini. Temuan ini memberikan wawasan tentang persepsi masyarakat dan dapat membantu dalam evaluasi kebijakan pendidikan.

Article History:

Received: August 8, 2024
 Revised: September 9, 2024
 Accepted: September 13, 2024
 Published: January 21, 2025

Kata Kunci — Analisis Sentimen; KIPK; Machine Learning; Support Vector Machine

DOI: 10.22441/jitkom.v9i1.001

I. PENDAHULUAN

Dalam beberapa tahun terakhir, penggunaan media sosial di Indonesia telah mengalami peningkatan yang signifikan, mencerminkan transformasi digital yang cepat dalam masyarakat [1]. Menurut laporan *We Are Social Digital 2024 Global Overview Report* yang dirilis pada Januari 2024, jumlah pengguna aktif media sosial di Indonesia mencapai 139 juta jiwa atau setara dengan 49,9% dari jumlah populasi di Indonesia [1]. Peningkatan penggunaan media sosial mempengaruhi penyebaran informasi dan opini publik. Kemudahan penulisan pesan dan status membuat media sosial populer di berbagai usia. Keberagaman usia pengguna menghasilkan berbagai jenis pesan seperti saran, kritik, dan komentar, yang memungkinkan observasi dan informasi lebih mendalam tentang topik di platform seperti X atau Facebook. Salah satu topik yang ramai dibicarakan di berbagai platform media sosial adalah permasalahan penerima KIPK. Analisis sentiment bisa digunakan untuk mengetahui opini publik tentang beasiswa KIPK.



Gambar 1. Post Viral KIPK 1



Gambar 2. Post Viral KIPK 2

Menurut data yang disampaikan oleh Badan Pusat Statistik pada tahun 2023, hanya 26,85% tingkat partisipasi sekolah penduduk Indonesia yang berusia antara 19 hingga 24 tahun (Badan Pusat Statistik, 2023). Pemerintah telah berupaya menangani tantangan besar terkait tingginya angka putus sekolah bagi siswa berprestasi serta mereka yang kesulitan menyelesaikan pendidikan mereka karena masalah keuangan. Salah satu langkah yang diambil pemerintah untuk mengatasi tingginya tingkat putus sekolah di kalangan siswa berprestasi dan mereka yang mengalami kesulitan finansial dalam menyelesaikan pendidikan adalah melalui program bantuan KIP Kuliah [2]. Program ini diperkenalkan pada tahun 2020 sebagai pengganti dari program beasiswa sebelumnya yang dikenal dengan nama Bidikmisi.

Tantangan dalam penentuan penerima KIP Kuliah yang tepat sasaran mendorong perlunya evaluasi ketat, dan analisis data serta umpan balik dari masyarakat membantu mengidentifikasi ketidaksesuaian dan memperbaiki proses seleksi tersebut [3]. Penelitian ini akan menerapkan n-grams dalam proses vektorisasi teks untuk meningkatkan akurasi model. Penggunaan n-grams bertujuan untuk menangkap konteks yang lebih luas dari teks, yang diharapkan dapat meningkatkan performa prediksi dibandingkan dengan metode sebelumnya yang tidak memanfaatkan fitur ini [4]. Dengan pendekatan ini, diharapkan hasil penelitian dapat memberikan wawasan yang lebih komprehensif tentang sentimen publik terhadap kebijakan work from home, serta menyediakan referensi yang lebih luas bagi pemerintah dan pihak terkait [5].

II. LITERATURE REVIEW

Studi literatur melibatkan pencarian teori-teori dari berbagai jurnal dan buku yang relevan dengan penelitian yang akan dilakukan, dengan tujuan memperoleh dasar teoritis yang sesuai dengan permasalahan yang akan diselidiki. Analisis sentimen

adalah metode untuk mengkaji opini publik, yang memungkinkan kita memahami suatu topik dari berbagai perspektif [6]. Dalam pelaksanaan analisis sentimen, beberapa aspek penting harus diperhatikan seperti topik yang dianalisis, sumber data yang relevan, dan algoritma yang digunakan untuk analisis. Salah satu algoritma yang dapat digunakan adalah *Support Vector Machine* (SVM), yang saat ini menunjukkan kinerja yang cukup baik dalam klasifikasi kalimat [7]. Metode SVM yang ditunjang dengan *feature extraction* dapat diadaptasi untuk analisis sentimen, menghasilkan tingkat akurasi yang jauh lebih tinggi dibandingkan dengan algoritma lain seperti *Naïve Bayes* [8]. Dengan mengimplementasikan algoritma SVM, sistem dapat memproses dan memahami teks dari media sosial dengan lebih baik, menangani nuansa bahasa dan konteks, serta memberikan hasil klasifikasi sentimen yang lebih akurat [9].

