

Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM

Mohammed Hafizh Al-Areef¹, Kana Saputra S.²

^{1,2}Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

Email: hafizh2608@mhs.unimed.ac.id, kanasaputras@unimed.ac.id

Email Penulis Korespondensi: hafizh2608@mhs.unimed.ac.id

Article History:

Received Aug 02th, 2023

Revised Aug 11th, 2023

Accepted Aug 27th, 2023

Abstrak

Platform media sosial Twitter menjadi salah satu platform yang banyak digunakan oleh masyarakat Indonesia untuk berkomunikasi, dan mengakses informasi dengan cepat. Hal ini menyebabkan banyak sekali sentimen masyarakat Indonesia yang dapat dijadikan sebagai studi kasus salah satunya mengenai calon presiden Indonesia tahun 2024. Beberapa tokoh publik seperti Ganjar Pranowo, Prabowo Subianto, dan Ridwan Kamil sudah mulai banyak dibicarakan oleh masyarakat sebagai calon presiden dalam beberapa sentimen pada platform Twitter. Sentimen mengenai para tokoh publik tersebut akan diklasifikasikan dengan algoritma *Long Short Term Memory* (LSTM) dengan label positif dan negatif. Data sentimen selanjutnya akan melewati proses *pre-processing*, pelabelan data dengan *textblob*, *word embedding* dengan *fasttext*, hingga *data balancing* dengan SMOTE sebelum akhirnya model akan dilatih dan diuji. Tahapan penentuan *hyperparameter tuning* dilakukan sebelum melatih model LSTM seperti penentuan jumlah unit, *learning rate*, *dropout*, batch size hingga jumlah epoch agar menghasilkan model latih yang baik. Hasil uji dan evaluasi performa untuk setiap model yang telah dilatih adalah 82% akurasi, 86% presisi, 92% *recall*, dan 89% *f1-score* pada model Ganjar Pranowo. 82% akurasi, 82% presisi, 96% *recall*, dan 89% *f1-score* pada model Prabowo Subianto. 87% akurasi, 91% presisi, 95% *recall*, 93% *f1-score* pada model Ridwan. 87%.

Kata Kunci : Analisis Sentimen, *Fasttext*, *Textblob*, SMOTE, LSTM, *Hyperparameter Tuning*

Abstract

The social media platform Twitter is one of the platforms widely used by Indonesians to communicate, and access information quickly. This has led to a lot of sentiments of Indonesian people that can be used as a case study, one of which is about Indonesian presidential candidates in 2024. Several public figures such as Ganjar Pranowo, Prabowo Subianto, and Ridwan Kamil have started to be discussed by the public as presidential candidates in several sentiments on the Twitter platform. Sentiments about these public figures will be classified with the *Long Short Term Memory* (LSTM) algorithm with positive and negative labels. Sentiment data will then go through a *pre-processing* process, data labeling with *textblob*, word embedding with *fasttext*, to data balancing with SMOTE before finally the model will be trained and tested. The *hyperparameter tuning* stage is carried out before training the LSTM model such as determining the number of units, *learning rate*, *dropout*, batch size to the number of epochs in order to produce a good training model. The test results and performance evaluation for each model that has been trained are 82% accuracy, 86% precision, 92% recall, and 89% *f1-score* on the Ganjar Pranowo model. 82% accuracy, 82% precision, 96% recall, and 89% *f1-score* on Prabowo Subianto model. 87% accuracy, 91% precision, 95% recall, 93% *f1-score* on Ridwan model. 87%.

Keyword : Sentiment Analysis, *Fasttext*, *Textblob*, SMOTE, LSTM, *Hyperparameter Tuning*

1. PENDAHULUAN

Internet adalah salah satu sarana untuk mengakses informasi dan komunikasi yang dapat diakses oleh seseorang dimana saja dan kapan saja. Seiring dengan berjalannya kemajuan teknologi, penggunaan internet sebagai sarana komunikasi juga bertumbuh sangat pesat. Indonesia adalah salah satu negara dengan pengguna internet terbesar di dunia. Berdasarkan hasil survey dari APJII (Asosiasi Penyelenggara Jasa Internet Indonesia) pada tahun 2022, jumlah penduduk

yang terkoneksi internet mencapai 210.026.769 jiwa dengan peningkatan persentase penetrasi internet sebesar 12,22% dari tahun 2018 [1].

Penggunaan internet semakin meluas dengan banyaknya platform yang menyediakan berbagai layanan kepada masyarakat untuk mengakses internet. Media sosial menjadi salah satu platform yang banyak digunakan untuk berkomunikasi, dan mengakses informasi dengan cepat. Informasi yang terdapat pada media sosial sangat beragam, seperti komentar, kritikan, hingga opini. Twitter menjadi salah satu platform yang banyak digunakan oleh masyarakat Indonesia. Dilansir dari We Are Social, pengguna Twitter di Indonesia mencapai 18,45 juta pada tahun 2022 [2]. Hal ini membuktikan bahwa terdapat banyak sekali sentimen masyarakat Indonesia yang terdapat pada platform Twitter yang dapat dijadikan sebagai studi kasus.

