

Sentimen Analisis Twitter Ibu Kota Negara Nusantara Menggunakan Long Short-Term Memory dan Lexicon Based

Saepul Aripriyanto ^{1*}, Tukino ², Ammar Sufyan ³, Riandi Nandaputra ³,

¹ Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

² Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Indonesia

³ Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

* saepul.aripriyanto@uinjkt.ac.id, ² tukino@ubpkarawang.ac.id, ³ ammar.sufyan20@mhs.uinjkt.ac.id, ⁴ riandi.np20@mhs.uinjkt.ac.id

ABSTRACT – In the 2020 APJII Survey, Indonesians who use Twitter for social media are 10% of the entirety of social media users in Indonesia (APJII 2020), the issue that is being discussed a lot both on social media and offline discussions, is the National Capital City (IKN) of the Archipelago, which is the new capital city of the Republic of Indonesia. The relocation of the capital city raises pros and cons. With these pros and cons, an analysis of public sentiment regarding the IKN issue becomes a necessity. In this research, the model that will be used to analyze sentiment analysis uses the Long Short-Term Memory (LSTM) algorithm and lexicon based on two scenarios, which is the scenario that uses 100 data of tweets and 5112 data of tweets. The results for the 100 tweets dataset scenario obtained 64% accuracy, 40% precision, 64% recall, and 79% F1-Score. Meanwhile, the results for the 5112 tweets data scenario obtained 79% accuracy, 82% precision, 79% recall, 79% F1-Score. The sentiment results obtained from the 5112 tweets data are 44.8% positive sentiment, 36.2% negative sentiment and 19.0% neutral sentiment. Based on this research, the number of datasets will affect the performance of deep learning models built using lexicon based and LSTM algorithms

Keywords: IKN Nusantara, Analysis Sentiment, LSTM, Lexicon Based, Twitter

ABSTRAK – Dalam Survei APJII 2020, masyarakat Indonesia yang menggunakan Twitter untuk media sosial adalah 10% dari total pengguna sosial media di Indonesia (APJII 2020), kemudian masalah yang sedang banyak diperbincangkan baik di media sosial ataupun diskusi luring adalah Ibu Kota Negara (IKN) Nusantara yang merupakan ibu kota baru Republik Indonesia. Dengan adanya pemindahan ibu kota, hal ini menimbulkan pro dan kontra. Dengan adanya pro dan kontra tersebut, diperlukan sebuah analisis mengenai sentimen masyarakat terkait isu IKN tersebut. Dalam penelitian ini, model yang akan digunakan untuk menganalisis sentimen analisis menggunakan algoritma Long Short-Term Memory (LSTM) dan lexicon based dengan dua skenario yaitu dengan skenario menggunakan data 100 tweet dan 5112 tweets. Hasil skenario tersebut untuk dataset 100 tweet diperoleh accuracy 64%, precision 40%, recall 64%, dan F1-Score 79%. Sementara itu, hasil dari 5112 data tweet diperoleh accuracy 79%, precision 82%, recall 79%, F1-Score 79%. Untuk hasil sentimen dari data 5112 tweet, diperoleh 44,8% sentimen positif, 36,2% sentiment negatif dan 19,0% netral. Berdasarkan penelitian ini, jumlah dataset akan berpengaruh pada performa model deep learning yang dibangun menggunakan lexicon based dan Algoritma LSTM.

Kata Kunci: IKN Nusantara, Sentimen Analisis, LSTM, Lexicon Based, Twitter

1. PENDAHULUAN

Dalam Survei APJII 2020, masyarakat Indonesia yang menggunakan Twitter untuk social media adalah 10% dari Total Pengguna Sosial Media di Indonesia (APJII 2020). Kemudian masalah yang sedang banyak diperbincangkan baik di sosial media dan diskusi Luring adalah Ibu Kota Negara (IKN) Nusantara yang merupakan Ibu Kota baru Republik Indonesia. dengan adanya pemindahan Ibu Kota hal ini menimbulkan pro dan kontra, ada Sebagian Masyarakat Indonesia yang tidak setuju dan Sebagian lain menyetujui dari Pemindahan Ibu Kota tersebut. Selain dari Pemindahan Ibu Kota juga Nama Ibu Kota yaitu “Nusantara” juga

ikut menuai pro dan kontra karena Nusantara dianggap daerah yang Luas yang sebanding dengan Indonesia [1].

