Tersedia Online : http://e-journals.unmul.ac.id/

ADOPSI TEKNOLOGI DAN SISTEM INFORMASI (ATASI)

Alamat Jurnal : http://e-journals2.unmul.ac.id/index.php/atasi/index

Analisis Sentimen Ulasan Jembatan Repo-Repo di Google Maps Menggunakan Metode SVM dengan Ekstraksi Fitur BERT

Muhammad Ikram [1)], Rudiman [2)*], Naufal Azmi Verdikha [3)]
Departemen Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Kalimantan Timur

E-Mail : 1911102441065@umkt.ac.id [1)]; rud959@umkt.ac.id [2)]; nav651@umkt.ac.id [3)]

ARTICLE INFO

Article history:
Received : 06-08-2025
Revised : 02-09-2025
Accepted : 06-11-2025
Available online : 29-04-2026

Keywords:
Sentiment Analysis
IndoBERT
Chi-Square
SVM
Google Maps

ABSTRACT

This study aims to improve sentiment analysis accuracy on reviews of the Repo-Repo Bridge in Tenggarong, Kutai Kartanegara, East Kalimantan, collected from Google Maps. Previous research employed the TF-IDF method and Naïve Bayes classification but achieved only 58% accuracy. To address this limitation, this study applied feature extraction using IndoBERT with the same dataset from prior research. Feature selection was performed using the Chi-Square method, and classification was conducted with Support Vector Machine (SVM). The results show that Chi-Square feature selection successfully reduced the feature dimensions from 768 to 100 and increased classification accuracy from 57% to 65%. In addition to accuracy, the model's performance also improved in other evaluation metrics, with precision increasing from 55% to 60%, recall from 57% to 65%, and f1-score from 56% to 61%. Furthermore, the SVM training process became more efficient, saving 0.1092 seconds. These findings demonstrate that the combination of IndoBERT, Chi-Square, and SVM is effective in enhancing text-based sentiment classification performance.

Kata Kunci :
Analisis Sentimen
IndoBERT
Chi-Square
SVM
Google Maps

ABSTRAK

Penelitian ini bertujuan untuk meningkatkan akurasi analisis sentimen pada ulasan Jembatan Repo-Repo Kabupaten Tenggarong Kutai Kartanegara, Kalimantan Timur, yang diambil dari Google Maps. Penelitian sebelumnya menggunakan metode TF-IDF dan klasifikasi Naïve Bayes, namun hanya mencapai akurasi 58%. Untuk mengatasi keterbatasan tersebut, penelitian ini menerapkan ekstraksi fitur menggunakan IndoBERT dengan data yang sama dari sumber penelitian sebelumnya. Seleksi fitur digunakan dengan metode Chi-Square, dan klasifikasi menggunakan Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa seleksi fitur Chi-Square mampu mereduksi dimensi fitur dari 768 menjadi 100, serta meningkatkan akurasi klasifikasi dari 57% menjadi 65%. Selain akurasi, performa model juga meningkat pada metrik evaluasi lainnya, yaitu precision dari 55% menjadi 60%, recall dari 57% menjadi 65%, dan f1-score dari 56% menjadi 61%. Selain itu, waktu pelatihan SVM juga menjadi lebih efisien dengan penghematan sebesar 0,1092 detik. Temuan ini membuktikan bahwa kombinasi IndoBERT, Chi-Square, dan SVM efektif dalam meningkatkan performa klasifikasi sentimen berbasis teks.

2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

1. PENDAHULUAN

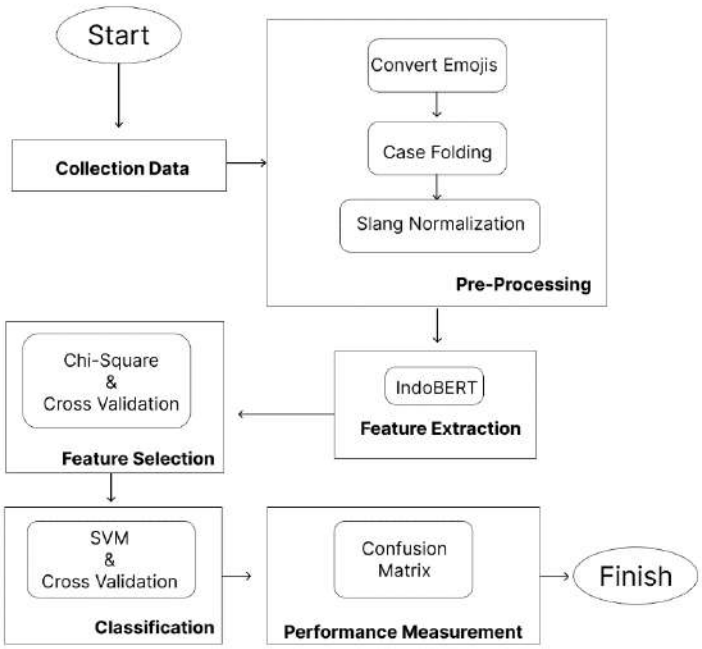
Perkembangan teknologi informasi yang pesat telah mengubah cara masyarakat memberikan penilaian terhadap suatu tempat, termasuk destinasi wisata. Salah satu platform yang banyak digunakan adalah Google Maps, yang tidak hanya menyediakan informasi lokasi, tetapi juga memungkinkan pengguna