Indonesia merupakan negara demokrasi yang mana peran rakyat sangat menentukan dalam memilih pemimpin negara yang dilaksanakan melalui pemilihan umum atau yang sering disebut pemilu. Pemilihan Presiden atau pilpres merupakan pesta rakyat yang selalu menjadi perbincangan hangat ketika menjelang pemilu. Banyak sekali opini pro dan kontra dari berbagai kalangan masyarakat mengenai calon presiden dan wakil presiden semasa menjelang pemilu. Tak luput juga pilpres yang akan dilaksanakan pada tahun 2024 mendatang. Dengan semakin dekatnya waktu pemilu, semakin jelas juga calon calon presiden dan wakil presiden yang akan melaksanakan kampanye.

Beberapa tokoh publik sudah mulai banyak diprediksi sebagai calon presiden dalam beberapa sentimen pada platform Twitter seperti Ganjar Pranowo, Prabowo Subianto, dan Anies Baswedan. Tokoh publik lainnya yang ikut diprediksi sebagai calon wakil presiden dalam Twitter yaitu Erick Thohir, Agus Harimurti Yudhoyono (AHY), dan Airlangga Hartanto. Nama-nama diatas akan dijadikan sebagai keyword dalam klasifikasi sentimen pengguna Twitter mengenai pilpres 2024 agar terlihat citra para tokoh publik tersebut apakah positif atau negatif di mata masyarakat berdasarkan sentimen masyarakat di Twitter.

Analisis sentimen sangat diperlukan dalam memantau media sosial. Terlebih lagi dalam memantau sentimen masyarakat mengenai pilpres 2024. Di media sosial, analisis sentimen berguna untuk mendapatkan gambaran umum dari suatu opini. Pemilihan presiden 2024 akan dilaksanakan di tahun mendatang, namun sudah banyak sekali bermacam-macam sentimen dari masyarakat baik itu sentimen positif maupun negatif untuk para calon presiden Indonesia. Namun bagi masyarakat yang awam terhadap politik maupun yang tidak mengikuti perkembangan di media sosial, mereka masih belum menemukan pilihan presiden yang tepat di hati untuk memimpin negara Indonesia.

Berdasarkan survei capres 2024 yang dilakukan oleh Lembaga Survei Nasional (LSN) pada tanggal 29 Oktober hingga 2 November 2022, didapatkan bahwa dari 1.230 responden 26,5% diantaranya masih tidak tahu atau tidak menjawab siapa calon presiden Indonesia 2024. Hasil survei terbaru lainnya yang dilakukan oleh Lembaga Survei Indonesia (LSI) pada tanggal 7-11 Januari 2023, dari 1.221 responden 17% diantaranya masih tidak tahu atau tidak menjawab siapa calon presiden Indonesia 2024 nantinya. Maka dari itu analisis sentimen diperlukan untuk menyediakan informasi bagi masyarakat mengenai mayoritas pilihan masyarakat terhadap calon presiden berdasarkan opini publik dari Twitter.

Sebelumnya, telah dilakukan penelitian mengenai analisis sentimen terhadap calon presiden Indonesia dengan berbagai metode. Diantaranya yaitu penggunaan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) dengan word embedding TF-IDF terhadap calon presiden 2019, dengan hasil akurasi pada SVM sebesar 76,5% dan hasil akurasi pada KNN sebesar 68,3% [3]. Penelitian lainnya yang menerapkan metode Naïve Bayes terhadap bakal calon presiden 2024, dengan hasil akurasi tertinggi sebesar 73,68% untuk dataset Ganjar Pranowo hingga akurasi terendah sebesar 60% untuk dataset Prabowo Subianto [4].

Sebelumnya juga telah dilakukan penelitian yang membahas analisis sentimen menggunakan metode Long Short-Term Memory (LSTM). Seperti pada penelitian analisis sentimen tweet mengenai COVID-19 menggunakan Word Embedding dan metode Long Short-Term Memory (LSTM) dengan hasil perbandingan dari metode LSTM, RNN dan Naïve Bayes. Didapatkan hasil akurasi metode LSTM sebesar 81%, akurasi metode RNN sebesar 74% dan akurasi metode Naïve Bayes sebesar 71% [5]. Penelitian lainnya yang terkait yaitu analisis sentimen publik twitter terhadap pemilu 2024 menggunakan LSTM, pelabelan dengan lexicon based feature dan tanpa melakukan hyperparameter tuning, didapatkan hasil akurasi sebesar 78% [6].

Melihat dari hasil penelitian terdahulu, algoritma LSTM memiliki hasil akurasi paling tinggi dibandingkan dengan algoritma lainnya. Ini dikarenakan LSTM memiliki kelebihan dalam melakukan analisis sentimen, yaitu mampu mengingat informasi dari data yang telah diproses dan menggunakan kemampuan memori tersebut untuk memahami data input lebih baik [7]. Kemampuan memori ini hanya dimiliki oleh algoritma LSTM yang merupakan pengembangan dari algoritma Recurrent Neural Network (RNN) dalam mengatasi masalah long-term dependency dalam mengingat data jangka panjang.