Metode Lexicon Based adalah metode yang bersifat *unsupervised learning* dimana prosesnya tidak membutuhkan data latih. Metode ini memerlukan kamus kata yang berisi kata positif dan kata negatif. Untuk menentukan orientasi sentimen dari suatu teks didasarkan atas fungsi positif dan negatif. Perbandingan metode *machine learning* klasik dan *deep learning* dilakukan untuk analisis sentimen pada *tweets* pada Twitter terkait kebijakan *New Normal*. *Long Short-Term Memory (LSTM)* merupakan metode dari *deep learning* yang dibandingkan dengan Naïve Bayes (NB).

Hasil penelitian ini menunjukkan bahwa LSTM memiliki akurasi, presisi dan *recall* paling baik sebesar 8,33% dibandingkan Naïve Bayes (NB) yang memperoleh



nilai akurasi, presisi dan recall sebesar 80% [2]. Kemudian jika LSTM dibandingkan dengan SVM dan KNN untuk memprediksi Bug hasil evaluasinya adalah untuk LSTM Accuracy 89,8%, F-Measure 89,2%, recall 89,7%, precision 87,6% , MCC 79,6%. Untuk SVM Accuracy 86,5%, F-Measure 86,5%, recall 86,5%, precision 86,6%, MCC 73,2%. Sedangkan untuk KNN Accuracy 74,1%, F-Measure 74,2%, recall 74,1%, precision 74,3%, MCC 48,5%, dari hasil tersebut LSTM-lah yang paling baik [3]. Kemudian jika LSTM dibandingkan dengan SVM untuk menganalisa pasar stok saham maka hasil keduanya adalah LSTM memberikan akurasi yang tinggi ketika ada menggunakan data set yang besar tetapi untuk melakukan prediksi harga saham memerlukan waktu yang lama dari pada SVM [4]. Sedangkan untuk memprediksi harga saham perbandingan antara LSTM dan SVM lebih baik akurasi LSTM dengan nilai 66,83% sedangkan akurasi SVM dengan nilai 65,20% [5].

2. DASAR TEORI

A. IKN Nusantara

Nusantara adalah Nama Ibu Kota Baru dari Republik Indonesia. Hal ini disebutkan oleh Menteri Perencanaan Pembangunan Nasional/Kepala Bappenas Suharso Monoarfa dalam rapat Panitia Khusus (Pansus) Rancangan Undang-Undang Ibu Kota Negara (RUU IKN) di Kompleks Parlemen, Jakarta, Senin (17/1/2022) setelah mendapatkan persetujuan dari Presiden Republik Indonesia [6].

B. Analisis Sentimen

Natural Language Processing (NLP) adalah sebuah algoritma dan sistem yang memungkinkan komputer dapat memahami serta melakukan tugas yang melibatkan bahasa manusia. NLP dikenal juga menjadi Linguistik Komputasi, Bahasa Komputer, serta Pemrosesan Bahasa. NLP dapat melakukan analisis bahasa manusia baik tertulis maupun lisan untuk mendapatkan informasi yang berguna. NLP bertujuan untuk memahami bahasa manusia alami yang digunakan untuk mengatasi masalah komputer [7].

Tugas dasar analisis sentimen adalah mengklasifikasikan teks ke dalam kalimat atau dokumen dan menentukan apakah pendapat yang diungkapkan dalam kalimat atau dokumen tersebut positif atau negatif. Analisis humor juga dapat mengekspresikan emosi seperti kesedihan, kegembiraan, dan kemarahan. Ekspresi atau perasaan mengacu pada fokus subjek tertentu. Satu pernyataan pada satu subjek mungkin memiliki arti yang berbeda dari yang lain [8].

