

CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Analisis Sentimen Pada Komentar Media Sosial Terkait Isu Joki Dengan Menggunakan Metode LSTM

Sentiment Analysis on Social Media Comments Regarding Jockey Issues Using the LSTM Method

Dimas Surya Prasetyo^{1*}, Karina Auliasari², Yosep Agus Pranoto³

^{1,2,3} Institut Teknologi Nasional Malang

Jalan Raya Karanglo km 2 Malang, Indonesia

Email: ¹diimprasyos@gmail.com, ²karina.auliasari86@gmail.com, ³yoaitn@gmail.com

*Corresponding Author

ABSTRAK

Di kalangan akademisi dan masyarakat, fenomena praktik perjokian dalam penyusunan skripsi telah menimbulkan kekhawatiran karena mengancam reputasi akademik. Media sosial menjadi sarana utama bagi masyarakat untuk mengekspresikan pendapat mereka tentang masalah ini, baik melalui kecaman maupun menjadikan kebiasaan baru. Namun, volume data yang besar, gaya bahasa yang beragam, dan penggunaan bahasa informal adalah masalah untuk menganalisis data media sosial. Untuk mencapai tujuan ini, penelitian ini menggunakan metode *Long Short-Term Memory* (LSTM) untuk menganalisis sentimen masyarakat. Metode ini dipilih karena kemampuan LSTM untuk memahami konteks kata dalam kalimat kompleks, yang membuatnya cocok untuk komentar panjang di media sosial. Hasil penelitian menunjukkan pengujian kinerja model untuk mengidentifikasi sentimen positif dan negatif, diperoleh akurasi sebesar 71%, recall sebesar 67%, dan precision sebesar 77%. Hal ini menunjukkan bahwa model mampu mengenali pola sentimen dengan cukup baik dari data yang ada.

Kata Kunci: *sentimen; perjokian; LSTM; analisa teks; integritas akademik; pemrosesan bahasa alami; opini masyarakat*

ABSTRACT

Among academics and society, the phenomenon of the practice of cheating in the preparation of theses has raised concerns because it threatens academic reputation. Social media has become the main means for people to express their opinions on this issue, either through criticism or creating new habits. However, large data volumes, diverse language styles, and the use of informal language are problems for analyzing social media data. To achieve this



This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license

goal, this research uses the Long Short-Term Memory (LSTM) method to analyze public sentiment. This method was chosen because of LSTM's ability to understand the context of words in complex sentences, which makes it suitable for long comments on social media. The research results show that testing the performance of the model to identify positive and negative sentiments, an accuracy of 71%, recall of 67%, and precision of 77% were obtained. This shows that the model is able to recognize sentiment patterns quite well from existing data.

Keywords: *sentiment, jockeying, LSTM, text analysis, academic integrity, natural language processing, public opinion.*

1. PENDAHULUAN

Sebagai seorang mahasiswa, untuk menyelesaikan studinya diwajibkan untuk menyusun skripsi. Tugas akhir skripsi ini tidak hanya menguji kemampuan siswa dalam memahami dan mengaplikasikan pengetahuan akademis tetapi juga mengasah kemampuan mereka untuk berpikir kritis, membuat kesimpulan, dan melakukan penelitian. Namun, praktik perjokian di mana pihak ketiga secara tidak sah membantu siswa menyelesaikan tugas akhir atau skripsi – menjadi lebih umum akhir-akhir ini. Di tengah masyarakat, berbagai reaksi muncul mengenai fenomena ini, terutama dari segi akademik, praktik ini merusak prinsip akademis dan mengurangi integritas pendidikan [1].

Banyak masyarakat, khususnya pengguna media sosial, aktif menyuarakan opini mereka tentang isu ini pada konten – konten yang membahas mengenai praktik perjokian. Ada yang mengecam perjokian karena dianggap melakukan kecurangan atau melanggar integritas akademik, sementara sebagian lainnya memberikan opini yang berbeda, baik karena melihat permasalahan di balik fenomena ini atau karena sudah dianggap sebagai sesuatu yang umum terjadi atau dinormalisasi oleh masyarakat [2].

Oleh karena itu, analisis yang lebih mendalam terhadap sentimen masyarakat mengenai isu perjokian menjadi sangat penting. Hasil analisis ini dapat dimanfaatkan untuk mengevaluasi sistem pembelajaran dan merancang kebijakan yang lebih efektif dalam meningkatkan integritas akademik. Namun, menganalisis opini masyarakat dari media sosial bukanlah tugas yang mudah, mengingat volume data yang besar, keberagaman gaya penulisan, serta penggunaan bahasa informal dan slang dalam banyak komentar.

Penelitian ini menggunakan metode *Long Short-Term Memory* (LSTM) dipilih sebagai solusi yang efektif dalam menganalisis sentimen masyarakat. LSTM adalah bagian dari algoritma jaringan saraf yang dikenal mampu memahami konteks kata dalam kalimat secara lebih baik. Karena kebanyakan pada data teks komentar di media sosial terdapat kalimat – kalimat panjang dan kompleks. Berdasarkan hal tersebut, metode ini dianggap penulis cocok untuk menganalisis sentimen masyarakat terhadap perjokian, terutama dengan variasi gaya bahasa dan penulisan. Kemudian hasil analisa ini dapat digunakan untuk evaluasi proses pembelajaran oleh institusi pendidikan dan pembuat kebijakan [3].

