

Prediksi Harga Bitcoin Menggunakan Transformer: Mengintegrasikan Harga, Sentimen, Tren, dan Volume dalam Analisis Deret Waktu Multivariat

Gilang Islamay Putra Djuharis¹, Muhaza Liebenlito², Nur Inayah³

islamayputra7@gmail.com, muhazaliebenlito@uinjkt.ac.id, nur.inayah@uinjkt.ac.id

^{1,2,3}UIN Syarif Hidayatullah Jakarta

Informasi Artikel	Abstrak
Diterima : Direview : Disetujui :	Perkembangan teknologi bersandingan langsung dengan berkembangannya mata uang kripto. Dengan bitcoin mengalami kenaikan sebesar 75% hanya dengan jangka satu tahun pada 2020, perkembangan mata uang kripto sudah tidak bisa dipungkiri lagi. Dengan volatilitasnya harga mata uang kripto maka terdoronglah pencarian model paling mutakhir untuk meramal harga mata uang kripto. Penelitian kali ini bereksperimen untuk mencari model mutakhir tersebut tidak hanya dari sisi peramalan tetapi juga dalam sisi Pemrosesan Bahasa Alami, dengan model transformers yang menggunakan mekanisme <i>attention</i> untuk keduanya. Diintegrasikan analisis sentimen, trend, dan volume ke dalam variabel model dan didapatkan model yang paling mutakhir adalah dengan hanya memasukan variabel sentimen serta trend sehingga mendapatkan model dengan RMSE sebesar 0.0256, sMAPE sebesar 3.0428%, serta MAPE sebesar 3.0565%. Yang mana melampaui model <i>Long-Short Term Memory</i> dengan RMSE sebesar 0.0262, sMAPE sebesar 3.0615%, serta MAPE sebesar 3.0887%
Kata Kunci	
Mata Uang Kripto, Bitcoin, Analisis Sentimen, Model Transformers, Prediksi Harga	

Keywords	Abstrak
Cryptocurrency, Bitcoin, Sentiment Analysis, Transformer Models, Price Prediction	<i>The development of technology goes hand in hand with the growth of cryptocurrency. With Bitcoin experiencing a 75% increase in just one year in 2020, the development of cryptocurrency cannot be denied. Due to the volatility of cryptocurrency prices, there is a drive to find the most advanced models for predicting cryptocurrency prices. This research experiment explores the search for such advanced models, not only from the forecasting perspective but also in Natural Language Processing, using transformer models with attention mechanisms for both aspects. Sentiment analysis, trends, and volume are integrated into the model variables, and the most advanced model is obtained by incorporating only sentiment and trend variables, resulting in an RMSE of 0.0256, sMAPE of 3.0428%, and MAPE of 3.0565%. Which surpass Long-Short Term Memory's model with an RMSE of 0.0262, sMAPE of 3.0615%, and MAPE of 3.0887%</i>



A. Pendahuluan

Perkembangan yang pesat pada bidang ilmu teknologi dan informasi, *Financial Tech* (Fintech) atau teknologi keuangan terus melakukan terobosan baru, mendorong peningkatan dan inovasi model keuangan serta membentuk kembali rantai pasok dan rantai nilai untuk industri keuangan[1]. Sebagai sistem pertukaran mata uang elektronik alternatif yang baru, *cryptocurrency* atau mata uang kripto telah diakui secara luas memiliki konsekuensi signifikan bagi pasar-pasar berkembang dan ekonomi global[2].

Menurut Google Trends, istilah "cryptocurrency" mencapai puncak popularitasnya pada bulan Mei 2021. Dengan popularitas mata uang kripto di kalangan masyarakat, banyak investor yang melihat peluang keuntungan. Salah satu mata uang kripto yang paling populer adalah Bitcoin[2]. Data dari coinmarketcap juga menunjukkan bahwa harga Bitcoin, salah satu mata uang kripto terpopuler, mengalami kenaikan sebesar 405% selama tahun 2020 dan 161% selama tahun 2021 dengan nilai market capitalization atau nilai total dari semua bitcoin yang beredar di pasar paling besar 1.28 trillion dollar pada 9 November 2021.

Banyak sekali hal menarik dari mata uang kripto seperti kemudahan penggunaan, keamanan, dan desentralisasi. Dengan dapat diakses melalui berbagai perangkat ditambah dengan identitas yang aman dan transparan, tetapi pada saat yang bersamaan bersifat anonim dengan setiap transaksi tercatat pada blockchain yang mendasarinya tanpa melibatkan peran perantara seperti bank[3].

Volatilitas harga yang tinggi pada mata uang kripto menjadi salah satu alasan mengapa beberapa investor enggan masuk ke pasar ini[4], harga Bitcoin sudah tepatnya 59 kali mengalami penurunan lebih dari 10% dalam sehari. Dengan penurunan yang paling besar dialami pada tanggal 13 Maret 2020 yaitu sebesar 35.19%. Oleh karena itu, diperlukan suatu model yang dapat memprediksi harga mata uang kripto di masa depan untuk membantu investor memaksimalkan keuntungan dan meminimalkan kerugian.

Banyak penelitian telah dilakukan mengenai peramalan harga Bitcoin menggunakan berbagai teknik. Sebagai contoh, penelitian yang dilakukan oleh Nasir et al. (2019) menggunakan data Google Trends dan algoritma machine learning seperti Vector Autoregression dan Random Forest untuk meramalkan pergerakan harga Bitcoin berdasarkan harga dan popularitasnya[5]. Penelitian tersebut menunjukkan bahwa data pencarian juga dapat menjadi variabel prediktor untuk aset investasi. Penelitian lain oleh Alghamdi et al. (2022) menemukan hubungan yang kuat antara harga Bitcoin dengan sentimen, dengan Mean Absolute Error (MAE) sebesar 0.245, Mean Square Error (MSE) sebesar 0.2528, dan Root Mean Squared Error (RMSE) sebesar 0.5028[6].

Penelitian-penelitian tersebut melihat bahwa penggunaan teknik-teknik machine learning dan analisis sentimen dapat memberikan kontribusi yang signifikan dalam meramalkan pergerakan harga Bitcoin. Namun, seiring dengan perkembangan teknologi, terutama di bidang kecerdasan buatan, model-model yang lebih canggih dan efektif telah muncul. Salah satu model paling mutakhir per-2023 adalah model transformer dengan menggunakan mekanisme self-attention yang juga semakin populer seiring dengan kemunculan Chat GPT (Chat Generative Pre-Trained Transformer) yang menarik perhatian dunia. Menurut Google Trends, kata kunci "Transformer Deep Learning" dan "Transformer Model" mengalami

peningkatan popularitas sejak awal tahun 2022 dan mencapai puncaknya pada bulan Maret dan Juni 2023.

