

**ANALISIS SENTIMEN PUBLIK TERHADAP PROGRAM KARTU INDONESIA
PINTAR KULIAH (KIP-K) MENGGUNAKAN METODE NAÏVE BAYES DAN
SUPPORT VECTOR MACHINE**

Nurul Azizah¹, Zuhajji², Andi Akram Nur Risal³

^{1,3}Jurusan Teknik Informatika dan Komputer, Fakultas Teknik,
Universitas Negeri Makassar

²Jurusan Pendidikan Teknik Elektro, Fakultas Teknik, Universitas Negeri
Makassar

¹nurulzizah210@gmail.com

²zuhajji@unm.ac.id

³akramandi@unm.ac.id

ABSTRACT

This study aims to classify public sentiment toward the Kartu Indonesia Pintar Kuliah (KIP-K) program based on comments collected from Twitter and Instagram social media platforms. The dataset consists of 3,214 public comments gathered via web scraping using relevant keywords during the period of January 2024 to May 2025. Data processing stages include preprocessing (cleaning, case folding, tokenizing, stopword removal, and stemming), automatic sentiment labeling using the InSet (Indonesian Sentiment Lexicon) lexicon-based approach, and feature extraction using TF-IDF. Two classification methods were compared: Naïve Bayes and Support Vector Machine (SVM). The evaluation results show that SVM outperforms Naïve Bayes in all metrics, achieving 80% accuracy with macro precision, recall, and F1-score of 0.81, 0.80, and 0.80, respectively. In contrast, Naïve Bayes achieved 60% accuracy with a macro score of 0.61. This study indicates that the majority of public sentiment toward KIP-K is positive, although there are significant proportions of neutral and negative sentiments related to targeting inaccuracies and fund misuse.

Keywords: sentiment analysis, kip-k, naïve bayes, support vector machine, tf-idf

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap program Kartu Indonesia Pintar Kuliah (KIP-K) berdasarkan komentar dari platform media sosial Twitter dan Instagram. Dataset terdiri dari 3.214 komentar publik yang dikumpulkan melalui web scraping menggunakan kata kunci relevan pada periode Januari 2024 hingga Mei 2025. Tahapan pemrosesan data meliputi preprocessing (cleaning, case folding, tokenizing, stopword removal, stemming), pelabelan sentimen otomatis berbasis leksikon InSet, dan ekstraksi fitur menggunakan TF-IDF. Dua metode klasifikasi dibandingkan: Naïve Bayes dan Support Vector Machine (SVM). Hasil evaluasi menunjukkan SVM unggul dalam semua metrik dengan akurasi 80%, macro precision, recall, dan F1-score masing-masing 0,81,

0,80, dan 0,80. Sementara Naïve Bayes mencapai akurasi 60% dengan macro score 0,61. Penelitian ini mengindikasikan bahwa mayoritas sentimen publik terhadap KIP-K bersifat positif, meskipun terdapat proporsi sentimen netral dan negatif yang signifikan terkait ketidaktepatan sasaran dan penyalahgunaan dana.

Kata Kunci: analisis sentimen, kip-k, naïve bayes, support vector machine, tf-idf.

A. Pendahuluan

Pendidikan merupakan faktor utama dalam menciptakan sumber daya manusia yang unggul dan berdaya saing tinggi. Namun, tidak semua individu memiliki kesempatan yang sama untuk mengakses pendidikan tinggi, terutama mereka yang berasal dari keluarga dengan keterbatasan ekonomi. Untuk mengatasi kesenjangan ini, pemerintah Indonesia melalui Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbudristek) menyediakan program Kartu Indonesia Pintar Kuliah (KIP-K). Program ini bertujuan untuk memberikan bantuan biaya kuliah dan biaya hidup kepada mahasiswa yang memenuhi kriteria ekonomi tertentu, agar dapat menyelesaikan pendidikan tinggi tanpa kendala finansial.

