

# Implementasi Analisis Sentimen dan Model Deep Learning Untuk Prediksi Harga Bitcoin

Andree Fernando Pratama<sup>1)</sup>, Tri Basuki Kurniawan<sup>2)</sup>, Misinem<sup>3)</sup>, Deshinta Arrova Dewi<sup>4)</sup>

<sup>1,2</sup> Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Bina Darma,  
Jl. Jenderal Ahmad Yani No.3, Kota Palembang, Sumatera Selatan 30111

<sup>3</sup> Program Studi Teknik Komputer, Fakultas Vokasi, Universitas Bina Darma, Jl. Jenderal  
Ahmad Yani No.3, Kota Palembang, Sumatera Selatan 30111

<sup>4</sup> Magister Ilmu Komputer, Universitas Nusa Putra, Sukabumi, Jawa Barat,  
e-mail: \*andreefnando@gmail.com, tribasukikurniawan@binadarma.ac.id,  
misinem@binadarma.ac.id, deshinta.arrovadewi@nusaputra.ac.id

## Abstrak

*Cryptocurrency adalah sebuah mata uang digital atau virtual yang dijamin oleh kriptografi. Dengan adanya kriptografi, mata uang digital ini menjadi hampir tidak mungkin dipalsukan. Adapun pencatatan semua transaksi yang dilakukan tersimpan pada blockchain. Blockchain ini tersebar luas antara satu komputer dengan komputer lain dan terkoneksi di dalam satu jaringan yang tersebar luas sehingga tidak terpusat pada satu tempat, atau dikenal dengan istilah desentralisasi. Analisis Sentimen digunakan untuk melihat reaksi masyarakat atau publik pada umumnya terhadap suatu isu yang sedang terjadi apakah reaksi publik bersifat positif maupun negatif, namun penentuan sentimen tersebut akan didasar pada label mana yang paling banyak muncul. Kemudian sentimen yang telah didapatkan ditentukan korelasinya terhadap kenaikan harga bitcoin sehingga sentimen tersebut dapat menjadi salah satu atribut yang dapat digunakan untuk meningkatkan akurasi prediksi harga bitcoin. Bitcoin kemudian diprediksi menggunakan model deep learning dengan layer LSTM.*

**Kata kunci**—Prediksi, Cryptocurrency, Analisis sentimen, Deep learning

## Abstract

*Cryptocurrency is a digital or virtual currency that is guaranteed by cryptography. With cryptography, this digital currency has become almost impossible to counterfeit. The recording of all transactions made is stored on the blockchain. This blockchain is widespread between one computer and another and is connected in a network that is widespread so that it is not centralized in one place, otherwise known as decentralization. Sentiment analysis is used to see the reaction of the community or the public in general to an ongoing issue, whether the public reaction is positive or negative. the determination of these sentiments will be based on which label appears the most. Then the sentiment that has been obtained is determined to correlate with the increase in bitcoin prices so that sentiment can be one of the attributes that can be used to increase the accuracy of bitcoin price predictions. Bitcoin is then predicted using a deep learning model with an LSTM layer.*

**Keywords**— Prediction, Cryptocurrency, Sentiment Analysis, Deep Learning

## 1. PENDAHULUAN

**C**ryptocurrency adalah sebuah mata uang digital atau virtual yang dijamin dengan menggunakan sistem *cryptography*. Dengan adanya *cryptography*, mata uang digital ini menjadi hampir tidak mungkin dipalsukan. Adapun pencatatan semua transaksi yang dilakukan tersimpan pada *blockchain*. *Blockchain* ini tersebar luas antara satu komputer dengan

komputer lain dan terkoneksi di dalam satu jaringan yang tersebar luas sehingga tidak terpusat pada satu tempat, atau dikenal dengan istilah desentralisasi.

Dengan kata lain, *cryptocurrency* adalah mata uang digital yang dapat digunakan untuk transaksi antar pengguna tanpa perlu melewati pihak ketiga. Selain menggunakannya sebagai alat transaksi, banyak pengguna yang memanfaatkan *cryptocurrency* sebagai instrumen investasi. Hal ini disebabkan oleh naik turunnya nilai *cryptocurrency*. Semakin banyak orang yang percaya pada *cryptocurrency*, maka nilainya akan semakin tinggi. *Cryptocurrency* muncul sebagai jawaban atas kendala sistem pembayaran saat ini yang sangat bergantung kepada pihak ketiga. Salah satu jenis *Cryptocurrency* adalah Bitcoin [1].

Perubahan nilai tukar *bitcoin* ini sangat cepat karena dapat dipengaruhi oleh banyak faktor seperti masalah politik, atau masalah ekonomi global. Sehingga untuk dapat melakukan prediksi yang akurat terhadap *bitcoin* tersebut tidak mudah. Selain karena masalah diatas, fluktuasi dari harga *bitcoin* sangat dinamis dan tinggi membuat banyak peneliti yang melakukan penelitian untuk memprediksi harga *cryptocurrency* tidak mendapatkan tingkat akurasi yang memuaskan.