2. TINJAUAN TEORI

2.1 Analisis Sentimen

Analisis sentimen adalah proses yang digunakan untuk memahami dan mengidentifikasi emosi atau pendapat yang terkandung dalam teks yang berkaitan dengan berbagai topik atau

entitas tertentu. Dalam beberapa tahun terakhir, analisis sentimen telah menjadi sub bagian dari text mining yang semakin penting. Fokus utamanya adalah mengekstrak opini, sentimen, dan emosi dari teks yang tidak terstruktur, seperti yang ditemukan di media sosial, dan menentukan apakah opini tersebut bersifat positif, negatif, atau netral. Analisis ini memberikan informasi penting yang dapat diolah untuk berbagai tujuan strategis dan analitis [4].

2.2 Media Sosial

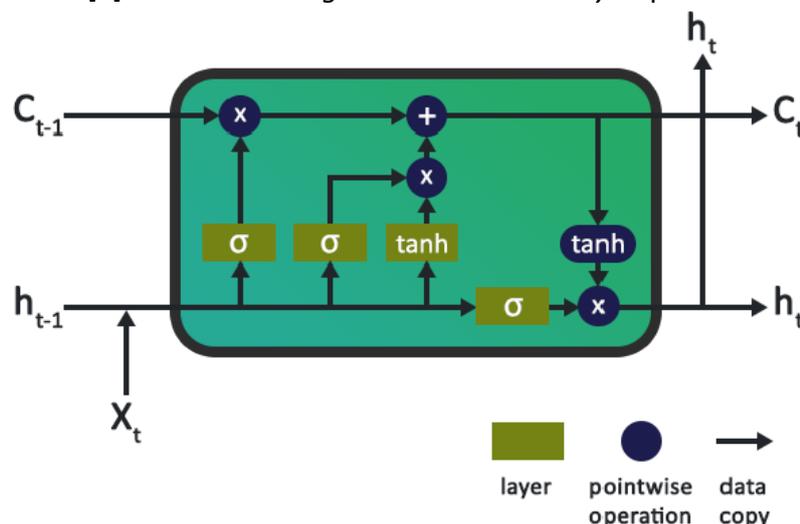
Media sosial, sebagai platform berbasis internet yang menyampaikan berbagai informasi dari seluruh dunia, memungkinkan individu dan kelompok berinteraksi dengan mudah dan efektif tanpa batasan geografis. Media sosial tidak hanya berfungsi sebagai alat komunikasi, tetapi juga berperan penting dalam mengubah perilaku, pola pikir, dan perspektif publik, sehingga internet, media sosial, dan penggunaannya menjadi satu kesatuan dalam kehidupan manusia [5].

2.3 Joki

Joki adalah definisi untuk Tindakan dimana seseorang membayar atau meminta bantuan orang lain untuk menggantikan atau melakukan pekerjaan atau tugas atas nama mereka. Tindakan ini banyak sekali terjadi pada beberapa tahun kebelakang, Orang yang melakukan tugas tersebut dikenal sebagai "penjoki" atau "joki". Dalam bidang akademik dan lainnya, tindakan ini dipandang tidak etis karena melanggar prinsip kejujuran, integritas dan keadilan [6].

2.4 Long Short-Term Memory

Long Short-Term Memory (LSTM) adalah salah satu metode deep learning yang dapat mengatasi masalah ketergantungan jangka panjang dalam pemrosesan data sekuensial. Dengan menambahkan struktur sel memori dan tiga gerbang (gate) utama seperti gerbang masukan (input gate), gerbang pelupaan (forget gate), dan gerbang keluaran (output gate), metode ini mengembangkan *Recurrent Neural Networks* (RNN). Fungsi utama LSTM adalah untuk mengabaikan atau menghapus informasi yang tidak diperlukan dan mempertahankan informasi yang relevan dalam memori jaringan untuk jangka waktu yang lama. Ini membuat metode ini sangat efektif dalam menangani masalah seperti "vanishing gradient" dan "exploding gradient" [7]. Arsitektur *Long Short-Term Memory* dapat diamati pada gambar 3.



Gambar 1. Arsitektur Metode *Long Short-Term Memory*[8]

LSTM (*Long Short-Term Memory*) memiliki struktur gerbang yang unik untuk menyaring informasi secara efisien dari urutan data. Model LSTM terdiri dari beberapa unit gerbang: gerbang input (input gate), gerbang lupa (forget gate), dan gerbang output (output gate). Berikut adalah persamaan dari setiap gerbang:

1) Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

2) Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

3) Cell State

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

4) Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

5) Hidden State

$$h_t = o_t \cdot \tanh(C_t) \quad (5)$$

Keterangan:

W_f = matriks bobot untuk forget gate

h_{t-1} = hidden state dari waktu sebelumnya

x_t = input pada waktu saat ini

b_f = bias untuk forget gate

σ = fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 hingga 1 untuk menentukan seberapa banyak informasi dari *cell state* sebelumnya yang perlu dihapus

\tanh = fungsi aktivasi tangen yang menghasilkan nilai antara -1 hingga 1 untuk secaraimbang mempertahankan dan menghapus informasi.