Attention is All You Need, adalah judul karya ilmiah diberikan oleh delapan orang dari google pada tahun 2017. Mereka mengusulkan model terbaru yang merupakan peningkatan dari model berbasis *recurrent* untuk *Natural language Processing* (NLP) dengan nama Transformer dengan menggunakan mekanisme *attention*[7]. Lalu pada tahun 2020 keluar karya ilmiah berjudul *A Transformer-Based Framework for Multivariate Time Series Representation Learning* mengusulkan penggunaan arsitektur yang sama untuk melakukan peramalan deret waktu dan mendapatkan evaluasi model transformer mengungguli model lainnya (Rocket, Long-Short Term Memory, XGBoost, dll)[8].

Sentimen analisis juga bukanlah hal yang baru, sebelum ditemukannya transformer telah banyak model sentimen untuk berbagai macam bidang. Dari bidang sosial, kesehatan, maupun politik. Meskipun demikian, masih sedikit model analisis sentimen yang tertuju pada bidang mata uang kripto khususnya menggunakan infrastruktur transformer. beberapa diantaranya adalah cryptobert oleh ElKulako, dimana model ini dilatih menggunakan 3,2 juta post sosial media seperti StockTwits, Telegram, Reddit, serta twitter mengenai mata uang kripto. Dan CryptoBERT oleh kk08.

ElKulako melatihnya dengan data training yang besar, namun ketika model di uji dengan data sentimen bitcoin yang sudah dilabeli secara manual. Model hanya mendapatkan akurasi sebesar 46%. Sedangkan model oleh kk08 mendapatkan akurasi sebesar 79%, akan tetapi modelnya hanya mendeteksi dua label yaitu positif dan negatif. Sehingga model menurun menjadi 50% saat mengkalkulasikan sentimen netral.

Beberapa penelitian, seperti penelitian oleh Eray Gemici dan Müslüm Polat tentang hubungan antara harga dan volume di pasar Bitcoin menunjukkan adanya korelasi antara harga dan volume[9]. Serta penelitian oleh Ettredge et al. yang menggunakan data pencarian web untuk memprediksi statistik macroeconomic, menemukan hubungan antara popularitas suatu topik dengan penjualan rumah dan statistik macroeconomic[10]. Selain itu, penelitian yang dilakukan oleh Choi dan Varian dalam penelitian mereka yang berjudul "Predicting the Present with Google Trends" juga menghasilkan kesimpulan yang sama[11]. Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa popularitas dan volume mata uang kripto memiliki korelasi dengan harga. Dalam penelitian ini, akan ditambahkan variabel sentimen media sosial untuk mengembangkan model prediksi yang lebih baru.

Dalam sisi analisis deret waktu multivariat, Mehtab dan Sen (2020) menemukan bahwa penggunaan model multivariat dalam memprediksi harga saham memberikan hasil yang lebih baik dibandingkan dengan pendekatan univariat[12]. Hansun et al. (2022) menerapkan analisis multivariat dalam prediksi mata uang kripto dengan membandingkan tiga pendekatan menggunakan recurrent neural network (RNN), yaitu Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), dan The Gated Recurrent Unit (GRU). Mereka menggunakan lima variabel, yaitu harga tutup, harga buka, harga tertinggi, harga terendah, dan volume dari lima mata uang kripto, termasuk Bitcoin, Ethereum, Cardano, Tether, dan Binance Coin. Hasil penelitian menunjukkan bahwa Bi-LSTM dan GRU memiliki

performa yang serupa dengan rata-rata Mean Absolute Percentage Error (MAPE) sebesar 0.0465712 untuk Bi-LSTM dan 0.0446512 untuk GRU, sedangkan LSTM memiliki MAPE sebesar 0.0529916. Meskipun LSTM lebih unggul dalam performa pada dataset USDT dan BNB, namun LSTM memiliki variansi yang lebih besar dibandingkan dengan Bi-LSTM dan GRU[13].

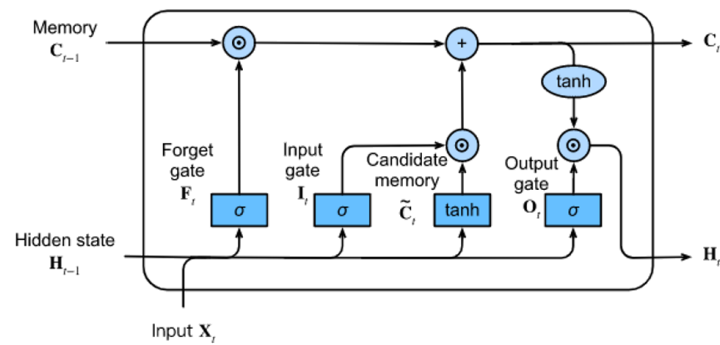
Model transformer telah digunakan dalam berbagai bidang, tidak hanya untuk menentukan skor sentimen dari kalimat, tetapi juga untuk peramalan data deret waktu. Zhao et al. (2022) melakukan penelitian mengenai kemampuan model transformer dalam meramalkan harga Bitcoin dan Ethereum menggunakan analisis sentimen, dan membandingkannya dengan model LSTM. Hasil penelitian menunjukkan bahwa model LSTM lebih unggul dibandingkan dengan model transformer, dengan model transformer dilatih menggunakan data Bitcoin memiliki MSE sebesar 0.00037, MAPE sebesar 0.05816, dan MAE sebesar 0.01432. Dibandingkan dengan LSTM yang memiliki MSE sebesar 0.00032, MAPE sebesar 0.04613, dan MAE sebesar 0.01346. Penelitian ini juga menemukan bahwa model yang dilatih dengan data Bitcoin memiliki peningkatan dalam memprediksi harga Ethereum dibandingkan dengan model yang dilatih dengan data Ethereum itu sendiri[14].

Terinspirasi oleh penelitian-penelitian yang telah disebutkan, penelitian ini bertujuan untuk menjelajahi lebih lanjut penggunaan model transformer dalam memprediksi harga Bitcoin dengan mempertimbangkan analisis sentimen dan popularitasnya menggunakan data dari Twitter, Reddit, dan Google Trends. Mengintegrasikan transformer dalam menganalisis sentimen hingga memprediksikan harga bitcoin itu sendiri dengan variabel variabel yang diperoleh. Adapun dikarenakan ada limitasi perangkat keras maka digunakan google colab untuk menjalankan programnya dan tokenisasi natural language processing dilimitasi sampai 256 token.