Pendapat masyarakat dan pandangan terhadap *cryptocurrency* juga memberikan dampak terhadap perubahan harga pada *bitcoin*. Seperti pada akhir 2020, Elon Musk yang merupakan pemilik perusahaan tesla tersebut mengkonfirmasi ketertarikannya terhadap *bitcoin*. Hal ini membuat harga *bitcoin* melambung tinggi dari yang awalnya 266jt rupiah menjadi 569jt rupiah. Hal ini dikarenakan pernyataan tersebut membuat sentimen masyarakat menjadi sangat positif dengan keberadaan *bitcoin* tersebut. Untuk meneliti minat dan pandangan masyarakat tersebut maka terciptalah bidang ilmu yang berfokus terhadap analisis sentimen.

Analisis Sentimen adalah suatu teknik mengolah data berupa teks dengan tujuan mengambil informasi mengenai sentimen berlabel positif, netral atau negatif. Analisis sentimen didapatkan dari para pengguna internet di media sosial yang memberikan sebuah penilaian atau opini pribadi terhadap suatu topik. Menurut Winarko, E. (2017). Analisis Sentimen adalah riset komputasional dari opini sentimen dan emosi yang diekspresikan secara tekstual. [2]

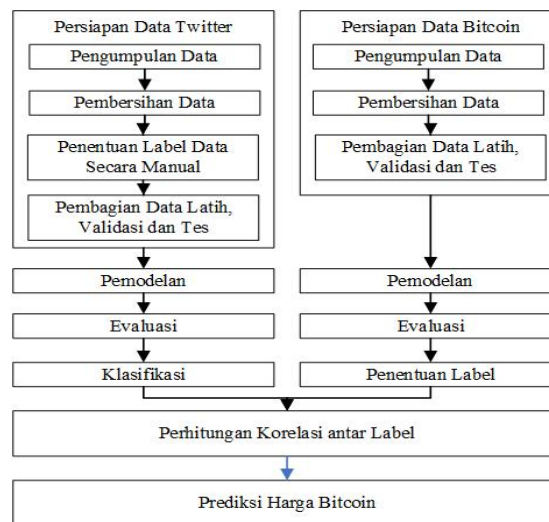
*Deep learning* merupakan subbidang *machine learning* yang algoritmanya terinspirasi dari struktur otak manusia. Struktur tersebut dinamakan *Artificial Neural Networks* atau disingkat ANN. Pada dasarnya, ini merupakan jaringan saraf yang memiliki tiga atau lebih lapisan ANN. ANN mampu belajar dan beradaptasi terhadap sejumlah besar data serta menyelesaikan berbagai permasalahan yang sulit diselesaikan dengan algoritma *machine learning* lainnya.

Karno, A. S. B., Hastomo, W., Arif, D., & Moreta, E. S. (2020). "Berawal dari model *forward neural network* sederhana kemudian berkembang menjadi *backpropagation neural network*, *recurrent neural network* (RNN). Karena ketidak mampuan RNN dalam menyimpan memori jangka panjang (terutama untuk data historis), dilakukan penyempurnaan yang menghasilkan bentuk sel dengan beberapa gate kompleks didalamnya, yaitu LSTM (*Long Short Term Memory*). Banyaknya gate kompleks dalam LSTM tentunya akan mempengaruhi waktu proses." [3]

Berdasarkan latar belakang diatas perlu dilakukan penelitian lanjutan untuk meningkatkan prediksi harga bitcoin dengan terlebih dahulu melihat di internet terhadap topik *bitcoin* tersebut apakah sentimen masyarakat sangat berpengaruh terhadap perubahan harga dari bitcoin sehingga bitcoin layak untuk dilakukan prediksi. Dalam hal ini apabila sentimen masyarakat memiliki korelasi yang positif terhadap perubahan harga *bitcoin* maka sentimen terus dapat dijadikan sebagai atribut tambahan dalam memprediksi harga bitcoin dikemudian hari. Peneliti juga melakukan prediksi harga menggunakan model *deep learning* dengan layer LSTM yang mampu menghasilkan sebuah model dengan tingkat akurasi prediksi yang baik dan dapat dimanfaatkan untuk melakukan prediksi harga bitcoin.