Sejak diperkenalkan sebagai Bidikmisi pada tahun 2010 dan disempurnakan menjadi KIP-Kuliah Merdeka pada tahun 2021, program ini telah membantu banyak mahasiswa mendapatkan pendidikan tinggi. Pada tahun 2023, lebih dari 700.000 mahasiswa telah menerima bantuan KIP-K (Yanuar, 2024). Namun, di balik manfaatnya yang besar, implementasi KIP-K masih diwarnai berbagai permasalahan serius, di antaranya ketidaktepatan sasaran dan penyalahgunaan dana, yang sering

menjadi sorotan publik melalui platform media sosial Twitter dan Instagram.

Analisis sentimen merupakan teknik pengolahan bahasa alami (Natural Language Processing/NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks sebagai positif, negatif, atau netral (Liu, 2022). Penelitian ini menggunakan dua metode klasifikasi yang populer, yaitu Naïve Bayes dan Support Vector Machine (SVM), untuk membandingkan kinerja keduanya dalam mengklasifikasikan sentimen publik terhadap program KIP-K di media sosial.

B. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan data berupa komentar pengguna media sosial Twitter (X) dan Instagram. Data dikumpulkan melalui teknik web scraping menggunakan kata kunci "KIPK", "PenerimaKIPK", "KIP Kuliah", "Penerima KIP-K", dan "Beasiswa KIP-K" pada periode Januari 2024 hingga Mei 2025, menghasilkan 3.214 komentar dalam bentuk file CSV.

Tahapan penelitian meliputi: (1) *Preprocessing data (cleaning, case folding, tokenizing, stopword removal, stemming)*; (2) Pelabelan sentimen otomatis menggunakan leksikon InSet (*Indonesian Sentiment Lexicon*)

dengan kategori positif (skor > 0), negatif (skor < 0), dan netral (skor = 0); (3) Splitting data dengan rasio 80% data latih (2.430 data) dan 20% data uji (608 data); (4) Ekstraksi fitur menggunakan TF-IDF dengan algoritma *CountVectorizer* dari *library Scikit-Learn* yang menghasilkan 5.044 fitur unik; (5) Pelatihan dan pengujian model Naïve Bayes dan SVM; (6) Evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score.

C. Hasil dan Pembahasan

1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini terdiri dari 3.214 komentar publik yang diambil dari media sosial Twitter dan Instagram menggunakan teknik web scraping. Hasil pengumpulan data dapat dilihat pada Gambar 1.

tanggal	nama	isi
23/01/2021 10:14	Andriani Kari	ada orang yang...
24/01/2021 10:24:49		Can mengah... (link)
25/01/2021 11:15 01/01/2021		Van... (link)
27/01/2021 08:05		... (link)
30/01/2021 01:10		... (link)
30/01/2021 08:12	David Ali	... (link)
30/01/2021 09:05		... (link)
31/01/2021 12:04		... (link)
01/02/2021 08:15	Arhan	... (link)
02/02/2021 12:45	Arhan	... (link)
02/02/2021 12:55	Arhan	... (link)
02/02/2021 13:17	Arhan	... (link)

Gambar 1. Hasil Pengumpulan Data

2. Analisis Data

a. Preprocessing Data

Preprocessing data dilakukan untuk membersihkan, mengubah, dan menyiapkan data mentah agar sesuai untuk analisis lebih lanjut. Setelah proses preprocessing, jumlah data bersih menjadi 3.033 data. Tabel 1 menunjukkan tahapan preprocessing yang meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.