$\tanh(C_t)$ = fungsi aktivasi tangen untuk mengubah cell state menjadi output tersembunyi dengan rentang -1 hingga 1

2.5 Evaluasi Klasifikasi

Proses yang sangat penting dalam pengembangan model machine learning dan deep learning adalah evaluasi. Ini dilakukan untuk menilai kinerja model yang telah dikembangkan dan untuk menentukan apakah model tersebut cukup baik atau memerlukan perbaikan. Beberapa metrik penting untuk evaluasi klasifikasi adalah *Recall*, *Precision*, dan *Accuracy*. Metrik-metrik ini dapat diperoleh dari Confusion Matrix, sebuah tabel yang digunakan untuk menunjukkan hasil klasifikasi model dan menunjukkan berapa banyak prediksi yang benar dan salah untuk setiap kelas [9].

Tabel 1. *Confusion Matrix*

Predicted Class	True Class	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Dari Keempat elemen confusion matrix diatas, selanjutnya akan digunakan untuk menghitung evaluasi yaitu:

1) Accuracy, mengukur seberapa baik model melakukan klasifikasi data secara keseluruhan.

Ini dapat dihitung dengan menggunakan rumus berikut:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (6)$$

2) Precision, mengukur seberapa akurat model dalam memprediksi kelas positif. Ini dapat dihitung dengan menggunakan rumus berikut:

$$Precision = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

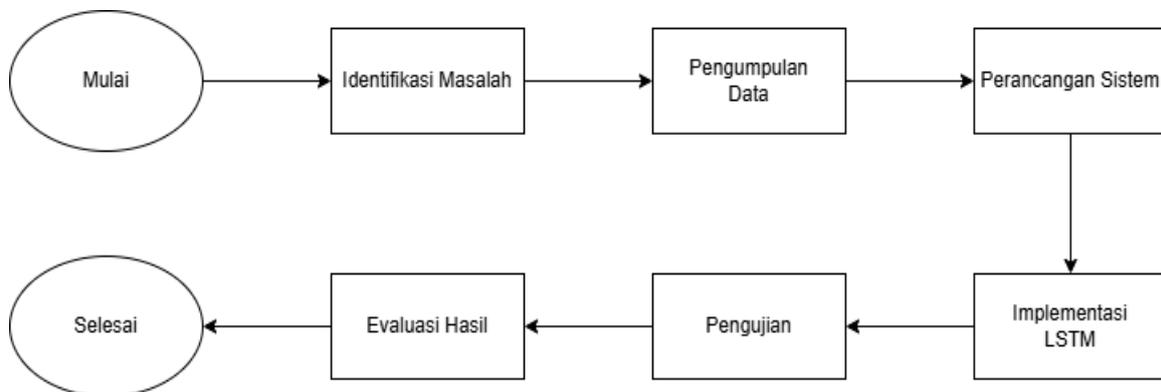
3) Recall, mengukur kemampuan model untuk mendeteksi kasus positif nyata. Ini dapat dihitung dengan menggunakan rumus berikut:

$$Recall = \frac{TP}{TP+FP} \times 100\% \quad (8)$$

3. METODE

3.1. Metode Penelitian

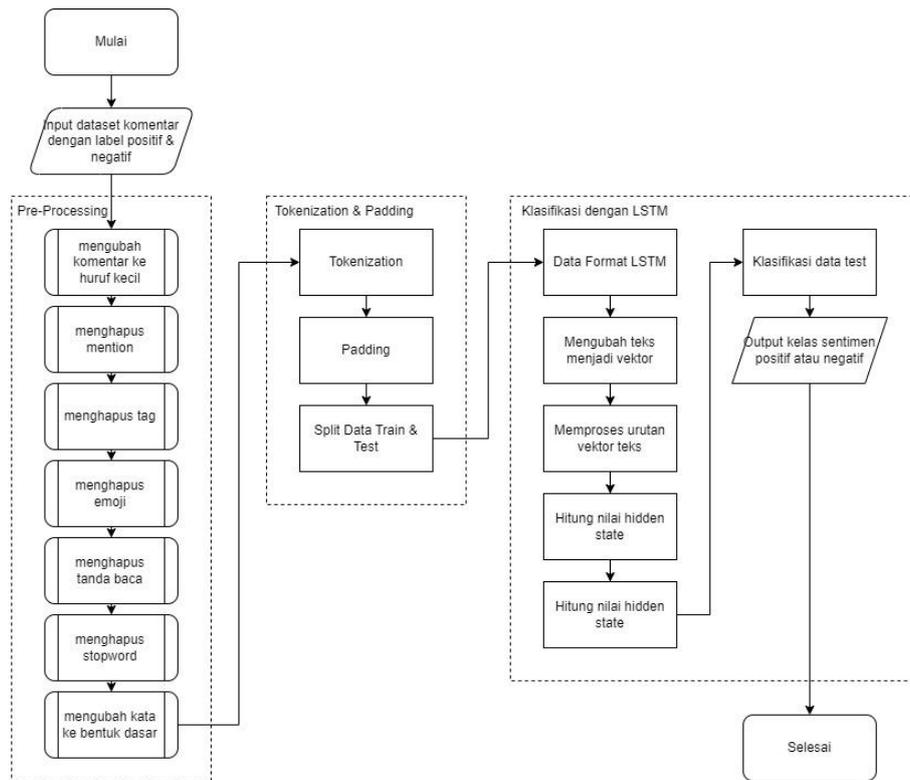
Metode penelitian ini dimulai dengan identifikasi masalah, di mana isu utama yang ingin diselesaikan diidentifikasi, diikuti oleh pengumpulan data yang relevan untuk mendukung penelitian, seperti data teks untuk analisis sentimen. Selanjutnya, dilakukan perancangan sistem secara konseptual untuk menunjukkan struktur dan alur kerja model LSTM. Setelah perancangan selesai, tahap implementasi LSTM dilakukan dengan melatih model menggunakan data yang dikumpulkan. Model yang sudah diimplementasikan kemudian diuji pada tahap pengujian untuk mengevaluasi performanya dengan data uji. Hasil pengujian tersebut dianalisis pada tahap evaluasi hasil untuk menilai keefektifan model berdasarkan metrik performa. Akhirnya, proses ditutup dengan tahap selesai, setelah evaluasi hasil dilakukan dan kesimpulan ditarik. Diagram alur metode dapat diamati pada gambar 1 [10].