B. Metode Penelitian

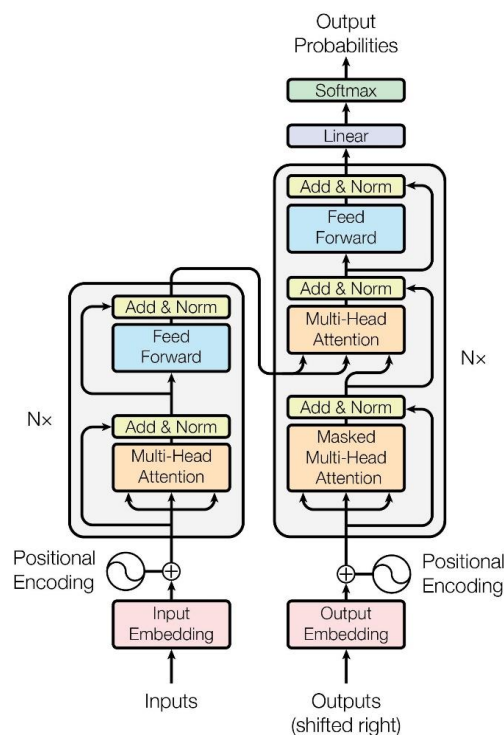
Penelitian ini mengadopsi pendekatan ekperimental dengan menginvestiasi kemampuan model Transformer dalam menganalisis sentimen dalam topik Bitcoin dan memprediksi harga bitcoin dengan mengintegrasikan faktor-faktor kunci seperti harga, sentimen, tren, serta volume. Pendekatan eksperimental dipilih karena memungkinkan pengujian secara langsung terhadap data waktu nyata untuk menganalisa kinerja model terhadap variabel-variabel yang digunakan.

Model dasar (*baseline*) yang digunakan adalah *Long-Short Term Memory*. Tipe *Recurrent Neural Network* (RNN) ini berkapabilitas untuk mempelajari korelasi jangka panjang antara variabel, karena itu model ini cocok untuk analisis deret waktu.



Gambar 1. Arsitektur LSTM

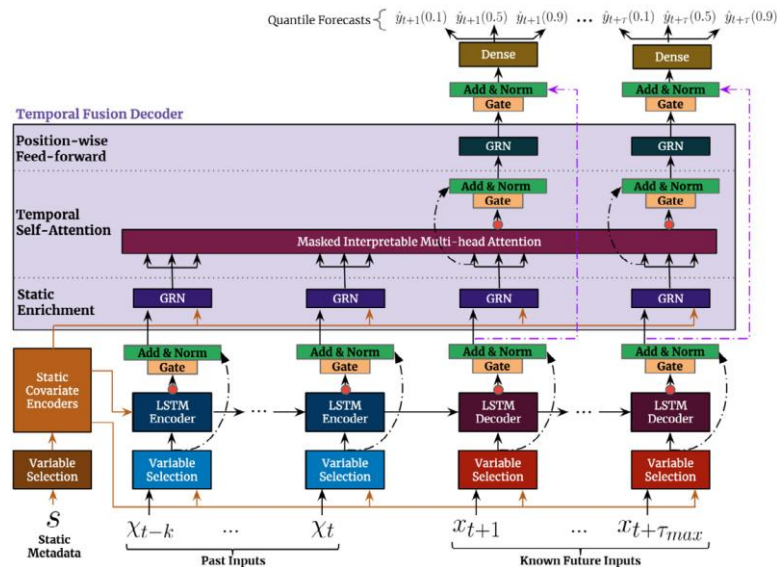
Mengikuti arsitektur RNN pada umumnya yaitu dengan struktur rantai, setiap *cell* dalam LSTM memiliki 4 gerbang. Yaitu yang pertama adalah *Forget Gate*, dimana dilakukan fungsi sigmoid untuk mendeterminasi seberapa banyak yang harus dilupakan pada ingatan jangka panjang (*Long-Term Memory*). Kedua *Input Gate* dengan fungsi sigmoid yaitu berapa banyak yang harus diingat serta yang ketiga *Candidate Memory* adalah apa saja yang harus diingat. Yang terakhir ada output gate, dimana pada gerbang ini dilakukan fungsi sigmoid untuk mendeterminasi output untuk cell selanjutnya.



Gambar 2. Arsitektur Transformers

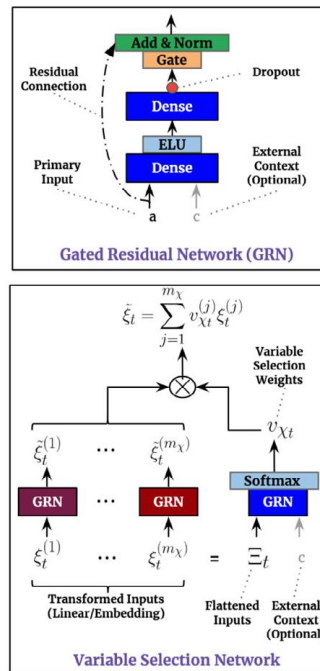
Mekanisme yang dikedepankan pada transformer adalah mekanisme *Self Attention*. Dengan mekanisme ini transformer dapat melihat hubungan antara satu kata dengan kata yang lainnya, dan paralelisasi dengan menumpukan *attention* (*Multi-Head Attention*) dapat dilakukan karena mengkalkulasikan hubungan suatu kata dengan kata lainnya tidak perlu mengetahui nilai kata lainnya.[7] Lalu pada

tahun 2021 Lim Brian, dkk mengeluarkan arsitektur transformer terbaru yang menggabungkan *Long-Short Term Memory* (LSTM) dengan mekanisme *attention* untuk peramalan deret waktu multivariat dengan kemampuan untuk menjelaskan dengan nama *Temporal Fusion Transformers* (TFT)[15].



Gambar 3. Arsitektur *Temporal Fusion Transformers*

Model TFT dibandingkan dengan model lainnya, mendapatkan performa yang bagus, dikarenakan penggunaan arsitektur LSTM pada Encoder serta Decodernya sehingga dapat menangkap korelasi jangka pendek ataupun panjang dari suatu kovariat, ditambah dengan *Multi-Head Attention* sehingga model dapat menimbang kovariat yang mana yang harus lebih difokuskan pada waktu tertentu untuk melakukan prediksi. TFT juga menerapkan mekanisme gerbang dan *Variable Selection Network* (VSN) untuk meminimalisir kontribusi variabel yang tidak berperan signifikan saat pelatihan model.



Gambar 4. Mekanisme VSN


TFT juga mengedepankan *explainability* atau kejelasan dari model itu sendiri, dibandingkan dengan model lain yang bersifat *black-box* dimana kejelasan dari perhitungan model tidak dapat dilihat yang menimbulkan masalah praktis dan etis[16]. TFT menyelesaikan masalah itu dengan mekanisme *attention*, TFT bisa memberikan penjelasan tentang dinamika temporal serta menganalisa relasi global sementara dari seluruh data (Seperti musim dan *lag effects*)[15].