Tabel 1. Hasil Preprocessing Data

Komentar	Processing data
Cara Mengajukan Kartu Indonesia Pintar Kuliah Gimana Kak 🙏	Cleaning Cara Mengajukan Kartu Indonesia Pintar Kuliah Gimana Kak Casefolding cara mengajukan kartu indonesia pintar kuliah gimana kak Tokenizing ['cara', 'mengajukan', 'kartu', 'indonesia', 'pintar', 'kuliah', 'bagaimana', 'kak'] Stopword ['mengajukan', 'kartu', 'indonesia', 'pintar', 'kuliah', 'kak'] Stemming aju kartu indonesia pintar kuliah kak

b. Pelabelan Data

Pelabelan data dilakukan menggunakan leksikon InSet (*Indonesian Sentiment Lexicon*). Setiap teks dianalisis dengan menghitung jumlah kata positif dan negatif, kemudian dihitung skor sentimennya. Jika skor > 0 maka label "positif", skor < 0 maka label "negatif", dan skor = 0 maka label "netral". Hasil pelabelan dataset dapat dilihat pada Gambar 2.

stemming data	Score	Sentimen
tinggal jelek	0	Netral
aku koma indonesia pintar kuliah kak	-1	Negatif
aku indonesia pintar daftar bidikmisi	0	Netral
izin paham kail masalah daftar kartu indonesia pintar kuliah orbi lobes pntara daftar suruh klaim akun aku data balas agustus bulat desember pas kema anak pns daftar kartu indonesia pintar kuliah	-5	Negatif
	-1	Negatif

Gambar 2. Hasil Pelabelan Data InSet Lexicon Based

c. Splitting Data

Splitting dataset dilakukan untuk membagi data menjadi data latih (80% = 2.430 data) dan data uji (20% = 608 data). Rasio ini dipilih karena umum digunakan dalam penelitian machine learning dan memberikan hasil pelatihan yang baik sekaligus

menyediakan data uji yang memadai (Pribadi & Ernawati, 2023).

d. Ekstraksi Fitur (TF-IDF)

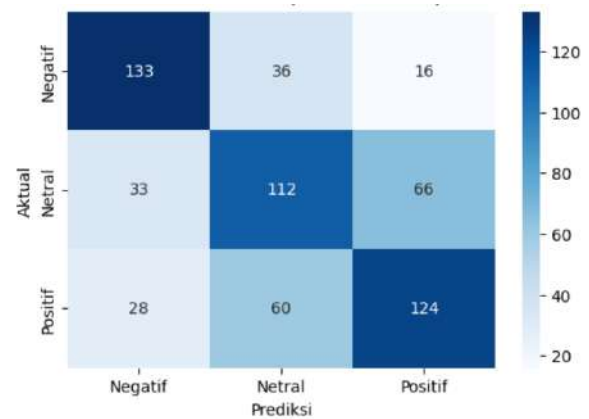
Ekstraksi fitur menggunakan algoritma CountVectorizer dari library Scikit-Learn untuk mengubah fitur teks menjadi representasi vektor. Data latih terdiri dari 2.430 dokumen yang direpresentasikan dalam vektor berdimensi 5.044 fitur, sedangkan data uji terdiri dari 608 dokumen dengan 5.044 fitur yang konsisten. Tabel 2 menunjukkan simulasi pembobotan kata dengan TF-IDF.

Tabel 2. Simulasi Pembobotan Kata dengan TF-IDF

Terms	TF		df	idf (log ₁₀ (D/DF))		tf-idf		W	
	d1	d2				d1	d2	d1	d2
KIPK	1	1	2		0	0	0	0	0
bantu	1	0	1	0,301029996	0,301	0	0,091	0	
kuliah	1	0	1	0,301029996	0,301	0	0,091	0	
gratis	1	0	1	0,301029996	0,301	0	0,091	0	
tidak	0	1	1	0,301029996	0	0,301	0	0,091	
tepat	0	1	1	0,301029996	0	0,301	0	0,091	
sasaran	0	1	1	0,301029996	0	0,301	0	0,091	

e. Metode Naïve Bayes

Metode Naïve Bayes digunakan untuk mengklasifikasikan komentar publik menjadi tiga sentimen: positif, netral, dan negatif. Model dilatih menggunakan 80% data (2.430 data) dan diuji dengan 20% data (608 data) menggunakan algoritma Multinomial Naïve Bayes dari pustaka Scikit-learn. Confusion Matrix hasil klasifikasi dapat dilihat pada Gambar 3.