## 2. METODE PENELITIAN

Penelitian ini menggunakan metode dekskriptif yang menggambarkan fakta-fakta dan informasi dalam situasi atau kejadian secara sistematis, faktual dan akurat untuk menggambarkan sentimen pengguna twitter terhadap bitcoin dan menggunakan metode studi pustaka dengan cara mengumpulkan data buku, jurnal penelitian, artikel dan lainnya yang berhubungan dengan prediksi harga cryptocurrency dengan model deep learning. Secara rinci penelitian terdiri dari beberapa tahapan dimana tahapan tersebut di bagi menjadi 2 yang membedakan tahapan dari pengelolaan data twitter yang dimulai dari pengumpulan data twit dari pengguna twitter, pembersihan data, penentuan label, pembagian data, pemodelan, evaluasi dan klasifikasi. Kemudian tahapan pengelolaan data harga bitcoin yang dimulai dari pengumpulan data harga bitcoin, pembersihan data, pembagian data, pemodelan, evaluasi dan prediksi. Setelah kedua proses dilalui, tahapan akhirnya adalah perhitungan korelasi diantara kedua label dari data tersebut sehingga dapat diketahui apakah sentimen pada twitter mempengaruhi perubahan harga dari bitcoin. Tahapan tersebut dapat dilihat ada Gambar 1.



Gambar 1. Tahapan Penelitian

## 3. HASIL DAN PEMBAHASAN

### 3.1 Persiapan Data

Pada tahapan ini, data twit pengguna twitter didapatkan dengan menggunakan library *tweepy* yang tersedia pada Bahasa pemrograman python dengan API yang telah disediakan oleh aplikasi twitter itu sendiri. Pada penggunaannya didapatkan data twit berbahasa Indonesia yang membahas topik mengenai bitcoin berjumlah 1749 data, dimulai dari tanggal 18 – 26 april 2022. Data yang didapatkan kemudian melalui tahapan pembersihan data yang tidak relevan dengan penelitian sehingga data mengalami pengurangan jumlah menjadi 1379 data. Data yang telah dibersihkan kemudian diberi label berdasarkan sentimen atau pendapat dari twit tersebut secara manual sehingga data berbentuk *supervised*, label terdiri dari 2 yaitu Positif dan Negatif. Data yang sudah diberi label kemudian dibagi menjadi 3 dengan perbandingan 70% data latih, 25% data validasi dan 5% data uji.

data harga bitcoin didapatkan dengan menggunakan *library* *yfinance* yang telah disediakan. Data yang didapatkan merupakan data *timeseries* berjenis *unsupervised* berjumlah 1827 dengan kurun waktu 5 tahun sejak 26 april 2022. Data yang didapatkan kemudian dibersihkan apabila terdapat *noise* atau kesalahan pada data. Data kemudian dibagi menjadi 3 dengan perbandingan yang sama pada data twitter, yaitu 70% data latih, 25% data validasi dan 5% data uji.

### 3.2 Pemodelan

Pada penelitian ini, ada 2 model yang digunakan. Model pertama digunakan untuk mengetahui sentimen pada data twitter yang telah diolah kemudian model kedua digunakan untuk memprediksi harga cyptocurrency.

Untuk mengetahui sentimen suatu kalimat berbahasa Indonesia, penelitian ini menggunakan model IndoNLU yang merupakan *benchmark* Indonesia pertama untuk *Natural Language Understanding (NLU)* yang terdiri dari kumpulan dua belas tugas yang beragam. Tugas utama dikategorikan berdasarkan input, seperti kalimat tunggal dan pasangan kalimat, dan tujuan, seperti tugas klasifikasi kalimat dan pelabelan urutan tugas. *Benchmark* ini dirancang untuk memenuhi berbagai gaya dalam bahasa Indonesia formal dan sehari-hari, yang sangat beragam. [4]

Untuk mengetahui sentimen analisis terhadap pengguna aplikasi twitter terhadap bitcoin, model IndoNLU yang digunakan adalah model yang telah dilatih menggunakan dataset SmSA. SmSA tingkat kalimat dataset Analisis sentimen ini adalah kumpulan komentar dan ulasan dalam bahasa indonesia diperoleh dari beberapa platform online. Teks diambil dan kemudian dijelaskan oleh beberapa ahli bahasa Indonesia untuk menyusun kumpulan data ini. Ada tiga kemungkinan sentimen pada dataset SmSA: positif, negatif, dan netral.

Data yang telah dipersiapkan sebelumnya kemudian dibagi dan diatur sesuai dengan peran masing masing seperti pada gambar 2.

```
[12] train_dataset_path = './dataset/smsa_doc-sentiment-prosa/Data Training Preprocessing.tsv'
valid_dataset_path = './dataset/smsa_doc-sentiment-prosa/Data Validasi Preprocessing.tsv'
test_dataset_path = './dataset/smsa_doc-sentiment-prosa/Data Testing Preprocessing.tsv'

train_dataset = DocumentSentimentDataset(train_dataset_path, tokenizer, lowercase=True)
valid_dataset = DocumentSentimentDataset(valid_dataset_path, tokenizer, lowercase=True)
test_dataset = DocumentSentimentDataset(test_dataset_path, tokenizer, lowercase=True)

train_loader = DocumentSentimentDataLoader(dataset=train_dataset, max_seq_len=512, batch_size=32, num_workers=16, shuffle=True)
valid_loader = DocumentSentimentDataLoader(dataset=valid_dataset, max_seq_len=512, batch_size=32, num_workers=16, shuffle=False)
test_loader = DocumentSentimentDataLoader(dataset=test_dataset, max_seq_len=512, batch_size=32, num_workers=16, shuffle=False)
```