Berdasarkan pada latar belakang dan hasil dari penelitian diatas penulis akan melakukan penelitian Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM. Penulis juga akan melakukan *data scraping* Twitter dengan menggunakan *library* SNScrape berbasis Python yang akan mengumpulkan seluruh data tweet dari seluruh pengguna Twitter yang nantinya akan dikumpulkan kedalam suatu dataframe agar data bisa diolah. Diharapkan dengan adanya penelitian ini dapat menghasilkan hasil akurasi yang lebih baik dengan mempertimbangkan beberapa atribut seperti *word embedding*, pelabelan teks hingga *hyperparameter tuning*.

2. METODOLOGI PENELITIAN

2.1 Pengumpulan Data

Data diperoleh dari platform Twitter dengan teknik *data scraping* dan menggunakan sebuah *tools* bernama SNScrape yang merupakan sebuah *library* python untuk mengumpulkan data tweet dari Twitter . Data diambil pada rentang waktu 1 Februari 2022 hingga 28 Februari 2023. Kata kunci yang digunakan dalam *data scraping* menggunakan SNScrape disesuaikan dengan nama tokoh publik calon presiden Indonesia 2024 yang banyak disebut masyarakat di Twitter, yaitu “Ganjar Pranowo”, “Prabowo Subianto”, “Anies Baswedan” “Puan Maharani” dan “Ridwan Kamil”. Masing-masing calon presiden akan memiliki data tweet tersendiri yang nantinya akan dihitung persentase sentimen positif dan negatif dari masing-masing calon presiden.

2.2 Pelabelan Data

Pada penelitian ini TextBlob akan memberi 2 label pada data, yaitu positif dan negatif. Pelabelan akan memerlukan proses terjemahan dari bahasa Indonesia ke bahasa Inggris. Setelah tahap terjemahan selesai, TextBlob akan memulai proses pelabelan pada data tweet dengan memproses data dan akan menghasilkan output berupa nilai polaritas [8]. Jika nilai polaritas lebih besar dari 0 (nilai polaritas > 0), maka hasil pelabelan akan positif. Namun jika sebaliknya, nilai polaritas kurang dari 0 (Nilai polaritas < 0), maka hasil pelabelan akan negatif [9].

2.3 Text Pre-Processing

Pada tahap *text pre-processing* data akan dikelola menjadi menjadi lebih terstruktur sebelum dilakukannya proses klasifikasi. Tujuan dari melakukan preprocessing adalah untuk mempersiapkan data agar siap digunakan dengan menghapus unsur-unsur yang dianggap tidak relevan [10]. Terdapat beberapa proses *text pre-pocessing* seperti *text cleansing*, *case folding*, *normalization*, *tokenization*, *stopword removal* dan *stemming* [11].

2.3.1 Text Cleansing

Tahapan awal dalam *text pre-processing* adalah *text cleansing*. Tahapan *text cleansing* bertujuan menghapus *noise* seperti URL, emoji, ataupun simbol-simbol yang ada tetapi tidak bermakna dalam proses klasifikasi teks.

2.3.2 Case Folding

Proses *case folding* bertujuan untuk menyamakan bentuk huruf pada teks data. Semua kalimat yang ada dalam data *tweet* yang ditulis dengan huruf besar akan ditransformasikan menjadi huruf kecil. Hal ini dilakukan agar tidak ada perbedaan bentuk huruf yang dapat mengganggu proses selanjutnya dalam analisis sentimen.

2.3.3 Stopword Removal

Proses selanjutnya adalah *stopword removal* yaitu menghilangkan kata-kata yang tidak bermakna seperti kata sambung yang nantinya dapat berpengaruh dalam kecepatan proses analisis sentimen. Proses ini membutuhkan sebuah daftar kamus sebagai acuan dalam penghapusan kata-kata tersebut. Dalam hal ini, *library* Sastrawi membantu dengan menyediakan daftar kata sambung bahasa Indonesia sebagai acuan dalam proses *stopword removal*.

2.3.4 Normalization

Tahapan berikutnya dalam *text pre-processing* adalah normalisasi. Pada tahap ini, kalimat yang memiliki ejaan kata yang disingkat atau diperpanjang akan diperbaiki menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) dengan cara memanfaatkan kamus *lexicon* bahasa Indonesia yang dapat digunakan untuk *mapping pattern* kata yang ingin dinormalisasikan.

2.3.5 Stemming

Stemming adalah proses menemukan kata dasar dari kata yang sudah melewati proses *stopword removal*. Kata akan direduksi menjadi kata dasar dengan menghilangkan imbuhan depan atau belakang. *Library* Sastrawi mampu memberikan daftar kata dasar dalam bahasa Indonesia untuk dijadikan referensi dalam proses stemming.

2.3.6 Tokenization

Tokenizing adalah proses memecah teks menjadi unit-unit yang lebih kecil, yang disebut token. Token biasanya dianggap sebagai kata-kata individu dalam teks. *Tokenizing* sangat penting dalam analisis teks karena memberikan cara untuk mengukur frekuensi kata-kata dalam teks dan memungkinkan untuk menganalisis konteks kata-kata. Dalam proses *tokenizing*, bukan hanya kata saja yang dipisahkan, tetapi juga angka dan tanda baca. Hal ini penting dilakukan karena angka dan tanda baca juga bisa mempengaruhi penilaian terhadap suatu kalimat atau kata [12].

