

Analisis Sentimen Masyarakat Terhadap Exchange Tokocrypto Pada Twitter Menggunakan Metode LSTM

Public Sentiment Analysis Against Tokocrypto Exchange on Twitter Using LSTM Method

Green Arther Sandag^{*1}, Jacqueline Waworundeng²

Universitas Klabat; Jln. Arnold Mononutu, 0431-891035

¹²Program Studi Informatika, Fakultas Ilmu Komputer UNKLAB, Airmadidi

e-mail: [1greensandag@unklab.ac.id](mailto:greensandag@unklab.ac.id), [2jacqueline.morlav@unklab.ac.id](mailto:jacqueline.morlav@unklab.ac.id)

Abstrak

Internet telah berperan penting dalam mempengaruhi semua aktivitas manusia di era teknologi saat ini. Dengan adanya internet dapat digunakan untuk berbagai tujuan, termasuk berbagi pengetahuan, bertransaksi, bersosialisasi, berbelanja, bisnis, pendidikan, dan banyak hal lain yang dapat dilakukan. Sementara internet semakin populer, berbagai macam transaksi digital terus berkembang, salah satunya yaitu pertukaran koin ke koin lain yang disebut cryptocurrency (mata uang kripto). Cryptocurrency adalah aset digital yang menggunakan kriptografi yang kuat untuk mengenkripsi transaksi keuangan, dan memverifikasi transfer aset. Salah satu exchange cryptocurrency untuk investasi di Indonesia yaitu Tokocrypto. Dengan begitu ramainya antusias dari cryptocurrency, banyak masyarakat Indonesia menggunakan media sosial seperti Twitter untuk mencari informasi, memberikan pendapat, juga informasi. Untuk mengklasifikasi tweet publik di Twitter ke dalam kategori positif dan negatif, diperlukan model analisis sentimen. Penelitian ini menggunakan metode Long Short Term Memory (LSTM), dimana LSTM merupakan salah satu pengembangan neural network yang dapat digunakan untuk pemodelan data time series terhadap tweet pengguna Twitter terhadap exchange Tokocrypto. Didapatkan sebanyak 2022 tweets positif, 1632 tweets negative, dan 1012 tweets netral.

Kata kunci—Internet, Mata Uang Kripto, Investasi, Pertukaran, Tokocrypto, LSTM

Abstract

The internet has played an important role in influencing all human activities in today's technological era. With the internet can be used for various purposes, including sharing knowledge, transacting, socializing, shopping, business, education, and many other things that can be done. While the internet is getting more and more popular, various kinds of digital transactions continue to develop, one of which is the exchange of coins for other coins which are called cryptocurrencies. Cryptocurrencies are digital assets that use strong cryptography to encrypt financial transactions, and verify asset transfers. One of the cryptocurrency exchanges for investment in Indonesia is Tokocrypto. With such enthusiasm for cryptocurrency, many Indonesians use social media such as Twitter to find information, provide opinions, as well as information. To classify public tweets on Twitter into positive and negative categories, a sentiment analysis model is needed. This study uses the Long Short Term Memory (LSTM) method, where LSTM is a neural network development that can be used for modeling time series data on Twitter users' tweets against the Tokocrypto exchange. There were 2022 positive tweets, 1632 negative tweets, and 1012 neutral tweets.

Keywords—Internet, Cryptocurrency, Investment, Exchange, Tokocrypto, LSTM

1 PENDAHULUAN

Perkembangan zaman yang begitu cepat saat ini membuat semua informasi dapat didapatkan dengan mudah melalui internet. Adanya internet sangat membantu manusia dalam melakukan aktivitas sehari-hari, begitu juga dengan digitalisasi dari sektor keuangan yang mana menerapkan pertukaran mata uang digital atau yang lebih dikenal *cryptocurrency* (mata uang kripto) sebagai alat investasi yang beresiko tinggi. Perdagangan mata uang digital menjadi terkenal akibat nilai tukarnya yang semakin meningkat. Pertukaran aset mata uang digital di Indonesia terjadi salah satunya di *platform exchange Tokocrypto*.