*) Corresponding Author
https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

memberikan ulasan dan penilaian secara langsung (Saiful Nur Budiman et al., 2024). Ulasan-ulasan ini dapat menjadi sumber data penting dalam menganalisis persepsi publik terhadap suatu lokasi wisata.
Jembatan Repo-Repo merupakan salah satu destinasi wisata yang terletak di Kalimantan Timur, yang dirancang khusus untuk pejalan kaki dan dikenal dengan keindahan pemandangan Sungai Mahakam (Hermawati, 2024). Meskipun demikian, persepsi pengunjung terhadap tempat ini beragam dan terekam melalui ulasan mereka di Google Maps. Untuk menggali lebih dalam pandangan publik tersebut, diperlukan analisis sentimen yang mampu mengelompokkan ulasan menjadi sentimen positif, negatif, atau netral.
Penelitian sebelumnya pernah melakukan analisis sentimen terhadap ulasan Jembatan Repo-Repo menggunakan metode Naïve Bayes dan ekstraksi fitur TF-IDF (Perdana, 2024). Namun, metode tersebut menghasilkan akurasi yang relatif rendah, yaitu sebesar 58%, karena keterbatasan TF-IDF dalam memahami konteks bahasa alami.
Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan baru dengan memanfaatkan model IndoBERT untuk ekstraksi fitur, Chi-Square untuk seleksi fitur, dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi. IndoBERT, sebagai model bahasa berbasis transformer yang dilatih khusus untuk Bahasa Indonesia, mampu menangkap konteks dan makna dalam kalimat dengan lebih baik dibandingkan pendekatan statistik tradisional (Aljabar et al., 2024). Sedangkan Chi-Square digunakan untuk meningkatkan efisiensi dengan menyaring fitur yang paling relevan terhadap label sentimen (Gusti et al., 2023).
Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi sentimen ulasan wisata Jembatan Repo-Repo dan membandingkannya dengan hasil dari metode sebelumnya. Dengan pendekatan yang lebih kontekstual dan efisien, diharapkan penelitian ini dapat memberikan kontribusi dalam pengembangan analisis sentimen berbahasa Indonesia, khususnya di bidang pariwisata digital

2. METODE PENELITIAN

Pada penelitian analisis sentimen ini diawali dengan pengumpulan data, data yang digunakan diperoleh dari data penelitian sebelumnya milik (Perdana, 2024). Kemudian dataset yang didapatkan harus melewati tahapan *pre-processing* yang meliputi *case folding*, *convert emoji* dan *slang normalization*, selanjutnya tahap ekstraksi fitur IndoBERT yang menghasilkan fitur berupa *embeddings*. Hasil dari eksraksi fitur akan diseleksi menggunakan Chi-Square untuk memilih fitur yang paling relevan terhadap label sentimen, selanjutnya model SVM dilatih menggunakan fitur dari IndoBERT untuk mengklasifikasikan sentimen ke dalam kelas Positif, Negatif, dan Netral. Tahap terakhir adalah evaluasi model menggunakan *confusion matrix* untuk melihat akurasi model dalam melakukan klasifikasi sentimen.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Penelitian ini menggunakan dataset yang sama dari penelitian sebelumnya milik (Perdana, 2024). Dataset tersebut bersumber dari ulasan jembatan Repo-Repo yang ada di Google maps, data dikumpulkan dengan menggunakan metode *crawling*.

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

B. *Pre-Processing*

Pre-processing adalah tahap dimana teks asli diolah dengan menerapkan tahapan dasar untuk mengubah atau menghilangkan elemen-elemen yang tidak relevan (Najjichah et al., 2019). Adapun tahapan pre-processing yang akan dilakukan pada penelitian ini adalah:

1) *Case folding* adalah proses normalisasi teks dengan mengubah semua karakter huruf besar menjadi huruf kecil.
2) *Convert emoji* adalah proses merubah emoji atau emoticon menjadi string yang merepresentasikan makna dari emoji atau emoticon tersebut.
3) *Slang normalization* adalah proses untuk mengubah kata-kata atau frasa slang (bahasa gaul, tidak baku, atau singkatan) menjadi bentuk standar atau baku.