2.4 Pembagian Data

Data latih dan data uji akan dibagi dengan rasio 80% untuk data latih dan 20% sisanya akan digunakan sebagai data uji untuk menguji performa model yang sudah dilatih.

2.5 Word Embedding

Setelah melalui tahap *data pre-processing* dan *data splitting*, tahap selanjutnya yaitu mengekstraksi fitur pada data dengan teknik *word embedding*. Pada penelitian ini *word embedding* yang digunakan adalah *fastText*. Proses *word embedding* memerlukan data *pre-trained* untuk melatih model *pre-trained* untuk vektorisasi kata. Data *pre-trained* pada penelitian ini akan menggunakan data *word vector* bahasa Indonesia dari situs resmi *fasttext* [10]. Selanjutnya hasil dari *pre-trained* model akan di simpan ke dalam *corpus* untuk nantinya akan digunakan oleh *fastText* untuk vektorisasi data *training*.

2.6 Data Balancing

Proses *data balancing* diterapkan untuk menghindari ketidakseimbangan data. Data balancing adalah proses memastikan bahwa jumlah sampel dalam setiap kelas pada data kita seimbang. Ketidakseimbangan kelas dapat mempengaruhi performa model *machine learning*, terutama dalam kasus di mana kelas minoritas memiliki jumlah sampel yang sangat sedikit.

Salah satu teknik yang dapat digunakan untuk mengatasi ketidakseimbangan kelas adalah SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE adalah metode *oversampling* yang menghasilkan sampel sintetis untuk kelas minoritas dengan menerapkan interpolasi antara sampel yang sudah ada [13]. Dengan menggunakan SMOTE, kita dapat meningkatkan jumlah sampel pada kelas minoritas tanpa mengulangi sampel yang sudah ada.

2.7 Long Short-Term Memory

Pemodelan sentimen dilakukan dengan memanfaatkan algoritma LSTM dan hasil dari proses pelatihan adalah sebuah model yang mampu mengklasifikasikan sentimen positif dan negatif terkait dengan calon presiden Indonesia 2024.

Langkah awal dalam membuat model LSTM adalah inialisasi parameter. Dalam tahap ini, beberapa parameter yang harus ditentukan meliputi jumlah unit, *batch size*, jumlah *epoch*, dan *optimizer* sebagai *learning rate*. Lalu pada layer kedua yaitu mengatur nilai *dropout* untuk menghindari *overfitting* pada model. Lalu pada *layer* terakhir adalah *dense layer* yang akan menghubungkan semua parameter pada *layer* sebelumnya dan mengolah *output* klasifikasi dari model. Pengklasifikasian pada *dense layer* akan menggunakan fungsi aktivasi sigmoid yang sesuai dengan data kelas biner yaitu kelas positif dan negatif.

2.8 Evaluasi Performa

Proses evaluasi performa dilakukan untuk mengetahui akurasi, presisi, dan recall hasil eksperimen. Evaluasi performa dilakukan dengan menggunakan *confusion matrix* yang mengukur tingkat *True Positive (TP rate)*, tingkat *True Negative (TN rate)*, tingkat *False Positive (FP rate)*, dan tingkat *False Negative (FN rate)*. *TP rate* menunjukkan persentase kelas positif yang benar diklasifikasi sebagai positif, sedangkan *TN rate* menunjukkan persentase kelas negatif yang benar diklasifikasi sebagai negatif. *FP rate* adalah kelas negatif yang salah diklasifikasi sebagai positif, dan *FN rate* adalah kelas positif yang salah diklasifikasi negatif [14].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Sumber data didapat dari hasil *scraping tweet* pada Twitter dengan menggunakan SNScrape. Data *tweet* diambil berdasarkan *keyword* masing masing bakal calon presiden yaitu Ganjar Pranowo 2024, Prabowo Subianto 2024, dan Ridwan Kamil 2024. Data yang didapat untuk masing-masing nama yaitu 11.984 data untuk Ganjar Pranowo, 14.325 data untuk Prabowo Subianto, dan 2.325 data untuk Ridwan Kamil. Data yang terkumpul selanjutnya akan disimpan kedalam format *comma separated value (csv)* untuk dilakukan *pre-processing*.

3.2 Text Pre-Processing

Tahap selanjutnya data hasil *scraping* akan melewati proses *pre-processing* yang bertujuan agar data teks dapat diolah dengan baik oleh mesin. Tabel 1 menunjukkan hasil *text pre-processing*.