Penelitian terkait analisis sentimen pada algoritma *Support Vector Machine* (SVM) sudah dilakukan, topik yang digunakannya juga bermacam-macam. Yakni Tokoh Publik, Aplikasi Transportasi Online, Maskapai Penerbangan, Calon Presiden, dll. Penelitian ini terinspirasi oleh studi "*Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization*" yang menganalisis sentimen terhadap kebijakan work from home selama pandemi COVID-19 di Indonesia dengan menggunakan metode *Support Vector Machine* (SVM) yang dioptimalkan melalui *Randomized Search Optimization* [10]. Studi tersebut menunjukkan bahwa mayoritas netizen mendukung kebijakan tersebut dan berhasil mencapai F1 Score sebesar 83.362% dengan *preprocessing* yang mencakup ekspansi akronim, terjemahan kata slang, dan terjemahan emoji.

III. METODOLOGI PENELITIAN

Metode yang digunakan pada penelitian ini dibagi menjadi dua tahap, yaitu metode pengumpulan data dan metode analisis data.

A. Crawling Data

Crawling data diambil dari twitter dengan katan kunci 'KIPK'. Data yang diambil berupa teks tweet dari pengguna twitter. Data diambil dari bulan April 2024 sampai dengan bulan juni 2024 berjumlah 2948 tweet.



Gambar 3. Hasil Crawling

B. Tweets Labeling

Setelah proses crawling data, tahap selanjutnya adalah pelabelan tweet secara manual. Pada tahap ini, setiap tweet yang telah dikumpulkan melalui proses *crawling* diberi label berdasarkan kategori tertentu yang relevan dengan penelitian.

Pelabelan manual ini penting untuk memastikan bahwa data memiliki kategori yang jelas dan akurat, sehingga dapat digunakan secara efektif dalam tahap selanjutnya[11].

full_text	label
Hoalah yg mampu ngapain sih minta klpk cok	negatif
@undpmerless 1 hari 1 penerima KIPK yang di spill	netral
anak klpk mana yg pake macbook anjrit???? no...	negatif
@undpmerless lho korban ki yo seng kudune oie...	negatif
guys serius trya ada klpk s2 ga ya? rela deh g...	negatif
taporin aja ini masak kullun yg bakemnya gaj...	negatif
Ngga semua penerima klpk hedon dengan uang...	positif
@undpmerless klo yg viral di bali sadar semua...	negatif
@hachway @dabstip @terapayoko @tanyafid di kampus...	netral
@hachway @dabstip @terapayoko @tanyafid di kampus...	netral
Guo benci lagi ngafat gilaan soalnya temen gua...	negatif
@undpmerless mau bangent ngeri yg sial KIPK...	negatif
@undpmerless temen aku jagg klpk tranya hp coach...	netral
@hachway @dabstip @terapayoko @tanyafid di kampus...	netral
@undpmerless Wkwk penerima KIPK eheh bgi...	negatif
@hachway @dabstip @terapayoko @tanyafid di kampus...	negatif
@undpmerless Gen aik no sahy mistal ni ada...	netral
Gu yg anak swasta non klpk borak? beli phone...	negatif
tidak membenarkan penilaisannya ngambil uang...	negatif
@undpmerless @undpmerless Temen aja yg mndm...	netral
@kemasukshelby @hachway @undpmerless Bulyng...	netral
@undpmerless ckrick manenya ya sial penerima...	negatif

Gambar 4. Contoh Pelabelan Manual

C. Text Preprocessing

Data set dimodifikasi melalui *text preprocessing* untuk membuatnya lebih mudah dikenali oleh sistem komputer. Proses ini juga mengubah data yang awalnya tidak teratur menjadi lebih rapi dan terstruktur. Tahapan dalam *text preprocessing* meliputi enam proses yaitu *cleaning*, *stopwords*, *tokenisasi*, *stemming*, *normalisasi*, dan *wordcloud*[11].

