

Analisis Sentimen Pada Pembatalan Tuan Rumah Indonesia Di Piala Dunia U-20 Menggunakan Fasttext *Embeddings* Dan Algoritma *Recurrent Neural Network*

Aan Evian Nanda

Universitas Pembangunan Nasional Veteran Jawa Timur

Andreas Nugroho Sihananto

Universitas Pembangunan Nasional Veteran Jawa Timur

Agung Mustika Rizki

Universitas Pembangunan Nasional Veteran Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294

Korespondensi penulis: aan.eviananda74@gmail.com

Abstract. Indonesia's golden opportunity to take part in a world-class soccer competition at the U-20 World Cup competition was wiped out, as FIFA gave the decision to revoke Indonesia's status as host of the U-20 World Cup. Indonesian netizens who felt disappointed expressed their opinions and trended on social media Twitter. This research focuses on sentiment analysis of tweets using a combination of FastText embeddings method for word vectorization and using LSTM type RNN algorithm for sentiment classification. The dataset used totals 9,645 data consisting of 4,141 positive data and 5,504 negative data taken from March 29, 2023 to April 05, 2023. The test results on the LSTM model provide the best performance with an accuracy value of 74.92%, precision 74.74%, recall 74.92%, and f1-score 74.78%. The conclusion of this research is that the majority of datasets have negative sentiments, which means that people are more likely to give negative opinions than to provide support to Indonesian football which is experiencing problems. It is hoped that with this conclusion in the future people will better control their opinions and provide positive opinions when Indonesia is experiencing problems.

Keywords: Sentiment, Twitter, U-20 World Cup, FastText, LSTM, RNN.

Abstrak. Peluang emas Indonesia untuk mengikuti kompetisi bola kelas dunia pada kompetisi Piala Dunia U-20 musnah, karena FIFA memberi keputusan pencabutan status Indonesia sebagai tuan rumah Piala Dunia U-20. Netizen Indonesia yang merasa kecewa mengungkapkan opini mereka dan menjadi *trending* di media sosial Twitter. Penelitian ini berfokus pada analisis sentimen *tweet* menggunakan kombinasi metode FastText *embeddings* untuk vektorisasi kata dan menggunakan algoritma RNN jenis LSTM untuk klasifikasi sentimen. Dataset yang digunakan berjumlah total 9.645 data yang terdiri dari 4.141 data positif dan 5.504 data negatif yang diambil pada rentang waktu 29 maret 2023 s.d 05 April 2023. Hasil pengujian pada model LSTM memberikan performa terbaik dengan nilai *accuracy* sebesar 74.92%, *precision* 74.74%, *recall* 74.92%, dan *f1-score* 74.78%. Kesimpulan dari penelitian ini adalah mayoritas dari *dataset* memiliki sentimen negatif yang berarti masyarakat lebih cenderung memberikan opini negatif daripada memberikan dukungan kepada sepak bola Indonesia yang sedang mengalami masalah. Diharapkan dengan adanya kesimpulan tersebut kedepannya masyarakat lebih mengontrol opini mereka dan memberikan opini yang positif saat Indonesia sedang mengalami masalah.

Kata kunci: Sentimen, Twitter, Piala Dunia U-20, FastText, LSTM, RNN.

LATAR BELAKANG

Sepak bola merupakan salah satu olahraga paling populer di dunia. Tidak ada bentuk budaya populer lain yang dapat menimbulkan gairah kebersamaan dalam perjalanan sejarah olahraga dunia kecuali sepak bola. Daya tarik lintas budaya sepak bola meluas dari Eropa dan Amerika Selatan ke Australia, Afrika, Asia bahkan Amerika Serikat. Organisasi sepak bola dunia adalah Federation of Football Association (FIFA) yang bermarkas di Swiss. Dalam sejarahnya sepak bola memiliki perjalanan yang panjang untuk sampai kepada konsep dan permainan yang *modern* seperti saat ini (Lukyani & Agustina, 2020).