Gambar 2. Diagram Alur Metode Penelitian

3.2. Flowchart Metode

Flowchart metode yang akan digunakan dalam sistem yang menggunakan *Long Short-Term Memory* dapat dilihat pada gambar 2 berikut.



Gambar 3. Flowchart Metode *Long Short-Term Memory*

Terdapat beberapa tahap dalam proses analisis sentimen ini yang tergambar dalam *flowchart* metode pada gambar 2 berikut adalah penjelasan prosesnya secara garis besar:

1) Pra-Pemrosesan

Data mentah yang dari komentar akan dibersihkan melalui berbagai proses. Ini akan membuat data siap untuk digunakan dalam pelatihan model. Mengubah teks menjadi huruf kecil, menghapus mention, tag, emoji, tanda baca, stopwords, dan mengubah kata-kata menjadi bentuk dasar (stemming atau lemmatization) adalah beberapa tindakan yang dilakukan pada proses ini.

2) *Tokenization & Padding*

Setelah proses pembersihan, teks diubah menjadi token, yaitu dibagi menjadi kata atau sub-kata. Padding kemudian diterapkan untuk menyesuaikan panjang setiap urutan data, memungkinkan model untuk bekerja lebih cepat.

3) *Split Data Train & Test*

Pada proses ini data akan dibagi menjadi dua bagian, yaitu data pelatihan (training) dan data pengujian (*testing*). Data pelatihan digunakan untuk mengajarkan model, sedangkan data pengujian digunakan untuk mengevaluasi seberapa baik model dalam mengklasifikasi pada data yang belum pernah dilihat sebelumnya.

4) Klasifikasi LSTM

Sebelum dimasukkan ke dalam model, data harus diformat untuk memenuhi persyaratan LSTM. Ini memastikan bahwa data dapat diproses dengan cara yang tepat. Karena model LSTM memerlukan input dalam bentuk angka, teks kemudian diubah menjadi vektor numerik. Setelah itu, LSTM memproses urutan vektor dalam teks untuk memahami konteks setiap kata yang didasarkan pada urutan kata dalam kalimat, yang membantu model dalam menentukan sentimen komentar. Selama proses pemrosesan, LSTM juga

menghitung nilai posisi tersembunyi untuk setiap kata yang menyimpan informasi penting tentang teks, yang memungkinkan model untuk mengingat detail penting dalam teks. Setelah model dilatih, sentimen komentar diklasifikasikan menjadi positif atau negatif dengan menggunakan data uji yang belum pernah dilihat sebelumnya. Klasifikasi akhir dibuat berdasarkan pola yang telah dipelajari.

3.3. Arsitektur Sistem

Arsitektur sistem menunjukkan proses klasifikasi sentimen komentar dari data media sosial seperti YouTube dan Instagram. Data komentar dan datetime yang diperoleh melalui scraping termasuk dalam input. Data melalui proses pre-processing, yang mencakup pemilihan fitur, normalisasi teks (menggunakan lowercase, menghapus tanda baca, tag, pernyataan, emoji, dan spasi berlebih), penghapusan stopwords, konversi slang ke kata baku, dan stemming. Selanjutnya, pemodelan dilakukan dengan memilih kolom teks dan label, menghilangkan label netral, tokenisasi teks, memasukkan layer neural network, dan mengembedding. Sebelum model dilatih, data dibagi menjadi training dan testing. Arsitektur sistem dapat diamati pada tabel