full_text	label	word count	char_count	avg_word	stopwords	hashtags	numeric	upper	cleaned
Hoalah yg mampu ngapain sih minta klpk cok	0	8	43	4.375000	2	0	0	0	hoalah yg mampu ngapain sih minta klpk cok
@undpmerless 1 hari 1 penerima KIPK yang di s...	1	9	50	4.666667	3	0	2	1	hari penerima klpk yang di spill
anak klpk mana yg pake macbook anjrit???? no...	0	27	140	4.185185	6	0	0	0	anak klpk mana yg pake macbook anjrit notebook hp g...
@undpmerless lho korban ki yo seng kudune oie...	0	32	168	4.281250	3	0	0	2	lho korban ki yo seng kudune oie hak tp diisak ka...
guys serius trya ada klpk s2 ga ya? rela deh g...	0	44	229	4.227273	8	0	0	0	guys serius trya ada klpk s2 ga ya? rela deh g...

Gambar 5. Hasil Cleaning

Tahap pertama adalah *cleaning*, atribut yang tidak diperlukan dan tidak bermakna, seperti *mention*, *hashtag*, *indikator retweet*, *URL*, karakter yang diulang tiga kali atau lebih, karakter non-alfanumerik kecuali spasi, dihapus. Selain itu, beberapa spasi yang berurutan diganti dengan satu spasi tunggal, dan spasi di awal serta akhir teks dihapus.

full_text	label	word count	char_count	avg_word	stopwords	hashtags	numeric	upper	cleaned
Hoalah yg mampu ngapain sih minta klpk cok	0	8	43	4.375000	2	0	0	0	hoalah yg mampu ngapain sih minta klpk cok
@undpmerless 1 hari 1 penerima KIPK yang di s...	1	9	50	4.666667	3	0	2	1	hari penerima klpk yang di spill
anak klpk mana yg pake macbook anjrit???? no...	0	27	140	4.185185	6	0	0	0	anak klpk mana yg pake macbook anjrit notebook hp g...
@undpmerless lho korban ki yo seng kudune oie...	0	32	168	4.281250	3	0	0	2	lho korban ki yo seng kudune oie hak tp diisak ka...
guys serius trya ada klpk s2 ga ya? rela deh g...	0	44	229	4.227273	8	0	0	0	guys serius trya ada klpk s2 ga ya? rela deh g...

Gambar 6. Hasil Stopwords

Proses selanjutnya adalah *removal stopwords*, proses ini bertujuan untuk menghilangkan kata-kata yang sering digunakan namun tidak mempengaruhi sentimen suatu kalimat[12]. Dalam penelitian ini, penghapusan *stopword* menggunakan kamus dari corpus *stopword* berbahasa Indonesia.

full_text	label	word count	char_count	avg_word	stopwords	hashtags	numeric	upper	cleaned
Hoalah yg mampu ngapain sih minta klpk cok	0	8	43	4.375000	2	0	0	0	[hoalah, yg, ngapain, sih, klpk, cok]
@undpmerless 1 hari 1 penerima KIPK yang di s...	1	9	50	4.666667	3	0	2	1	[penerima, klpk, spill]
anak klpk mana yg pake macbook anjrit???? no...	0	27	140	4.185185	6	0	0	0	[anak, klpk, yg, pake, macbook, anjrit, notebook, hp, g...]
@undpmerless lho korban ki yo seng kudune oie...	0	32	168	4.281250	3	0	0	2	[lho, korban, ki, yo, seng, kudune, oie, hak, tp, de...]
guys serius trya ada klpk s2 ga ya? rela deh g...	0	44	229	4.227273	8	0	0	0	[guys, serius, trya, klpk, s, ga, ya, rela, de...]

Gambar 7. Hasil Tokenize

Selanjutnya, proses *tokenisasi* adalah memecah kalimat menjadi potongan-potongan kata atau token. Proses ini bertujuan untuk mengidentifikasi dan melacak asal kemunculan setiap kata dalam teks.

0	[hoalah, yg, ngapain, sih, klpk, cok]
1	[penerima, klpk, spill]
2	[anak, klpk, yg, pake, macbook, anjrit, notebo...]
3	[lho, korban, ki, yo, seng, kudune, hak, tp, s...]
4	[guys, serius, trya, klpk, s, ga, ya, rela, de...]