Tahun 2023, Indonesia memiliki peluang emas untuk mengikuti kompetisi sepak bola kelas dunia pada ajang Piala Dunia Under 20 (U-20) yang diadakan oleh FIFA. Negara Indonesia berhasil mengatasi persaingan Brasil dan Peru sebagai pesaing tuan rumah Piala Dunia U-20, sehingga Indonesia terpilih sebagai tuan rumah turnamen tersebut. Pengumuman resmi Indonesia sebagai tuan rumah dilakukan oleh presiden FIFA Gianni Infantino dalam FIFA *council meeting* yang diadakan pada tanggal 24 November 2019 di Shanghai, China. Namun, mendekati pelaksanaan Piala Dunia U-20 beberapa pihak menolak keikutsertaan tim nasional Israel untuk bermain di negara Indonesia. Suara penolakan ini mendorong FIFA untuk membatalkan *drawing* piala dunia yang akan dilaksanakan di Bali pada tanggal 31 Maret 2023, munculnya permasalahan tersebut membuat FIFA menganggap Indonesia tidak mampu untuk menjadi tuan rumah Piala Dunia U-20. Akibat munculnya permasalahan tersebut FIFA secara resmi memberi keputusan pencabutan status Indonesia sebagai tuan rumah Piala Dunia U-20 melalui *website* resmi fifa.com pada hari Rabu tanggal 29 Maret 2023, keputusan yang dibuat menjelang dua bulan sebelum turnamen dimulai membuat netizen Indonesia merasa kecewa dan menjadi *trending* Twitter (Khairil Rahim, 2023).

Twitter merupakan salah satu media yang digunakan untuk memberikan opini melalui *tweet* (Pradana dkk., 2020). Inti dari Twitter adalah *tweet*, sejak tahun 2017 *tweet* telah diperluas menjadi 280 karakter. Tweet dapat berisikan komentar, opini, ataupun percakapan secara online untuk menyebarkan sebuah informasi. Seringkali pengguna Twitter sulit untuk dapat memahami informasi yang disampaikan oleh pengguna lain karena informasi yang dibagikan hanya dalam bentuk teks. Meskipun teks dapat dijadikan sebagai representasi dari opini yang ingin disampaikan oleh penulis, sering terjadi ketimpangan antara informasi yang sedang dibahas dan informasi yang diinterpretasikan oleh pembaca. Oleh sebab itu, banyak penelitian pembelajaran mesin yang dilakukan untuk menghasilkan sebuah program yang dapat membantu menganalisis hal tersebut. Salah satunya yaitu menggunakan metode FastText *embeddings* dan Algoritma Recurrent Neural Network (RNN).

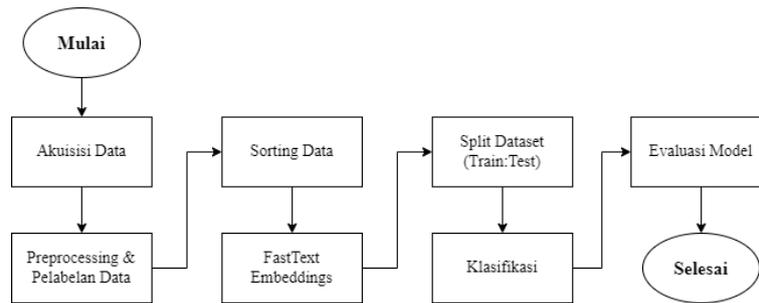
Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini (Arsi & Waluyo, 2021). Informasi sentimen yang dihasilkan dapat berupa opini positif, opini negatif, ataupun netral.