Tabel 2. Arsitektur Sistem

Input	Proses	Output
1) Full text 2) label	<p>Text Pre-processing:</p> <ul style="list-style-type: none"> - Seleksi fitur(drop semua kolom yang tidak digunakan, kecuali kolom komentar) - Lowercase (semua kata dijadikan huruf kecil) - Menghapus tanda baca - Menghapus tag dan mention - Menghapus emoji - Menghapus spasi berlebih - Mengganti kata slang/gaul menjadi kata baku - Stopword (menghapus kata yang dianggap tidak penting (dan, di , dengan, saya, dll)) - Stemming (mengubah kata menjadi bentuk sederhananya) <p>Modelling:</p> <ul style="list-style-type: none"> - Memilih kolom teks dan label - Menghilangkan label netral - Melakukan tokenisasi untuk mengubah teks menjadi urutan angka (contoh: "terima kasih" menjadi [12,45]) - menjadi vektor (embedding layer) - Memproses data sekuensial teks (LSTM layer) - Menghasilkan keluaran klasifikasi (dense layer) - Mengkonversi label menjadi numerik (encoding) - Memisahkan data train dan test - Melatih model 	<ul style="list-style-type: none"> 1) Klasifikasi Positif atau Negatif tiap komentar 2) Distribusi positif negatif disajikan dalam bentuk grafis 3) Frekuensi kata pada sentimen positif dan negatif (wordcloud)

- Evaluasi:
- Memvisualisasi kinerja model dengan metrik (accuracy dan loss function)

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pada penelitian ini pada proses pengumpulan data menggunakan metode scraping dan metode klasifikasi *Long Short-Term Memory* dalam analisis sentimen. Pengumpulan data dilakukan scraping dari platform YouTube dan Twitter. Jumlah komentar yang di scraping ada sejumlah 4055 data.

Tabel 3. Data hasil scraping pada YouTube dan twitter

username	full_text	created_at
@username1	saya tau banyak jasa joki skripsi di medsos tapi.. apa iya selaku itu gimana testimoni alumni yg bisa sukses di kerjaan meski pakai joki tugas akhir? https://t.co/F7wqZxHAMm	2024-06-05T05:15:45Z
@username1	secara kognitif mengingat itu baru jadi pondasi bangku kuliah bahkan Tugas Akhir itu mengasah kognitif sampai tingkat analyzing evaluating dan creating tanpa bantuan joki lulusan S1 masih jadi beban banget Higher-Order Thinking Skills nya apalagi pada pake joki https://t.co/yYosRiA9sT	2024-07-23T16:03:23Z
@username2	Nih baca baik-baik. Walaupun beberapa permintaan klo khusus buat KTI yang biasanya jumlah halaman dibatasi itu artinya kalian harus meringkas sendiri tanpa ngurangin nilai masalah. Klo skripsi mah makin panjang makin ahoy iya enggak? https://t.co/JiTiZ1ho4b	2020-01-01T06:44:32Z
@username3	Tidak semua pemakai joki itu adalah korban sistem . Sebagian (besar mungkin?) adalah karena memang ga punya integritas. Trus cari2 alasan pembenaran A sd Z. Kalau ada yg ga sanggup nulis skripsi tapi mayoritas temannya bisa artinya ybs memang ga cocok kuliah.	2024-07-22T16:46:05Z
@username4	Mahasiswa pakai joki skripsi itu masalah. Dosen nebeng nama di artikel karya mahasiswa itu masalah. Banyaknya jurnal predator itu juga masalah. Kalau ditarik pangkalnya sama: sistem yang kurang menghargai proses dan terlalu fokus pada pencapaian metrik-metrik belaka.	2024-07-22T02:20:28Z

4.2 Pra – Pemrosesan

Setelah data dikumpulkan dan kemudian dilakukan pra – pemrosesan data dimana data dibersihkan dari tanda baca, link, emoji, menghapus stopword, stemming, hingga proses tokenisasi. sehingga data yang sudah dikumpulkan agar data siap digunakan untuk proses modelling. Pra - pemrosesan ini dapat diamati pada proses berikut:

- 1) Mengubah teks menjadi huruf kecil

Langkah pertama adalah mengubah teks menjadi huruf kecil. Ini dilakukan untuk mengurangi perbedaan kata yang disebabkan oleh kapitalisasi yang berbeda, seperti pada tabel 4.

Tabel 4. Proses lowercasing

Sebelum	Sesudah
Cuma mau ngingetin bahwa joki itu ada ancaman pidana. Soalnya diatur di UU Sisdiknas 2003 dan KUHP. https://t.co/Oz56vXHorf	cuma mau ngingetin bahwa joki itu ada ancaman pidana. soalnya diatur di uu sisdiknas 2003 dan kuhp. https://t.co/oz56vxhorf

2) Menghapus mention atau username

Komentar yang berisi username yang diawali dengan @ akan dihapus dari teks karena informasi ini biasanya tidak relevan untuk analisis isi teks seperti pada tabel 5.

Tabel 5. Proses menghapus mention atau username

Sebelum	Sesudah
@username1 hmm saya tidak setuju ya. skripsi terlalu kompleks kalo dibandingin sama tugas2 lainnya. di skripsi kamu gunain semua pengetahuan yang udah kamu pelajarin jadi kerangka berpikir kamu dalam membuat suatu tulisan ilmiah. inget skripsi yang baik adalah skripsi yang selesai.	hmm saya tidak setuju ya. skripsi terlalu kompleks kalo dibandingin sama tugas2 lainnya. di skripsi kamu gunain semua pengetahuan yang udah kamu pelajarin jadi kerangka berpikir kamu dalam membuat suatu tulisan ilmiah. inget skripsi yang baik adalah skripsi yang selesai.

