

IMPLEMENTASI *TEXT MINING* UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI

I Komang Dharmendra¹⁾ I Gusti Ngurah Ady Kusuma²⁾ Ida Ayu Mirah Cahya Dewi³⁾
Edwar⁴⁾

Program Studi Sistem Informasi¹⁾²⁾³⁾ Program Studi Manajemen Informatika⁴⁾
Fakultas Informatika dan Komputer¹⁾²⁾³⁾, Fakultas Bisnis dan Fokasi⁴⁾
Institut Teknologi dan Bisnis STIKOM Bali¹⁾²⁾³⁾⁴⁾

dharmendra@stikom-bali.ac.id¹⁾, ady_kusuma@stikom-bali.ac.id²⁾, mirahcahyadewi@stikom-bali.ac.id³⁾,
edwar.ridwan@stikom-bali.ac.id⁴⁾

ABSTRACT

The research aims to analyze the sentiment of alumni opinions using the Support Vector Machine (SVM) method. Analyzing the sentiment of alumni opinions is an important factor in evaluating the quality of educational institutions. The SVM method is utilized to classify opinions with high accuracy, employing TF-IDF for weighting and vectorization. The research results indicate an accuracy of 0.873, with precision, recall, and F1-score values of 0.877, 0.803, and 0.823, respectively. These findings demonstrate that SVM can be an effective choice in analyzing alumni sentiment. The research outcomes provide valuable insights for educational institutions to understand and enhance alumni satisfaction, as well as identify areas that require improvement. In the context of decision-making, the analysis of alumni sentiment can influence strategies and the development of educational programs. Future studies may consider expanding the sample size and exploring other natural language processing techniques to enhance the performance of analyzing alumni sentiment.

Keywords: *Text Mining, Sentiment analysis, Alumni Opinions, TF-IDF, Support Vector Machine*

ABSTRAK

Penelitian bertujuan untuk menganalisis sentimen opini alumni menggunakan metode *Support Vector Machine* (SVM). Analisis sentimen opini alumni merupakan faktor penting dalam evaluasi kualitas institusi pendidikan. Metode SVM digunakan untuk mengklasifikasikan opini dengan tingkat keakuratan yang tinggi, dengan menggunakan TF-IDF untuk pembobotan dan vektorisasi. Hasil penelitian menunjukkan akurasi sebesar 0.873, dengan nilai precision, recall, dan F1-Score berturut-turut sebesar 0.877, 0.803, dan 0.823. Temuan ini menunjukkan bahwa SVM dapat menjadi pilihan yang efektif dalam analisis sentimen opini alumni. Hasil penelitian memberikan wawasan penting bagi institusi pendidikan untuk memahami dan meningkatkan kepuasan alumni, serta mengidentifikasi aspek-aspek yang perlu diperbaiki. Dalam konteks pengambilan keputusan, hasil analisis sentimen opini alumni dapat memengaruhi strategi dan pengembangan program pendidikan. Penelitian selanjutnya dapat mempertimbangkan perluasan sampel dan eksplorasi teknik pemrosesan bahasa alami lainnya untuk meningkatkan performa analisis sentimen opini alumni.

Kata Kunci : *Text Mining, Analisis Sentimen, Opini Alumni, TF-IDF, Support Vector Machine.*

PENDAHULUAN

Tingkat kepuasan dan opini yang dimiliki oleh alumni merupakan faktor penting dalam evaluasi kualitas sebuah institusi pendidikan (Dharmendra, Saputra, and Pramaita 2019). *Sentiment analysis*, atau analisis sentimen, telah menjadi metode yang signifikan dalam memahami dan menganalisis opini dan

sentimen yang terkandung dalam teks (Buntoro 2019). Dengan memanfaatkan teknik pemrosesan bahasa alami dan pembelajaran mesin, *Sentiment analysis* memungkinkan para peneliti dan praktisi untuk menggali pandangan dan preferensi alumni terhadap berbagai aspek pengalaman mereka di perguruan tinggi.