Pengumpulan data dilakukan dengan mengambil data twitter melalui Kaggle dan *scraping* sosial media reddit dilakukan dengan mengambil komen yang terdapat pada daily discussion di r/bitcoin, guna mengisi kekosongan data dan meluaskan cangkupan data sentimen analisis itu sendiri menjadi 03/12/2017 hingga 30/06/2023. Serta data historis bitcoin diambil melalui API yang disediakan oleh binance. Serta data google trends diambil secara manual dalam jangka tiap 270 hari dari 03/12/2017 hingga 30/06/2023 lalu data dinormalisasi dengan mencocokkan tanggal yang berkesinambungan. Ini dilakukan guna mengambil data harian google trends.

Hasil yang optimal dalam analisis teks dibutuhkan beberapa langkah. Yaitu tokenisasi teks dan normalisasi teks.[17] Dimana tokenisasi melakukan pemisahan kalimat menjadi kata kata lalu dinormalisasi dengan penghapusan karakter spesial (contoh: #, @), mengubah seluruh kalimat menjadi huruf kecil, penghapusan *stopwords* (contoh: yang, di), dan lain lain[18]. Pengaplikasian pra pemrosesan data dilakukan dengan menghapus link di kalimat (jika ada), mengubah seluruh kalimat untuk menjadi huruf kecil, menghapus tanda baca, tokenisasi, menghapus *stopwords*, menghapus nomor dan karakter spesial, lemmatisasi, dan menghapus spasi yang tidak dibutuhkan.

Pra pemrosesan juga dilakukan dengan menghapus tweet yang bersifat spam. Mata uang kripto yang melonjak kepopularitasannya juga menyebabkan banyaknya tweet yang bersifat spam dan dibuat oleh robot.

Tabel 1. Contoh tweet spam

<i>date</i>	<i>Text</i>
2021-07-03 20:53:51	The \$BTC price is at \$34655.23 right now.  Compared to the last tweet, the price has dropped by \$72.77 (-0.21%).  In the last 24 hours the price has increased by \$1488.36 (4.49%).
2021-10-19 09:16:20	#Bitcoin #BTC https://t.co/5wDhcU31UB   Bitcoin Whale Alert: [TX: 6e2f93751abd08b63306d5351da016187927ac712f442d6709d23d10b01e83c8]-[ADDR: 1DuhtLa8TtCC547WfSNLHZvT91b9PQQeDD]-[#BTC: 16.68212472]-[BLOCK_DATE: 2021-10-19 10:37:49] #btc_whale_alert #bitcoin- BTC_Whale_Alert
2021-06-21 07:43:39	WhaleTrades:   \$2,500,000 #bitcoin LONGED @\$32,805.8359 [21/06/21 07:34:55]  BitMEX \$XBTUSD
2022-04-16 15:59:12	 I'd take a fast nickel over a slow dime - buyerofblood  Market Cap. Swap  What would one #Ecash \$XEC cost if it had the market capitalization of #Bitcoin \$BTC? One #XEC would be worth \$0.0401409.     More: https://t.co/fSCkKGmhck... https://t.co/rmygNBMIhr
2021-02-10 23:10:02	Bitcoin BTC Current Price: \$45,161.73 1 Hour: 1.21% 24 Hours: -4.05% 7 Days: 21.40% #btc #bitcoin

Pelabelan data manual juga dilakukan untuk melatih model analisis sentimen, dengan total 3.241 data yang telah diberi label secara manual dengan sentimen negatif terdapat sebanyak 488 data, sentimen netral sebanyak 1.177 data, dan sentimen positif sebanyak 1.516 data. Selanjutnya, untuk memperoleh data tambahan, diambil 562 data dari website SurgeAI, dengan distribusi sentimen negatif sebanyak 260 data dan sentimen positif sebanyak 302 data. Meskipun demikian, jumlah total data sebanyak 3.803 dapat diperbanyak lebih lanjut melalui augmentasi data.

Augmentasi data banyak dilakukan pada *computer vision*[19], akan tetapi pada bidang NLP masih sedikit yang membahas[20]. *Back-translation* adalah teknik augmentasi data yang sangat efektif untuk terjemahan mesin neural[21], yang mana pendekatan ini menjanjikan karena memiliki kemampuan mempertahankan label yang baik secara alami dan kemampuan memparafrase yang sangat berharga[22]. *Back-Translation* dilakukan dengan teks asli dikembalikan ke dalam bahasa aslinya setelah dua kali terjemahan. Teks asli S_1 diterjemahkan ke dalam bahasa lain (seperti Bahasa Indonesia) sebagai S_2 , dan kemudian diartikan kembali ke dalam bahasa aslinya sebagai S_3 . [23]

Back-Translation pada data teks untuk mengklasifikasikan suatu kalimat juga sudah pernah dilakukan. Augmentasi data menghasilkan peningkatan sekitar 5% pada bobot rata-rata F1 makro.[24] Berdasarkan semua data eksperimental, sebagian besar metrik model setelah augmentasi data menggunakan *Back-Translation* mengalami peningkatan.

Layanan API Google Translate digunakan pada penelitian kali ini untuk melakukan *Back-Translation*, dengan berdasarkan akurasi dari *An Updated Evaluation of Google Translate Accuracy* oleh Milam Aiken yang menganalisis akurasi google translate dengan 50. Dan pada penelitian kali ini diambil 20 bahasa paling akurat[25]

Tabel 2. Akurasi Google Translate

No.	language	bleu2	bleu3	bleuMean
1	Italian	100	90	95
2	French	89	88	88,5
3	Swedish	85	86	85,5
4	Danish	84	82	83
5	Portuguese	75	91	83
6	Indonesian	81	82	81,5
7	Polish	79	84	81,5
8	Croatian	83	77	80
9	Bulgarian	79	80	79,5
10	Finnish	82	77	79,5
11	Norwegian	83	75	79
12	Russian	74	84	79
13	Spanish	78	80	79
14	Dutch	71	84	77,5
15	Afrikaans	71	83	77
16	German	72	81	76,5
17	Slovak	68	83	75,5
18	Czech	64	86	75
19	Latvian	73	77	75
20	Albanian	70	80	75

dimana blue2 adalah translasi dari bahasa inggris ke bahasa tersebut, dan blue3 adalah translasi dari bahasa asal ke bahasa inggris.

Augmentasi juga dilakukan menggunakan model translasi yang terdapat di huggingface dengan model model yang telah dilatih oleh HelsinkiNLP yaitu Universitas Helsinki yang terdapat pada Finland. Model translasi yang diambil antara lain adalah Chinese, Spanyol, Russia, Jepang, Jerman, Perancis, Italia, dan Indonesia. Ini dilakukan guna memperluas cakupan dan membandingkan model yang diaugmentasikan dengan Google Translate dan HelsinkiNLP. Setelah augmentasi, *undersampling* atau mengambil jumlah data sebanyak data yang paling minoritas dilakukan agar pelatihan model tidak condong ke satu sentimen.