Tokocrypto adalah *platform cryptocurrency* yang menawarkan layanan kepada masyarakat umum untuk membeli dan menjual *cryptocurrency*. *Tokocrypto* bertujuan untuk menjadi bursa dalam pertukaran aset digital terkemuka di Asia Tenggara dengan menyediakan *platform* yang cepat, nyaman, instan, dan nyaman bagi konsumen untuk melakukan transaksi, serta pusa komunitas tempat dunia *crypto* dapat berkumpul, berbagi informasi dan keterampilan teknologi, dan memperjuangkan ide-ide baru[1]. *Exchange* (pertukaran) adalah tindakan yang melibatkan penawaran atau penjualan untuk memperoleh intensif yang bernilai serupa untuk mendapatkan sesuatu yang diinginkan atau dibutuhkan. Keadaan yang terjadi saat ini berupa kecemasan yang berlebihan dan merasakan ketakutan akan ketinggalan tren yang sedang berlangsung atau sering disebut *FOMO (Fear of Missing Out)*, sehingga banyak masyarakat Indonesia sering menggunakan media sosial seperti *Twitter* untuk memberikan pendapat terkait kritik, isu, saran, dan opini-opini publik mengenai sebuah tren. Banyak orang yang beramai-ramai mencari informasi untuk mendapatkan keuntungan maksimal terutama menggunakan *exchange* yang terkenal seperti *Tokocrypto*. *Tokocrypto* sendiri telah menyediakan berbagai platform, baik untuk *website* dan juga *mobile* sehingga membantu *user* agar lebih mudah dalam melakukan transaksi. Akibat dari digitalisasi yang begitu pesat terutama adanya mata uang digital resmi yang diluncurkan di Indonesia berupa *Toko Token (TKO)* yang dimiliki oleh *Tokocrypto* membuat masyarakat begitu antusias terhadap perkembangan dalam aset *digital*.

Analisis sentimen, juga dikenal sebagai penambangan opini, adalah metode untuk memahami, mengekstrak, dan menganalisis data tekstual secara otomatis untuk mendapatkan informasi sentimen dari sebuah kalimat, baik positif maupun negatif[2]. Analisis sentimen ini dapat membantu untuk mencari solusi terhadap masalah sebelumnya. Dalam penelitian ini akan digunakan metode *Long Short Term Memory (LSTM)*. Salah satu bentuk RNN (*Recurrent Neural Network*) adalah LSTM, yang merupakan sistem penyimpanan pola data. Karena LSTM akan mempelajari data mana yang harus disimpan dan data mana yang harus dibuang, serta sangat cocok untuk data teks, video, dan deret waktu. Memori di setiap neuron diatur oleh beberapa gerbang di LSTM [3].

Dalam penelitian [3] mengenai prediksi harga *Bitcoin* menggunakan metode *Long Short Term Memory Neural Network*, peneliti membangun sistem yang dibuat terdiri dari *Preprocessing Data*, inialisasi parameter, *training LSTM Network*, komposisi data, jumlah *hidden neuron*, *max epoch*, dan melakukan uji terhadap data *testing*. Pengujian tersebut hasil yang terbaik yaitu dengan komposisi data latih sebanyak 70% dan data uji 30%, parameter 1 pola *time series*, jumlah neuron *hidden 25*, *max epoch 100* dengan akurasi rata-rata latih 95,36% dan data *testing 93,5%*.

Penelitian lain [4] yaitu tentang analisis sentimen pada *Twitter* berbahasa Indonesia dengan metode *GloVe (Global Vector)*. Metode ini menggunakan *word embedding* yang terkait dengan model LSTM untuk membentuk model analisa sentimen. *Training* model *GloVe* menggunakan beberapa dimensi vektor yaitu dimensi 50, 100, 150, 200, 250, 300, dan 350, sedangkan model LSTM digunakan 64 unit LSTM dan *dropout 0.5* dengan lapisan aktivasi *ReLU 250* unit dan 1 unit aktivasi sigmoid. Hasil terbaik yang didapati untuk performa klasifikasi yaitu dengan tingkat akurasi 73%.

Referensi [5] tentang sentimen masyarakat terhadap *Covid-19* di *Twitter* yang menggunakan algoritma LSTM dan algoritma pembelajaran mesin tradisional. Diperoleh hasil

akurasi yang mencapai lebih dari 90%. Pada penelitian ini juga membandingkan dengan algoritma pembelajaran mesin lain seperti Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, dan K-Nearest Neighbor. Hasil akurasi algoritma pembelajaran mesin tradisional yang paling bagus ada algoritma SVM dengan accuracy sebesar 88%, recall 88% dan precision 88%.

2 METODE PENELITIAN

2.1 Analisis Data

Data untuk penelitian ini berasal dari tweet berbahasa Indonesia di *Twitter*. Tweet yang digunakan adalah tweet yang mengungkapkan pendapat orang Indonesia tentang *exchange Tokocrypto*. Total ada 4666 tweet yang digunakan sebagai data. Data dipilih secara manual dari tweet berbahasa Indonesia. Sentimen yang ditemukan di tweet terdapat tiga kategori sentimen yaitu positif, negatif, dan netral.