Sentiment analysis memiliki implikasi yang luas dalam pengambilan keputusan stakeholder di perguruan tinggi. Keputusan strategis, perencanaan program, pengembangan kurikulum, dan perbaikan layanan pendidikan dapat dipengaruhi oleh pemahaman yang akurat tentang sentimen alumni. Dengan memahami sentimen yang terkandung dalam opini alumni, perguruan tinggi dapat merancang strategi yang sesuai untuk meningkatkan kualitas pendidikan, menyesuaikan program yang ada, dan memperbaiki aspek-aspek tertentu yang kurang memuaskan bagi alumni (Dharmendra et al. 2022).

Namun, analisis sentimen opini alumni juga memiliki tantangan tersendiri. Salah satunya adalah kompleksitas dalam memproses bahasa manusia yang sarat dengan variasi, ambigu, dan konteks tersirat (Fauziyyah 2020). Selain itu, penting untuk mengatasi permasalahan seperti identifikasi subjek, perbedaan dalam gaya penulisan, serta pengaruh dari faktor eksternal yang dapat memengaruhi sentimen. Untuk mengatasi tantangan ini, penelitian ini akan menggunakan algoritma *Support Vector Machine* (SVM) dalam melakukan analisis sentimen. SVM telah terbukti efektif dalam tugas klasifikasi teks dan mampu menghadapi permasalahan kompleks dalam pemrosesan bahasa alami (Wati and Ernawati 2021).

Rencana penelitian ini bertujuan untuk melakukan analisis sentimen opini alumni menggunakan metode SVM. Evaluasi model akan dilakukan menggunakan matriks evaluasi yang mencakup akurasi, precision, recall, dan F1-Score (Fitri, Andreswari, and Hasibuan 2019). Dengan memanfaatkan matriks evaluasi ini, penelitian ini bertujuan untuk memberikan wawasan yang mendalam tentang sentimen alumni terhadap perguruan tinggi dan memberikan rekomendasi untuk perbaikan dan pengembangan institusi tersebut.

TINJAUAN PUSTAKA

Text Mining

Text mining mengacu pada proses mendapatkan informasi bernilai tinggi dari teks. Informasi bernilai tinggi umumnya diperoleh dengan menganalisis pola dan tren melalui metode seperti pembelajaran statistik. Text mining melibatkan langkah-langkah

seperti mengorganisir teks input (seperti parsing) dengan menambahkan beberapa fitur linguistik dan menghilangkan yang tidak relevan, serta menyimpannya dalam basis data (Ardiada, Sudarma, and Giriantari 2019). Selanjutnya, pola-pola dalam data yang terstruktur diidentifikasi, kemudian hasilnya dievaluasi dan diinterpretasikan. Proses text mining umumnya mencakup kategorisasi teks, pengelompokan teks, ekstraksi konsep/entitas, pembuatan taksonomi yang rinci, analisis sentimen, penyimpulan dokumen, dan pemodelan hubungan antara entitas (misalnya, pembelajaran tentang hubungan antara entitas yang disebutkan dalam teks) (Saputra et al. 2022).

Sentiment analysis

Sentiment analysis, juga dikenal sebagai analisis sentimen, adalah suatu teknik dalam bidang pengolahan bahasa alami yang bertujuan untuk menentukan dan memahami sentimen atau opini yang terkandung dalam teks (Dharmendra et al. 2022). Pendekatan ini melibatkan penggunaan metode komputasional dan algoritma untuk mengklasifikasikan teks menjadi kategori sentimen yang berbeda, seperti positif, negatif, atau netral. Dalam konteks ilmiah, *Sentiment analysis* menjadi penting dalam banyak aplikasi, termasuk pengenalan pola konsumen, pengawasan media sosial, dan analisis umpan balik pelanggan (Buntoro 2019).

Sentiment analysis dapat diterapkan pada berbagai jenis teks, mulai dari ulasan produk, media sosial, hingga komentar di platform daring. Untuk melakukan analisis sentimen, langkah-langkah umumnya meliputi *preprocessing teks*, di mana teks tersebut dibersihkan dari karakter khusus, diubah menjadi bentuk yang dapat diproses oleh komputer, dan kemudian dilakukan ekstraksi fitur untuk mengidentifikasi pola penting yang berkaitan dengan sentimen. Setelah itu, metode pembelajaran mesin seperti klasifikasi atau pendekatan berbasis aturan digunakan untuk mengklasifikasikan teks ke dalam kategori sentimen yang sesuai.