Gambar 2. Pembagian data latih, validasi dan tes

Model yang telah diatur dicoba dengan kalimat sederhana untuk mengetahui performa sementara sebelum model dilakukan training. Percobaan kalimat dapat dilihat pada gambar 3.

```
text = 'Harga bitcoin hari ini gimana?'
subwords = tokenizer.encode(text)
subwords = torch.LongTensor(subwords).view(1, -1).to(model.device)

logits = model(subwords)[0]
label = torch.topk(logits, k=1, dim=-1)[1].squeeze().item()

print(f'Text: {text} | Label : {12w[label]} ({{F.softmax(logits, dim=-1).squeeze()[label] * 100:.3f}}%)')

Text: Harga bitcoin hari ini gimana? | Label : positive (38.231%)
```

Gambar 3. Penggunaan model dengan kalimat sederhana

Deep learning merupakan subbidang machine learning yang algoritmanya terinspirasi dari struktur otak manusia. Struktur tersebut dinamakan Artificial Neural Networks atau disingkat ANN. Berawal dari model forward neural network sederhana kemudian berkembang menjadi backpropagation neural network, recurrent neural network (RNN). Karena ketidakmampuan RNN dalam menyimpan memori jangka panjang (terutama untuk data historis), dilakukan penyempurnaan yang menghasilkan bentuk sell dengan beberapa gate kompleks didalamnya, yaitu LSTM (Long Short Term Memory). [5]

Jaringan LSTM [6] Merupakan tipe yang berbeda *Recurrent Neural Networks*. LSTM mampu mempelajari ketergantungan jangkakan Panjang dan sangat populer untuk digunakan dengan data beruntun seperti *time-series* data. LSTM cell menggunakan *input gate*  $i_t$ ; a *forget gate*  $f_t$ ; and dan *output gate*  $o_t$ . *Gate* inilah yang akan mendefinisikan apakah data bisa lewat ataupun tidak berdasarkan dari priotias data. *Gates* juga membuat jaringan mampu mempelajari apa yang harus disimpan, dilupakan, diingat, diperhatikan dan juga dihasilkan. *cell state* and

*hidden state* digunakan untuk mengumpulkan data yang akan digunakan untuk state selanjutnya. Pada gambar.x. dapat dilihat model LSTM yang telah dibuat untuk memprediksi harga bitcoin.

### 3.3 Evaluasi

Pada tahapan evaluasi, kualitas dan efektivitas dari hasil klasifikasi sentimen pada data twitter dan hasil prediksi harga pada data bitcoin dapat diketahui. Evaluasi pada kualitas model sentimen terdiri dari 4, yaitu akurasi, presisi, recall dan nilai F1 dimana Accuracy merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data, Precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif, Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif dan F1 Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan. [7].

Precision dan Recall digunakan untuk mengukur kinerja sistem. Precision adalah kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan. Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Accuracy adalah tingkat kedekatan antara nilai yang didapat terhadap nilai sebenarnya. [8]

Untuk Mengetahui kualitas prediksi dari harga bitcoin. Evaluasi dilakukan dengan menggunakan 2 penilaian. Yaitu RMSE dan MAPE. RMSE mengevaluasi seberapa baik model dapat memprediksi nilai kontinu. Manfaat utama menggunakan RMSE adalah kemampuannya dalam mengetahui tingkat kesalahan juga menskalakan skor dalam unit yang sama dengan nilai perkiraan [9]. Rumus RMSE dapat dilihat pada persamaan (1) dimana  $N$  merupakan jumlah observasi,  $x_i$  adalah sebagai sebenarnya dan  $\hat{x}_i$  sebagai nilai prediksi

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (1)$$

MAPE menggambarkan persentase kesalahan dalam estimasi atau peramalan hasil aktual selama periode tertentu dengan perhitungannya dapat dilihat pada persamaan (2)

$$MAPE = \frac{100}{n} \sum \frac{|A_t - F_t|}{A_t} \quad (2)$$

Berdasarkan Lewis 1982 [10], Nilai MAPE dapat diartikan atau diinterpretasikan menjadi 4 kategori, yaitu:

- <10% = Sangat akurat
- 10-20% = bagus
- 20-50% = Wajar
- > 50% = Tidak akurat

Semakin kecil nilai MAPE maka semakin kecil error pada hasil estimasi.