C. Metode Lexicon Based

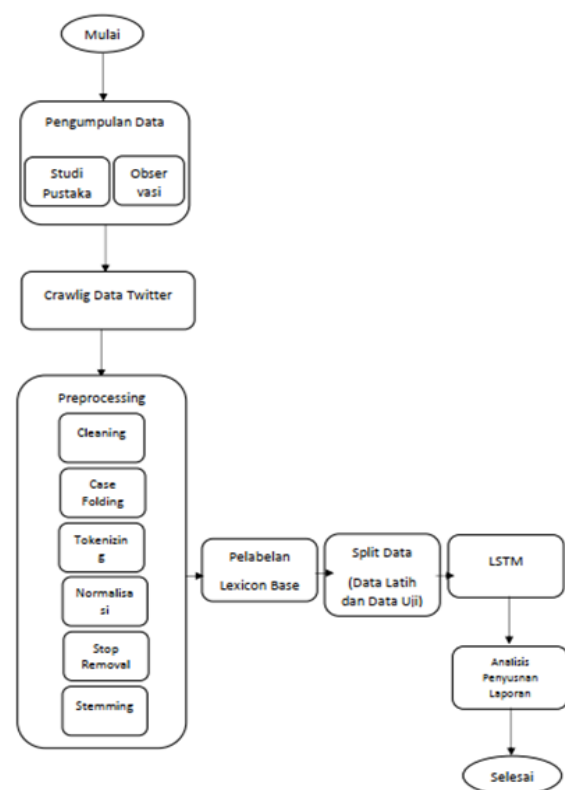
Ada teknik berbasis kamus yang terkenal dalam analisis sentimen (*lexicon based*). Secara umum, proses analisis sentimen berbasis kamus (*lexicon-based*) terdiri dari pengumpulan data, pemuatan kamus, pra-pemrosesan, dan ekstraksi kata kunci [9].

D. Long Short-Term Memory

Hochreiter dan Schmidhuber (1997) memperkenalkan memori jangka pendek panjang (LSTM) untuk memecahkan masalah difusi gradien dari jaringan saraf berulang; *Recurrent Neural Network (RNN)*. LSTM adalah varian dari RNN yang dirancang untuk menghindari masalah mengingat informasi jangka panjang di RNN [10].

3. METODOLOGI

Metodologi penelitian dilakukan dalam beberapa tahap, mulai dari pengumpulan data hingga pembuatan model dan evaluasi model. Langkah-langkah pengolahan data Twitter hingga analisis data menggunakan algoritma LSTM dengan menggunakan Lexicon Based, Gambar 1.



Gambar 1. Metodologi Penelitian

A. Crawling Data Twitter

Pengambilan data dari Twitter menggunakan Twitter API dan menggunakan bahasa pemrograman Python. Kata kunci yang digunakan adalah IKN Nusantara. Kata kunci tersebut dipilih karena mewakili dari *tweet* masyarakat Indonesia dalam mengomentari IKN Nusantara. Penarikan data di Twitter dibatasi hanya dari Bulan Juli 2022 ke belakang (*backdate*). Untuk jumlah data terdapat 2 kelompok yaitu data yang berjumlah 100 data (*tweet*) dan 5112 data (*tweet*) hal ini dimaksud untuk menganalisis apakah berpengaruh jumlah data terhadap *performance* LSTM dan Lexicon Based.

B. Pre-Processing

Proses penarikan data dari Twitter dengan menggunakan API Twitter yaitu Tweepy dan dengan kata kunci IKN Nusantara. Setelah dilakukan penarikan data langkah selanjutnya adalah langkah *preprocessing* yaitu: (1) *Data Cleaning*, yaitu membersihkan *tweets* yang tidak sesuai sehingga menjadi data menjadi sesuai, (2) *Case Folding*, yaitu proses mengubah *tweet* menjadi huruf kecil semua sehingga tidak ada huruf kapital, (3) *Tokenization*, yaitu pembagian *tweet* yang berupa kalimat ke dalam beberapa kata, (4) *Stop Removal*, yaitu proses penyaringan, pemilihan kata yang dianggap penting dari hasil token yaitu kata-kata apa saja yang di gunakan untuk mewakili dokumen, (5) *Stemming*, yaitu proses mendapatkan kata dasar dengan menghapus awalan dan akhiran, pada proses ini menggunakan *Library* Sastrawi yang merupakan *library* Python untuk merubah kata ke bentuk kata dasarnya.