Terdapat beberapa penelitian terdahulu yang juga mengangkat analisis sentimen sebagai topik penelitian. Pada tahun 2020, Faisal Faturhman, Budi Irawan dan Cesi Setianingsih melakukan penelitian untuk menganalisis sentimen pada BPJS Kesehatan menggunakan RNN. Dataset diambil dari media sosial Twitter sejumlah 1400 data, pemberian bobot dataset dilakukan menggunakan metode Word2Vec dimana setiap kata mempunyai nilai *vector* masing-masing. Setelah setiap kata memiliki bobot *dataset* akan dilakukan proses klasifikasi dengan menggunakan Algoritma RNN menggunakan tipe *Long Short-Term Memory* (LSTM). Hasil dari klasifikasi berupa 700 data berlabel positif dan 700 data berlabel negatif. Hasil pengujian didapatkan partisi 90% data latih dan 10% data uji dengan tingkat akurasi paling tinggi mencapai 86.67% (Faturhman dkk., 2020).

Penelitian lain yang membahas tentang analisis sentimen dilakukan oleh Meytry Petronella Purba dan Yuliagnis Transver Wijaya pada tahun 2022. Purba dan rekannya menguji seberapa akurat Metode LSTM dan FastText untuk menganalisis *basic emotion* masyarakat pada masa Pandemi Covid-19. Pada penelitian tersebut, data yang digunakan sebanyak 4527 diambil dari Twitter dengan cara di-*scraping* menggunakan kata kunci “kesehatan mental” dan “kesehatan jiwa”, selanjutnya dilakukan *preprocessing*. Setelah data diuji dengan menggunakan metode LSTM dan menerapkan FastText didapatkan tingkat akurasi sebesar 74,48% hanya dengan 5 *epoch* (Purba & Wijaya, 2022).

Berdasarkan paparan diatas, maka peneliti akan melakukan analisis sentimen dengan topik pembatalan tuan rumah Indonesia di Piala Dunia U-20, dari data *tweet* yang diperoleh akan dilakukan *preprocessing* kemudian di vektorisasi menggunakan metode FastText *embeddings* dan diklasifikasi menggunakan algoritma RNN jenis LSTM yang akan menghasilkan nilai akurasi sentimen positif dan negatif.

METODE PENELITIAN



Gambar 1. Metode Penelitian

Akuisisi Data

Akuisisi Data merupakan tahapan pertama dalam penelitian, dimana akuisisi data dilakukan menggunakan *library* Python yakni Snscape dengan topik pembatalan tuan rumah indonesia di piala dunia U-20 dan menggunakan tiga *keyword* yaitu “pembatalan piala dunia”, “piala dunia u 20”, dan “fifa batal u-20” dalam rentang waktu postingan mulai tanggal 29 maret 2023 s.d. 05 April 2023. Hasil dari *scraping* akan dimasukkan ke dalam file dengan format csv. Data yang akan digunakan dalam penelitian ini adalah data *tweet* berbahasa Indonesia. Data yang didapat nantinya memiliki berbagai kolom yaitu *date*, *content*, *username*, *follower*, *language*, dan *url tweet*.

Preprocessing & Pelabelan

Dikarenakan data yang berhasil dikumpulkan melalui proses *scraping* tergolong sebagai raw data atau data mentah yang merupakan data yang belum melalui proses pengolahan apapun. Sehingga data *tweet* yang dikumpulkan tersebut masih mengandung hal-hal yang tidak diperlukan dalam proses pelatihan model. Agar data *tweet* tersebut menjadi lebih bersih dan tidak mengganggu proses analisis atau *modeling*, maka diperlukan beberapa langkah *preprocessing* data (Purba & Wijaya, 2022).

Tahapan pada proses ini meliputi pembersihan text (*cleaning*), mengubah huruf menjadi huruf kecil (*case-folding*), mengubah kalimat dalam teks menjadi potongan kata (*tokenizing*), memperbaiki kesalahan penulisan pada kata (*normalization*), menghilangkan kata-kata yang dianggap tidak berpengaruh pada kalimat seperti kata ganti, depan dan sambung (*stopword removal*), mengubah kata menjadi bentuk kata dasar (*stemming*).

Langkah selanjutnya melakukan pelabelan data, data yang digunakan untuk pelabelan adalah data hasil dari *normalization* yang kemudian diubah menjadi bahasa Inggris, karena proses pelabelan ini menggunakan *library* VADER yang hanya *support* teks berbahasa Inggris (Alfyando dkk., 2024).