Gambar 3. Confusion Matrix Heatmap Naïve Bayes

Confusion Matrix 3×3 pada Gambar 3 menunjukkan bahwa model Naïve Bayes mengklasifikasikan secara benar sebesar 124 data positif, 112 data netral, dan 133 data negatif. Tingkat akurasi model Naïve Bayes adalah 60%. Tabel 3 menunjukkan contoh cara kerja metode Naïve Bayes.

Tabel 3. Contoh Cara Kerja Metode Naïve Bayes

Tanggapan	Total kata	Kata unik	Denominator
Saya senang dapat bantuan kipk (positif)	5	Saya, senang, dapat, bantuan, kipk	15
Program kipk tidak adil (negatif)	4	kipk, program, tidak, adil	14
Kipk adalah bantuan mahasiswa (netral)	4	tidak, adil, adalah, mahasiswa (10)	14

Tabel 4. Hasil Prediksi Naïve Bayes

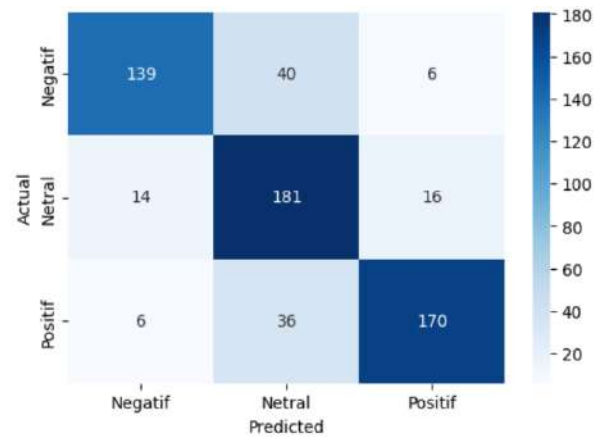
No.	Komentar	Kelas Awal	Kelas Akhir
1.	Hai student info kating putar kipk pra registrasi biaya didik klik lembaga pendonor beasiswa tinggal isi dokumen dukung pakai kartu serta kipk	Positif	Positif
2.	Semoga tahan untuk negara pendidikan	Positif	Positif
3.	Ya prioritas kategori	Positif	Positif
4.	Kipk bermanfaat untuk mahasiswa	Positif	Positif
5.	Berani niat survei rumah online offline lihat langsung sesuai nyata sasar tolong kuota kipk jokowi nadiem makarim puslapdik kipkuliah tambah kuota kip kuliah	Positif	Negatif
6.	Min kip kuliah ku daftar email hilang lupa sandinomor telfon atas apakh tunggu klim akun	Negatif	Netral
7.	Tanya kan temenku anak kipk tapi uang dibeliin ip kak	Negatif	Netral

Tabel 5. Hasil Uji Klasifikasi Naïve Bayes

	Presisi	Recall	F1 Score
Positif	0,60	0,58	0,59
Netral	0,54	0,53	0,53
Negatif	0,69	0,72	0,7

f. Metode Support Vector Machine (SVM)

Support Vector Machine melakukan prediksi kelas positif, negatif, dan netral dengan mempelajari pengetahuan yang terkandung dalam data latih. Confusion Matrix hasil klasifikasi SVM dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix Heatmap SVM
 Confusion Matrix 3×3 pada Gambar 4 menunjukkan bahwa model SVM mengklasifikasikan secara benar sebesar 170 data positif, 181 data netral, dan 139 data negatif dari total 608 data uji. Tingkat akurasi SVM adalah 80%.