### 3.4 Hasil Evaluasi

Pada tahapan ini, data yang telah dilakukan pelatihan pada tahapan pemodelan kemudian dievaluasi pada tahapan evaluasi sehingga menghasilkan nilai terukur yang akan digunakan untuk menentukan kualitas dan efektivitas model. Pada gambar 4. merupakan model dari IndoNLU yang telah dilatih dengan data latih dan divalidasi dengan data validasi. Hasilnya adalah model memiliki nilai evaluasi sebagai berikut;

- Akurasi : 78%
- Nilai F1 : 61%
- Recall : 61%
- Presisi : 70 %

```
+ Code + Text
list_hyp += batch_hyp
list_label += batch_label
metrics = document_sentiment_metrics_fn(list_hyp, list_label)

pbar.set_description("VALID LOSS:{:.4f} {}".format(total_loss/(i+1), metrics_to_string(metrics)))

metrics = document_sentiment_metrics_fn(list_hyp, list_label)
print("Epoch {} VALID LOSS:{:.4f} {}".format(epoch+1,
total_loss/(i+1), metrics_to_string(metrics)))

(Epoch 1) TRAIN LOSS:0.9524 LR:0.00000300: 100%|██████████| 31/31 [00:11<00:00, 2.73it/s]
(Epoch 1) TRAIN LOSS:0.9524 ACC:0.56 F1:0.39 REC:0.41 PRE:0.44 LR:0.00000300
| 0/11 [00:00<?, ?it/s]/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision is ill-defin
warn_prf(average, modifier, msg_start, len(result))
VALID LOSS:0.8458 ACC:0.65 F1:0.45 REC:0.48 PRE:0.44: 100%|██████████| 11/11 [00:02<00:00, 5.24it/s]
(Epoch 1) VALID LOSS:0.8458 ACC:0.65 F1:0.45 REC:0.48 PRE:0.44
(Epoch 2) TRAIN LOSS:0.8437 LR:0.00000300: 100%|██████████| 31/31 [00:08<00:00, 3.70it/s]
(Epoch 2) TRAIN LOSS:0.8437 ACC:0.65 F1:0.45 REC:0.49 PRE:0.43 LR:0.00000300
VALID LOSS:0.7833 ACC:0.69 F1:0.48 REC:0.51 PRE:0.46: 100%|██████████| 11/11 [00:02<00:00, 5.36it/s]
(Epoch 2) VALID LOSS:0.7833 ACC:0.69 F1:0.48 REC:0.51 PRE:0.46
(Epoch 3) TRAIN LOSS:0.7505 LR:0.00000300: 100%|██████████| 31/31 [00:08<00:00, 3.63it/s]
(Epoch 3) TRAIN LOSS:0.7505 ACC:0.69 F1:0.50 REC:0.54 PRE:0.65 LR:0.00000300
VALID LOSS:0.6958 ACC:0.73 F1:0.51 REC:0.55 PRE:0.49: 100%|██████████| 11/11 [00:02<00:00, 5.45it/s]
(Epoch 3) VALID LOSS:0.6958 ACC:0.73 F1:0.51 REC:0.55 PRE:0.49
(Epoch 4) TRAIN LOSS:0.6697 LR:0.00000300: 100%|██████████| 31/31 [00:08<00:00, 3.54it/s]
(Epoch 4) TRAIN LOSS:0.6697 ACC:0.72 F1:0.54 REC:0.57 PRE:0.73 LR:0.00000300
VALID LOSS:0.6413 ACC:0.76 F1:0.61 REC:0.61 PRE:0.72: 100%|██████████| 11/11 [00:02<00:00, 5.28it/s]
(Epoch 4) VALID LOSS:0.6413 ACC:0.76 F1:0.61 REC:0.61 PRE:0.72
(Epoch 5) TRAIN LOSS:0.5801 LR:0.00000300: 100%|██████████| 31/31 [00:08<00:00, 3.60it/s]
(Epoch 5) TRAIN LOSS:0.5801 ACC:0.78 F1:0.66 REC:0.66 PRE:0.74 LR:0.00000300
VALID LOSS:0.5606 ACC:0.78 F1:0.61 REC:0.61 PRE:0.70: 100%|██████████| 11/11 [00:02<00:00, 5.24it/s](Epoch 5) VALID LOSS:0.5606 ACC:0.78 F1:0.61 REC:0.61 PRE:0.70
```

Gambar 4. Hasil Evaluasi Model IndoNLU

Dalam evaluasi model prediksi harga bitcoin diketahui model memiliki nilai RMSE 2033.28 dan nilai MAPE 3.53%. perhitungan nilai dapat dilihat pada gambar 5.

```
### Calculate RMSE performance metrics
import math
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error
#math.sqrt(mean_squared_error(y_train, train_predict))
rmse = mean_squared_error(datatest, datapred, squared=False)

mape = mean_absolute_percentage_error(datatest, datapred)
mape_percentage = mape*100
print('Result Root Mean Square Error Prediction Model :', rmse)
print('Result Mean Absolute Percentage Error Prediction Model : ', float(np.round(mape_percentage, 2)), '%')

Result Root Mean Square Error Prediction Model : 2033.284647544704
Result Mean Absolute Percentage Error Prediction Model : 3.53 %
```