Dalam pengembangan *Sentiment analysis*, tantangan utama adalah mengatasi ambiguitas, ironi, dan kompleksitas bahasa manusia. Para

peneliti berupaya untuk meningkatkan akurasi analisis sentimen dengan memanfaatkan teknik-teknik yang lebih canggih, termasuk penggunaan jaringan saraf tiruan dan pemodelan bahasa yang lebih kompleks. Selain itu, perkembangan dalam bidang analisis emosi juga berkontribusi pada pengembangan *Sentiment analysis*, di mana pendekatan yang lebih rinci dan nuansa emosi dapat diperhitungkan. *Sentiment analysis* memiliki potensi besar dalam memberikan wawasan tentang opini dan persepsi manusia secara massal, yang dapat berguna dalam pengambilan keputusan bisnis, analisis pasar, dan pemantauan reputasi merek.

Praproses Teks

Struktur data yang efisien dapat mempermudah otomatisasi proses komputer. Dalam konteks analisis teks, dokumen teks harus disiapkan sebelum dapat digunakan dalam proses utama (Acheampong, Wenyu, and Nunoo-Mensah 2020). Persiapan awal dokumen teks atau dataset mentah ini dikenal sebagai praproses teks (text preprocessing) (Bayari and Bensefia 2021). Praproses teks bertujuan untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur. Praproses melibatkan beberapa tahapan, termasuk pengubahan kecil huruf, tokenisasi, penyaringan, dan stemming.

TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode statistik yang digunakan untuk mengevaluasi pentingnya sebuah kata dalam suatu dokumen dalam kumpulan dokumen. TF-IDF terdiri dari dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF) (Simatupang and Utomo 2019). Term Frequency mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Semakin sering kata muncul, semakin penting kata tersebut dalam konteks dokumen tersebut. Namun, Term Frequency tidak memberikan bobot yang seimbang terhadap kata-kata umum yang mungkin muncul di banyak dokumen. IDF, di sisi lain, mengukur seberapa jarang sebuah kata muncul dalam seluruh kumpulan dokumen. Kata-kata yang jarang muncul,

seperti kata-kata yang spesifik atau unik, cenderung memberikan informasi yang lebih berharga daripada kata-kata umum. IDF memberikan bobot yang lebih tinggi pada kata-kata jarang ini.

Dengan menggabungkan TF dan IDF, TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul dalam dokumen tertentu tetapi jarang muncul di dokumen-dokumen lain. Ini memungkinkan kita untuk mengidentifikasi kata-kata kunci yang memiliki signifikansi dan relevansi tinggi dalam suatu dokumen.

Penerapan TF-IDF dapat digunakan dalam berbagai tugas analisis teks, termasuk klasifikasi dokumen, pengelompokan topik, dan pencarian informasi (Buntoro 2019). Bobot TF-IDF dapat digunakan sebagai fitur untuk melatih model pembelajaran mesin dalam tugas-tugas ini. Dengan menggunakan TF-IDF, kita dapat mengidentifikasi kata-kata penting yang mempengaruhi pemahaman dan makna dokumen, dan dengan demikian meningkatkan kualitas analisis teks yang dilakukan.

Support Vector Machines (SVM)

Support Vector Machine (SVM) adalah sebuah metode klasifikasi dalam pembelajaran mesin yang digunakan untuk mengklasifikasikan data dengan membangun sebuah hyperplane dalam ruang multidimensi (Wati and Ernawati 2021). SVM memaksimalkan jarak antara hyperplane dan titik-titik data yang berbeda kelasnya, sehingga menciptakan batas keputusan yang optimal. Dalam analisis sentimen, SVM dapat digunakan sebagai alat untuk mengklasifikasikan teks ke dalam kategori sentimen yang berbeda, seperti positif, negatif, atau netral (Dharmendra et al. 2022).