Tabel 6. Hasil Prediksi Metode SVM

NO	Komentar	Kelas Awal	Kelas Akhir
1	harap perintah tambah anggar kipk utuh biaya hidup komponen penting kipkuliah tambahkuotakipkuliah	Positif	Positif
2	kuliah gratis keluarga kurang mampu	Positif	Positif
3	bantu mahasiswa kurang mampu sangat bermanfaat	Positif	Positif
4	beri bantuan dana penerima tepat sasaran	Positif	Positif
5	ngehack lihat akun kipk sinkron lolos sinkron lolos	Positif	Negatif
6	terima dana beli hp ip salahh	Negatif	Netral
608	kuliah jamin orang sukses kuliah beri didik bekal dunia kerja kipkuliah tambah kuotakip kuliah pusdatindikbud nadiem makarim	Netral	Netral

Tabel 7. Hasil Uji Model SVM

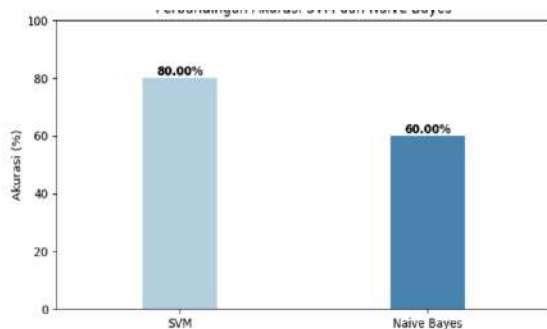
	Presisi	Recall	F1 Score
Positif	0,88	0,80	0,84
Netral	0,70	0,85	0,77
Negatif	0,87	0,77	0,80

3. Analisis Hasil Perbandingan

Setelah dilakukan pelatihan dan pengujian terhadap kedua algoritma, diperoleh hasil evaluasi kinerja model sebagaimana disajikan pada Tabel 8.

Tabel 8. Hasil Perbandingan Model NB dan SVM

Metrik Evaluasi	Naive Bayes	SVM
Macro Precision	0,61	0,81
Macro Recall	0,61	0,80
Macro F1-score	0,61	0,80



Gambar 5. Perbandingan Akurasi SVM dan Naive Bayes

4. Pembahasan

a. Analisis Hasil Naive Bayes

Naive Bayes memperoleh nilai akurasi sebesar 60%, dengan skor makro untuk precision, recall, dan F1-score masing-masing sebesar 0,61. Performa yang relatif rendah ini disebabkan oleh karakteristik data teks media sosial yang mengandung berbagai bentuk bahasa informal, singkatan, serta ambiguitas makna. Hal ini sesuai dengan temuan Wibowo & Wahyuni (2022) yang menyatakan bahwa Naive Bayes cenderung kurang

efektif dalam menangani data teks yang memiliki korelasi antar kata yang kuat. Confusion matrix menunjukkan bahwa hanya 124 dari 212 data berlabel positif yang terklasifikasi secara benar.

b. Analisis Hasil Support Vector Machine

SVM menghasilkan performa yang jauh lebih baik dengan tingkat akurasi 80%. Nilai macro precision, recall, dan F1-score masing-masing sebesar 0,81, 0,80, dan 0,80. Seperti diungkapkan Rahman & Utami (2021), SVM dikenal memiliki kemampuan unggul dalam mengklasifikasikan data teks high-dimensional seperti hasil ekstraksi TF-IDF. Kinerja SVM yang konsisten pada ketiga kategori sentimen mengindikasikan kemampuan model dalam membedakan fitur linguistik dari masing-masing kelas, bahkan dalam kondisi data yang tidak seimbang.

c. Perbandingan Kinerja

Berdasarkan metrik evaluasi makro, SVM unggul secara signifikan dalam semua aspek. Precision makro SVM mencapai 0,81 dibandingkan Naive Bayes yang hanya 0,61. Keunggulan SVM dapat dijelaskan oleh karakteristik algoritmanya yang berfokus pada pencarian hyperplane optimal dan kemampuannya menangani data non-linear melalui kernel trick. Sementara Naive Bayes bekerja berdasarkan asumsi independensi antar fitur yang sering tidak terpenuhi dalam teks natural (Setiawan & Lestari, 2020).