Gambar 5. Hasil Evaluasi Model Deep Learning Dengan Layer LSTM

### 3.5 Klasifikasi Sentimen

Pada tahapan ini, model IndoNLU yang telah diinisialisasi dan dilatih kemudian digunakan untuk mengklasifikasikan data yang telah disediakan sebelumnya. Setelah pengklasifikasian, data twitter yang sudah memiliki label didapatkan. Data dapat dilihat pada gambar 6.

	text	Label	Persentase
0	Lumayan Sudah 2x WD dapat 20 Ikuti mining BTC ...	positive	45.431%
1	Luna Foundation Guard LFG mengumumkan bahwa me...	neutral	59.282%
2	apa ini ribut2 bitcoin dkk di TL baru turun di...	negative	93.966%
3	Apa Bitcoin mau ikutan lebaran juga karena yan...	negative	40.148%
4	Siapa aja nih yg mining Bitcoin cuma pake brow...	positive	44.992%
...	...	...	...
1374	klo bs kembali kemasala lalu q mw beli bitcoin...	positive	44.011%
1375	KRIPTONESIA KOMUNITAS mengajarkan kita dan me...	neutral	57.603%
1376	mak aiii sampai bitcoin terus hahahahhahabab	negative	87.245%
1377	Ababil Pantasan bitcoin naik udah pada terima ...	negative	92.881%
1378	Pdhl sbml nya jg agnez tambil di miami bitcoin...	negative	91.308%

Gambar 6. Hasil Klasifikasi Sentimen

Hasil dari model tersebut merupakan sentimen terhadap bitcoin dimulai sejak tanggal 18 hingga 26 april yang dapat dilihat pada tabel 1. Dimana data sudah dikelompokkan perhariannya, dihitung setiap labelnya dan ditentukan sentimennya. semua sentimen terhadap bitcoin pada tanggal tersebut adalah negatif.



Tabel 1. Tabel Perhitungan klasifikasi perhari

Tanggal	Positif	Negatif	Sentimen
18 April 2022	161	140	Positif
19 April 2022	95	119	Negatif
20 April 2022	30	30	Negatif
21 April 2022	56	87	Negatif
22 April 2022	48	117	Negatif
23 April 2022	20	97	Negatif
24 April 2022	95	114	Negatif
25 April 2022	42	101	Negatif
26 April 2022	11	16	Negatif

### 3.6 Prediksi Bitcoin

Di tahapan ini, harga bitcoin diprediksi menggunakan model LSTM yang telah dibuat sebelumnya. Harga bitcoin yang digunakan adalah harga tanggal 17 hingga tanggal 28 april. Hal ini dikarenakan perlu adanya pelabelan harga. apakah harga tersebut naik maupun turun berdasarkan dari harga sebelumnya. Tabel 2 menunjukkan harga bitcoin dan juga pelabelan harga tersebut.

Tabel 2. Klasifikasi harga bitcoin secara manual

Tanggal	Harga	Label
17 April 2022	39716.9531	-
18 April 2022	40826.2148	Naik
19 April 2022	41502.75	Naik
20 April 2022	41374.3789	Turun
21 April 2022	40527.3633	Turun
22 April 2022	39740.3203	Turun
23 April 2022	39486.7305	Turun
24 April 2022	39469.293	Turun
25 April 2022	40458.3086	Naik
26 April 2022	38117.4609	Turun

### 3.7 Korelasi Antar label

Data sentimen yang telah diberikan label dan data harga bitcoin yang telah diprediksi kemudian diberikan label secara manual, kemudian dicari korelasi diantara labelnya.

Tabel Kontingensi merupakan tabel yang digunakan untuk mengukur hubungan (asosiasi) antara dua variabel kategorik dimana tabel tersebut merangkum frekuensi bersama dari observasi pada setiap kategori variable [11].

Tabel kontingensi dasarnya setara dengan kubus OLAP, yang melemparkan tradisional database relasional sebagai kubus berdimensi tinggi dengan dimensi yang sesuai dengan atribut [12]. Tabel kontingensi antara label sentimen dan harga bitcoin dapat dilihat pada tabel 3.