SVM memiliki keunggulan dalam analisis sentimen karena kemampuannya mengatasi masalah pada data yang tidak linier atau data dengan dimensi yang tinggi. Dalam konteks analisis sentimen, SVM dapat mengidentifikasi pola kompleks yang mungkin sulit ditangkap oleh metode klasifikasi lainnya. Namun, SVM juga memiliki beberapa kelemahan, seperti sensitivitas terhadap skala data dan keterbatasan dalam mengatasi data dengan ukuran yang sangat besar. Oleh karena itu,

pilihan pengaturan parameter dan teknik penyesuaian yang tepat sangat penting dalam implementasi SVM untuk analisis sentimen yang akurat. Persamaan untuk algoritma SVM bisa dilihat pada persamaan 1.

$$y = f(x) \dots \dots \dots (1)$$

Keterangan:

y : kelas prediksi

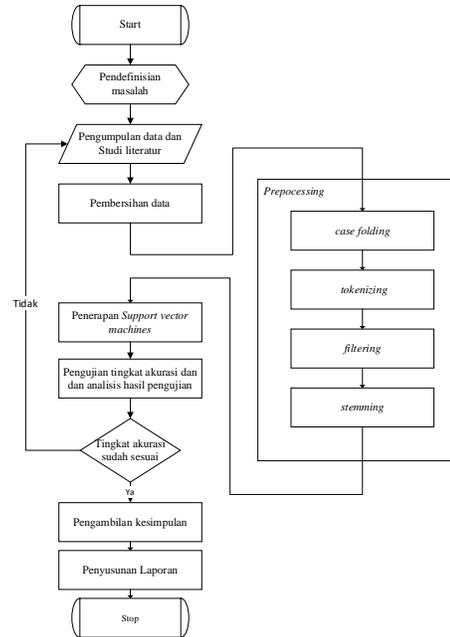
f : fungsi keputusan

x : vektor fitur input

Dengan menggunakan SVM dalam analisis sentimen, dapat dilakukan klasifikasi teks dengan tepat ke dalam kategori sentimen yang sesuai, seperti positif, negatif, atau netral. Hal ini memberikan wawasan berharga dalam memahami opini dan pandangan yang terkandung dalam teks. Penerapan SVM dalam analisis sentimen telah memberikan manfaat dalam berbagai bidang, termasuk analisis umpan balik pelanggan, pengawasan media sosial, dan pengambilan keputusan bisnis yang didasarkan pada sentimen konsumen. Dengan memanfaatkan kemampuan SVM dalam mengklasifikasikan sentimen, dapat memperoleh pemahaman yang lebih mendalam tentang preferensi dan persepsi pengguna (Anggraini, Kusri, and Fatta 2023).

METODE PENELITIAN

Kerangka Berpikir



Gambar 1. Kerangka Berpikir

Gambar 1 menunjukkan penerapan SVM dalam *Sentiment analysis* melibatkan beberapa tahapan. Pertama, teks yang akan dianalisis harus dipreproses untuk menghilangkan karakter khusus, mengubah teks menjadi bentuk yang dapat diproses oleh komputer, dan mengekstraksi fitur penting yang berkaitan dengan sentimen. Fitur-fitur ini dapat mencakup kata-kata kunci, frasa, atau penghitungan frekuensi kata. Setelah itu dilakukan pembobotan menggunakan TF-IDF untuk memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul dalam opini. Ini dilakukan untuk mengidentifikasi kata-kata kunci yang memiliki signifikansi dan relevansi tinggi dalam opini.

Setelah dilakukan pembobotan, dilanjutkan dengan pembangunan model, dimana SVM dilatih menggunakan dataset yang telah diklasifikasikan secara manual untuk mengenali pola-pola sentimen. SVM kemudian dapat digunakan untuk mengklasifikasikan teks baru berdasarkan pola-pola yang telah dipelajari.