C. Ekstraksi Fitur IndoBERT

BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model bahasa yang efektif dalam memahami konteks bahasa alami. Dalam analisis sentimen, BERT dapat menghasilkan representasi vektor kata yang sangat mendalam dan mampu menangkap konteks yang rumit dalam teks (Aljabar et al., 2024). Maka dari itu BERT dapat menangkap makna semantik dan struktur sintaksis dari kata-kata dalam teks, sehingga sangat bermanfaat untuk tugas-tugas seperti klasifikasi sentimen (Maringka et al., 2025). Dalam penelitian ini menggunakan ekstraksi fitur IndoBERT, IndoBERT adalah sebuah model bahasa berbasis BERT yang telah dilatih secara khusus untuk Bahasa Indonesia dan digunakan untuk berbagai tugas pemrosesan bahasa alami (Tobing et al., 2025). Berikut tahapan dasar dalam penggunaan model IndoBERT untuk ekstraksi fitur:

1) Persiapan Dataset: persiapkan kolom review dalam dataset yang telah melalui tahapan *pre-processing.*
2) Melakukan tokenisasi khusus untuk model BERT.
3) Mengirim input teks yang sudah ditokenisasi ke model IndoBERT lalu menerima representasi (fitur) dari model tersebut.
4) Ambil Representasi Token [CLS].
5) mempersiapkan fitur dari IndoBERT: merubah dimensi token [CLS] menjadi *array* agar dapat masuk ke model SVM.

D. Seleksi Fitur Chi-Square

Dalam analisis sentimen seleksi fitur berfungsi memilih fitur-fitur (kata atau frasa) yang paling relevan dan informatif untuk memprediksi sentimen suatu teks, seleksi fitur membantu mengidentifikasi kata-kata kunci yang paling berkontribusi pada keputusan klasifikasi (Amrullah et al., 2020). Dalam penelitian ini seleksi fitur menggunakan metode Chi-Square, berperan penting dalam menyaring fitur-fitur yang paling relevan dengan sentimen, sehingga dapat meningkatkan akurasi dan efisiensi model. Berikut tahapan dalam penggunaan Chi-Square:

1) Mempersiapkan data fitur dan data label sentimen.
2) Membuat skala ke rentang 0 hingga 1 dari fitur IndoBERT.
3) Menentukan rentang jumlah fitur yang akan diuji (nilai k).
4) Mencari nilai K terbaik.
5) Mengambil nilai K terbaik.

E. Klasifikasi SVM

Algoritma yang digunakan untuk klasifikasi data pada penelitian ini menggunakan algoritma SVM (*Support Vector Machine*). SVM merupakan sebuah metode dalam pembelajaran mesin yang diterapkan untuk tugas klasifikasi dan regresi, SVM beroperasi dengan membentuk hyperplane pada ruang berdimensi tinggi untuk memisahkan data berdasarkan kelas sentimen. Tujuan utamanya adalah mencari hyperplane yang optimal dengan memaksimalkan margin antar kelas, sehingga hasil klasifikasi lebih akurat (Hayati et al., 2024). Berikut rumus klasifikasi SVM dengan kernel linier:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i \, y_i K(x, x_{i)} + b\right)$$

Keterangan:

a. $f(x)$ adalah fungsi keputusan untuk memprediksi kelas dari data uji x,
b. N adalah jumlah sampel data latih,
c. αi adalah bobot dari vektor pendukung,
d. yi adalah label kelas dari data latih,
e. $K(x,xi)$adalah fungsi kernel yang menghitung kesamaan antara x dan xi
f. b adalah bias

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

F. Evaluasi

Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi. Dalam bahasa yang lebih sederhana, confusion matrix membantu memahami seberapa baik model klasifikasi dalam memprediksi suatu kelas atau kategori (Dwiki et al., 2021). *Confusion matrix* menghasilkan beberapa ukuran evaluasi penting seperti akurasi, *precision*, *recall*, *dan f1-score* yang digunakan untuk menilai kinerja model klasifikasi secara lebih menyeluruh. (Albin Pranata et al., 2024). Rumus berikut digunakan untuk menghitung evaluasi *confusion matrix:*

1) Akurasi: seberapa banyak prediksi model yang benar dibandingkan total data.

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Jumlah\ total\ data} x\ 100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

2) *Precision:* dari semua data yang diprediksi model sebagai kelas tertentu, berapa banyak yang benar-benar termasuk kelas itu.

$$Precision = \frac{Jumlah\ data\ yang\ diprediksi\ benar\ pada\ suatu\ kelas}{Jumlah\ semua\ data\ yang\ diprediksi\ pada\ kelas\ itu} x\ 100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

3) *Recall*: dari semua data yang sebenarnya milik suatu kelas, berapa banyak yang berhasil diprediksi benar oleh model.

$$Recall = \frac{Jumlah\ data\ yang\ diprediksi\ benar\ pada\ suatu\ kelas}{Jumlah\ seluruh\ data\ sebenarnya\ pada\ kelas\ itu} x\ 100\% \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

4) *F1-Score*: nilai rata-rata antara presisi dan recall yang seimbang.