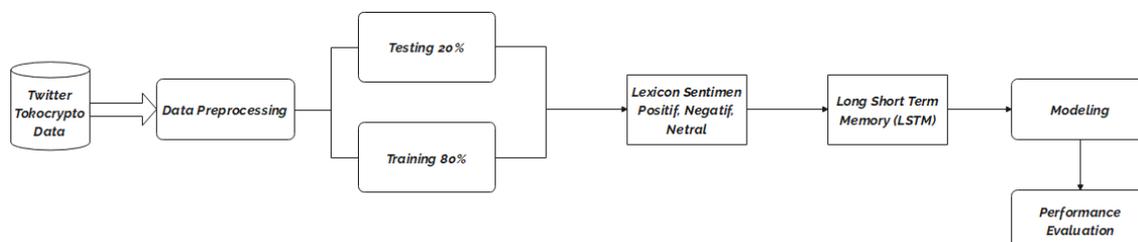
Tabel 1 Jenis Sentimen

Jenis Sentimen Tweet	Positif	Negatif	Netral
Tokocrypto	2022	1632	1012

Dari tabel diatas, diketahui data yang mengandung kata positif berjumlah 2022 data, data yang mengandung kata negatif berjumlah 1632 data, dan data yang mengandung kata netral berjumlah 1012 data. Data tersebut telah melalui proses *preprocessing* data dan menghapus tweet yang *double*.

2.2 Desain Penelitian

Gambar 1 memperlihatkan proses penerapan model LSTM. Dimana proses yang pertama kali dilakukan yaitu mengambil data *Twitter* (tweet) tentang *Tokocrypto*, dilanjutkan dengan data *preprocessing* untuk mengelola data yang ada. Kemudian *dataset* dilakukan *testing* dan *training* sebelum dilanjutkan pada tahap selanjutnya. Penelitian ini menggunakan *Lexicon Sentimen* terhadap kata positif, negatif, dan netral dari data tweet yang ada, lalu dilanjutkan pada tahap penerapan model *Long Short Term Memory*. Proses terakhir yang terjadi yaitu akan dibuat *modeling* dan dievaluasi data yang ada.

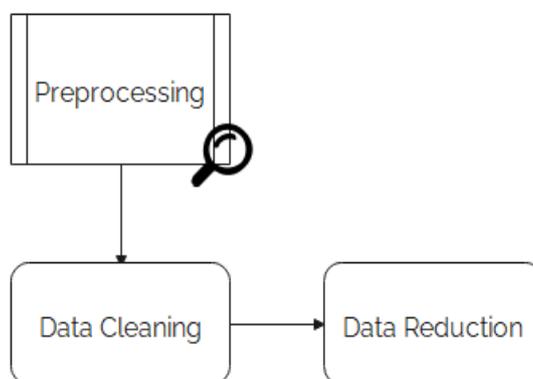


Gambar 1 Desain Penelitian

2.2.1 Data Preprocessing

Cleaning data dan data *reduction* adalah dua bagian dari data *preprocessing*. Proses pembersihan data yang hilang pada atribut dalam sebuah *dataset* agar lebih koheren dikenal dengan istilah *cleaning* data. Sedangkan data *reduction* adalah cara menghilangkan data dari atribut yang kurang penting dengan tujuan meminimalkan data dengan tetap menjaga keakuratannya. Langkah-langkah dalam *Preprocessing* :

- Step 1 : Menghapus *mentions*, *hashtag*, RT, *link*, dan angka.
 Step 2 : Membuat semua huruf menjadi huruf kecil.
 Step 3 : Memisahkan *string* menjadi kumpulan token.
 Step 4 : Melakukan normalisasi untuk kata alay/slang menjadi kata baku bahasa Indonesia.
 Step 5 : Menghapus *stopwords*.
 Step 6 : Membuat semua kata menjadi kata baku.
 Step 7 : Menyambung kembali token-token menjadi satu kalimat.



Gambar 2 Proses *Data Preprocessing*

Pada *Step 1* dilakukan dengan menghapus *mentions*, *hashtag*, RT, *link* dan angka agar *tweets* yang nantinya diproses lebih baik, kemudian semua huruf dibuat menjadi huruf kecil, juga memisahkan *string* menjadi kumpulan token. Pada *step* selanjutnya setiap kata dinormalisasikan untuk kata-kata alay/slang untuk menjadi kata baku bahasa Indonesia, dan menghapus *stopwords*, kemudian setiap kata diubah menjadi kata baku. Setelah semua proses sudah selesai, token yang ada itu akan disambung kembali menjadi satu kalimat. Hal ini dilakukan juga agar kata yang akan diproses sesuai dengan ketentuan yang ada. *Data preprocessing* yang telah selesai didapati sebanyak 4665 data *tweets*.

2.2.2 *Lexicon Sentiment*

Lexicon based sentiment classification adalah klasifikasi berdasarkan istilah positif, negatif, atau netral yang digunakan dalam *tweet* yang sudah dibersihkan. Istilah-istilah dalam kamus *lexicon* Indonesia sudah dicocokkan dengan klasifikasi ini, dan juga diterapkan *slang word* dari *Twitter COVID-19 Indonesia Sentiment Analysis* [6]. Sentimen *tweet* positif digambarkan sebagai kumpulan *tweet* yang berisi istilah-istilah positif. Sentimen *tweet* negatif digambarkan sebagai *tweet* yang mengandung istilah negatif. Namun, dalam kasus tertentu, jika kedua istilah ini memiliki nilai yang sama, *tweet* tersebut didefinisikan sebagai netral. Gambar 1 menunjukkan diagram alur untuk klasifikasi ini. Dari 5000 *tweet*, 2022 sentimen positif, 1632 sentimen negatif, dan 1012 sentimen netral diidentifikasi menggunakan pendekatan *Lexicon Based*. Pendekatan berbasis *lexicon* akan digunakan untuk mengklasifikasikan kalimat [7].