C. Pelabelan Lexicon Based

Pelabelan dilakukan dengan menggunakan kamus data Liu yang telah dilakukan penerjemahan ke dalam Bahasa Indonesia. Kamus Liu sendiri terdapat *lexicon based* positif yang mempunyai bobot positif (>0) dan kamus Lexicon Negatif (<0) Pelabelan digunakan untuk menentukan mana kata yang bernilai positif, negatif dan netral.

D. Long Short-Term Memory (LSTM)

Pada Tahapan ini kita akan melakukan membuat sebuah model *deep learning* dengan menggunakan LSTM. Hal pertama yang dilakukan adalah membuat Model dengan parameter *embed_dim = 16*, *hidden_unit = 16*, *dropout_rate = 0.2*, *optimizers = Adam*, *learning_rate = 0.001*. setelah Model siap maka Langkah selanjutnya adalah melakukan Hyperparameter Tuning hal ini

dilakukan untuk mendapatkan hyperparameter yang terbaik yang akan dijadikan sebagai alat untuk melatih model sehingga performance model akan lebih optimal.

4. HASIL DAN PEMBAHASAN

A. Preprocessing Data

Data yang diambil dalam penelitian ini adalah tweet twitter dengan kata Kunci IKN Nusantara. Untuk data yang diambil dilakukan skenario yaitu skenario dengan data set sedikit yaitu sekitar 100 *tweets* dan data skenario banyak yaitu 5112 *tweets*. Setelah data di *crawling* dengan menggunakan IP Twitter pada *social media* Twitter kemudian data dilakukan *preprocessing* dengan melakukan *step Cleansing Data*, *Case Folding*, *Tokenizing*, *Stop Removal* dan *Stemming*. Berikut merupakan tahapan dari masing – masing *Preprocessing*.

Tabel 1 menggambarkan data hasil *crawling* Twitter yang kemudian dilakukan *step Cleansing Data* dimana data dihilangkan dari tanda baca special, nomor, dan #.

Berdasarkan Tabel 2 dengan melalui *step Case Folding* data dirubah menjadi huruf kecil semua sehingga dalam data tidak terdapat huruf besar.

Berdasarkan Tabel 3 data yang sudah dilakukan *case folding* selanjutnya dilakukan *step Tokenizing* yaitu memisahkan data yang sebelumnya kalimay menjadi kata – perkata.

Pada Tabel 4 yaitu masuk kedalam *Stop Removal* proses *filtering*, pemilihan kata-kata penting dari hasil token yaitu kata-kata apa saja yang di gunakan untuk mewakili dokumen, kata yang tidak penting akan dilakukan penghapusan.

Pada Tabel 5 menjelaskan bahwa setelah dilakukan *stop removal* kemudian baru dilakukan *Stemming data* yaitu Proses untuk mendapatkan kata dasar dengan menghilangkan afiks dan sufiks. Untuk proses *Stemming* pada penelitian ini menggunakan *Library* Sastrawi.

Tabel 1. Proses *Cleansing* Data

No	Tweets	Text Clean
1	jika suatu program negara sudah dijalankan, uang yg sudah dikeluarkan pun bernilai triliunan, lalu saat ganti presiden program itu dihentikan, maka yg rugi siapa? yg rugi kita semua, maka, tidak ada jalan lain, selain kita tuntaskan program ini. #iknnusantara #ibukotanegara	jika suatu program negara sudah dijalankan, uang yg sudah dikeluarkan pun bernilai triliunan, lalu saat ganti presiden program itu dihentikan, maka yg rugi siapa? yg rugi kita semua. maka, tidak ada jalan lain, selain kita tuntaskan program ini.
2	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini ! https://t.co/y6tlctctbh #mataindonesia #pembangunanikn #iknnusantara #mengacupadapeta #petadasar	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini !
3	ini ya perkembangannya ikn. #ibukotanegara #iknnusantara #ikn https://t.co/elikzff96a	ini ya perkembangannya ikn.