3) Menghapus link atau tautan

Link atau tautan dihapus dari teks karena tidak memberikan informasi yang relevan mengenai opini masyarakat seperti pada tabel 6.

Tabel 6. Proses menghapus link atau tautan

Sebelum	Sesudah
cuma mau ngingetin bahwa joki itu ada ancaman pidana. soalnya diatur di uu sisdiknas 2003 dan kuhp. https://t.co/oz56vxhorf	cuma mau ngingetin bahwa joki itu ada ancaman pidana. soalnya diatur di uu sisdiknas 2003 dan kuhp.

4) Menghapus karakter non – huruf dan angka

Dalam langkah ini, karakter seperti tanda baca, simbol, dan angka dihilangkan dari teks seperti pada tabel 7.

Tabel 7. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
cuma mau ngingetin bahwa joki itu ada ancaman pidana. soalnya diatur di uu sisdiknas 2003 dan kuhp.	cuma mau ngingetin bahwa joki itu ada ancaman pidana soalnya diatur di uu sisdiknas dan kuhp

5) Menghapus spasi berlebih

Dalam langkah ini, spasi berlebihan diantara kata – kata pada teks akan dihapus agar jarak antar teks konsisten seperti pada tabel 8.

Tabel 8. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
cuma mau ngingetin bahwa joki itu ada ancaman pidana soalnya diatur di uu sisdiknas dan kuhp	cuma mau ngingetin bahwa joki itu ada ancaman pidana soalnya diatur di uu sisdiknas dan kuhp

6) Mengganti kata tidak baku

Dalam langkah ini, kata – kata tidak baku dan kata slang atau gaul akan diganti menjadi kata baku seperti pada tabel 9.

Tabel 9. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
jangan sok tahu deh lu soal joki lihat video baru biar lu paham	jangan sok tahu deh kamu soal joki lihat video baru biar kamu paham

7) Mengganti kata tidak baku

Dalam langkah ini, kata – kata tidak baku dan kata slang atau gaul akan diganti menjadi kata baku seperti pada tabel 10.

Tabel 10. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
jangan sok tahu deh lu soal joki lihat video baru biar lu paham	jangan sok tahu deh kamu soal joki lihat video baru biar kamu paham

8) Menghapus stopwords

Dalam langkah ini proses menghapus kata-kata umum (seperti dan, di, itu, yang) yang tidak tidak relevan untuk analisis teks seperti pada tabel 11.

Tabel 11. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
cuma mau mengingatkan bahwa joki itu ada ancaman pidana soalnya diatur di uu sisdiknas dan kuhp	cuma mau mengingatkan joki ancaman pidana soalnya diatur uu sisdiknas kuhp

9) Stemming

Stemming adalah proses menghilangkan imbuhan dan mengubah kata menjadi bentuk dasarnya seperti pada tabel 12.

Tabel 12. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
cuma mau mengingatkan joki ancaman pidana soalnya diatur uu sisdiknas kuhp	cuma mau ingat joki ancam pidana soal atur uu sisdiknas kuhp

10) Tokenisasi

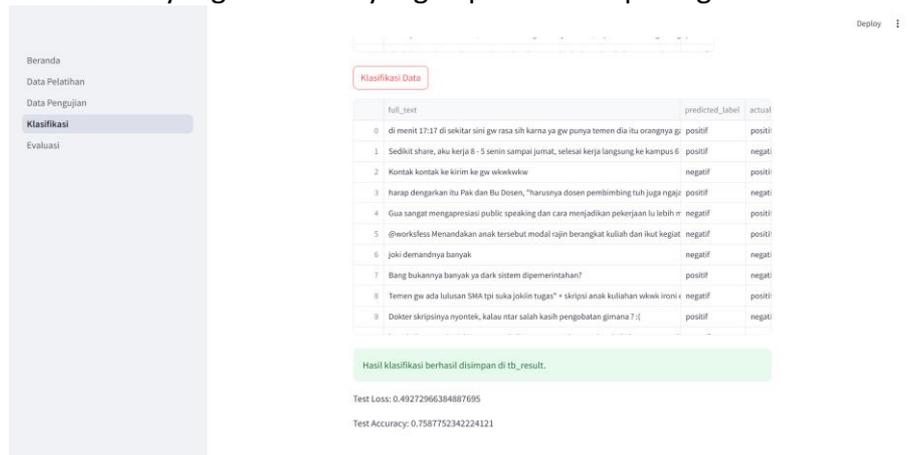
Proses tokenisasi adalah pembagian teks menjadi bagian-bagian kecil yang disebut token, seperti kata atau sub-kata, dan kemudian mengubahnya menjadi representasi numerik yang berbasis indeks. seperti pada tabel 13.