Name: cleaned, dtype: object

Gambar 8. Hasil Stemming

Proses selanjutnya *stemming*, proses mengubah kata-kata ke bentuk dasarnya dengan menghapus imbuhan atau akhiran yang tidak perlu. Proses ini bertujuan untuk menyederhanakan kata-kata sehingga variasi bentuk kata yang sama dapat dikelompokkan bersama dan memudahkan analisis teks.

full_text	label	word count	char_count	avg_word	stopwords	hashtags	numeric	upper	cleaned
Hoalah yg mampu ngapain sih minta klpk cok	0	8	43	4.375000	2	0	0	0	hoalah yang ngapain sih klpk cok
@undpmerless 1 hari 1 penerima KIPK yang di s...	1	9	50	4.666667	3	0	2	1	terima klpk spill
anak klpk mana yg pake macbook anjrit???? no...	0	27	140	4.185185	6	0	0	0	anak klpk yang pake macbook notebook hp g...
@undpmerless lho korban ki yo seng kudune oie...	0	32	168	4.281250	3	0	0	2	lho korban ki ya seng kudune hak tapi saud ka...
guys serius trya ada klpk s2 ga ya? rela deh g...	0	44	229	4.227273	8	0	0	0	guys serius trya klpk s2k ya rela aku la...

Gambar 9. Hasil Normalisasi

Kemudian proses *normalisasi* untuk mengubah teks ke dalam bentuk yang konsisten dan seragam. Proses ini melibatkan beberapa tindakan, seperti menghapus kata, dan menyamakan format teks (misalnya, mengubah singkatan atau variasi penulisan ke bentuk standar). Tujuannya untuk menghilangkan perbedaan format dan variasi dalam teks sehingga memudahkan analisis dan perbandingan data.

D. Feature Extraction TF-IDF

	Mean_TFIDF
anak	0.023837
dapat	0.023435
banget	0.022019
salah	0.021228
kamu	0.021149
nya	0.021024
emang	0.019523
tapi	0.017763
gimana	0.017699
sih	0.017604
kuliah	0.017246
pake	0.016851
hedonis	0.016300
tau	0.014371
uang	0.014287
tuh	0.014260
sasar	0.014205
butuh	0.013789
kaya	0.013730
hidup	0.013718

Gambar 10. Hasil Mean TF IDF 1

Pada tahap ini, teks yang telah diproses diubah menjadi representasi numerik dengan menghitung bobot masing-masing kata berdasarkan frekuensinya dalam dokumen serta seberapa umum kata tersebut dalam keseluruhan kumpulan dokumen.

E. Proses *Validation Cross*

Cross-validation scores: [0.59795571 0.64735945 0.61669506 0.64906303 0.64054514]
Average cross-validation score: 0.6303236797274276

Gambar 11. Validation cross

Nilai cross-validation bervariasi antara sekitar 59.80% hingga 64.91% pada tiap fold, yang menunjukkan bahwa performa model relatif stabil namun tidak terlalu tinggi.

Rata-rata akurasi model adalah 63.03%, yang memberikan gambaran bahwa model dapat memprediksi dengan benar sekitar 63% dari data dalam cross-validation. Ini membantu mengukur seberapa baik model akan bekerja pada data baru tanpa overfitting atau underfitting pada data pelatihan.

F. Pemodelan *Support Vector Machine*

Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin terawasi yang digunakan untuk klasifikasi [13]. Algoritma ini memanfaatkan kernel untuk memetakan data ke ruang berdimensi lebih tinggi agar data yang tidak dapat dipisahkan secara linier dapat diklasifikasikan. Penggunaan kernel memerlukan pengaturan parameter yang dapat disesuaikan sesuai kebutuhan, yang disebut *hyperparameters*. Dalam studi ini, diterapkan *kernel Radial Basis Function* (RBF). Dengan demikian, *hyperparameters* yang dioptimalkan adalah C dan gamma. Parameter C berfungsi sebagai parameter regulasi, sedangkan parameter gamma adalah koefisien yang digunakan dalam kernel RBF, polinomial, atau sigmoid. Selain itu n-grams juga dimanfaatkan untuk menunjang akurasi.