Sumber: Hasil Penelitian (2022)

Tabel 2. Proses *Case Folding*

No	Tweets	Text Clean
1	jika suatu program negara sudah dijalankan, uang yg sudah dikeluarkan pun bernilai triliunan, lalu saat ganti presiden program itu dihentikan, maka yg rugi siapa? yg	jika suatu program negara sudah dijalankan, uang yg sudah dikeluarkan pun bernilai triliunan, lalu saat ganti presiden program itu



No	Tweets	Text Clean
	rugi kita semua, maka, tidak ada jalan lain, selain kita tuntaskan program ini. #iknnusantara #ibukotanegara	dihentikan, maka yg rugi siapa? yg rugi kita semua. maka, tidak ada jalan lain, selain kita tuntaskan program ini.
2	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini ! https://t.co/y6tlctctbh #mataindonesia #pembangunanikn #iknnusantara #mengacupadapeta #petadasar	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini !
3	ini ya perkembangannya ikn. #ibukotanegara #iknnusantara #ikn https://t.co/elikzff96a	ini ya perkembangannya ikn.

Sumber: Hasil Penelitian (2022)

Tabel 3. Proses *Tokenizing*

No	Tweets	Text Clean
1	jika suatu program negara sudah dijalankan uang yg sudah dikeluarkan pun bernilai triliunan lalu saat ganti presiden program itu dihentikan maka yg rugi siapa yg rugi kita semua maka tidak ada jalan lain selain kita tuntaskan program ini	['jika', 'suatu', 'program', 'negara', 'sudah', 'dijalankan', 'uang', 'yg', 'sudah', 'dikeluarkan', 'pun', 'bernilai', 'triliunan', 'lalu', 'saat', 'ganti', 'presiden', 'program', 'itu', 'dihentikan', 'maka', 'yg', 'rugi', 'siapa', 'yg', 'rugi', 'kita', 'semua', 'maka', 'tidak', 'ada', 'jalan', 'lain', 'selain', 'kita', 'tuntaskan', 'program', 'ini']
2	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini	['pembangunan', 'ikn', 'harus', 'mengacu', 'ke', 'peta', 'dasar', 'flash', 'news', 'baca', 'selengkapnya', 'disini']
3	ini ya perkembangannya ikn	['ini', 'ya', 'perkembangannya', 'ikn']

Sumber: Hasil Penelitian (2022)

Tabel 4. Proses *Stop Removal*

No	Tweets	Text Clean
1	jika suatu program negara sudah dijalankan uang yg sudah dikeluarkan pun bernilai triliunan lalu saat ganti presiden program itu dihentikan maka yg rugi siapa yg rugi kita semua maka tidak ada jalan lain selain kita tuntaskan program ini	['program', 'negara', 'dijalankan', 'uang', 'yg', 'dikeluarkan', 'bernilai', 'triliunan', 'ganti', 'presiden', 'program', 'dihentikan', 'yg', 'rugi', 'yg', 'rugi', 'jalan', 'tuntaskan', 'program']
2	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini	['pembangunan', 'ikn', 'mengacu', 'peta', 'dasar', 'flash', 'news', 'baca', 'selengkapnya']
3	ini ya perkembangannya ikn	['ya', 'perkembangannya', 'ikn']

Sumber: Hasil Penelitian (2022)

Tabel 5. Proses *Steeming*

No	Tweets	Text Clean
1	jika suatu program negara sudah dijalankan uang yg sudah dikeluarkan pun bernilai triliunan lalu saat ganti presiden program itu dihentikan maka yg rugi siapa yg rugi kita semua maka tidak ada jalan lain selain kita tuntaskan program ini	['program', 'negara', 'jalan', 'uang', 'yg', 'keluar', 'nila', 'triliun', 'ganti', 'presiden', 'program', 'henti', 'yg', 'rugi', 'yg', 'rugi', 'jalan', 'tuntask', 'program']
2	pembangunan ikn harus mengacu ke peta dasar flash news baca selengkapnya disini	['bangun', 'ikn', 'acu', 'peta', 'dasar', 'flash', 'news', 'baca', 'lengkap']
3	ini ya perkembangannya ikn	['ya', 'kembang', 'ikn']

Sumber: Hasil Penelitian (2022)