$$F1 - Score = \frac{2\ x\ (Precission\ x\ Recall)}{Precission + Recall} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

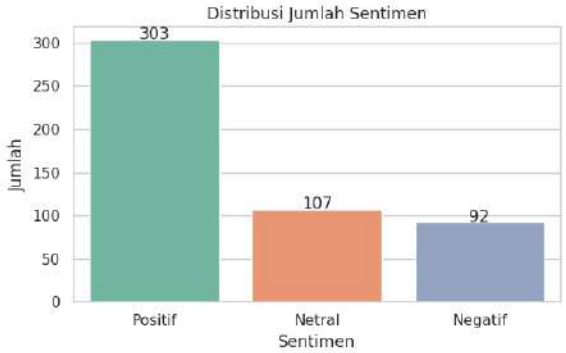
3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan adalah data dari penelitian sebelumnya milik (Perdana, 2024), data berisikan ulasan pengunjung Jembatan Repo-Repo yang diambil dari Google Maps. Data yang berhasil dikumpulkan sebanyak 502 ulasan, dan disimpan dalam format CSV. Dari penelitian sebelumnya (Perdana, 2024) proses pengumpulan data dilakukan dengan metode *crawling* selama 10 menit pada tanggal 18 April 2024, dari pukul 15:45 WITA hingga pukul 15:55 WITA.

Tabel 1. Dataset

No	Review	Sentimen
1	Bagus jembatannya tapi tutup dimalam hari, tapi bisa nongkrong dan minum kopi di pinggir sungai mahakam dan bisa mancing juga..jembatan ini jalam mau kalai ke pulau bisa juga	Positif
2	Mantapppp,,,, rawat dan jaga aset kembangkan untuk pariwisata dan identitas budaya, Jembatan adalah aset kebanggaan masyarakat Kab. Tenggarong	Positif
….	….	….
501	Suasananya romantis bgt	Positif
502	Enak sejuk bisa lewat menuju ke pulau kumala	Positif



Gambar 2. Jumlah sentimen

B. *Pre-Processing*

Tujuan pre-processing adalah untuk memperoleh data yang sudah dibersihkan sebelum masuk proses ekstraksi fitur. Tahapan yang dilakukan antara lain convert emoji, case folding, slang normalization. Berikut hasil tahapan pre-processing yang telah diproses:

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

1) *Case Folding*

Case folding adalah proses di mana semua huruf dalam sebuah teks diubah menjadi huruf kecil. Hasil dari proses ini akan disajikan dalam bentuk tabel, seperti yang ditunjukkan pada Tabel 2.

Tabel 2. *Case Folding*

No	Sebelum	Sesudah
1	Bagus jembatannya tapi tutup dimalam hari, tapi...	bagus jembatannya tapi tutup dimalam hari, tapi...
2	Mantapppp,,,, rawat dan jaga aset kembangkan u...	mantapppp,,,, rawat dan jaga aset kembangkan u...
3	Pas kesini lg ada pengecatan bagus banget liat...	pas kesini lg ada pengecatan bagus banget liat...

2) *Convert Emoji*

Convert emoji adalah proses merubah emoji menjadi teks agar dapat dibaca oleh model ekstraksi fitur IndoBERT. Hasil dari proses ini akan disajikan dalam bentuk tabel, seperti yang ditunjukkan pada Tabel 3.

Tabel 3. *Convert Emoji*

No	sebelum	sesudah
1	..... jalan kaki, pas cuaca panas , sangat terasa sekali 😋	..... jalan kaki, pas cuaca panas , sangat terasa sekali:smirking_face:
2	..... pagi hari udaranya yang segar sinar matahari yg hangat😀 😄	..... pagi hari udaranya yang segar sinar matahari yg hangat:grinning_face_with_big_eyes::grinning_face_with_smiling_eyes:
3	..... tengah jembatan, tidak bisa foto bebas 🖖	..... tengah jembatan, tidak bisa foto bebas :victory_hand::victory_hand:

3) *Slang Normalization*

Case folding adalah proses di mana semua huruf dalam sebuah teks diubah menjadi huruf kecil. Hasil dari proses ini akan disajikan dalam bentuk tabel, seperti yang ditunjukkan pada Tabel 4.