Teknik Pengumpulan Data

Adapun jenis pengumpulan data yang digunakan dalam penelitian ini sebagai berikut didapat dari opini yang disampaikan oleh mahasiswa yang akan lulus, dan diisi pada halaman cdc.stikom-bali.ac.id. Dengan data yang digunakan berjumlah 591. Tabel 1 menunjukkan jumlah data yang digunakan

Tabel 1. Ratio penggunaan data

	<i>Jumlah</i>
<i>Positif</i>	197
<i>Negatif</i>	197
<i>Netral</i>	197
<i>Total</i>	591

HASIL PENELITIAN

Tahap Praproses

Pada tahap awal dilakukan tahap praproses yang merupakan langkah penting dalam mempersiapkan data sebelum dilakukan analisis sentimen menggunakan metode *Support Vector Machine (SVM)*.

Data yang telah dikumpulkan sering kali mengandung karakter khusus, tanda baca, dan link yang tidak relevan. Oleh karena itu, langkah ini melibatkan pembersihan data dari karakter-karakter yang tidak diperlukan.

Selain itu, proses pembersihan juga dapat mencakup penghapusan atau penggantian karakter yang mengganggu seperti emotikon, hashtag, atau mention.

Kemudian dilanjutkan dengan tokenisasi, yang merupakan proses memisahkan teks menjadi unit-unit yang lebih kecil, seperti kata-kata atau frasa. Dalam tokenisasi, teks yang panjang dipecah menjadi token-token yang dapat dijadikan unit analisis. Misalnya, kalimat "Saya suka makan nasi goreng" akan dipecah menjadi token-token seperti "Saya", "suka", "makan", "nasi", "goreng".

Setelah tokenisasi dilanjutkan dengan proses menghilangkan stopwords, stopwords adalah kata-kata yang umum digunakan dalam bahasa tetapi tidak memiliki nilai signifikan dalam analisis sentimen, seperti "dengan", "ia", "yang", dan sebagainya.

Dalam tahap menghilangkan stopwords, stopwords diidentifikasi dan dihapus dari teks untuk mengurangi noise dan meningkatkan relevansi analisis sentimen.

Proses terakhir praproses adalah stemming, stemming dan lemmatisasi adalah proses mengubah kata-kata ke bentuk dasarnya.

Stemming menghilangkan afiks atau akhiran kata untuk mendapatkan bentuk dasar kata. Contohnya, kata-kata seperti "berlari", "berlari", dan "berlari" akan diubah menjadi "laris". Pada penelitian ini menggunakan stemming sastrawi sebagai model stemming ("Sastrawi/Sastrawi: High Quality Stemmer Library for Indonesian Language (Bahasa)" n.d.). Tabel 2 menampilkan contoh kalimat sebelum proses praproses dan setelah proses praproses.

Tabel 2 Praproses

Praproses	Kalimat
Sebelum Praproses	Dosen memberikan materi dengan jelas, datang tepat waktu
Setelah Praproses	dosen beri materi dengan jelas datang tepat waktu

TF-IDF

Setelah proses praproses selesai, dilanjutkan dengan melakukan pembobotan dan vektorisasi menggunakan TF-IDF untuk membantu dalam mengidentifikasi kata-kata kunci yang memiliki pengaruh besar terhadap opini dan membantu mengurangi dampak kata-kata umum yang biasanya tidak memberikan banyak informasi yang berarti.

SVM

Setelah proses preproses data selesai dilakukan, dilanjutkan dengan proses ekstraksi fitur pada data latih untuk pengambilan ciri opini yang dapat menggambarkan karakteristik dari jenis opini yang ada. Dilanjutkan dengan membangun model SVM untuk analisa sentimen dengan menggunakan *kernel linear*.

Setelah pembentukan model SVM akan dilatih menggunakan fitur dan label latihan yang diberikan, kemudian dilanjutkan melakukan prediksi untuk melakukan evaluasi performa model SVM, misalnya dengan membandingkan