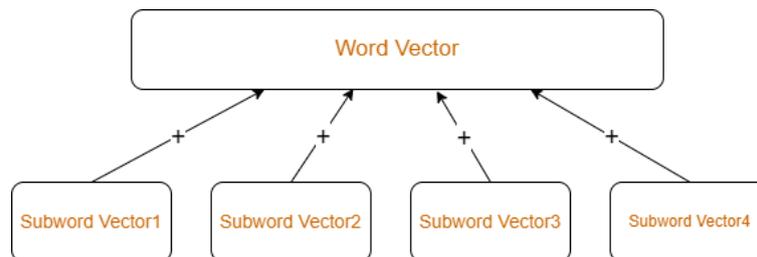
Sorting Data

Pada tahap ini dilakukan proses menghapus *tweet* duplikat karena dapat menghindari bias yang disebabkan oleh duplikasi informasi dalam *dataset*. Dengan menghapus *tweet* duplikat, dapat memastikan bahwa setiap *tweet* hanya diwakili satu kali dalam *dataset*, sehingga mencegah terjadinya peningkatan berlebihan pada sentimen tertentu dan memastikan representasi yang lebih seimbang.

Selanjutnya data disortir kembali, Jika jumlah kata dalam kalimat tersebut kurang dari 5 kata, maka dihapus dari *dataset*. Sehingga hanya kalimat dengan jumlah kata lebih dari sama dengan 5 yang diambil menjadi *dataset*. Hal ini dilakukan karena kalimat dengan jumlah kata lebih dari 4, lebih dapat merepresentasikan makna dari sebuah kalimat. Terakhir dilakukan penghapusan *tweet* berlabel netral karena pada penelitian ini *dataset* yang digunakan hanya berfokus pada sentimen negatif dan sentimen positif.

FastText Embeddings

Proses pembelajaran fitur yang dilakukan pada penelitian ini adalah dengan menggunakan FastText *word embedding*. Dimensi vektor/ *embedding size* yang digunakan berjumlah 300 dimensi. FastText memiliki kinerja yang baik, dapat melatih model pada *dataset* yang besar dengan cepat dan dapat memberikan representasi kata yang tidak muncul dalam data latih. Jika kata tidak muncul selama pelatihan model, kata tersebut dapat dipecah menjadi n-gram untuk mendapatkan *embedding* vektornya. Kelebihan inilah yang menjadikan FastText dapat menangani kata yang tidak pernah muncul di kamus, yang mana dalam Word2Vec hal seperti ini akan menghasilkan *error* (Vaclav Kosar, 2022). Gambar 2 merupakan gambaran n-gram pada FastText.



Gambar 2. Arsitektur N-gram FastText

FastText dapat melakukan perubahan kata maupun kalimat menjadi sebuah vektor secara otomatis. Hal tersebut terjadi karena FastText memiliki sebuah kamus yang telah terlatih sebelumnya. Pada penggunaan FastText sendiri, proses n-grams tersebut hanya dilakukan ketika kata yang diubah menjadi vektor tidak terdapat pada kamus yang telah terbuat. Namun kata *input* tersebut telah terdapat pada kamus. Maka hasil *return* tersebut sesuai dengan vektor

dari model yang telah dilatih oleh FastText. Gambar 3 merupakan vektorisasi kalimat pada FastText.

input sentence

vector size	1	2	3	...	300
piala	0.2	-0.1	0.5	...	0.7
dunia	0.4	0.3	-0.2	...	-0.4
batal	-0.1	0.6	-0.4	...	0.6
average	0.1	0.2	-0.3	...	0.3