Tabel 3. tabel kontingensi

		Sentimen Data Twitter	
		Positif	Negatif
Harga Bitcoin	Naik	1	2
	Turun	0	6

Koefisien phi ( $\phi$ ) diusulkan oleh George Utney Yule pada tahun 1912 sebagai indeks hubungan antara dua variabel dikotomis. Statistik sangat berguna bagi banyak peneliti karena merupakan korelasi produk-momen yang sudah dikenal ketika dihitung untuk data dari

distribusi binomial bivariat. Wearden, S. (2010). [13]. Perhitungan dari nilai koefisien phi dapat dilihat pada persamaan (3)

$$\varphi = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \quad (3)$$

koefisien phi merupakan symmetrical statistic, yang berarti variable independent dan variabel dependen dapat saling bertukaran. Cara menggambarkan koefisien phi sama dengan korelasi koefisien pearson. yang memiliki jarak diantara -1 sampai 1, dimana:

- 0, tidak saling berhubungan
- 1, memiliki hubungan positif yang kuat dimana labelnya saling berhubungan
- -1. Memiliki hubungan berlawanan yang kuat dimana labelnya berlawanan satu sama lain

The Political Science Department at Quinnipiac University memberikan gambaran terkait cara perhitungan yang efisien mengenai korelasi pearson dan dapat digunakan untuk menghitung nilai efisien dari phi korelasi. Nilai tersebut dapat dilihat pada tabel 4.

Tabel 4. Tabel penjelasan mengenai nilai koefisien phi

Nilai	Keterangan
+ .70 atau lebih	Hubungan Positif yang sangat Kuat
+ .40 hingga + .69	Hubungan Positif Yang kuat
+ .30 hingga + .39	Memiliki Hubungan positif
+ .20 hingga + .29	Memiliki Hubungan Positif yang lemah
+ .01 hingga + .19	Hampir tidak ada hubungan positif
0	Tidak Berhubungan
- .01 hingga - .19	Hampir tidak ada hubungan berlawanan
- .20 hingga - .29	Memiliki Hubungan Berlawan yang lemah
- .30 hingga - .39	Memiliki Hubungan berlawanan
- .40 hingga - .69	Memiliki Hubungan Berlawanan yang kuat
- .70 atau lebih	Memiliki Hubungan Berlawanan yang sangat kuat

Berdasarkan persamaan dan nilai pada tabel 3, nilai korelasi antara sentimen dan prediksi sudah dapat dihitung. Hasil perhitungannya dapat dilihat pada persamaan (4).

$$\begin{aligned} \varphi &= \frac{6 - 0}{\sqrt{(1+2)(0+6)(1+0)(2+6)}} \\ \varphi &= \frac{6}{\sqrt{(3)(6)(1)(8)}} \\ \varphi &= \frac{\sqrt{144}}{6} \\ \varphi &= \frac{12}{6} \\ \varphi &= 0,5 \end{aligned} \quad (4)$$

koefisien phi yang dihasilkan dengan data sentimen dan harga bitcoin dari tanggal 18 april hingga 26 april adalah 0,5 dimana pada tabel 4 nilai tersebut menyatakan bahwa label memiliki tingkat hubungan positif yang kuat.

### 3.8 Prediksi Harga Bitcoin

Dengan diketahui adanya korelasi antara sentimen dan kenaikan harga dari bitcoin. Tahapan selanjutnya adalah memprediksi harga bitcoin dengan model deep learning yang telah

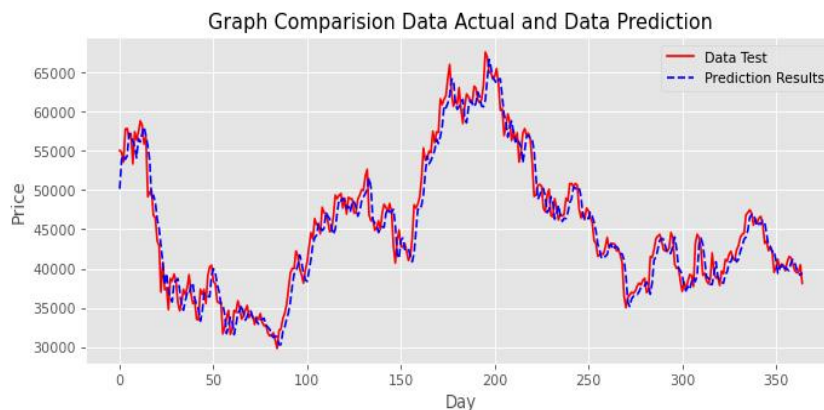


dievaluasi sebelumnya. Hasil perbandingan harga asli dan prediksi tersebut dapat dilihat pada gambar 7.

	Data Test	Prediction Results
355	39716.95313	40187.167969
356	40826.21484	39877.578125
357	41502.75000	39750.988281
358	41374.37891	40710.187500
359	40527.36328	41124.785156
360	39740.32031	40771.730469
361	39486.73047	39954.140625
362	39469.29297	39344.816406
363	40458.30859	39168.722656
364	38117.46094	39468.511719

Gambar 7. Perbandingan Harga Asli dan Prediksi Bitcoin

Kemudian, grafik yang membandingkan harga data tes dan harga prediksi dapat dilihat pada gambar 8.