Tabel 4. *Slang Normalization*

No	Sebelum	Sesudah
1	phobia ketinggian jadi pas mau liat ke bawah tu lemes lutut	phobia ketinggian jadi pas mau melihat ke bawah tu lemas lutut
2	Arena wisata sekalian unk joging,,,alangkah baiknya kalau di pulau kumala	arena wisata sekalian untuk joging,,,alangkah baiknya kalau di pulau kumala
3	didalam gak ada yg jual jadi pastikan sebelum lewat jembatan	didalam tidak ada yang jual jadi pastikan sebelum lewat jembatan

C. Ekstraksi Fitur BERT

Ekstraksi fitur dilakukan menggunakan IndoBERT dengan model "mdhugol/indonesia-bert-sentiment-classification". Model ini telah dilatih khusus untuk Bahasa Indonesia, terutama pada teks informal seperti unggahan media sosial. Proses ini bertujuan untuk memperoleh representasi vektor dari setiap kalimat ulasan pengguna Jembatan Repo-Repo di Google Map. Berikut tahapan ekstraksi fitur IndoBERT:

1) Tokenisasi dan Pembatasan Panjang

Setelah melakukan pre-processing, teks dimasukkan ke dalam tokenizer (tokenisasi khusus IndoBERT) dari IndoBERT. Tokenizer memiliki empat tahapan yaitu memecah tiap kata menjadi token, Menambahkan token spesial [CLS] di awal dan [SEP] di akhir, merubah tiap token menjadi token ID (angka), dan terakhir Padding & truncation yaitu menambahkan token kosong (token [PAD]) jika jumlah token pada satu kalimat kurang dari 256 token lalu membatasi agar jumlah token tidak melebihi 256 karena jumlah kata dalam kalimat pada dataset maksimal memiliki 205 kata. Ini bertujuan agar setiap kata dapat masuk ke model ekstraksi fitur IndoBERT. Berikut hasil dari tokenizer IndoBERT.

Tabel 5. Hasil tokenisasi IndoBERT

Index	Token	Token_ID
0	[CLS]	2
1	Bagus	1305
2	Jembatan	5113
3	##nya	57
9	,	30468
25	hari	406
36	Bisa	166

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

Index	Token	Token_ID
37	[SEP]	3
….	….	….
254	[PAD]	0
255	[PAD]	0

Terlihat pada tabel 5 index ke-0 diisi oleh token [CLS] dan diakhir kalimat diakhiri token [SEP], token [PAD] mengisi kekosongan sampai pada index ke-255. Setiap kata dalam kalimat akan dipisahkan, kata "jembatannya" akan terbagi menjadi dua token yang berbeda yaitu "jembatan" dan "##nya" yang menandakan dalam proses tokenisasi IndoBERT dapat memisahkan kata dasar dengan kata imbuhan. Setelah itu proses merubah setiap token menjadi token ID, setiap token akan dirubah menjadi token ID, terlihat setiap token memiliki ID masing-masing akan tetapi token [PAD] memiliki nilai 0 dikarenakan token [PAD] hanyalah token yang mengisis kekosongan untuk mencapai 255 jumlah token, maka dari itu token [PAD] tidak memiliki nilai. Semua ini dilakukan agar data dapat masuk ke ekstraksi fitur IndoBERT karena model IndoBERT hanya bekerja berdasarkan ID bukan teks langsung.

2) Representasi Vektor [CLS]

Setelah tokenisasi data akan masuk ke dalam model IndoBERT. Model ini akan menghasilkan vektor yang merupakan representasi dari setiap token. Setelah itu token [CLS] akan memiliki vektor yang merangkum seluruh makna dalam kalimat, dan digunakan sebagai fitur utama untuk representasi kalimat. Setiap token [CLS] memiliki 768 dimensi yang berupa bilangan float. Berikut vektor dimensi dari token [CLS].

Tabel 6. Representasi vektor token [CLS]

No	Teks	10 dimensi pertama dari token [CLS]
1	bagus jembatannya tapi tutup dimalam hari, tapi bisa nongkrong dan minum kopi di pinggir sungai mahakam dan bisa mancing juga.. jembatan ini jalam mau kalai ke….	[1.4325, 0.1797, 0.1200, 1.9341, 0.5162, -0.6534, 1.4394, 1.2315, 1.5528, 1.3409]
2	jembatan penyebrangan ke pulau kumala. pas desember 2022 kemarin ke sana sedang direnovasi, dicat kembali. pemandangan cukup bagus... mudahan dimajukan kembali...	[-0.0322, 0.7993, -0.7336, 0.6621, -0.1367, -0.4003, 2.4210, 1.7730, 1.7547, 1.2607]
3	mantapppp,,,, rawat dan jaga aset kembangkan untuk pariwisata dan identitas budaya, jembatan adalah aset kebanggaan masyarakat kab. tenggarong	[0.6652, 0.5971, -0.8172, 1.4567, 0.4907, 0.2707, 1.0996, 1.5608, 1.2120, 1.5874]

3) Hasil Ekstraksi Fitur

Setiap kalimat dalam dataset akan dirubah menjadi vektor yang berdimensi 768, Hasil ekstraksi ini disimpan dalam *array NumPy*. Vektor ini kemudian digunakan sebagai input untuk tahapan seleksi fitur dan klasifikasi sentimen. Berikut hasil dari ekstraksi fitur IndoBERT.