Tabel 13. Proses menghapus karakter non – huruf dan angka

Sebelum	Sesudah
cuma mau ingat jokiancam pidana soal atur uu sisdiknas kuhp	[45, 17, 196, 1, 1184, 976, 85, 310, 735, 1706, 2318]

4.3 Klasifikasi

Dengan menggunakan model LSTM, data komentar diklasifikasikan menjadi sentimen positif dan negatif. Hasil klasifikasi ini disimpan dalam kolom predicted_label, yang menunjukkan apakah komentar tersebut dikategorikan sebagai sentimen positif atau negatif. Data sentimen aktual dari setiap komentar ditemukan di kolom actual_label, yang merupakan hasil pelabelan manual yang dilakukan yang dapat diamati pada gambar 4.



Gambar 4. Tampilan halaman Klasifikasi

Berdasarkan informasi evaluasi model, pada data uji (test data), model mencapai Test Accuracy sebesar 74.3%, yang berarti sekitar 74.3%, dari prediksi model sesuai dengan label sebenarnya. Selain itu, model memiliki Test Loss sebesar 0.49, yang menunjukkan seberapa jauh prediksi model dari label sebenarnya, dengan nilai loss yang lebih rendah menunjukkan kinerja yang lebih baik.

4.4 Evaluasi Hasil

Pada halaman evaluasi hasil klasifikasi ditampilkan hasil dari prediksi mesin menggunakan metode LSTM pada kolom predicted_label dan pelabelan manual yang ada pada kolom actual_label dapat diamati pada gambar 5.

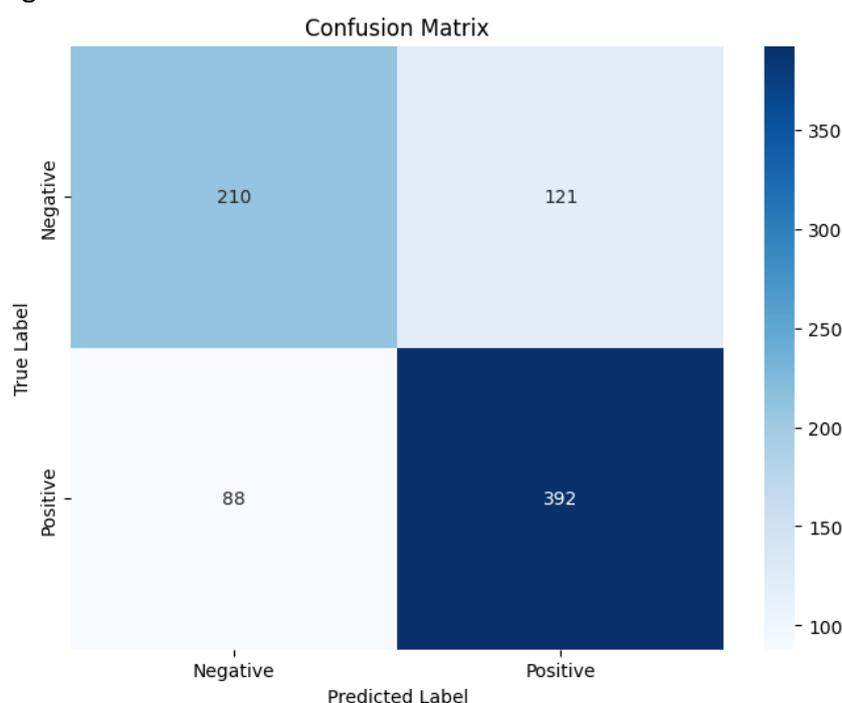


Gambar 5. Tampilan halaman Klasifikasi

Dengan menggunakan model LSTM, hasil klasifikasi sentimen pada data komentar ditunjukkan oleh Kolom Full_text berisi teks komentar yang diklasifikasikan, Predicted_label adalah label sentimen yang diprediksi oleh model, dan Actual_label adalah label sentimen yang sebenarnya, yang dihasilkan dari pelabelan manual. Juga ditampilkan wordcloud atau kata yang sering muncul pada data komentar positif dan negatif.

4.5 Pengujian Klasifikasi

Pengujian performa metode *Long Short-Term Memory* menggunakan *Confusion Matrix* menunjukkan bahwa model ini memiliki kemampuan untuk mengenali sampel positif dengan baik, tetapi masih ada kesalahan dalam mengidentifikasi sampel negatif, yang memiliki tingkat keakuratan yang lebih rendah.



Gambar 6. Tampilan halaman Klasifikasi

Berikut adalah penjelasan lebih detail mengenai grafik confusion matrix pada gambar 5:

- 1) True Negatives (TN): Sebanyak 210 sampel Negative diprediksi dengan benar sebagai Negative.
- 2) False Positives (FP): Sebanyak 121 sampel Negative diprediksi secara salah sebagai Positive.
- 3) False Negatives (FN): Sebanyak 88 sampel Positive diprediksi secara salah sebagai Negative.
- 4) True Positives (TP): Sebanyak 392 sampel Positive diprediksi dengan benar sebagai Positive.