Gambar 5. Jumlah Data Latih

Teknik augmentasi data juga diusulkan oleh Wei dan Zou dengan judul *Easy Data Augmentation* (EDA). Diterapkan operasi penggantian sinonim, penyisipan, pertukaran, dan penghapusan secara acak.[26] Analisis akurasi model paling akurat setelah data diaugmentasi menggunakan EDA juga dilakukan pada penelitian kali ini.

Data analisis sentimen telah dibersihkan lalu dilanjutkan dengan membangun model analisis sentimennya. Dikarenakan membangun model dari awal membutuhkan kekuatan komputasi serta data yang sangat besar, maka dilakukan *fine tuning* dari model yang sudah ada. Dengan memilih model yang paling besar akurasi terhadap data yang telah dilabeli manual, pelabelan data teks dilakukan serta diambil rata rata tiap harinya dengan rumus

$$\sum_{i=1}^n \frac{s(x_i)}{n} \quad (1)$$

dimana $s(x_i)$ adalah skor sentimen dari teks x_i dalam interval satu hari, dan n adalah banyak data teks dalam interval tersebut.[27]

Multiple Season-Trend Decomposition using LOESS (MSTL) digunakan untuk menganalisis serta mengekstrak tren musiman. Metode ini memungkinkan kita untuk memisahkan komponen musiman dan tren dari data deret waktu, membantu mengidentifikasi pola yang berkaitan dengan faktor-faktor musiman tertentu yang dapat mempengaruhi harga Bitcoin.[28] Dan *Time Lag Plot* (TLP) untuk mengeksplorasi korelasi antara variabel-variabel yang diamati dan harga Bitcoin serta melihat tendensi *outlier* atau nilai pencilan pada dataset.[29]

Nilai pencilan dalam data rangkaian waktu khususnya harga bitcoin yang memang memiliki sifat tidak stabil harus ditangani dengan hati hati, oleh karena itu metode Isolation Forest digunakan dalam pendeteksian nilai pencilan pada data ini. Penghapusan sekitar 10% dari nilai pencilan meningkatkan performa model untuk kebanyakan machine learning[30], maka pengaruh nilai pencilan terhadap kinerja modelpun dipelajari dengan tidak menghapus dan menghapus sekitar 10% dari nilai pencilan. Teknik yang digunakan untuk menggantikan nilai pencilan adalah *Moving Average* atau rata rata berjalan.[31]

$$y'_t = \frac{\sum_{i=0}^{n-1} y_{t-i}}{n} \quad (2)$$

dimana y'_t adalah pengganti titik observasi, y_t adalah titik observasi, dan n adalah besar jendela observasi.[32]

Tes Augmented Dickey-Fuller[33] juga dilakukan untuk mengatasi fluktuasi dan volatilitas tinggi harga serta volume bitcoin, data yang tidak stasioner diubah menjadi stasioner dengan menerapkan *detrending* atau penghapusan tren.[34] Teknik *detrending* yang dilakukan pada penelitian kali ini adalah *differencing transformation* atau dengan menghasilkan deret waktu baru di mana nilai baru y'_{t_i} pada waktu t_i dikalkulasikan dengan perbedaan diantara observasi orisinil dan observasi $y_{t_{i-1}}$ pada langkah waktu sebelumnya. Rumusnya dapat dijelaskan sebagai berikut.

$$y'_{t_i} = y_{t_i} - y_{t_{i-1}} \quad (3)$$

Normalisasi data juga dilakukan guna meningkatkan performa model[34], digunakan normalisasi Min-Max sehingga nilai nilai di dalam data dipetakan dalam rentang (0, 1) dengan rumus sebagai berikut

$$y_{t_i} = \frac{y_{t_i} - y_{min}}{y_{max} - y_{min}} \quad (4)$$

dimana y_{min} adalah nilai minimal dari y , dan y_{max} adalah nilai maksimal dari y . Dan untuk menghindari kebocoran pada data test, maka *fitting* skala data hanya diambil dari data latih.

Model tranformer dari website hugging face digunakan untuk model analisis sentimen serta website darts digunakan untuk peramalan deret waktu, dengan hyperparameter serta rentang pencarian sebagai berikut

Tabel 3. Hyperparameter dan Rentang Pencarian

Hyperparameter	Rentang Pencarian	
	Analisis Sentimen	Peramalan
input chunk length	(7, 30)	16
output chunk length	(7, 30)	32
learning rate	(0.00001, 0.00005)	(0.00001, 0.00005)
d model	[32, 64, 128]	-
nhead	[2,4,8]	-
num_encoder_layers	(2, 8)	-
num_decoder_layers	(2, 8)	-
dim_feedforward	[256, 512, 1024]	-
dropout	(0.0, 0.3)	-
batch_size	(7, 365)	-
include_year	[True, False]	-
warmup_steps	-	[0, 250, 500]
weight_decay	-	(0.0, 0.3)

Evaluasi dari kedua model juga penting untuk dilakukan, guna membandingkan hasil dari model dengan model yang sudah ada. Evaluasi dari model analisis sentiment menggunakan Akurasi, *F-1 Score*, *Precision*, dan *Recall*. Dengan akurasi dilakukan menggunakan rumus

$$Akurasi = \frac{\text{Total prediksi label yang benar}}{\text{Total prediksi}} \quad (5)$$

selanjutnya dengan *Precision* serta *Recall* menggunakan rumus sebagai berikut

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

dimana TP adalah *True Positive*, FP adalah *False Positive*, dan FN adalah *False Negative*. Dan yang terakhir adalah *F-1 Score*

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

Evaluasi peramalan peramalan deret waktu digunakan *Root Mean Square error* (RMSE), *Mean Absolute Percentage Error* (MAPE), serta *Symmetric Mean Absolute Percentage Error* (sMAPE) untuk evaluasinya.

$$RMSE = \sqrt{\sum_{t=1}^n \frac{(A_t - F_t)^2}{n}} \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|} \quad (10)$$

$$sMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{(|A_t| + |F_t|)/2} \quad (11)$$

dimana A_t adalah nilai sebenarnya pada waktu t , F_t adalah nilai prediksi pada waktu t , dan n adalah total data yang diprediksi.