Gambar 3. Embedding Vektor Kalimat Pada FastText

Split Dataset

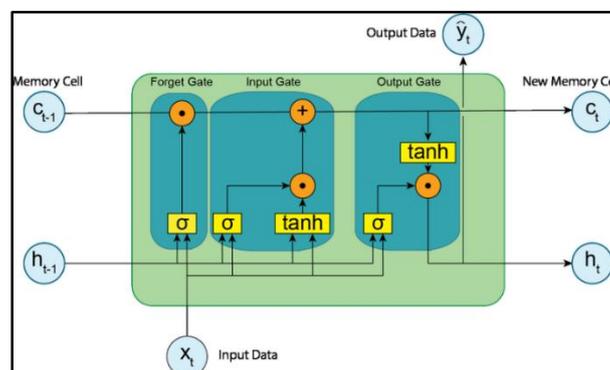
Dalam tahap split *dataset*, data dibagi menjadi dua jenis yaitu data latih merupakan data yang digunakan dalam pelatihan model, dan data uji merupakan data yang digunakan untuk pengujian model. Pada penelitian ini peneliti menggunakan 3 rasio *split data* seperti yang tertera pada Tabel 1.

Tabel 1. Rasio Split Data

No	Data Latih	Data Uji
1	70%	30%
2	80%	20%
3	90%	10%

Klasifikasi

LSTM adalah sebuah tipe dari RNN yang telah dimodifikasi dengan menambahkan sel memori. LSTM memiliki beberapa *gate* atau gerbang seperti *filter* yang berguna untuk menghapus atau menambah suatu informasi. Terdiri dari *input gate*, *forget gate*, dan *output gate* (Adinda Siti Shalehah, 2022). Gambar 4 merupakan gambaran Arsitektur LSTM.



Gambar 4. Arsitektur LSTM

Alur kerja LSTM dimulai dari *forget gate* untuk menentukan informasi yang tidak digunakan agar dikeluarkan dari *memory cell*. Output dari *forget gate* berupa nilai antara 0 dan

1, dimana nilai 1 menunjukkan bahwa informasi masih memiliki tingkat kemiripan data dan perlu disimpan, sedangkan nilai 0 menunjukkan informasi tidak memiliki kemiripan dan harus dihapus. Untuk menghitung *forget gate* digunakan rumus seperti persamaan 1.

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

Keterangan:

f_t = nilai *forget gate*

σ = fungsi aktivasi *sigmoid*

W_f = nilai bobot dari *forget gate*

$[h_{t-1}, x_t]$ = operasi konkatenasi (menambah baris dari x_t dengan baris dari h_{t-1})

h_{t-1} = nilai *output* dari waktu ke $t-1$

x_t = nilai *Input* pada waktu ke t

b_f = nilai bias pada *forget gate*

Selanjutnya menetapkan informasi yang bakal diperbarui melalui *input gate*. *Input gate* adalah bagian dari sel LSTM yang digunakan untuk mengontrol seberapa banyak informasi yang akan dimasukkan ke dalam state dan akan disimpan. *Input gate* menghasilkan vektor skalar antara 0 dan 1 untuk setiap elemen input yang diberikan. Nilai 1 menunjukkan bahwa seluruh informasi dari input diteruskan ke proses selanjutnya, sedangkan nilai 0 menunjukkan bahwa tidak ada informasi dari input yang diteruskan. Untuk menghitung *input gate* digunakan rumus seperti persamaan 2 dan 3

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

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

Keterangan:

i_t = nilai *input gate*

σ = fungsi aktivasi *sigmoid*

W_i = nilai bobot untuk nilai input pada waktu ke t

$[h_{t-1}, x_t]$ = operasi konkatenasi (menambah baris dari x_t dengan baris dari h_{t-1})

h_{t-1} = nilai *output* dari waktu ke $t-1$

x_t = nilai *input* pada waktu ke t

b_i = nilai bias pada *input gate*

\tilde{C}_t = nilai kandidat *cell state*

\tanh = fungsi aktivasi tanh (Hyperbolic Tangent)

W_c = nilai bias pada *input gate* pada *cell* ke *cell state*

b_c = nilai bias pada *cell* ke *cell state*

Terakhir menentukan perhitungan *output gates* dengan menjalankan aktivasi *sigmoid* pada *cell* LSTM. Output yang dihasilkan dari proses ini akan melalui tahap *filter* terlebih

dahulu. Proses *filter* menggunakan *layer sigmoid* yang menentukan nilai mana yang harus diunggulkan untuk *output*. Selanjutnya, *output* akan menjadi nilai dalam rentang antara -1 dan 1. Untuk menghitung *output gate* digunakan rumus seperti persamaan 4 dan 5.