Gambar 8. Grafik Perbandingan Harga Asli Dan Prediksi Bitcoin

#### 4. KESIMPULAN

Penggunaan model IndoNLU yang merupakan model *natural language understanding* bahas Indonesia pertama sangat berdampak untuk mempermudah penelitian mengenai analisis sentimen dengan sumber media sosial berbahasa Indonesia. Hal ini juga membuat penelitian tidak perlu melakukan penerjemahan Bahasa terlebih dahulu kebahasa Inggris untuk menggunakan *library* yang hanya mengerti Bahasa tersebut. Selain itu, diketahui bahwa Implementasi model IndoNLU dalam Analisis sentimen untuk menentukan pandangan masyarakat terhadap bitcoin memiliki dampak yang besar. Setelah dievaluasi, model IndoNLU Menghasilkan nilai akurasi 78%, nilai f1 61%, recall 61% dan presisi 70%. Setelah dilakukan perhitungan koefisien phi tingkat kenaikan harga bitcoin memiliki korelasi positif yang kuat terhadap sentimen yang sedang beredar dimedia social dengan nilai koefisien phi 0.50. Hal ini juga yang membuat analisis sentimen dapat menentukan tingkat kelayakan prediksi harga bitcoin. Setelah mendapatkan tingkat korelasi yang positif, peneliti juga mendapatkan akurasi yang baik pada saat memprediksi harga bitcoin menggunakan model deep learning dengan layer LSTM dengan nilai RMSE 2033.28 dan nilai MAPE 3.53%.

---

## 5. SARAN

Masih banyak kekurangan yang ada pada penelitian terutama terkait dengan data. Data yang didapatkan pada API yang telah disediakan oleh twitter sangat terbatas dan adanya kesulitan untuk mendapatkan data yang sudah lebih dari 1 bulan. Disisi lain penelitian ini juga hanya menggunakan data dari satu sumber media sosial yaitu twitter, dengan menambahkan banyaknya data dan lamanya jarak waktu penelitian. Model pasti akan dapat menghasilkan hasil korelasi yang lebih tinggi dan terkait prediksi harga bitcoin. Perubahan layer ataupun penambahan atribut pada data dapat meningkatkan tingkat akurasi dari model sehingga model dapat memprediksi harga bitcoin dengan lebih tepat.

## DAFTAR PUSTAKA

- [1] E. Dourado and J. Brito, 'Cryptocurrency. The New Palgrave Dictionary of Economics', *Online Edition*, 2014.
- [2] E. Winarko, 'Sentimen analisis tweet berbahasa Indonesia dengan deep belief network', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 11, no. 2, pp. 187–198, 2017.
- [3] A. S. B. Karno, W. Hastomo, D. Arif, and E. S. Moreta, 'OPTIMASI PORTOFOLIO DAN PREDIKSI CRYPTOCURRENCY MENGGUNAKAN DEEP LEARNING DALAM BAHASA PYTHON', *Prosiding SeNTIK*, vol. 4, no. 1, pp. 193–202, 2020.
- [4] B. Wilie *et al.*, 'IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding', *arXiv preprint arXiv:2009.05387*, 2020.
- [5] A. S. B. Karno, W. Hastomo, D. Arif, and E. S. Moreta, 'OPTIMASI PORTOFOLIO DAN PREDIKSI CRYPTOCURRENCY MENGGUNAKAN DEEP LEARNING DALAM BAHASA PYTHON', *Prosiding SeNTIK*, vol. 4, no. 1, pp. 193–202, 2020.
- [6] M. De Caux, F. Bernardini, and J. Viterbo, 'Short-Term Forecasting in Bitcoin Time Series Using LSTM and GRU RNNs', in *Anais do Symposium on Knowledge Discovery, Mining and Learning (KDMiLe 2020)*, Brasil, Oct. 2020, pp. 97–104. doi: 10.5753/kdmile.2020.11964.
- [7] R. Arthana, 'Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning', May 05, 2019. <https://rey1024.medium.com/mengenal-accuracy-precision-recall-dan-specificity-seerta-yang-diprioritaskan-b79ff4d77de8> (accessed Aug. 30, 2022).
- [8] V. Armando, 'Sistem Rekomendasi Pembelian Telepon Genggam dengan Metode Content-based Filtering', 2017.
- [9] N. J. Salkind, *Encyclopedia of research design*, vol. 1. Sage, 2010.
- [10] C. D. Lewis, *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann, 1982.
- [11] G. U. Yule, 'On the Methods of Measuring Association Between Two Attributes', *Journal of the Royal Statistical Society*, vol. 75, no. 6, p. 579, May 1912, doi: 10.2307/2340126.
- [12] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar, 'Privacy, accuracy, and consistency too: a holistic solution to contingency table release', 2007, pp. 273–282.
- [13] S. Wearden, 'Phi coefficient', *The Corsini encyclopedia of psychology*, pp. 1–1, 2010.