2.2.3 *Recurrent Neural Network (RNN)*

RNN merupakan teknik *deep learning* yang menggunakan arsitektur neural network untuk merepresentasikan data sekuensial. RNN dapat menyimpan informasi dari keadaan sebelumnya untuk menentukan kemungkinan menghasilkan keluaran berdasarkan keadaan tersebut. Namun, RNN memiliki cacat yang menyebabkannya mengalami gradien lenyap atau gradien yang meledak ketika terlalu banyak urutan yang diselesaikan. Istilah "gradien menghilang atau meledak" mengacu pada situasi di mana nilai gradien mungkin sangat kecil atau nol, serta

sangat tinggi. Hasilnya, metode yang disebut LSTM dibangun untuk mengatasi kekurangan ini dengan memperkenalkan sistem gerbang (*Long Short Term Memory*) [8].

2.2.4 *Long Short Term Memory (LSTM)*

LSTM merupakan pengembangan dari model RNN yang juga dapat digunakan untuk mengelola data yang bersifat sekuensial. Pada awalnya dikembangkan oleh *Hochreiter & Schmidhuber* (1997) untuk mengatasi permasalahan pada model RNN yakni *vanishing gradient* [3].

Persamaan variabel LSTM adalah x_t sebagai vektor *input*, F_t sebagai vektor aktivasi *forget gate*, I_t sebagai vektor aktivasi *input gate*, O_t untuk vector aktivasi *output gate*, H_t untuk vektor *hidden state*, juga dikenal sebagai unit keluaran dari unit LSTM), C_t untuk *state cell vector*, dan $W-U-b$ untuk matriks bobot dan parameter vektor bias yang harus diselidiki selama pelatihan. Berikut ini adalah versi persamaan LSTM yang disederhanakan [9]:

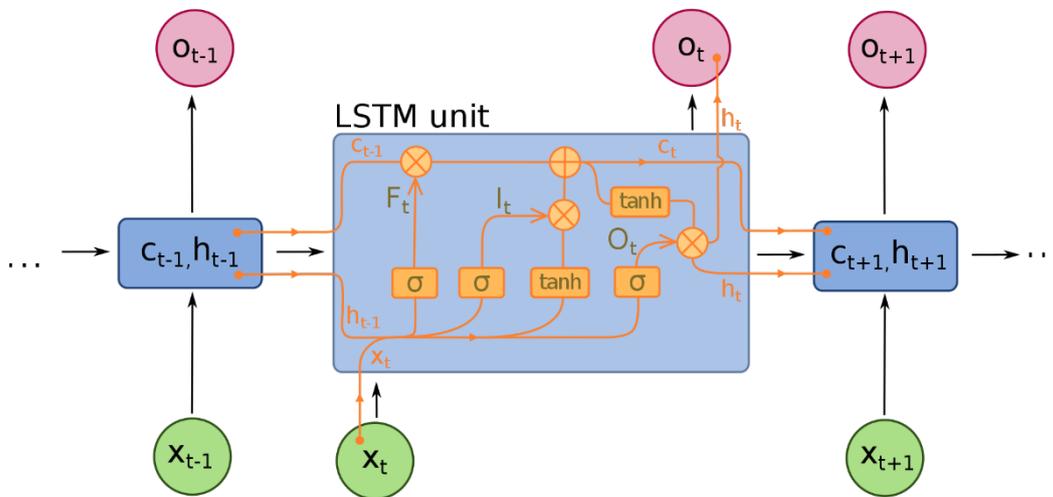
$$F_t = \sigma g (W_f x_t + U_f h_{t-1} + b_f) \tag{1}$$

$$I_t = \sigma g (W_i x_t + U_i h_{t-1} + b_i) \tag{2}$$

$$O_t = \sigma g (W_o x_t + U_o h_{t-1} + b_o) \tag{3}$$

$$C_t = f_t \circ c_{t-1} + i_t \circ \sigma g (W_c x_t + U_c h_{t-1} + b_c) \tag{4}$$

$$H_t = O_t \circ \sigma h (c_t) \tag{5}$$



Gambar 3 *Unit LSTM*

2.2.5 *Confusion Matrix*

Matriks penilaian yang disebut matriks ketidakpastian adalah salah satu cara untuk menguji hasil analisis sentimen. Ada banyak item yang dapat diukur dari hasil klasifikasi analisis sentimen dalam matriks ketidakpastian, antara lain akurasi, presisi, sensitivitas, dan *f-score* yang merupakan gabungan dari dua uji sensitivitas dan presisi [10], [11], [12].