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

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

Keterangan:

o_t = nilai *output gate*

σ = fungsi aktivasi *sigmoid*

W_o = bobot untuk nilai input pada waktu ke t

$[h_{t-1}, x_t]$ = operasi konkatenasi (menambah baris dari x_t dengan baris dari h_t)

h_{t-1} = nilai *output* dari waktu ke t-1

x_t = nilai *input* pada waktu ke t

b_o = nilai bias pada *output gate*

\tanh = fungsi aktivasi tanh (Hyperbolic Tangent)

C_t = nilai *cell state* yang terbaru

Evaluasi Model

Evaluasi model merupakan proses analisis sentimen yang dilakukan dalam penelitian ini. Hasil evaluasi disajikan dalam bentuk nilai-nilai tertentu yang dapat mewakili kinerja model berkaitan dengan pengujian parameter yang telah disiapkan sebelumnya. Hasil pengujian performa model ditampilkan menggunakan *confusion matrix* (Rizki dkk., 2022). Selanjutnya, akan didapatkan *classification report* yang mengandung *precision*, *recall*, *f1-score* dari setiap kelas dan *accuracy* dari hasil klasifikasi tersebut (Arif Widiasan Subagio dkk., 2024).

HASIL DAN PEMBAHASAN

Hasil dari proses akuisisi data didapat data sebanyak 14.807 *tweet* tentang permasalahan pembatalan tuan rumah indonesia di Piala Dunia U-20 dan akan dijadikan sebagai *dataset* mentah pada penelitian.

Setelah berhasil mendapatkan *dataset* mentah, kemudian dilakukan proses *preprocessing* dan pelabelan data. Proses *preprocessing* dilakukan dengan tujuan mempersiapkan *dataset* agar siap untuk masuk ke tahap pelatihan. Untuk melakukan analisis sentimen, hanya diperlukan data teks pada kolom *content* yang sudah dilabeli. Oleh karena itu, dilakukan penghapusan pada kolom yang tidak relevan dengan proses ini. Hasil dari proses *preprocessing* yang nantinya akan digunakan untuk pelabelan data tertera pada Tabel 2.

Tabel 2. Hasil Preprocessing

Sebelum Preprocessing	Sesudah Preprocessing
Gagalnya perhelatan piala dunia U-20 di Indonesia menjadikan suatu pengalaman yg mendewasakan kita semua. @erickthohir	['gagal', 'helat', 'piala', 'dunia', 'indonesia', 'jadi', 'alam', 'dewasa']

Selanjutnya dilakukan proses merubah *tweet* ke dalam bahasa Inggris untuk dilakukan proses pelabelan otomatis menggunakan *library* VADER. Proses pelabelan menggunakan data teks dari hasil normalisasi kata. Skor sentimen komposit untuk setiap teks dihitung dengan menggunakan metode *polarity_scores*. Jika skor lebih besar atau sama dengan 0.05, dianggap sebagai sentimen positif, jika skor kurang atau sama dengan -0.05, dianggap sebagai sentimen negatif dan jika di antara kedua nilai tersebut dianggap sebagai sentimen netral. Setelah semua *tweet* mendapatkan label sentimen dilakukan tiga proses *sorting data* yaitu menghapus *tweet* duplikat, menghapus *tweet* kurang dari lima kata dan menghapus *tweet* berlabel netral. Informasi dari hasil *dataset final* tertera pada Tabel 3.