Tabel. 1 *Text Pre-Processing*

Tahap <i>Pre-Processing</i>	Hasil
Data Awal	@emhamka08 @bengkeldodo @prabowo @BANGSAygSUJUD @DarekaRendy @florielicious @prmtaa_indah @Gerindra @are_inismyname @UmarSandi5 Pilpres 2019 sy pilih Prabowo Sandi. Pilpres 2024 mayoritas rakyat ingin Presiden yg mampu melakukan Perubahan yg lebih baik
<i>Text Cleansing</i>	Pilpres 2019 sy pilih Prabowo Sandi. Pilpres 2024 mayoritas rakyat ingin Presiden yg mampu melakukan Perubahan yg lebih baik

<i>Case Folding</i>	pilpres 2019 sy pilih prabowo sandi. pilpres 2024 mayoritas rakyat ingin presiden yg mampu melakukan perubahan yg lebih baik
<i>Stopword Removal</i>	pilpres sy pilih prabowo sandi pilpres mayoritas rakyat presiden yg mampu melakukan perubahan yg lebih baik
<i>Normalization</i>	pilpres saya pilih prabowo sandi pilpres mayoritas rakyat presiden yang mampu melakukan perubahan yang lebih baik
<i>Stemming</i>	pilpres saya pilih prabowo sandi pilpres mayoritas rakyat presiden yang mampu laku ubah yang lebih baik
<i>Tokenization</i>	['pilpres', 'saya', 'pilih', 'prabowo', 'sandi', 'pilpres', 'mayoritas', 'rakyat', 'presiden', 'yang', 'mampu', 'laku', 'ubah', 'yang', 'lebih', 'baik']

3.3 Pelabelan Data

Setelah data melewati proses *pre-processing*, selanjutnya data akan diberi label dengan menggunakan *library textblob* secara otomatis. Namun sebelum melakukan pelabelan, teks harus diterjemahkan terlebih dahulu kedalam bahasa inggris mengingat *textblob* hanya mampu melabeli teks berbahasa inggris saja [15].

Tabel 2. Pelabelan Data

Tweet	Terjemahan	Polaritas	Label
pilpres saya pilih prabowo sandi pilpres mayoritas rakyat presiden yang mampu melakukan perubahan yang lebih baik	presidential election I choose prabowo sandi presidential election majority of the people president who is able to make better changes	0.56	Positif
propaganda kotor ini sangat bodoh politisi gerindra sedang mengatakan prabowo kalah pilpres lawan	This dirty propaganda is so stupid gerindra politicians are saying prabowo lost the presidential election to opponent	-0.42	Negatif

3.4 Word Embedding

Proses *word embedding* memanfaatkan *fasttext* dan *pre-trained model* bahasa Indonesia cc.id.300.bin yang didapatkan dari situs resmi fasttext untuk melakukan ekstraksi fitur pada data teks. Setiap kata yang divektorisasi akan menjadi fitur bagi model untuk melakukan analisis sentimen. Berikut adalah contoh hasil vektorisasi kata pada kalimat “siap dukung ridwan kamil pilpres”.

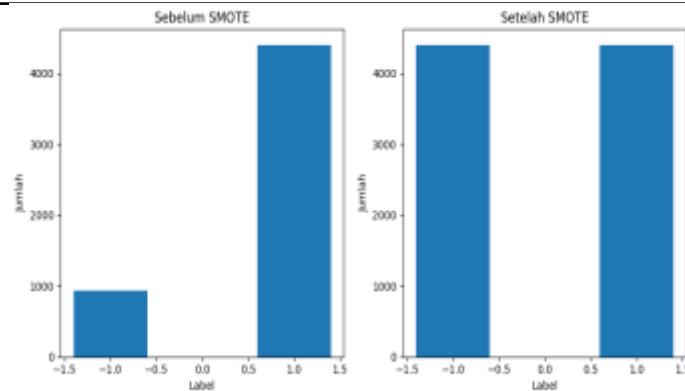
Tabel 3. Word Embedding

Kata	Dimensi				
	1	2	3	...	300
Siap	-6.5508	-1.8687	3.7531	...	1.7679
dukung	0.0473	-0.0429	0.0149	...	0.0925
ridwan	0.0925	-3.7221	7.1045	...	2.7389
kamil	0.0155	-0.0466	0.1281	...	-0.0429
pilpres	-0.0429	2.5110	-2.6334	...	-4.8905

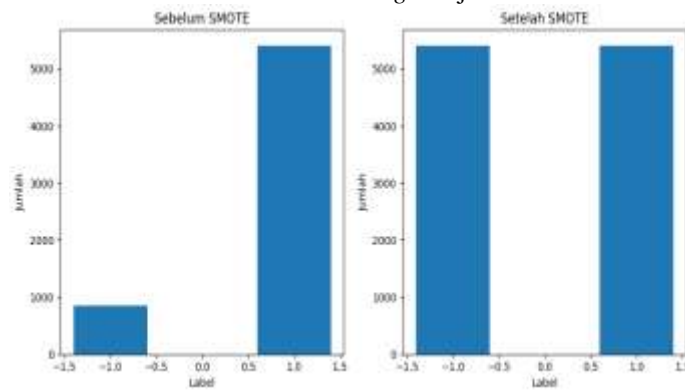
3.5 Data Balancing

Setelah melewati proses pelabelan, didapatkan perbandingan jumlah label positif dan negatif sangat jauh berbeda. Maka dari itu diperlukan *data balancing* untuk menyeimbangkan data dan menghindari model dari *overfitting* [16]. Proses *data balancing* dilakukan dengan menggunakan teknik *oversampling* SMOTE. SMOTE bekerja dengan menerapkan *sampling* ulang pada kelas minoritas lalu menghasilkan sampel sintetis baru sehingga dapat meningkatkan jumlah sampel pada kelas minoritas tanpa mengulangi sampel yang sudah ada.