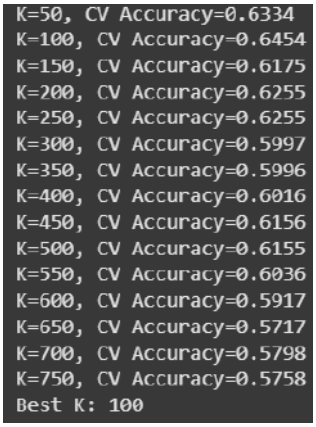
Tabel 7. Hasil ekstraksi fitur

No	Teks	10 dimensi pertama dari token [CLS]
1	bagus jembatannya tapi tutup dimalam hari, tapi bisa nongkrong dan minum kopi di pinggir sungai mahakam dan bisa mancing juga.. jembatan ini jalam mau kalai ke….	[1.4325166 0.17968872 0.11998307 1.9341043 0.5162436 -0.6534227 1.4393603 1.2314743 1.5527774 1.340939]
2	jembatan penyebrangan ke pulau kumala. pas desember 2022 kemarin ke sana sedang direnovasi, dicat kembali. pemandangan cukup bagus... mudahan dimajukan kembali...	[-0.03218023 0.79930896 -0.7335891 0.66210115 -0.13674192 -0.40029648 2.4210021 1.7729535 1.7546638 1.2606839]
3	mantapppp,,,, rawat dan jaga aset kembangkan untuk pariwisata dan identitas budaya, jembatan adalah aset kebanggaan masyarakat kab. tenggarong	[0.6651605 0.5971499 -0.8172071 1.4566596 0.49068442 0.27071485 1.0995876 1.5608149 1.2119578 1.587419]

D. Chi-Square dan *Cross Validation*

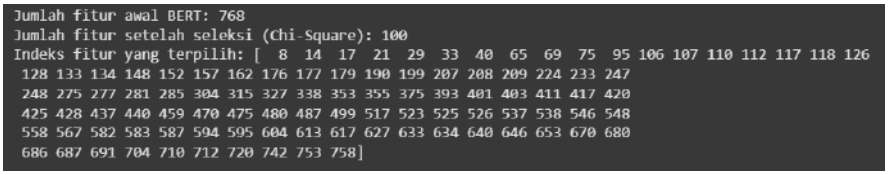
Pada tahap ini, dilakukan proses pemilihan fitur dan validasi model untuk memperoleh kombinasi jumlah fitur terbaik yang mampu memberikan performa klasifikasi optimal, untuk mengetahui jumlah fitur (nilai K) yang paling optimal, dilakukan pengujian dengan beberapa nilai K, mulai dari 50 hingga 750 dengan interval 50. Setiap nilai K akan digunakan untuk melatih model SVM dengan kernel linear. Akurasi dari setiap nilai K dievaluasi menggunakan 5-fold Cross-Validation, kemudian dibandingkan untuk memilih nilai K terbaik. Berikut fitur (nilai K) terbaik setelah melalui proses cross-validation.

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.



Gambar 3. hasil Chi-Square dan *cross validation*

Berdasarkan hasil pada gambar 3 diketahui akurasi tertinggi diperoleh pada K = 100, dengan akurasi sebesar 0.6454. Oleh karena itu, nilai K = 100 dipilih sebagai jumlah fitur optimal yang akan digunakan dalam proses pelatihan pada model klasifikasi SVM.



Gambar 4. jumlah fitur IndoBERT setelah melewati Chi-Square

Berdasarkan hasil pada gambar 4 menunjukkan hasil seleksi fitur menggunakan metode Chi-Square terhadap vektor representasi yang dihasilkan oleh IndoBERT, yang awalnya memiliki 768 dimensi fitur. Setelah proses seleksi, hanya 100 fitur terbaik yang dipilih berdasarkan nilai Chi-Square tertinggi, yaitu fitur-fitur yang dianggap paling relevan terhadap label sentimen.

E. Klasifikasi SVM

Dalam penelitian ini ada dua skema pengujian yaitu, tanpa menggunakan Chi-Square dan dengan menggunakan Chi-Square. Berikut adalah hasil akurasi menggunakan model klasifikasi SVM.

Tabel 8. Perbandingan waktu proses SVM dengan fitur yang sudah dioptimalkan

Fitur yang digunakan	Waktu *training* SVM
Keseluruhan fitur IndoBERT	0.1375 detik
Memakai 100 fitur yang sudah optimal	0.0284 detik
Penghematan waktu training dengan Chi-Square	0.1092 detik

Berdasarkan Tabel 8, Pengujian waktu pelatihan SVM menunjukkan bahwa penggunaan fitur hasil seleksi Chi-Square (k=100) membuat proses training lebih efisien. Tanpa Chi-Square, waktu pelatihan memakan waktu 0.1375 detik, sedangkan dengan Chi-Square hanya 0.0284 detik. Artinya, terdapat penghematan waktu sebesar 0.1092 detik, hasil ini membuktikan bahwa pengurangan fitur dapat mempercepat proses pelatihan model SVM.