Tabel 3. Dataset Final

Jumlah Positif	Jumlah Negatif	Total Data
4.141	5.504	9.645

Sebelum masuk ke proses *training*, data *tweet* akan diubah dahulu dari data teks menjadi data numerik menggunakan metode FastText *embeddings*, hal ini bertujuan agar data dapat diproses oleh model LSTM. Selanjutnya, data akan melalui pelatihan dengan rasio *split data* yang digunakan ada 3 yaitu 70:30, 80:20, dan 90:10. Tabel 4 menunjukkan *tuning hyperparameter* dari pengujian model LSTM yang digunakan.

Tabel 4. Tuning Hyperparameter Model LSTM

Model	Split Data	Learning Rate	Epoch	Optimizer	Unit
LSTM-1	70:30	0.001	15	Adam	128
LSTM-2	80:20	0.001	15	Adam	128
LSTM-3	90:10	0.001	15	Adam	128

Berdasarkan skenario tersebut, data *tweet* akan dilatih dengan data latih yang terpisah, dan kemudian diprediksi dengan data uji. Hasil pengujian model LSTM-1, LSTM-2, dan LSTM-3 ditunjukkan pada Tabel 5.

Tabel 5. Hasil Pengujian LSTM

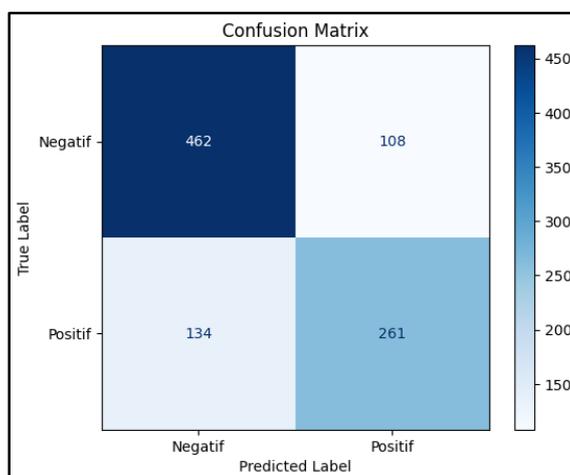
Model	Precision	Recall	F1-Score	Accuracy
LSTM-1	0.7284	0.7301	0.7289	0.7301
LSTM-2	0.7158	0.7107	0.7120	0.7107
LSTM-3	0.7474	0.7492	0.7478	0.7492

Berdasarkan tabel hasil pengujian pada perbandingan 3 *split data* yang ada, terlihat bahwa nilai akurasi terendah ditemukan pada model LSTM-2 dengan nilai *accuracy* 71.07%, *precision* 71.58%, *recall* 71.07%, *f1-score* 71.20%. Sedangkan hasil terbaik ditemukan pada model LSTM-3 dengan nilai *accuracy* 74.92%, *precision* 74.74%, *recall* 74.92%, dan *f1-score* 74.78%. Hasil terbaik ini dicetak tebal untuk menunjukkan performa terbaik dalam perbandingan tersebut. Hasil *classification report* dari model LSTM terbaik (LSTM-3) ditunjukkan pada Tabel 6.

Tabel 6. Classification Report Model LSTM-3

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
0	0.7752	0.8105	0.7925	570
1	0.7073	0.6608	0.6832	395
Weighted Avg	0.7474	0.7492	0.7478	965
Accuracy		0.7492		965

Seperti yang ditunjukkan pada tabel *classification report*, model mengenali kelas sentimen negatif (0) lebih baik daripada kelas sentimen positif (1). Model berhasil melakukan prediksi terhadap data uji sebanyak 965 dimana 723 data (74.92%) dapat diprediksi sesuai dengan label data aktual dan 242 data (25.08%) diprediksi berbeda dengan label data aktual. Hasil *confusion matrix* dari model LSTM-3 ditunjukkan pada Gambar 5.