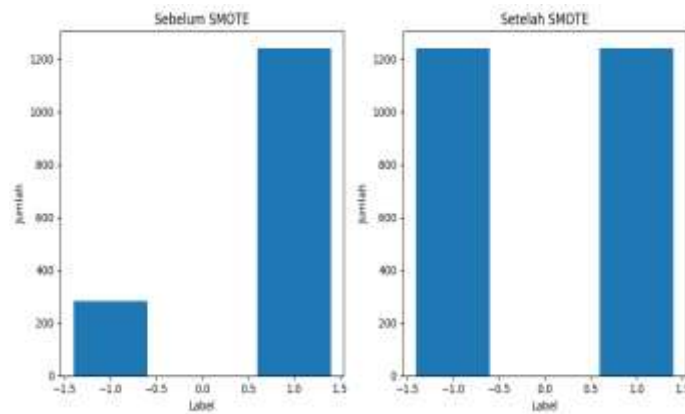
Grafik dibawah menunjukkan perbandingan label positif dan negatif sebelum dan sesudah dilakukan data balancing pada masing masing data latih calon presiden.



Gambar 1. Data Balancing Ganjar Pranowo



Gambar 2. Data Balancing Prabowo Subianto

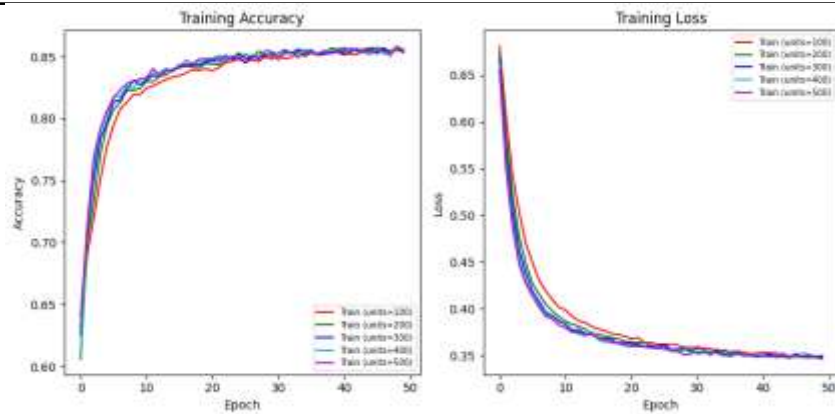


Gambar 3. Data Balancing Ridwan Kamil

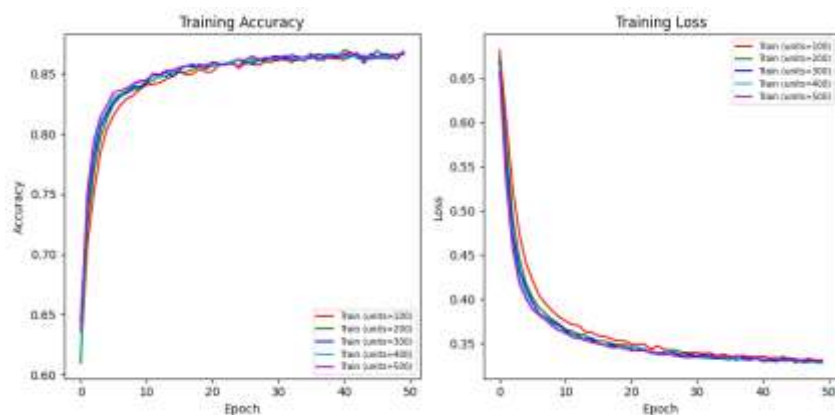
3.6 Implementasi LSTM

Setelah data latih melewati proses *balancing* selanjutnya yaitu menguji data latih pada pemodelan LSTM. Pada *layer* pertama model menggunakan variasi jumlah unit yang berbeda demi menghasilkan akurasi model yang baik. Variasi jumlah unit yang diuji coba adalah 100, 200, 300, 400, dan 500. Untuk menghindari *overfitting* pada model diperlukan *dropout* yang terletak pada *layer* kedua model. Nilai *dropout* yang digunakan sebesar 0.3 yang artinya 30% dari jumlah unit akan dipilih secara acak dan akan dinonaktifkan agar model tidak terlalu kompleks mempelajari data. Pada *layer* terakhir yaitu *dense layer* berisi nilai unit sebesar 1 yang berfungsi untuk menghasilkan output biner dan fungsi aktivasi yang digunakan adalah sigmoid. Lalu penulis juga menetapkan nilai *learning rate* sebesar 0.001 untuk optimasi model dengan adam.