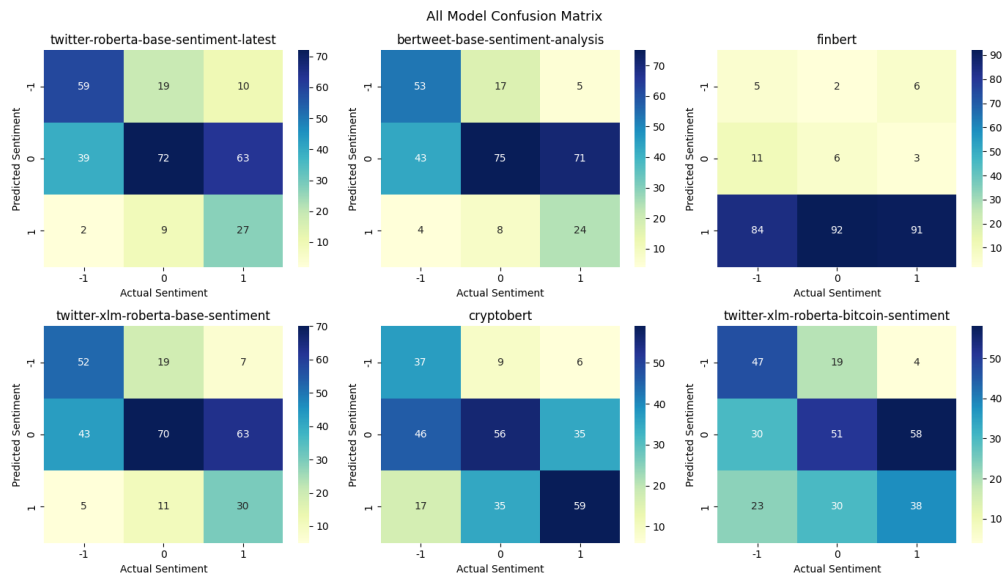
C. Hasil dan Pembahasan

Menjelajahi website Huggingface, ditemukan 6 kandidat model untuk sentimen analisis dengan 2 diantara untuk topik khusus bitcoin. Eksperimen-pun dilakukan untuk mengambil model yang terbaik untuk dilakukan pelatihan dengan menguji ketiga model dengan data yang sudah dilabeli manual. Berikut merupakan confusion matrix serta akurasi model kandidat

Tabel 4. Akurasi Model Kandidat

Model	Acc Negative Sentiment	Acc Neutral Sentiment	Acc Positive Sentiment	Overall Accuracy
-------	---------------------------	--------------------------	---------------------------	------------------

twitter-roberta-base-sentiment-latest	0,59	0,72	0,27	0,526666667
bertweet-base-sentiment-analysis	0,53	0,75	0,24	0,506666667
finbert	0,05	0,06	0,91	0,34
twitter-xlm-roberta-base-sentiment	0,52	0,7	0,3	0,506666667
cryptobert	0,37	0,56	0,59	0,506666667
twitter-xlm-roberta-bitcoin-sentiment	0,47	0,51	0,38	0,453333333



Gambar 6. Confusion Matrix Model Kandidat

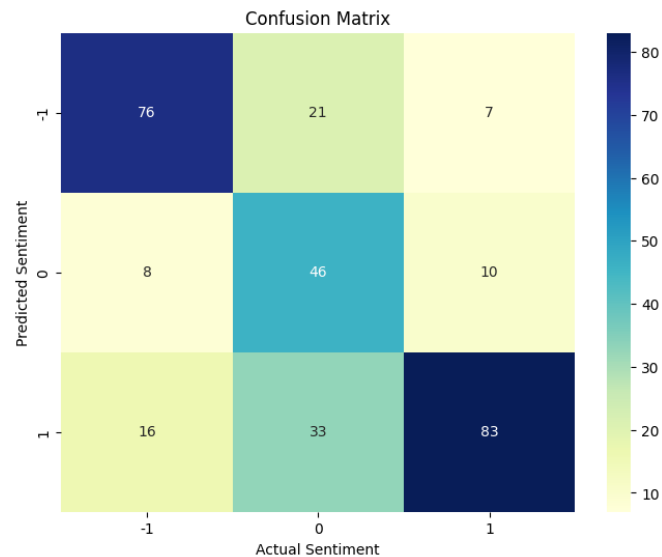
Berdasarkan data yang terdapat dalam *confusion matrix* dan tabel di atas, dapat diketahui terdapat 4 kandidat model dengan tingkat akurasi total terbesar. Maka dari itu 4 model ini-pun dilatih lebih lanjut dengan data latih. Hasil dari ke-4 model tersebut setelah dilatih dengan *hyperparameter* paling optimal adalah sebagai berikut

Tabel 5. Hasil Setelah Fine-Tuning

Model	Murni	GT	GTH	T
bertweet-base-sentiment-analysis	0.626667	0.666666	0.666666	0.656667
Cryptobert	0.666666	0.673333	0.676667	0.646667
twitter-roberta-base-sentiment-latest	0.653333	0.65	0.646667	0.64
twitter-xlm-roberta-base-sentiment	0.606667	0.596667	0.616667	0.62

Dataset murni atau yang tidak melewati augmentasi data mendapatkan akurasi paling besar pada model cryptobert, akan tetapi saat dilakukan augmentasi data peningkatan sebesar 1% pada data yang diaugmentasi dengan GTH. Dan ketika dataset murni dan GTH dilakukan *Easy Data Augmentation* menggunakan model cryptobert tidak didapatkan perkembangan dalam akurasi. Maka dipilih model yang paling akurat GTH dengan model Cryptobert yang memiliki *hyperparameter* *learning_rate* 2.16491656079216e-05, *per_device_eval_batch_size* sebesar 32, *per_device_train_batch_size* sebesar 16, *warmup steps* sebesar 500, dan *weight*

decay sebesar 0.1295835055926347. (Model dapat diakses melalui <https://huggingface.co/AfterRain007/cryptobertRefined>) Berikut merupakan confusion matrix dan hasil evaluasinya