B. Lexicon Based

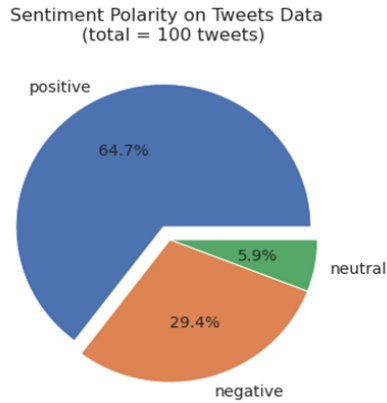
Setelah Data dilakukan *preprocessing* maka selanjutnya data *tweets* dilakukan pelabelan dengan menggunakan Lexicon based ke dalam 3 label yaitu Positif, Negatif dan Netral. Pelabelan dengan Lexicon based dilakukan dengan menghitung jumlah kata positif dan negatif di setiap *tweet*, jika jumlah kata positif lebih banyak maka dilabeli *tweet* positif dan jika jumlah kata negatif yang lebih banyak maka dilabeli *tweet* negatif dan jika jumlah kata positif dan negatif sama maka dilabeli dengan label

Netral. Kemudian untuk menentukan sebuah kata positif atau negatif lexicon based mempunyai kamus kata, dimana dalam penelitian ini kamus kata yang digunakan adalah Kamus Liu yang sudah diterjemahkan ke dalam Bahasa Indonesia.

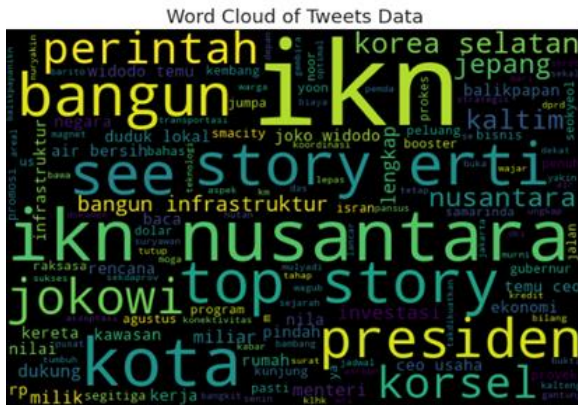
Dari hasil *preprocessing* 100 data *tweets* dihasilkan data bersih 51 data dimana sentimen positifnya 33, negatifnya 15 dan netral 3 atau dalam persen bisa dilihat pada Gambar 2. Dari Gambar 2 didapat untuk presentasi sentimen positif adalah 64,7%, sentimen negatif 28,4%



dan sentimen netral sebanyak 5,9% yang berarti sentimen positif tentang IKN jumlah banyak dari pada sentimen lainnya. Kemudian kata yang paling banyak muncul dalam data 100 tweets adalah pada Gambar 3.



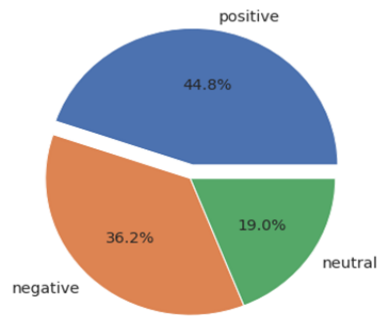
Gambar 2. Presentase Pengklasifikasian Lexicon Based pada 100 data tweets



Gambar 3. Kata yang sering Muncul dalam 100 Tweets

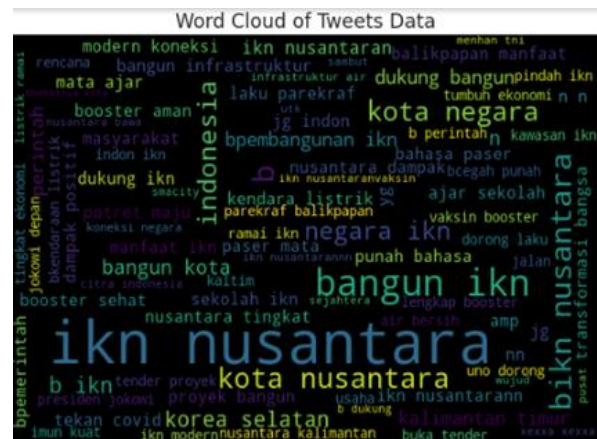
Berdasarkan Gambar 3 untuk 5 besar kata-kata yang sering muncul pada 100 data tweets adalah kata IKN, IKN Nusantara, kota, bangun dan story. Kemudian dari hasil preprocessing 5112 data tweets dihasilkan data bersih 1844 data dimana sentimen positif 826, negatif 668 dan netral 350 atau dalam presentase bisa dilihat pada Gambar 4.