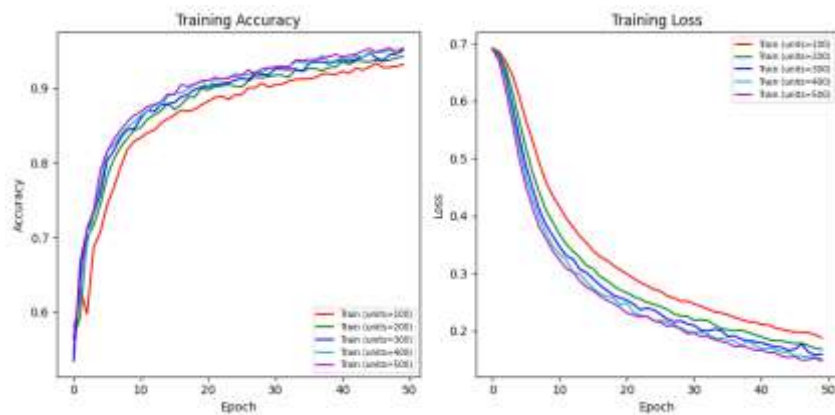
Setelah parameter model sudah diterapkan, selanjutnya menyusun *layer* yang telah dibangun sebelumnya untuk dilakukan pelatihan dengan pengukuran *loss* dan akurasi. Model dijalankan dengan menggunakan *batch size* 32 dan jumlah *epoch* 50. Banyaknya jumlah *epoch* mempengaruhi akurasi model dalam mempelajari data [17]. Berikut grafik perbandingan hasil akurasi dan *loss* pada setiap jumlah unit.



Gambar 4. Grafik Akurasi dan Loss Ganjar Pranowo



Gambar 5. Grafik Akurasi dan Loss Prabowo Subianto



Gambar 6. Grafik Akurasi dan Loss Ridwan Kamil

Tabel 4. Hasil Akurasi dan Loss

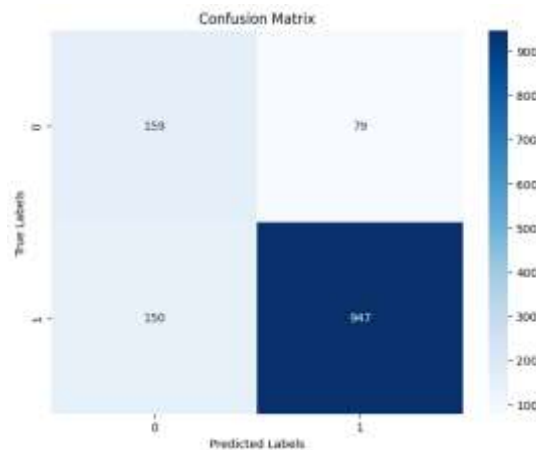
Data	Jumlah Unit	Akurasi	Loss
Ganjar Pranowo	100	0.8548	0.3492
	200	0.8553	0.3496
	300	0.8533	0.3500
	400	0.8562	0.3474
	500	0.8583	0.3465
Prabowo Subianto	100	0.8674	0.3313

	200	0.8672	0.3278
	300	0.8663	0.3296
	400	0.8656	0.3284
	500	0.8686	0.3284
	100	0.9316	0.1883
	200	0.9432	0.1679
Ridwan Kamil	300	0.9505	0.1597
	400	0.9529	0.1518
	500	0.9537	0.1476

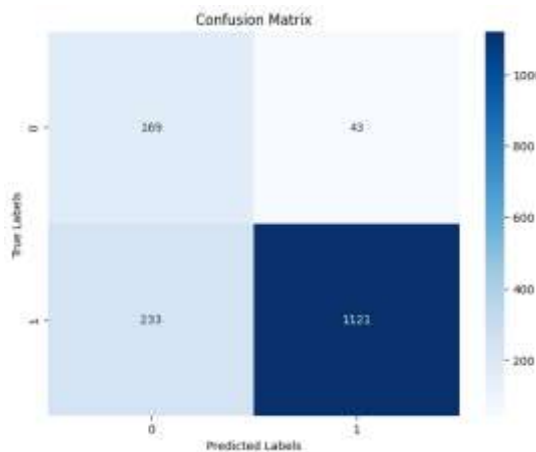
Berdasarkan tabel diatas, terlihat jumlah unit 500 mampu menghasilkan nilai akurasi yang baik dan nilai loss yang rendah pada setiap data. Maka dari itu model dengan jumlah unit 500 disimpan dan akan digunakan untuk menguji data uji nantinya.

3.7 Uji Model dan Evaluasi Performa

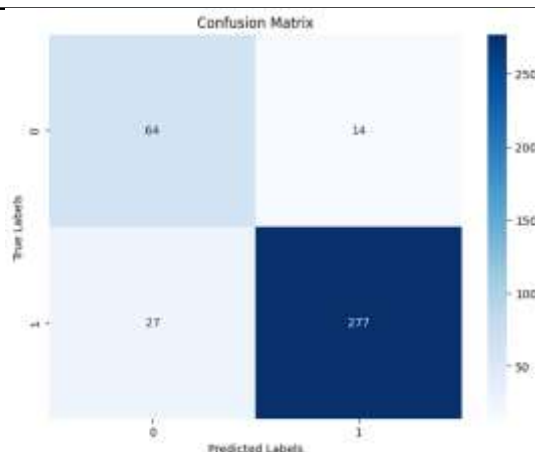
Tahap terakhir yaitu menguji model dengan data uji dan mengevaluasi performa model dengan menggunakan *confusion matrix*. Berikut adalah grafik *heatmap confusion matrix* untuk setiap model calon presiden.