Tabel 2 *Confusion Matrix*

	Prediksi Positif	Prediksi Negatif
Positif	TP	FN
Negatif	FP	TN

Tabel 2 menunjukkan contoh tabel matriks ketidakpastian dengan hasil klasifikasi dan nilainya masing-masing, yaitu *TP (True Positive)*, *FN (False Negative)*, *FP (False Positive)*, dan *TN (True Negative) (True Negative)*.

3 HASIL DAN PEMBAHASAN

Tahap pengumpulan data, tahap *preprocessing text*, tahap klasifikasi, dan tahap validasi adalah beberapa tahapan dalam *review* ini. Berikut hasil dari setiap tahap yang diselesaikan :

3.1 Pengumpulan Data

Metode yang dikenal sebagai *crawling* data digunakan selama tahap pengumpulan data. Bahasa pemrograman *Python* dan library *twint (Twitter Intelligence Tool)*. *Twint* adalah *twitter scraping* yang memungkinkan untuk mengambil *tweet* dari profil *twitter* tanpa menggunakan *API twitter* [13]. Prosedur ini digunakan untuk mengambil data *tweet* dari pengguna *twitter* yang memiliki subjek *Tokocrypto* di dalam profil mereka. Kata kunci '*tokocrypto*' digunakan untuk mengumpulkan data.

	id	username	created_at	tweet	replies_count	retweets_count	likes_count
0	1388332171245416454	tokocrypto	2021-05-01 03:19:18 UTC	Trading Hari Raya TKO Week 3 Leaderboard 1 Mei...	0	1	3
1	1388331258875285504	cryptofess_	2021-05-01 03:15:40 UTC	crypto! ini solusinya gimana ya? di tokocrypto...	1	0	0
2	1388330424582033409	wisnu118336	2021-05-01 03:12:21 UTC	@pangxuekai @tkobytokocrypto @Tokocrypto @cz_b...	0	0	0
3	1388329240798453762	xzc0013	2021-05-01 03:07:39 UTC	@pangxuekai @tkobytokocrypto @Tokocrypto kapan...	0	0	0
4	1388328701570404359	rendyricardo1	2021-05-01 03:05:31 UTC	@Tokocrypto knp gagal terus y https://t.co/Fp...	0	0	0
...
9995	1381240257400631300	moeccchan_100	2021-04-11 13:38:34 UTC	@dny_adr @iSafePal @Tokocrypto Selamat bang GG...	1	0	0
9996	1381239573653549057	sube_kp	2021-04-11 13:35:51 UTC	@Tokocrypto Wkwkwkw	0	0	0
9997	1381238893580705802	sak2ewae	2021-04-11 13:33:09 UTC	@KambingPrancis @Tokocrypto nah ini w lebih su...	0	0	0
9998	1381238851713179648	iang_bastian	2021-04-11 13:32:59 UTC	@KambingPrancis @Tokocrypto Cuma bisa bantu re...	0	0	0
9999	1381238778082168835	anton43397583	2021-04-11 13:32:41 UTC	@KambingPrancis @Tokocrypto Gue juga pernah wa...	0	0	0

Gambar 4 *Tweet Data*

3.2 Text Preprocessing

Tahap selanjutnya yaitu *text preprocessing*, beberapa proses seperti *case folding*, *tokenizing*, *stopword elimination*, dan *stemming* akan dilakukan pada data yang ada saat ini. *Case folding* adalah langkah pertama dalam prosedur *preprocessing text*. Semua huruf besar dalam dokumen akan dihapus *mentions*, *hashtag*, *RT*, *link*, dan angka. Proses penggantian baris baru dengan spasi juga dilakukan, menghapus tanda baca, mengubah semua huruf menjadi huruf kecil, memisahkan *string* menjadi kumpulan *token*, menghapus *stop words*, membuat semua kata menjadi kata baku, dan menyambung kembali token-token kata menjadi satu kalimat.