Tabel 9 Pengujian akurasi

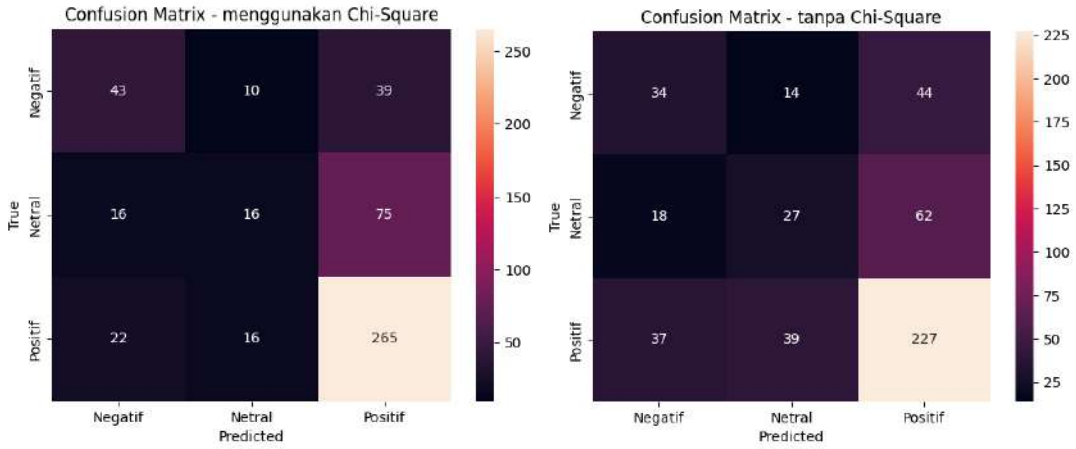
Chi-Square	Akurasi	*Precision*	*recall*	*F1-score*
Tanpa Chi-Square	57%	55%	57%	56%
Dengan Chi-Square	65%	60%	65%	61%

Berdasarkan Tabel 9, terlihat bahwa penerapan metode seleksi fitur Chi-Square memberikan peningkatan kinerja model secara signifikan dibandingkan tanpa menggunakan Chi-Square. Pada pengujian tanpa Chi-Square, model hanya mampu mencapai akurasi sebesar 57%, precision 55%, recall 57%, dan f1-score 56%. Namun, setelah dilakukan reduksi fitur dengan Chi-Square, kinerja model meningkat dengan akurasi mencapai 65%, *precision* 60%, *recall* 65%, dan f1-*score* 61%. Hal ini menunjukkan bahwa Chi-Square berhasil menyeleksi fitur yang lebih relevan, sehingga tidak hanya mengurangi dimensi data tetapi juga meningkatkan kemampuan model dalam mengklasifikasikan sentimen secara lebih tepat dan efisien.

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

F. Evaluasi *Confusion Matrix*

Untuk mengevaluasi performa dari model SVM, *confusion matrix* digunakan untuk memberikan gambaran klasifikasi benar dan salah dari masing-masing kelas sentimen.



Gambar 5. *Confusion matrix*

Pada Gambar 5, perbandingan confusion matrix antara model tanpa Chi-Square dan dengan Chi-Square menunjukkan adanya peningkatan kinerja klasifikasi setelah melakukan seleksi fitur Chi-Square. Pada confusion matrix tanpa Chi-Square, jumlah prediksi benar untuk kelas Negatif dan Netral masih relatif rendah masing-masing 34 dan 27 data, sementara kelas Positif lebih dominan terklasifikasi dengan benar 227 data. Namun setelah diterapkan Chi-Square, terlihat peningkatan yang cukup signifikan, terutama pada kelas Positif yang naik menjadi 265 data benar, serta kelas Negatif meningkat menjadi 43 data benar. Meskipun kelas Netral masih menunjukkan kesalahan klasifikasi yang cukup besar, hasilnya tetap lebih baik dengan 16 data benar dibandingkan distribusi sebelumnya. Hal ini membuktikan bahwa penerapan Chi-Square membantu mereduksi fitur yang kurang relevan sehingga model dapat mengenali pola data dengan lebih baik, menghasilkan prediksi yang lebih akurat, terutama pada kelas mayoritas. Nilai akurasi dapat dihitung sebagai berikut:

$$Akurasi = \frac{43 + 16 + 265}{43 + 10 + 39 + 16 + 16 + 75 + 22 + 16 + 265} = \frac{324}{502} X\ 100\% = 65\%$$

4. KESIMPULAN

Penelitian ini berhasil meningkatkan performa klasifikasi sentimen dibandingkan dengan penelitian sebelumnya yang hanya mencapai akurasi sebesar 58% menggunakan metode TF-IDF dan Naïve Bayes. Dengan menerapkan ekstraksi fitur menggunakan IndoBERT, seleksi fitur Chi-Square, dan klasifikasi SVM, akurasi model meningkat menjadi 65%. Selain itu, nilai evaluasi lainnya juga menunjukkan perbaikan, dengan precision sebesar 60%, recall sebesar 65%, dan f1-score sebesar 61%. Hasil ini membuktikan bahwa kombinasi IndoBERT, Chi-Square, dan SVM tidak hanya mampu meningkatkan akurasi, tetapi juga menghasilkan keseimbangan yang lebih baik antara presisi dan kemampuan model dalam mengenali data yang benar.