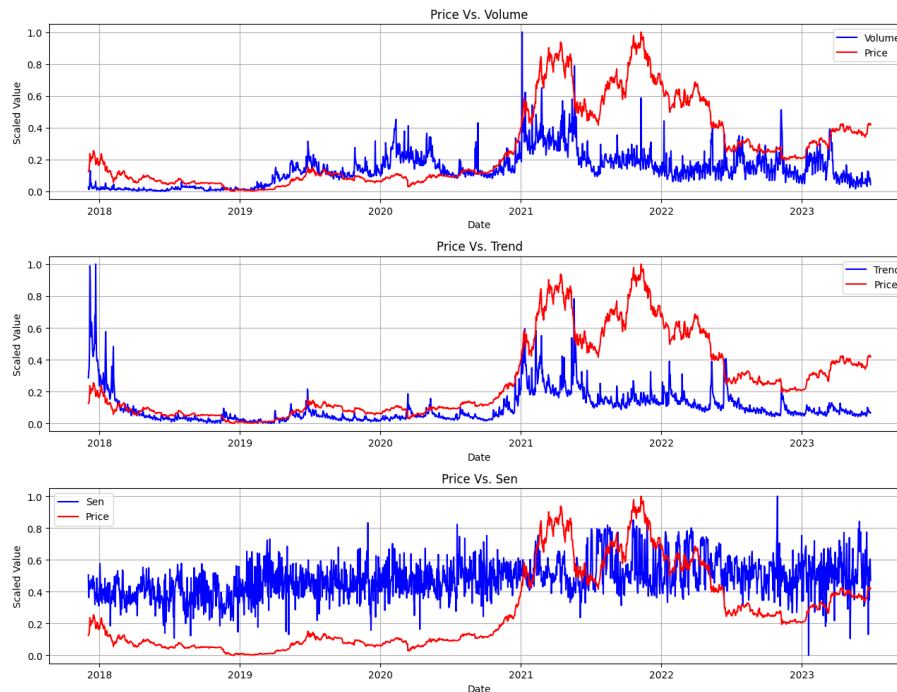


Gambar 7. Confusion Matrix Model Paling Akurat

Tabel 6. Hasil Evaluasi Model Paling Akurat

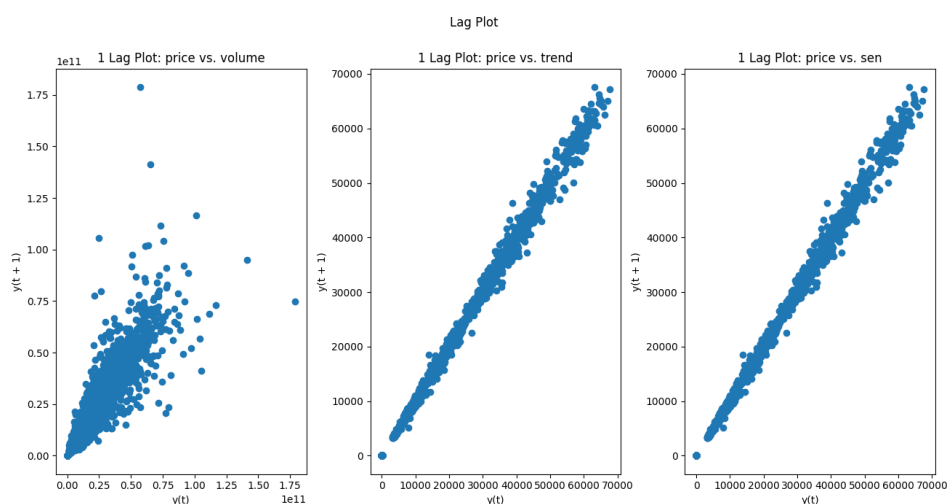
Sentiment Score	Precision	Recall	F1-Score	n
-1	0,73	0,76	0,75	100
0	0,72	0,48	0,56	100
1	0,63	0,83	0,72	100
Accuracy		0,68		300
Macro Avg	0,69	0,68	0,67	300
Weighted Avg	0,69	0,68	0,67	300

Setelah mendapatkan model analisis sentimen, dilakukan labeling dengan model yang telah dilatih kepada data teks dari twitter serta reddit dan diambil nilai rata rata sentimennya tiap hari. Setelah data di normalisasikan, maka didapatkan perbandingan data harga dengan variabel lain.



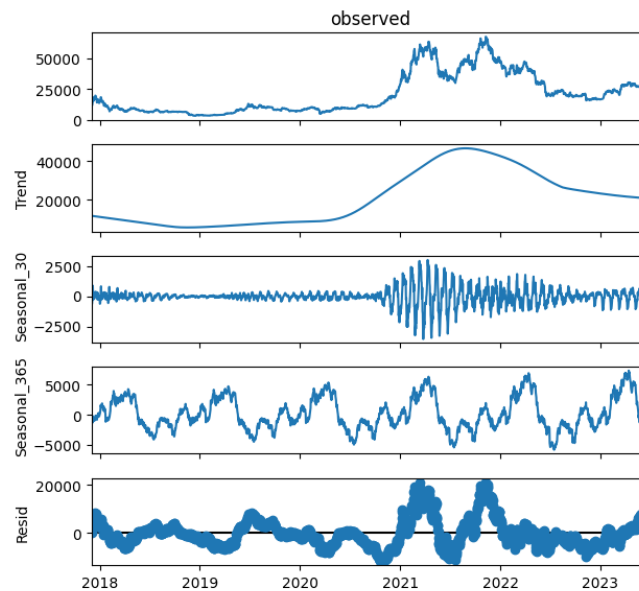
Gambar 8. Harga dibandingkan Semua Variabel

Sekilas, dari tabel diatas dapat dilihat harga dan sentiment tidak memiliki korelasi sama sekali dari tahun ke tahun. Akan tetapi beda hal dengan tren serta volume, dari tahun ke tahun tren sangat amat berkorelasi dengan harga. Tren dan volume yang naik seringkali bersandingan dengan harga bitcoin yang naik, akan tetapi bukan berarti tren dan volume yang turun bersandingan dengan harga yang turun. Sering kali Tren dan Volume naik bersandingan dengan harga yang turun. Akan tetapi diperlukan analisis mendalam untuk mendeterminasi terkait korelasi antara variabel dengan *time lag plot*. Berikut merupakan TLP harga dibandingkan dengan variabel lainnya, menggunakan lag = 1,



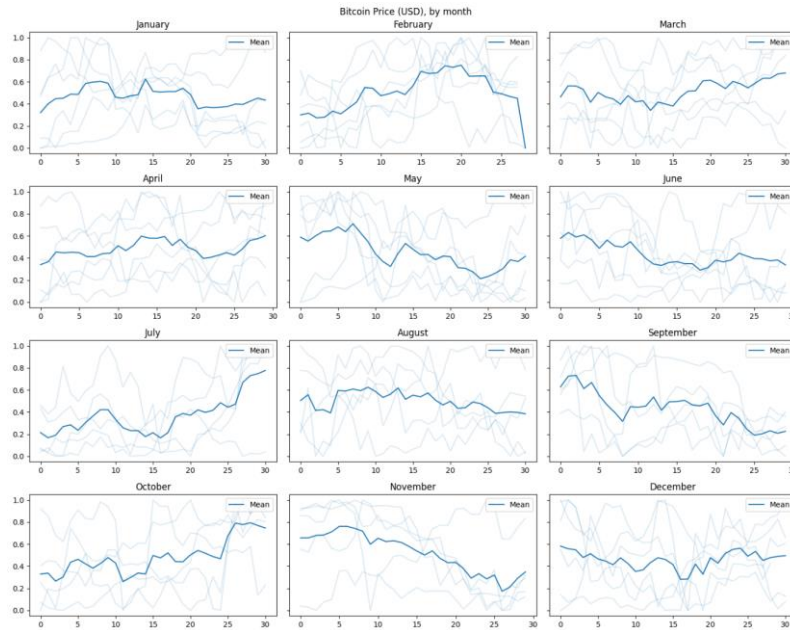
Gambar 9. Lag Plot harga dibandingkan Variabel Lain