Pelabelan otomatis berbasis leksikon InSet memiliki keterbatasan

dalam menangani konteks kalimat secara menyeluruh, bahasa gaul, singkatan, serta aspek sarkasme atau ironi. Keterbatasan ini berpotensi menurunkan kualitas data latih dan uji, sehingga memengaruhi performa model klasifikasi, khususnya pada metode Naïve Bayes.

Hasil klasifikasi menunjukkan bahwa mayoritas komentar publik terhadap KIP-K memiliki sentimen positif, mencerminkan apresiasi masyarakat terhadap program ini. Namun, keberadaan komentar negatif dan netral mengindikasikan masih adanya ketidakpuasan terkait ketidaktepatan sasaran dan penyalahgunaan dana. Temuan ini memperkuat peran media sosial sebagai sumber data opini publik yang representatif untuk evaluasi kebijakan secara *real-time*.

D. Kesimpulan

Penelitian ini membandingkan kinerja metode Naïve Bayes dan *Support Vector Machine* (SVM) dalam analisis sentimen publik terhadap program KIP-K berdasarkan komentar dari media sosial Twitter dan Instagram. Hasil penelitian menunjukkan bahwa: (1) SVM memberikan performa lebih baik dengan akurasi 80% dibandingkan Naïve Bayes yang mencapai 60%; (2) Macro precision, recall, dan F1-score SVM (0,81; 0,80; 0,80) secara konsisten lebih tinggi dibanding Naïve Bayes (0,61; 0,61; 0,61); (3) Mayoritas sentimen publik terhadap program KIP-K bersifat positif, meskipun terdapat proporsi negatif dan netral yang

signifikan terkait isu ketidaktepatan sasaran dan penyalahgunaan dana. Penelitian ini merekomendasikan penggunaan SVM dengan fitur TF-IDF untuk analisis sentimen pada teks media sosial berbahasa Indonesia.

DAFTAR PUSTAKA

- Amelia, I., & Sarimole, F. M. (2024). Analisis sentimen tanggapan pengguna media sosial x terhadap program beasiswa kip-kuliah dengan menggunakan algoritma support vector machine (SVM). *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(3), 2994-3003.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis sentimen zoom cloud meetings di play store menggunakan naïve bayes dan support vector machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293.
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis sentimen terhadap penggunaan aplikasi shopee menggunakan algoritma support vector machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32-35.
- Liu, B. (2022). *Sentiment Analysis and Opinion Mining*. Springer Nature.
- Pribadi, H., & Ernawati, D. (2023). Optimasi model pembelajaran mesin menggunakan rasio data latih dan uji yang beragam. *Jurnal Teknologi Informasi dan Komputer*, 11(2), 77-85.

- Rahman, R., & Utami, F. (2021). Efektivitas support vector machine dalam klasifikasi sentimen teks. *Jurnal Sistem Cerdas*, 10(2), 63-70.
- Salsabila, N. A. (2022). Analisis sentimen pada media sosial twitter terhadap tokoh gus dur menggunakan metode naïve bayes dan support vector machine (SVM). Bachelor's thesis. UIN Syarif Hidayatullah Jakarta.
- Setiawan, B., & Lestari, R. (2020). Analisis performansi naïve bayes pada data teks bahasa indonesia. *Jurnal Ilmu Komputer dan Informatika*, 8(3), 21-29.
- Wibowo, A., & Wahyuni, S. (2022). Perbandingan metode naïve bayes dan svm dalam analisis sentimen twitter terkait vaksin covid-19. *Jurnal Informatika*, 9(1), 45-52.
- Widodo, B. K., Matondang, N. H., & Prasvita, D. S. (2022). Penerapan algoritma naïve bayes untuk analisis sentimen penggunaan aplikasi jobstreet. *Techno.Com*, 21(3), 523-533.
- Yanuar. (2024). Fakta-fakta KIP Kuliah 2023. Diakses dari <https://puslapdik.kemdikbud.go.id/fakta-fakta-kip-kuliah-2023>.