DAFTAR PUSTAKA

- [1.] Acheampong, Francisca Adoma, Chen Wenyu, and Henry Nunoo-Mensah. 2020. "Text-Based Emotion Detection: Advances, Challenges, and Opportunities." *Engineering Reports* 2 (7): e12189. <https://doi.org/10.1002/eng2.12189>.
- [2.] Anggraini, Resti Kusuma, K. Kusri, and Hanif Al Fatta. 2023. "Metode *Support Vector Machine* Pada Klasifikasi Pengaduan Masyarakat." *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)* 8 (1): 110–16. <https://doi.org/10.30645/jurasik.v8i1.547>.
- [3.] Ardiada, I Made Dwi, Made Sudarma, and Dwi Giriantari. 2019. "Text Mining Pada Sosial Media Untuk Mendeteksi Emosi Pengguna Menggunakan Metode *Support Vector Machine* Dan K-Nearest Neighbour." *Majalah Ilmiah Teknologi Elektro* 18 (1): 55. <https://doi.org/10.24843/mite.2019.v18i01.p08>.
- [4.] Bayari, Reem, and Ameer Bensefia. 2021. "Text Mining Techniques for Cyberbullying Detection: State of the Art." *Advances in Science, Technology and Engineering Systems Journal* 6 (1): 783–90. <https://doi.org/10.25046/aj060187>.
- [5.] Buntoro, Ghulam Asrofi. 2019. "Sentiments Analysis for Governor of East Java 2018 in Twitter." *Sinkron* 3 (2): 49. <https://doi.org/10.33395/sinkron.v3i2.10025>.
- [6.] Dharmendra, I. Komang, Ni Nym Utami Januhari, I. Putu Ramayasa, and I. Made Agus Wirahadi Putra. 2022. "Uji Komparasi *Sentiment analysis* Pada Opini Alumni Terhadap Perguruan Tinggi." *Jurnal Teknik Informatika UNIKA Santo Thomas*, May, 1–6. <https://doi.org/10.54367/jtiust.v7i1.1748>.
- [7.] dharmendra, komang, Komang Oka Saputra, and I Nyoman Pramaita. 2019. "Analisa Sentiment Untuk Opini Alumni Perguruan Tinggi." *Majalah Ilmiah Teknologi Elektro* 18 (2): xxxx–xxxx. <https://doi.org/10.24843/MITE.2019.V18I02.P11>.
- [8.] Fauziyyah, Anni Karimatul. 2020. "ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON." *Jurnal Ilmiah SINUS* 18 (2): 31–42. <https://doi.org/10.30646/sinus.v18i2.491>.
- [9.] Fitri, Veny Amilia, Rachmadita Andreswari, and Muhammad Azani Hasibuan. 2019. "Sentiment analysis of Social Media Twitter with Case of Anti-LGBT Campaign in Indonesia Using Naïve Bayes, Decision Tree, and Random Forest Algorithm." *Procedia Computer Science, The Fifth Information Systems International Conference, 23-24 July 2019, Surabaya, Indonesia*, 161 (January): 765–72. <https://doi.org/10.1016/j.procs.2019.11.181>.
- [10.] Saputra, Irwansyah, Taufik Djatna, Riki Ruli A. Siregar, Dinar Ajeng Kristiyanti, Hasbi Rahma Yani, and Andri Agung Riyadi. 2022. "Text Mining of PeduliLindungi Application Reviews on Google Play Store." *Faktor Exacta* 15 (2). <https://doi.org/10.30998/faktorexacta.v15i2.10629>.
- [11.] "Sastrawi/Sastrawi: High Quality Stemmer Library for Indonesian Language (Bahasa)." n.d. Accessed January 9, 2022. <https://github.com/sastrawi/sastrawi>.
- [12.] Simatupang, Meylita Putri, and Dito Putro Utomo. 2019. "ANALISA TESTIMONIAL DENGAN MENGGUNAKAN ALGORITMA TEXT MINING DAN TERM FREQUENCY- INVERSE DOCUMENT FREQUENCY (TF-

IDF) PADA TOKO ALLMEEART.”
*KOMIK (Konferensi Nasional
Teknologi Informasi Dan Komputer)*
3 (1).
<https://doi.org/10.30865/komik.v3i1.1697>.

- [13.] Wati, Risa, and Siti
Ernawati. 2021. “Analisis Sentimen
Persepsi Publik Mengenai PPKM
Pada Twitter Berbasis SVM
Menggunakan Python.” *Jurnal Teknik
Informatika UNIKA Santo Thomas*,
November, 240–47.
<https://doi.org/10.54367/JTIUST.V6I2.1465>.