	id	username	created_at	replies_count	retweets_count	likes_count	text_clean	text_preprocessed
0	1388332171245416454	tokocrypto	2021-05-01 03:19:18 UTC	0	1	3	trading hari raya tko week leaderboard mei ...	[trading, raya, tko, week, leaderboard, mei, w...
1	1388331258875285504	cryptofess_	2021-05-01 03:15:40 UTC	1	0	0	crypto ini solusinya gimana ya di tokocrypto g...	[crypto, solusi, iya, tokocrypto, gin]
2	1388330424582033409	wisnu118336	2021-05-01 03:12:21 UTC	0	0	0	binance menginspirasi	[binance, inspirasi]
3	1388329240798453762	xzc0013	2021-05-01 03:07:39 UTC	0	0	0	kapan distribusi ke pak	[distribusi]
4	1388328701570404359	rendyricardo1	2021-05-01 03:05:31 UTC	0	0	0	knp gagal terus y	[gagal, iya]
...
4661	1381238357133455363	san69692	2021-04-11 13:31:01 UTC	0	0	0	sekali kali amal broo ikhlasin v	[kali, amal, broo, ikhlasin, v]
4662	1381238019030540296	kambingprancis	2021-04-11 13:29:40 UTC	0	0	0	oke bang tadi udh isi lagi kalo sampe minggu d...	[oke, bang, isi, minggu, gadiproses, datengin...
4663	138833378678939650	muchammadulim2	2021-05-01 03:25:41 UTC	0	0	0	dulu yoh di gaskeun	[yoh, gaskeun]
4664	1388159592803143681	flashvictory2	2021-04-30 15:53:32 UTC	1	0	0	pake tokocrypto aja ka lebih lengkap aku juga ...	[pakai, tokocrypto, kakak, lengkap, pakai]
4665	1384068001146695680	sdharm	2021-04-19 08:55:00 UTC	1	0	0	tinggal nunggu gabung dengan tokocrypto sekali...	[tinggal, nunggu, gabung, tokocrypto, wkwkwkw]

Gambar 5 *Text Preprocessing*

Setelah beberapa proses dilalui, akan dilaksanakan penentuan polaritas sentimen dari *tweet* yang didapatkan menggunakan *lexicon* sentimen untuk bahasa Indonesia.

	id	username	created_at	replies_count	retweets_count	likes_count	text_clean	text_preprocessed	polarity_score	polarity
0	1388332171245416454	tokocrypto	2021-05-01 03:19:18 UTC	0	1	3	trading hari raya tko week leaderboard mei ...	[trading, raya, tko, week, leaderboard, mei, w...	-4	negative
1	1388331258875285504	cryptofess_	2021-05-01 03:15:40 UTC	1	0	0	crypto ini solusinya gimana ya di tokocrypto g...	[crypto, solusi, iya, tokocrypto, gin]	3	positive
2	1388330424582033409	wisnu118336	2021-05-01 03:12:21 UTC	0	0	0	binance menginspirasi	[binance, inspirasi]	-3	negative
3	1388329240798453762	xzc0013	2021-05-01 03:07:39 UTC	0	0	0	kapan distribusi ke pak	[distribusi]	3	positive
4	1388328701570404359	rendyricardo1	2021-05-01 03:05:31 UTC	0	0	0	knp gagal terus y	[gagal, iya]	-4	negative
...
4661	1381238357133455363	san69692	2021-04-11 13:31:01 UTC	0	0	0	sekali kali amal broo iklasin v	[kali, amal, broo, iklasin, v]	4	positive
4662	1381238019030540296	kambingprancis	2021-04-11 13:29:40 UTC	0	0	0	oke bang tadi udh isi lagi kalo sampe minggu d...	[oke, bang, isi, minggu, gadproses, datengin...	-1	negative
4663	1388333778678939650	muchammadulinn2	2021-05-01 03:25:41 UTC	0	0	0	dulu yoh di gaskeun	[yoh, gaskeun]	0	neutral
4664	1388159592803143681	flashvictory2	2021-04-30 15:53:32 UTC	1	0	0	pake tokocrypto aja ka lebih lengkap sku juga ...	[pakai, tokocrypto, kakak, lengkap, pakaj]	-1	negative
4665	1384068001146695680	sdnarm	2021-04-19 08:55:00 UTC	1	0	0	tinggal tunggu gabung dengan tokocrypto sekali...	[tinggal, tunggu, gabung, tokocrypto, wkwkwkw]	-6	negative

Gambar 6 Lexicon Sentimen Bahasa Indonesia

3.3 Analisa dan Eksplorasi Data secara Visual

Proses analisa akan dicari kata-kata yang digunakan oleh *Top 10 Tweet* yang paling banyak diperbincangkan (berdasarkan jumlah *reply*). Pertama dibuat salinan dari *dataframes tweets*, lalu mengurutkan berdasarkan jumlah *reply* dan mengambil 10 *record* paling atas.



Gambar 7 Top 10 Tweet Berdasarkan Jumlah Reply

Proses analisa akan dicari kata-kata yang digunakan oleh *Top 10 Tweet* yang paling banyak diperbincangkan (berdasarkan jumlah *reply*). Pertama dibuat salinan dari *dataframes tweets*, lalu mengurutkan berdasarkan jumlah *reply* dan mengambil 10 *record* paling atas. Setelah mendapat *Top 10 Tweet* yang paling banyak diperbincangkan, akan dilakukan perbandingan polaritas sentimen pada data *tweet* “*tokocrypto*”. Hasil yang didapatkan sebagai berikut:

sebanyak (4666, 36). Proses data klasifikasi berikutnya adalah menyandikan (*encode*) data target ke dalam nilai numerik dan memisahkan data dengan komposisi data *train* 80% dan data *test* 20%. Hasil yang diperoleh yaitu (3732, 36) (3732,) (934, 36) (934,). Lanjutan dari proses sebelumnya yaitu membuat fungsi model dengan nilai *hyperparameter default*, model *KerasClassifier* akan digunakan dengan model *parameters*, yang dimana *dropout_rate* = 0.4, *embed_dim* = 64, *hidden_unit* = 32, *optimizers* = *RMSprop*, *learning_rate* = 0.001, *fit* parameters dengan *epoch* = 1500, *batch_size* = 512, dan memulai data validasi sebesar 10% data *train* sebesar 0.1. Diperoleh model *prediction* yang sama dengan model *x_train* dan *y_train* sebagai berikut:

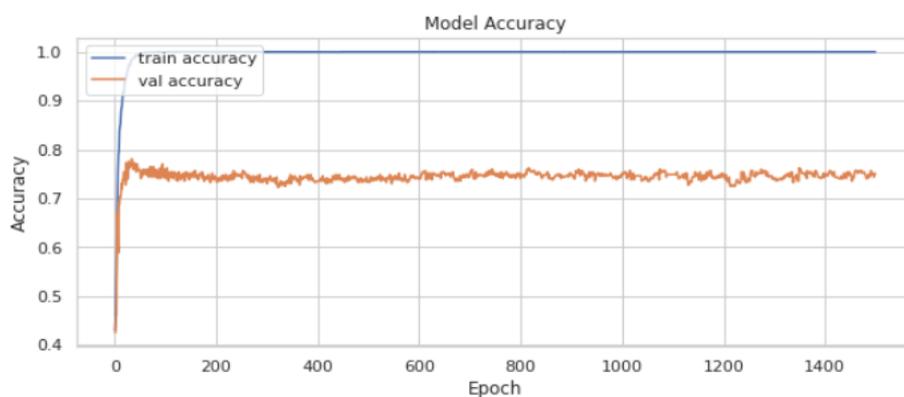
```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 36, 64)	320000
lstm (LSTM)	(None, 32)	12416
dropout (Dropout)	(None, 32)	0
dense (Dense)	(None, 3)	99

Total params: 332,515
 Trainable params: 332,515
 Non-trainable params: 0

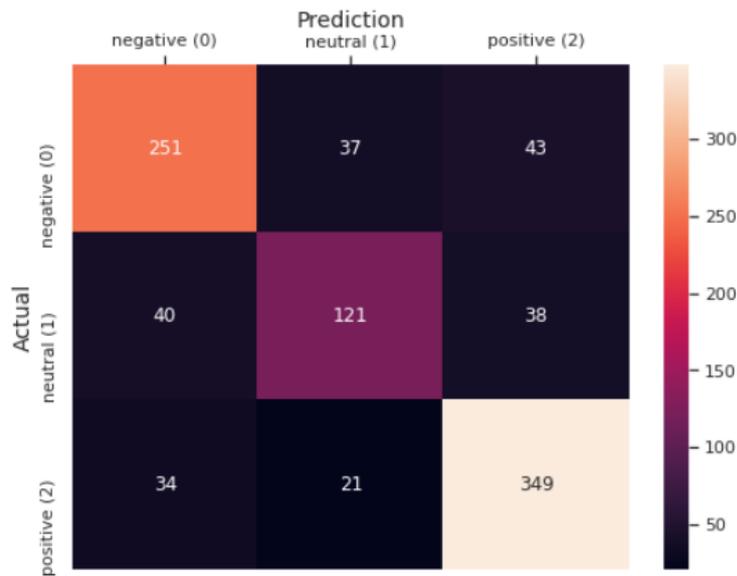
Gambar 10 Model Prediction

Setelah itu, kami kembangkan model dan melatihnya kemudian menyesuaikan dengan menggunakan data pelatihan yang dimiliki. Proses memvisualisasikan keakuratan pelatihan dan keakuratan validasi setelah pelatihan selesai untuk mengidentifikasi (*overfitting*). *Overfitting* merupakan fenomena dimana model kita akan mempelajari data *training* dengan tujuan mengoptimalkan nilai akurasi *training*, daripada mempelajari *training* secara garis besar. Pada hasil ini terlihat bahwa model yang dibuat ternyata *overfitting* dilihat dari hasil *train accuracy* lebih tinggi dengan *validation accuracy*. Hal ini terjadi karena dataset yang *imbalanced*. Visualisasi model *accuracy* berupa *train* dan *val accuracy* akan ditampilkan setelahnya pada gambar 10.



Gambar 11 Model Accuracy

Tahap terakhir yang akan dilalui yaitu memprediksi sentimen pada pengujian data dengan menggunakan model yang sudah dibuat, lalu divisualisasikan sebuah *confusion matrix*. Akurasi skor yang diperoleh nantinya akan ditampilkan pada *Model Accuracy on Test Data*. Setelahnya akan ditampilkan hasil dari sentimen prediksi pada pengujian data pada gambar 12.



Gambar 12 Prediction Sentiment Tweet Data and Visualize a Confusion Matrix

	text	polarity
0	ngakak anjir	Negative
1	gas ikutan biar dapet thr dari tokocrypto abdul	Negative
2	mending binance atau tokocrypto nik	Positive
3	bacaan enak buat yang mau lebih paham tentang solana nih sol	Positive
4	tokocrypto bagus gak buat pemula shm	Negative
...
929	beli tko pas launchpad tu keknya	Negative
930	mudah mudahan dapat	Positive
931	ndang bro mrcll	Neutral
932	selamat untuk tokocrypto semoga kedepannya semakin baik dan menjadi platform trading yang terbaik salam to the moon	Positive
933	menelusuri daya tarik solana bagi project defi sol	Positive

Gambar 13 Result from Prediction Sentiment on Data Test

4 KESIMPULAN

Berdasarkan dari hasil dan pembahasan dalam penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa penggunaan model *Long Short-Term Memory* (LSTM) bekerja dengan baik dalam mengklasifikasikan jumlah *tweets* tentang *tokocrypto* yang ada, juga menggunakan *Lexicon* Sentimen Bahasa Indonesia untuk diambil kata positif, negatif, dan netral. Dari 5000 *tweets* yang ada mengenai keyword *tokocrypto* dalam bahasa Indonesia, ditemukan sebanyak 2022 kata positif, 1632 kata negatif, dan 1012 kata netral.

5 SARAN