Gambar 5. Hasil Confusion Matrix Model LSTM-3

Pada Gambar 5 dapat ditunjukkan hasil *confusion matrix* sebagai berikut:

- Nilai True Positive (TP) yaitu model memprediksi dengan benar kelas positif (hasil prediksi dan data aktual positif) sejumlah 261 data.
- Nilai True Negative (TN) yaitu model memprediksi dengan benar kelas negatif (hasil prediksi dan data aktual negatif) sejumlah 462 data.
- Nilai False Positive (FP) yaitu model memprediksi salah kelas negatif (data aktual sentimen negatif yang diprediksi positif) sejumlah 108 data.

- d. Nilai False Negative (FN) yaitu model memprediksi salah kelas positif (data aktual sentimen positif yang diprediksi negatif) sejumlah 134 data.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis sentimen yang dilakukan terhadap masalah pembatalan tuan rumah Indonesia di Piala Dunia U-20, dari 9.645 *tweet* yang diposting di media sosial Twitter. Mayoritas dari data ini memiliki sentimen negatif sebanyak 5.504 *tweet*, sedangkan 4.141 *tweet* memiliki sentimen positif. Model LSTM-3 memberikan performa terbaik pada *tuning hyperparameter* yaitu *ratio split data* 90:10, *learning rate* 0.001, jumlah *epoch* sebanyak 15, menggunakan *optimizer* Adam, dan jumlah unit 128. Dengan nilai *accuracy* sebesar 74.92%, *precision* 74.74%, *recall* 74.92%, dan *f1-score* 74.78%.

Penelitian selanjutnya disarankan untuk menggunakan *dataset* yang jauh lebih valid, karena *dataset* yang digunakan pada penelitian ini adalah hasil pelabelan dari model kecerdasan buatan, sehingga memungkinkan terjadinya kesalahan label sentimen. Menambahkan algoritma selain RNN jenis LSTM untuk membandingkan nilai *accuracy* dan kinerja modelnya, sehingga dapat menghasilkan analisis sentimen yang lebih akurat.

DAFTAR REFERENSI

- Adinda Siti Shalehah. (2022). Analisa Kinerja RNN Menggunakan FastText Embedding terhadap Ulasan PeduliLindungi di Masa COVID-19 [Universitas Mercu Buana]. <https://repository.mercubuana.ac.id/70595/>
- Alfyando, M., Anggraeny, F. T., & Sihananto, A. N. (2024). Perbandingan Algoritma Random Forest dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store. *Jurnal Sistem Informasi dan Ilmu Komputer*, 2(1), 77–86.
- Arif Widiasan Subagio, Anggraini Puspita Sari, & Andreas Nugroho Sihananto. (2024). Klasifikasi Lexicon-Based Sentiment Analysis Tragedi Kanjuruhan pada Twitter Menggunakan Algoritma Convolutional Neural Network. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 4(1), 166–177. <https://doi.org/10.55606/juisik.v4i1.759>
- Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147.
- Faturohman, F., Irawan, B., & Setianingsih, C. (2020). Analisis Sentimen Pada Bpjs Kesehatan Menggunakan Recurrent Neural Network. *eProceedings of Engineering*, 7(2).
- Pradana, Y. R. Y., Astiningrum, M., & Hani'ah, M. (2020). Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter. *Seminar Informatika Aplikatif Polinema*, 35–39.

- Purba, M. P., & Wijaya, Y. T. (2022). Analisis Basic Emotion Masyarakat Pada Masa Pandemi COVID-19 di Media Sosial Twitter Dengan Metode LSTM-FastText. Seminar Nasional Official Statistics, 2022(1), 643–654.
- Rizki, A. M., Yuliasuti, G. E., Puspaningrum, E. Y., & Nurlaili, A. L. (2022). Klasifikasi Jenis Kelamin Berdasarkan Ciri Fisik Menggunakan Algoritma Neural Network. Scan: Jurnal Teknologi Informasi dan Komunikasi, 17(3), 18-22.
- Vaclav Kosar. (2022, Juni 12). *FastText Vector Norms And Out of Vocabulary (OOV) Words*. vaclavkosar.com. <https://vaclavkosar.com/ml/FastText-Vector-Norms-And-OOV-Words>