Sentiment Polarity on Tweets Data (total = 5112 tweets)



Gambar 4. Presentase Pengklasifikasian Lexicon Based pada 5112 data tweets

Dari Gambar 4 didapat untuk presentasi sentimen positif adalah 44,8%, sentimen negatif 36,2% dan sentimen netral sebanyak 19,0% yang berarti sentimen positif tentang IKN jumlah banyak dari pada sentimen lainnya. Kemudian kata yang paling banyak muncul dalam data 5112 tweet pada Gambar 5.



Gambar 5. Kata yang sering Muncul dalam 5112 Tweets

Berdasarkan Gambar 5 untuk 5 besar kata-kata yang sering muncul pada 5112 data tweets adalah kata IKN Nusantara, bangun IKN, kota Nusantara, Indonesia dan kota negara.

```
# Create the model with the best hyperparameter which has been determined
model = KerasClassifier(build_fn = create_model,
                        #Model Parameters
                        dropout_rate = 0.2,
                        embed_dim = 32,
                        hidden_unit = 16,
                        optimizers = RMSprop,
                        learning_rate = 0.001,
                        # Fit Parameters
                        epochs=10,
                        batch_size=128,
                        # Initiate validation data, which is 10% data from data train. It's used for evaluation model
                        validation_split = 0.1)

model_prediction = model.fit(X_train, y_train)
```

Gambar 6. LSTM Model

C. LSTM Model

Berdasarkan Gambar 6 untuk model LSTM terdiri dari `embed_dim` adalah 16, `hidden_unit` adalah 16, `dropout_rate` adalah 0.2, `optimizers` adalah Adam, `learning_rate` adalah 0.001. Kemudian setelah model terbentuk sebelum dilakukan evaluasi maka model dilakukan *training* dengan menggunakan epoch 10.

D. Evaluasi Model

Sebelum melakukan Evaluasi Model dilakukan *splitting data* 80% data latih dan 20% data uji. Setelah dilakukan *splitting data* maka dilakukan evaluasi model dengan menggunakan *Confusion Matrix* dengan 2 skenario yaitu dengan 100 data dan 5112 data.

```
from sklearn.metrics import classification_report
print (classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	4
2	0.64	1.00	0.78	7
accuracy			0.64	11
macro avg	0.32	0.50	0.39	11
weighted avg	0.40	0.64	0.49	11

Gambar 7. Hasil *Confusion Matrix* dengan Jumlah Data 100

Berdasarkan Gambar 7 didapat hasil *Confusion Matrix* dengan menggunakan 100 *tweets* adalah *precision* 40%, *recall* 64%, *F1-Score* 49% sedangkan untuk *accuracy* adalah 64%.