G. *Hyperparameter Tuning dengan Grid Search*

Hyperparameters adalah parameter yang ditetapkan sebelum proses pembelajaran dimulai. Parameter ini dapat disesuaikan dan secara langsung mempengaruhi performa model. Untuk menemukan *hyperparameters* yang paling optimal, berbagai strategi diperlukan untuk penyesuaian, dan salah satu cara yang paling praktis adalah dengan mencoba berbagai kombinasi *hyperparameters*. Seiring waktu, berbagai pendekatan telah diajarkan untuk mengoptimalkan *hyperparameters*, seperti *Grid Search* dan *Random Search*. *Grid Search* adalah metode tradisional untuk tuning *hyperparameters*. Metode ini melakukan pencarian menyeluruh melalui subset tertentu dari ruang *hyperparameter* sebuah algoritma pembelajaran. Algoritma *Grid Search* diuji menggunakan beberapa metrik kinerja, biasanya dengan metode *cross-validation* pada data pelatihan.

IV. HASIL DAN ANALISA

A. Optimasi *Hyperparameter* menggunakan *GridSearch CV*

```
from sklearn.svm import SVC
from sklearn.pipeline import Pipeline
```

```
from sklearn.model_selection import
GridSearchCV
```

Mengimpor SVC untuk klasifikasi menggunakan *Support Vector Machine*, *Pipeline* untuk menggabungkan langkah-langkah pemrosesan data dan pelatihan model dalam satu objek, serta *GridSearchCV* untuk melakukan pencarian hiperparameter secara menyeluruh dengan *cross-validation*. Alat-alat ini digunakan untuk membangun, menyusun, dan mengoptimalkan model *machine learning* secara efektif.

```
parameters = {
'vectorizer__min_df': [1, 5],
'vectorizer__max_df': [0.7, 0.8],
'vectorizer__ngram_range': [(1, 1), (1, 2)],
'svc__C': [0.1, 1, 10],
'svc__kernel': ['linear', 'rbf']
}
# Create the pipeline
pipeline = Pipeline([
('vectorizer',
TfidfVectorizer(sublinear_tf=True,
use_idf=True)),
('svc', SVC())
])
```

Mendefinisikan parameter untuk pencarian *hyperparameter* grid, termasuk pengaturan untuk *TfidfVectorizer* dan *SVC*. Parameter mencakup nilai untuk *min_df*, *max_df*, *ngram_range* pada *TfidfVectorizer*, serta nilai untuk C dan kernel pada *SVC*. Selain itu, kode ini membuat *pipeline* yang menggabungkan *TfidfVectorizer* untuk mengubah teks menjadi fitur numerik dan *SVC* untuk klasifikasi, memungkinkan pelatihan dan optimasi model secara terstruktur.

```
grid_search_vectorizer = GridSearchCV(
estimator=pipeline,
param_grid=parameters,
scoring='accuracy',
cv=5,
verbose=0
)
# Fit GridSearchCV
grid_search_vectorizer.fit(dftrain_copy['clea
ned'], dftrain_copy['label'])
# Print the best parameters and best score
print("Best Parameters:",
grid_search_vectorizer.best_params_)
print("Best Score:",
grid_search_vectorizer.best_score_)
```

Menggunakan *GridSearchCV* untuk mengoptimalkan model dengan mencoba berbagai kombinasi parameter pada pipeline yang mencakup *TfidfVectorizer* dan *SVC*. Proses ini melibatkan pelatihan model dengan data teks dan label, serta menggunakan *cross-validation* dengan 5-fold untuk memilih parameter terbaik berdasarkan akurasi.

```
import time
from sklearn import svm
from sklearn.metrics import classification_report
import
# Perform classification with SVM, kernel=rbf
and C=100
classifier_rbf = svm.SVC(kernel='rbf', C=100)
t0 = time.time()
classifier_rbf.fit(train_vectors,
dftrain_copy['label'])
t1 = time.time()
y_pred = classifier_rbf.predict(test_vectors)
t2 = time.time()
time_rbf_train = t1 - t0
time_rbf_predict = t2 - t1
```

Model *Support Vector Machine* (SVM) dengan *kernel Radial Basis Function* (RBF) dan parameter C=100, serta mengukur waktu yang diperlukan untuk pelatihan dan prediksi. Setelah melatih model dengan data fitur dan label pelatihan, waktu pelatihan dihitung, dan waktu prediksi diukur setelah model digunakan untuk memprediksi data fitur pengujian.

```
import warnings
warnings.filterwarnings('always')
```