Gambar 7. *Confusion Matrix* Ganjar Pranowo



Gambar 8. *Confusion Matrix* Prabowo Subianto



Gambar 9. Confusion Matrix Ridwan Kamil

Tabel 5. Hasil Pengujian Model Ganjar Pranowo

Akurasi	82.85%
Presisi	86.33%
Recall	92.30%
F1-Score	89.21%

Tabel 6. Hasil Pengujian Model Prabowo Subianto

Akurasi	82.38%
Presisi	82.79%
Recall	96.79%
F1-Score	89.04%

Tabel 7. Hasil Pengujian Model Ridwan Kamil

Akurasi	89.27%
Presisi	91.12%
Recall	95.19%
F1-Score	93.11%

Berdasarkan hasil pengukuran *confusion matrix* diatas, model memiliki akurasi yang cukup baik yang artinya mampu mengklasifikasi sentimen dengan baik. Model juga memiliki nilai presisi dan *recall* yang cukup tinggi yang menandakan model mempunyai sensitivitas sangat baik dalam mengenali sentimen positif yang merupakan sentimen yang dominan pada dataset. Sedangkan nilai pada *f1-score* mengindikasikan bahwa presisi dan *recall* memiliki keseimbangan yang baik dalam klasifikasi.

4. KESIMPULAN

Penelitian analisis sentimen ini menggunakan data yang diperoleh dari Twitter dengan teknik data scraping menggunakan SNScrape. Data yang didapatkan yaitu 11.984 data Ganjar Pranowo, 14.352 data untuk Prabowo Subianto, 16.646 data untuk Anies baswedan, 2.325 data Ridwan Kamil dan 10.984 data Puan Maharani. Lalu data melewati proses pre-processing, data labelling, word embedding, dan data balancing. Kemudian melakukan pemodelan klasifikasi LSTM dengan mempertimbangkan hyperparameter tuning seperti variasi jumlah unit, learning rate, dropout, batch size, hingga jumlah epoch agar menghasilkan model yang terbaik. Model yang terbaik diuji coba dan dievaluasi dengan confusion matrix. Hasil evaluasi performa setiap model calon presiden adalah 82% akurasi, 86% presisi, 92% recall, dan 89% f1-score pada model Ganjar Pranowo. 82% akurasi, 82% presisi, 96% recall, dan 89% f1-score pada model Prabowo Subianto. 89% akurasi, 91% presisi, 95% recall, dan 93% f1-score pada model Ridwan Kamil.

DAFTAR PUSTAKA

- [1] APJII, "Survei Profil Internet Indonesia 2022," 2022. <https://apjii.or.id/gudang-data/hasil-survei>.
- [2] We Are Social, "Digital 2022: Another Year Of Bumper Growth," 2022. <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/>.
- [3] F. Firmansyah, "Analisis sentimen terhadap hasil Pilpres 2019 dengan membandingkan algoritma Support Vector Machine (SVM) dan algoritma K-Nearest Neighbor (KNN)," UIN Sunan Gunung Jati Bandung, 2019.
- [4] M. R. Fais Sya' bani, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan

-
- Algoritme Naïve Bayes,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [5] M. Z. Rahman, Y. A. Sari, and N. Yudistira, “Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5120–5127, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [6] D. A. Firdlous and R. Andrian, “Analisis Sentimen Publik Twitter terhadap Pemilu 2024 menggunakan Model Long Short Term Memory,” *Sist. J. Sist. Inf.*, vol. 12, no. 1, pp. 52–60, 2023.
- [7] P. M. Sosa, “Twitter Sentiment Analysis using combined LSTM-CNN Models,” *Eprint Arxiv*, pp. 1–9, 2017.
- [8] E. Yolanda Talahaturuson, L. Junaedi, and A. Bimo Gumelar, “Analisis Sentimen Pergerakan Harga Mata Uang Kripto (Cryptocurrency) Menggunakan TextBlob-NLTK (Natural Language Toolkit),” *Jlk*, vol. 5, no. 2, pp. 1–7, 2022.
- [9] I. G. S. Mas Diyasa, N. M. I. Marini Mandenni, M. I. Fachrurrozi, S. I. Pradika, K. R. Nur Manab, and N. R. Sasmita, “Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1125, no. 1, p. 012034, 2021, doi: 10.1088/1757-899x/1125/1/012034.
- [10] S. Mujahidin, B. Prasetyo, and M. C. C. Utomo, “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes,” *Voteteknika (Vocational Tek. Elektron. dan Inform.)*, vol. 10, no. 3, p. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [11] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques*, Third Edit. Unites States of America: Morgan Kauffman, 2012.
- [12] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, “Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [13] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, “Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012102.
- [14] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” *INTEGER J. Inf. Technol.*, vol. 2, no. 1, pp. 32–41, 2017, doi: 10.31284/j.integer.2017.v2i1.95.
- [15] R. Azhar, A. Surahman, and C. Juliane, “Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, 2022.
- [16] C. Cahyaningtyas, Y. Nataliani, and I. R. Widasari, “Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE,” *Aiti*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [17] M. Wasil, H. Harianto, and F. Fathurrahman, “Pengaruh Epoch pada Akurasi menggunakan Convolutional Neural Network untuk Klasifikasi fashion dan Furniture,” *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 53–61, 2022, doi: 10.29408/jit.v5i1.4393.