Untuk penelitian lebih lanjut di masa mendatang, dapat digunakan berbagai metode klasifikasi dan perbandingan terhadap metode lain, juga algoritma yang lain dalam menemukan algoritma dengan prediksi terbaik terhadap kata-kata positif, negatif, dan netral, serta mengoptimalkan efisiensi dan meminimalkan kesalahan dalam proses pemodelan atau *training* suatu data.

DAFTAR PUSTAKA

- [1] P. Ladita, “Analisis penerapan aplikasi android tokocrypto menggunakan pendekatan design thinking dibantu dengan platform design toolkit v.2,” *J. Teknol.*, 2020.
- [2] A. V. Sudiantoro *et al.*, “Analisis Sentimen Twitter Menggunakan Text Mining Dengan,” vol. 10, no. 2, pp. 398–401, 2018.
- [3] M. W. P. Aldi, Jondri, and A. Aditsania, “Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin,” *J. Inform.*, vol. 5, No, no. 2, p. 3548, 2018, [Online]. Available: <http://openlibrarypublications.telkomniversity.ac.id>.
- [4] D. W. Hariyanto and W. Maharani, “Analisis Sentimen Pada Media Sosial Twitter Berbahasa Indonesia Dengan Metode Glove,” *eProceedings ...*, vol. 7, no. 3, pp. 9711–9720, 2020.
- [5] G. A. Sandag, A. M. Manueke, and M. Walean, “Sentiment Analysis of COVID-19 Vaccine Tweets in Indonesia Using Recurrent Neural Network (RNN) Approach,” *3rd Int. Conf. Cybern. Intell. Syst. ICORIS 2021*, 2021, doi: 10.1109/ICORIS52787.2021.9649648.
- [6] A. Bagus Sasmita, B. Rahayudi, and L. Muflikhah, “Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 1208–1214, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [7] G. N. Aulia and E. Patriya, “Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019,” *J. Ilm. Inform. Komput.*, vol. 24, no. 2, pp. 140–153, 2019, doi: 10.35760/ik.2019.v24i2.2369.
- [8] R. Majid and H. A. Santoso, “Conversations Sentiment and Intent Categorization Using Context RNN for Emotion Recognition,” *2021 7th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2021*, pp. 46–50, 2021, doi: 10.1109/ICACCS51430.2021.9441740.
- [9] F. Long, K. Zhou, and W. Ou, “Sentiment analysis of text based on bidirectional LSTM with multi-head attention,” *IEEE Access*, vol. 7, pp. 141960–141969, 2019, doi: 10.1109/ACCESS.2019.2942614.
- [10] G. Arther Sandag, F. Gara, and Irfan, “Android Application Market Prediction Based on User Ratings Using KNN,” *2020 2nd Int. Conf. Cybern. Intell. Syst. ICORIS 2020*, 2020, doi: 10.1109/ICORIS50180.2020.9320812.
- [11] Y. D. W. G. Made, “Optimasi Model Percakapan Bahasa Indonesia Berbasis Sequence To Sequence (Seq2seq),” *SI Thesis , UAJY*, pp. 12–22, 2019, [Online]. Available: <https://docplayer.info/57060813-Bab-iii-landasan-teori-a-proyek.html>.
- [12] A. Giachanou and F. Crestani, “Like it or not: A survey of Twitter sentiment analysis methods,” *ACM Comput. Surv.*, vol. 49, no. 2, 2016, doi: 10.1145/2938640.
- [13] “TWINT - Twitter Intelligence Tool.” <https://github.com/twintproject/twint>.