Mengimpor pustaka warnings dan mengatur pengaturan peringatan dengan `warnings.filterwarnings('always')`, yang memastikan bahwa semua peringatan ditampilkan setiap kali muncul. Ini berguna untuk melihat dan menangani peringatan yang mungkin diabaikan secara default selama eksekusi kode.

```
#results
print("Training time: %fs; Prediction time:
%fs" % (time_rbf_train, time_rbf_predict))
report =
classification_report(dftrain_copy['label'],
y_pred, output_dict=True)
```

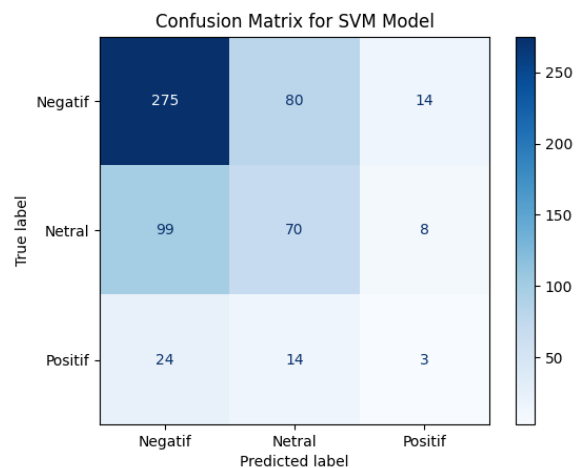
Mencetak waktu pelatihan dan prediksi model, serta menghasilkan laporan evaluasi klasifikasi yang membandingkan label sebenarnya dengan label yang diprediksi.

B. Evaluasi Model menggunakan SVM

Pengujian yang dilakukan pada 941 data menggunakan metode *Support Vector Machine* menghasilkan nilai akurasi sebesar 59%. Hasil evaluasi tersebut dapat dilihat lebih jelas melalui *confusion matrix* yang disajikan pada Tabel 1.

Tabel 1. Hasil Analisis Sentimen

	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Negatif	275	80	14
Netral	99	70	8
Positif	24	14	3



Gambar 12. Confusion Matrix

Dengan akurasi mencapai 0.59, dapat disimpulkan bahwa model klasifikasi sangat efektif dalam mengidentifikasi sentimen pada data pengujian. Untuk memahami performa metode klasifikasi dalam konteks setiap kelas, dapat dilihat pada nilai precision, recall, F1-score, dan support yang disajikan secara rinci pada Tabel 2.

Tabel 2. Hasil Analisis Sentimen

	Precision	Recall	F1-Score	Support
Negatif	0.69	0.75	0.72	369
Netral	0.43	0.40	0.41	177
Positif	0.12	0.07	0.09	41
Accuracy			0.59	587
Macro AVG	0.41	0.40	0.41	587
Weighted AVG	0.57	0.59	0.58	587

Meski akurasi adalah metrik utama dalam evaluasi, penting juga untuk mempertimbangkan metrik evaluasi lainnya seperti precision, recall, F1-score, dan support untuk mendapatkan gambaran yang lebih lengkap tentang kinerja model. Memahami sentimen pengguna memungkinkan untuk mengambil langkah-langkah dalam meningkatkan pengalaman pengguna dan menangani masalah yang mungkin muncul. Namun, mengingat adanya variasi dalam akurasi antara iterasi, ada kemungkinan bahwa model dapat ditingkatkan lebih lanjut melalui penyesuaian tambahan atau dengan menerapkan teknik pemrosesan teks yang lebih canggih.

V. KESIMPULAN

Berdasarkan hasil penelitian, penulis berhasil menerapkan Algoritma SVM yang ditunjang dengan ekstraksi feature TF-IDF untuk analisis sentimen terhadap data tweet dengan kata kunci "KIPK", "Mahasiswa KIPK", "Penerima KIPK". Implementasi ini terbukti efektif, dengan hasil akurasi pada data tweet yang berjumlah 941, di mana 587 data digunakan untuk pelatihan dan 354 data untuk pngujian. Kami menggunakan Hyperparameter Optimization dengan metode untuk mencari kombinasi terbaik dari hyperparameter untuk model machine

learning agar mendapatkan performa terbaik. Hasilnya menunjukkan bahwa *Kernel RBF* memiliki presisi terbaik. Dalam penelitian ini, analisis sentimen terhadap penerima KIPK di media sosial X menggunakan algoritma *Support Vector Machine* (SVM) berhasil mencapai tingkat akurasi sebesar 59%.