Precision, Accuracy, dan Recall metrik digunakan untuk mengevaluasi performa model *Long Short-Term Memory* (LSTM) untuk klasifikasi sentimen. Nilai-nilai dari *Confusion Matrix* digunakan sebagai dasar untuk evaluasi ini. Berikut ini adalah penjelasan tentang perhitungan untuk masing-masing metrik:

$$Accuracy = \frac{392+210}{392+210+121+88} \times 100\% = 74.3\% \quad (9)$$

$$Precision = \frac{392}{392+121} \times 100\% = 76.4\% \quad (10)$$

$$Recall = \frac{392}{392+88} \times 100\% = 81.7\% \quad (11)$$

Model LSTM menunjukkan akurasi 74.3%, precision 76.4%, dan recall 81.7%, yang menunjukkan kinerja yang cukup baik.

5. KESIMPULAN

Dengan menggunakan metode *Long Short-Term Memory* (LSTM), penelitian ini menghasilkan model klasifikasi sentimen yang berhasil untuk menganalisis sentimen positif dan negatif pada 4.055 data komentar di media sosial. Model yang dihasilkan menunjukkan kemampuan yang cukup baik untuk menyelesaikan tugas klasifikasi sentimen, dengan akurasi sebesar 74,3%, ketepatan sebesar 76,4%, dan recall sebesar 81,7%. Namun, masih ada kesalahan dalam menemukan sentimen negatif.

Hasil penelitian ini mendukung pengembangan analisis sentimen di bidang akademik dan praktis. Penelitian ini menunjukkan metode LSTM efektif dalam mengolah data teks yang memiliki fitur informal dan kompleks dari media sosial. Secara praktis, hasil analisis ini dapat digunakan oleh institusi pendidikan dan pembuat kebijakan untuk menilai apakah mereka dapat meningkatkan kredibilitas akademik dengan membuat program untuk mencegah perjokian atau memperbaiki sistem pendidikan.

Untuk meningkatkan akurasi dan kemampuan model untuk membedakan sentimen dengan lebih baik, disarankan untuk menerapkan teknik penyesuaian hyperparameter, memperluas dataset latih, atau menggunakan model arsitektur yang lebih kompleks, seperti transformer atau bidirectional LSTM. Selain itu, penelitian ini dapat diperluas dengan melakukan analisis sentimen multikategori atau menggunakannya dalam konteks lain yang tidak berkaitan dengan masalah pendidikan.

REFERENSI

- [1] Sari, E. A. P., & Kurniawan, D. J. (2023). Faktor – Faktor Yang Mempengaruhi Penggunaan Jasa Joki Tugas Oleh Pelajar Dan Mahasiswa. *KOLONI*, 2(2), 93–101.
- [2] Baihaqi, M. H., Ni'mah, E. R. A., Rohmah, F. N., Husna, A. F. S., Amarthani, N. T., & Nabilla, S. Y. (2024). Persepsi Mahasiswa Universitas Negeri Semarang terhadap Jasa Joki Tugas. *Jurnal Mediasi*, 3(1), 25–34.
- [3] Ghilmany, M. H. (2024). Analisis sentimen pada ceramah Islam berbahasa Indonesia menggunakan metode long short-term memory. Skripsi. Etheses UIN Malang.
- [4] Deltania, D., Garno, G., & Jamaludin, A. (2024). Analisis Sentimen Publik Terhadap Invasi Zionis Kepada Hamas Menggunakan Support Vector Machine (SVM). *Jurnal Mahasiswa Teknik Informatika*. Vol. 8 No. 4, pp. 4465-4472.
- [5] Peni, P., Suarna, N., & Prihartono, W. (2024). Analisis Sentimen Generasi Z Terhadap Pengetahuan Tentang Kehidupan Gay Menggunakan Algoritma Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika*. Vol. 8 No. 3, pp. 2901-2907.

- [6] Indra, F., Andreina, K., Kania, N. S., & Valensky, S.(2023). Peran Joki Dalam Perkuliahan Terhadap Etika: Tinjauan Dari Perspektif Mahasiswa Dan Dosen. Jurnal Bangun Manajemen. Vol. 2 No. 1, pp. 113-119.
- [7] Riyantoko, P. A., Fahrudin, T. M., Prasetya, D. A., Trimono, T., & Timur, T. D. (2022). Analisis Sentimen Sederhana Menggunakan Algoritma LSTM dan BERT untuk Klasifikasi Data Spam dan Non-Spam. Prosiding Seminar Nasional Sains Data. Vol 2 No 1, pp. 103-111.
- [8] Fahira, Fahira & Prianto, Cahyo. (2023). Prediksi Pola Kedatangan Turis Mancanegara dan Menganalisis Ulasan Tripadvisor dengan LSTM dan LDA. Jurnal Tekno Insentif. 17. 69-83. 10.36787/jti.v17i2.1096.
- [9] Raflizar D. Y. (2024). Analisis Sentimen Untuk Deteksi Ujaran Kebencian Pada Media Sosial Terkait Pemilu 2024 Menggunakan Metode Support Vector Machine. Skripsi thesis. Institut Teknologi Nasional Malang.
- [10] Rani, M., Prawira D., & Mutiah N. (2023). Analisis Sentimen TerhadapVaksin COVID-19 Menggunakan Naive Bayes Classifier, Support Vector Machine dan K-Nearest Neighbors. Journal of ComputerEngineering, System and Science. 8(1). 1-11.