Penambahan metode seleksi fitur Chi-Square juga terbukti dapat meningkatkan akurasi model SVM dari 57 % menjadi 65%. Hal ini menunjukkan bahwa Chi-Square mampu menyaring fitur-fitur hasil ekstraksi BERT yang paling relevan terhadap label sentimen. Selain meningkatkan akurasi, metode ini juga membuat proses pelatihan model SVM menjadi lebih efisien dan efektif karena mengurangi dimensi fitur yang digunakan dari 768 menjadi 100, sehingga dapat menghemat waktu training model SVM sebanyak 0,1092 detik.

5. DAFTAR PUSTAKA

Albin Pranata, R., Rudiman, & Azmi Verdikha, N. (2024). METODE PEMBOBOTAN TF-IDF UNTUK KLASIFIKASI TEKS QUICK COUNT PEMILIHAN WAKIL PRESIDEN INDONESIA 2024 PADA X TWITTER DENGAN METODE SVM. 18(2), 126. https://doi.org/10.47111/JTI

Aljabar, A., Karomah, B. M., Kunci, K., & Bert, : (2024). Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film. In Journal of System and Computer Engineering (JSCE) ISSN (Vol. 5, Issue 1).

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.

Amrullah, A. Z., Sofyan Anas, A., Adrian, M., & Hidayat, J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. Jurnal, 2(1). https://doi.org/10.30812/bite.v2i1.804

Dwiki, A., Putra, A., & Juanita, S. (2021). Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN. 8(2). http://jurnal.mdp.ac.id

Gusti, I., Ngurah, A., Semadi, R., Samsudin, M., & Dharmendra, K. (2023). Perbandingan Metode Seleksi Fitur Pada Analisis Sentimen (Studi Kasus Opini PILKADA DKI 2017). Journal of Informatics, 8(1), 11–18. https://doi.org/https://doi.org/10.51211/itbi.v8i1.2408

Hayati, T. N., Fatimah, N. S., Fitria, L., & Agustin, S. (2024). Klasifikasi Lahan Perkebunan Kelapa Sawit Pada Citra Foto Udara Menggunakan Metode Local Binary Pattern dan Klasifikasi SVM. SABER: Jurnal Teknik Informatika, Sains Dan Ilmu Komunikasi, 2(3), 138–146. https://doi.org/10.59841/saber.v2i3.1399

Hermawati, A. D. (2024, June 23). Telan Anggaran Rp 29,8 Miliar, Jembatan di Kaltim Ini Bolehkan Pengunjung Memasang Gembok Sebagai Simbol Cinta. AYOBANDUNG.COM.

Maringka, R. C., Justino, R., Makarawung, N., Kunci, K., Bert, :, & Kebencian, U. (2025). OPTIMALISASI ANALISIS UJARAN KEBENCIAN ULASAN E-COMMERCE BERBASIS BERT DAN FAISS. In Journal of Information System Management (JOISM) e-ISSN (Vol. 7, Issue 1).

Najjichah, H., Syukur, A., & Subagyo, H. (2019). PENGARUH TEXT PREPROCESSING DAN KOMBINASINYA PADA PERINGKAS DOKUMEN OTOMATIS TEKS BERBAHASA INDONESIA. In Jurnal Teknologi Informasi (Vol. 15, Issue 1). http://research.

Perdana, M. R. (2024). Analisis Sentimen Ulasan Jembatan Repo-Repo di Google Maps Menggunakan Metode Naive Bayes Dengan Fitur Ekstrasi TF-IDF. Universitas Muhammadiyah Kalimantan Timur.

Saiful Nur Budiman, Sri Lesanti, & Erwan. (2024). Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes. JAMI: Jurnal Ahli Muda Indonesia, 5(2), 114–126. https://doi.org/10.46510/jami.v5i2.311

Tobing, C. J. L., IGN Lanang Wijayakusuma, & Luh Putu Ida Harini. (2025). Perbandingan Kinerja IndoBERT dan MBERT Untuk Deteksi Berita Hoaks Politik dalam Bahasa Indonesia. JST (Jurnal Sains Dan Teknologi), 14(1), 114–123. https://doi.org/10.23887/jstundiksha.v14i1.92126

https://doi.org/10.30872/atasi.v5i1.3588
2026 Adopsi Teknologi dan Sistem Informasi (ATASI) with CC BY SA license.