Hasil ini mengindikasikan efektivitas SVM dalam mengklasifikasikan dan menganalisis sentimen dari data teks informal yang diperoleh dari pengguna media sosial. Berbagai sumber data, baik primer maupun sekunder, telah memberikan landasan yang kuat untuk pemahaman mendalam terhadap respons Masyarakat terhadap penerima KIPK. Sentimen negatif mendominasi dengan tingkat akurasi yang signifikan, menunjukkan adanya ketidakpuasan yang tinggi terhadap penerima KIPK tertentu di kalangan pengguna media sosial X. Meskipun demikian, adanya tingkat akurasi yang lebih rendah untuk kelas sentimen lainnya menunjukkan kompleksitas dalam menganalisis bahasa informal dan variasi respons masyarakat.

DAFTAR PUSTAKA

- [1] A. Fadli, "Transformasi Digital dan Moderasi Beragama: Memperkuat Ummatan Wasathan di Indonesia," *Schemata: Jurnal Pascasarjana UIN Mataram*, vol. 12, no. 1, pp. 1–14, 2023, [Online]. Available: <https://journal.uinmataram.ac.id/index.php/schemata/article/view/7773>
- [2] Q. S. Arsandy, Lisda Karolin, Ribka Dameria Sinuhaji, K. H. Lubis, Dimas Rumaharbo, and K. F. Samuel, "Pengaruh Beasiswa Kip-Kuliah Terhadap Motivasi Dan Prestasi Belajar Mahasiswa," *Jurnal Peluang*, vol. 11, no. 2, pp. 67–77, 2023.
- [3] F. S. Jasmine, "Pengaruh Beasiswa KIP-K Terhadap Prestasi Belajar Mahasiswa Manajemen Pendidikan Angkatan 2021 Universitas Negeri Surabaya," *JPBB : Jurnal Pendidikan, Bahasa dan Budaya*, vol. 2, pp. 61–70, 2023, doi: 10.55606/jpbb.v2i2.1437.
- [4] R. Kumar. Dwivedi and A. Kr. Saxena, *Proceedings of the 2019 8th International Conference on System Modeling & Advancement in Research Trends : SMART-2019 : (22nd-23rd November, 2019)*. Prof. Rakesh Kumar Dwivedi : Principal College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, 2019.
- [5] Fatimah Rahmadayana and Yuliant Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, 2021, doi: 10.29207/resti.v5i5.3457.
- [6] A. P. Nardilasari, et. Al., "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *Journal of Information Technology and Computer Science*, vol.8, no.1, pp. 11 – 18, Maret 2026.
- [7] M. A. Saddam, E. K. Dewantara, and A. Solichin, "Sentiment Analysis of Flood Disaster Management in Jakarta on Twitter Using Support Vector Machines," *Sinkron*, vol. 8, no. 1, pp. 470–479, Jan. 2023, doi: 10.33395/sinkron.v8i1.12063.
- [8] T. R. I. W. A. Widyanto, "Komparasi Naive Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter," *SINTECH JOURNAL*, vol. 6, 2023, doi: <https://doi.org/10.351598>.
- [9] P. S. Hutapea and W. Maharani, "Sentiment Analysis on Twitter Social Media towards Shopee E-Commerce through Support Vector Machine (SVM) Method," *JINAV: Journal of Information and Visualization*, vol. 4, no. 1, pp. 7–17, Jan. 2023, doi: 10.35877/454ri.jinav1504.
- [10] F. Rahmadayana and Y. Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.
- [11] N. Yadav, A. Shitole, O. Kudale, A. Rao, and S. Gupta, "Twitter Sentiment Analysis using Supervised Machine Learning," 2020.
- [12] H. Setiawan, E. Utami, and S. Sudarmawan, "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 1, pp. 43–51, Jul. 2021, doi: 10.31603/komtika.v5i1.5189.
- [13] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naive Bayes dan Support Vector Machine," *CESS (Journal of Computer Engineering System and Science)*, vol. 5, no. 2, pp. 293–293, Jul. 2020, doi: <https://doi.org/10.24114/cess.v5i2.18186>.