```
from sklearn.metrics import classification_report
print (classification_report(y_test, y_pred))
```

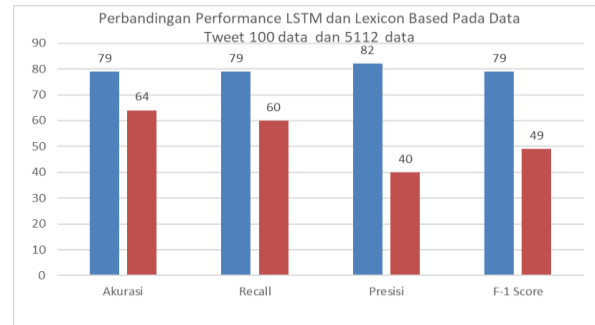
	precision	recall	f1-score	support
0	0.69	0.94	0.80	141
1	0.76	0.62	0.68	63
2	0.94	0.73	0.82	165
accuracy			0.79	369
macro avg	0.80	0.76	0.77	369
weighted avg	0.82	0.79	0.79	369

Gambar 8. Hasil *Confusion Matrix* dengan Jumlah Data 5112

Berdasarkan Gambar 8 hasil *Confusion Matrix* yang didapat dengan menggunakan 5112 *tweets* adalah *precision* 82%, *recall* 79%, *F1-Score* 79% dan *accuracy* adalah 79%.

Berdasarkan 2 skenario dengan jumlah data *tweets* yang berbeda diperoleh hasil perbandingan *accuracy*, *recall*, *precision* dan *F-1 Score* pada Gambar 9.

Dari hasil pada Gambar 9 diperoleh bahwa dengan menggunakan dataset 5112 diperoleh *Precision* 82%, *Recall* 79%, *F1-Score* 79% dan *Model Accuracy* adalah 79% dimana *performance* tersebut lebih baik dari pada menggunakan dataset 100 *tweets* yang menghasilkan *Precision* 40%, *Recall* 64%, *F1-Score* 49% sedangkan untuk *Accuracy* adalah 64%.



Gambar 9. Perbandingan *Performance Model* pada 100 Data *Tweets* dan 5112 Data *Tweets*

5. KESIMPULAN

Kesimpulan pada penelitian ini adalah dataset berpengaruh pada penerapan kombinasi algoritma LSTM dan *Lexicon Based* dalam sentimen analisis di sosial media Twitter mengenai IKN Nusantara. Semakin besar dataset, diperoleh *performance* yang akan lebih baik.

Saran kedepannya, penelitian dapat dilakukan dengan menggunakan algoritma dan ekstraksi fitur yang berbeda serta jumlah dataset juga berbeda, sehingga dapat lebih baik lagi mengamati pengaruh jumlah dataset pada *Machine Learning* maupun *Deep Learning*.

6. UCAPAN TERIMA KASIH

Penelitian ini dapat berlangsung dengan baik karena banyak pihak yang mendukung, salah satunya adalah lembaga di UIN Syarif Hidayatullah Jakarta yaitu Pusat Penelitian dan Publikasi (PUSLITPEN) yang membantu secara moral dan material.

DAFTAR PUSTAKA

- [1] S. Rasyid, "Tuai Pro Kontra, Ini Kata Pakar UGM Terkait Nama 'Nusantara' sebagai IKN Baru," *Merdeka.com*, 2022. <https://www.merdeka.com/jateng/tuai-pro-kontra-ini-kata-pakar-ugm-terkait-nama-nusantara-sebagai-ikn-baru.html> (accessed Jan. 20, 2022).
- [2] A. Rahman *et al.*, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 2, pp. 299–303, 2022, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/54704>
- [3] H. Bani-Salamah, M. Sallam, and B. al shboul, "A deep-learning-based bug priority prediction using RNN-LSTM neural networks," *E-Informatica Software Engineering Journal*, vol. 15, no. 1, pp. 29–45, 2021, doi: 10.37190/E-INF210102.
- [4] G. B. Srushti, "Comparison of Support Vector Machine and Long Short-Term Memory for Stock Market Analysis," pp. 2768–2770, 2020.

- [5] H. Ahuja, S. Mehra, and Y. Bhavaraju, "Stock Price Prediction Using SVM and LSTM," *International Research Journal of Engineering and Technology*, no. July, pp. 4963–4968, 2020, [Online]. Available: www.irjet.net
- [6] N. R. Aditya, "Ketuk Palu, DPR Sahkan RUU Ibu Kota Negara Jadi UU," *Kompas.com*, 2022. <https://nasional.kompas.com/read/2022/01/18/13511191/ketuk-palu-dpr-sahkan-ruu-ibu-kota-negara-jadi-uu> (accessed Jan. 18, 2022).
- [7] I. Huda, "Implementasi Natural Language Processing (NLP) untuk Aplikasi Pencarian Lokasi," *Jurnal Nasional Teknologi Terapan (JNTT)*, vol. 3, no. 2, p. 15, 2021, doi: 10.22146/jntt.35036.
- [8] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining tentang Masalah Obesitas di Indonesia," *Journal of Information Systems for Public Health*, vol. 4, no. 1, pp. 28–36, 2019.
- [9] Y. Azhar, "Metode Lexicon-Learning Based untuk Identifikasi Tweet Opini Berbahasa Indonesia," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 6, no. 3, p. 237, 2018, doi: 10.23887/janapati.v6i3.11739.
- [10] P. A. Qori, D. S. Oktafani, and I. Kharisudin, "Analisis Peramalan dengan Long Short Term Memory pada Data Kasus Covid-19 di Provinsi Jawa Tengah," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 5, pp. 752–758, 2022.