Harga terhadap tren dan sentimen dapat dilihat dari grafik di atas bahwa keduanya memiliki sifat autokorelasi positif yang tinggi, serta harga terhadap volume memiliki autokorelasi yang bersifat moderat. Dan dari ketiga grafik di atas dapat diketahui bahwa variabel variabel memiliki tendensi nilai pencilan (dijelajahi lebih lanjut pada bagian *isolation forest*). Lalu untuk tren musiman dilakukan MSTL, dikarenakan data yang digunakan bersifat harian maka ditetapkan 2 jangka waktu yaitu per-bulan (Seasonal_30) dan per-tahun (Seasonal_365).



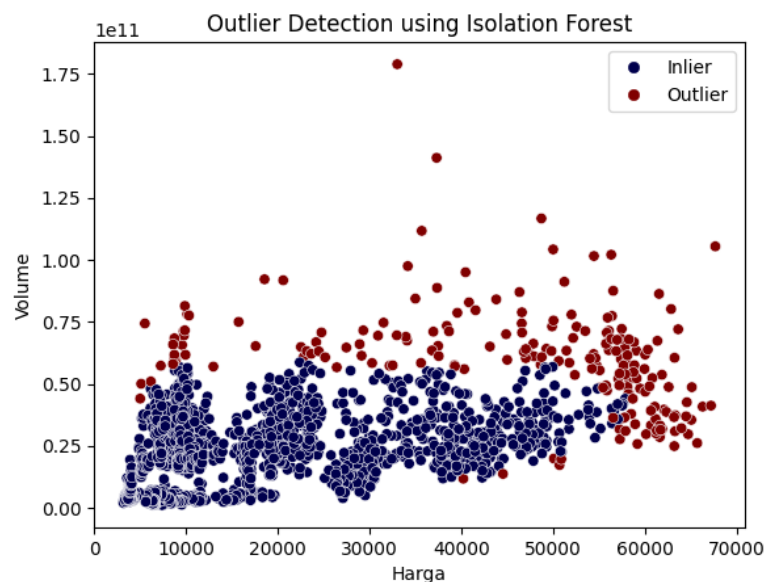
Gambar 10. MSTL Harga

Dapat dilihat dari grafik di atas, komponen residual menunjukkan bahwa ada jumlah variasi yang signifikan yang tidak dapat dijelaskan dalam harga Bitcoin itu sendiri melainkan disebabkan oleh berbagai faktor lain. Dapat dilihat juga dari grafik per tahun, terdapat tendensi lebih tinggi pada bulan Desember dan rendah pada bulan Juni. Lalu untuk grafik per bulan sulit untuk diuraikan karena ukuran terlalu kecil, maka dari itu grafik bulanan dipecahkan menjadi 12 untuk tiap bulan.



Gambar 11. Harga per Bulan

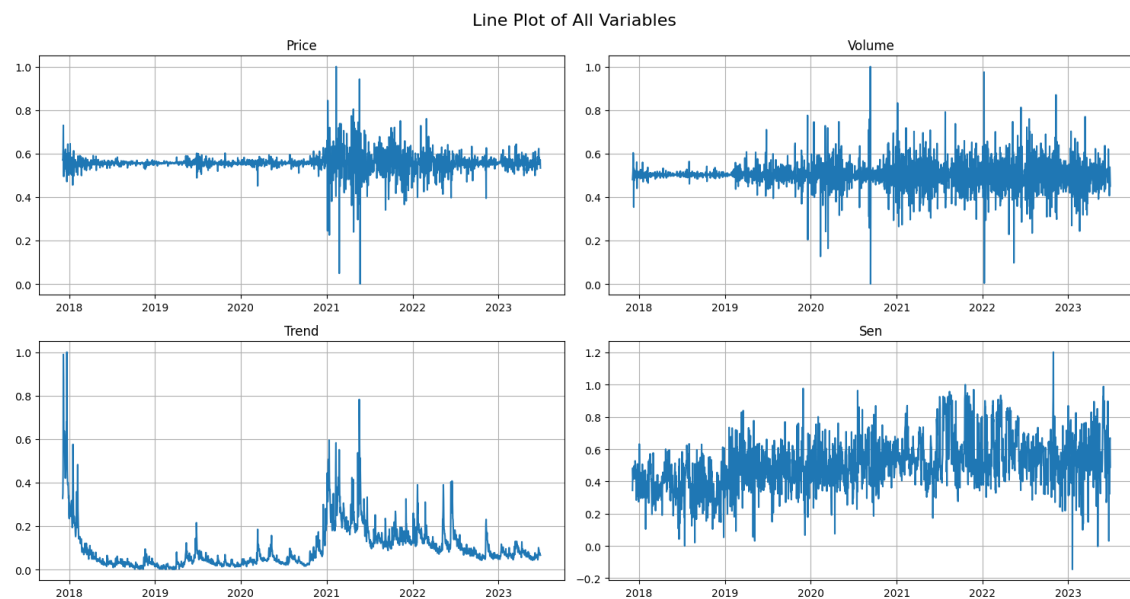
Dapat dilihat pula harga bitcoin seringkali tidak memiliki tendensi untuk naik ataupun turun berdasarkan bulannya, akan tetapi pada Juli dan Oktober memiliki tendensi untuk naik. Dan pada bulan September, serta November harga memiliki tendensi untuk turun. Perlu diingat bahwa tanggal 29 Februari rata rata harga turun drastis dikarenakan tanggal 29 hanya sekali dilewati (yaitu pada tahun kabisat). Lalu sebelum dilakukan peramalan deret waktu, yang terakhir dilakukan adalah penanganan nilai pencilan.



Gambar 12. Grafik Nilai Pencilan

Berdasarkan *Isolation Forest* terdapat 204 nilai pencilan berdasarkan harga dan volume, penanganan nilai pencilan dilakukan dengan mengambil sepuluh

persen dari total banyaknya nilai pencilan yang memiliki score paling tinggi dan di transformasi dengan menggantikan nilai pencilan menggunakan rata rata berjalan.



Gambar 13. Data Variabel

dan beberapa eksperimen dijalankan dengan menghapus satu ataupun lebih dari data kovariat.

Tabel 7. Hasil Peramalan Model *Temporal Fusion Transformers*

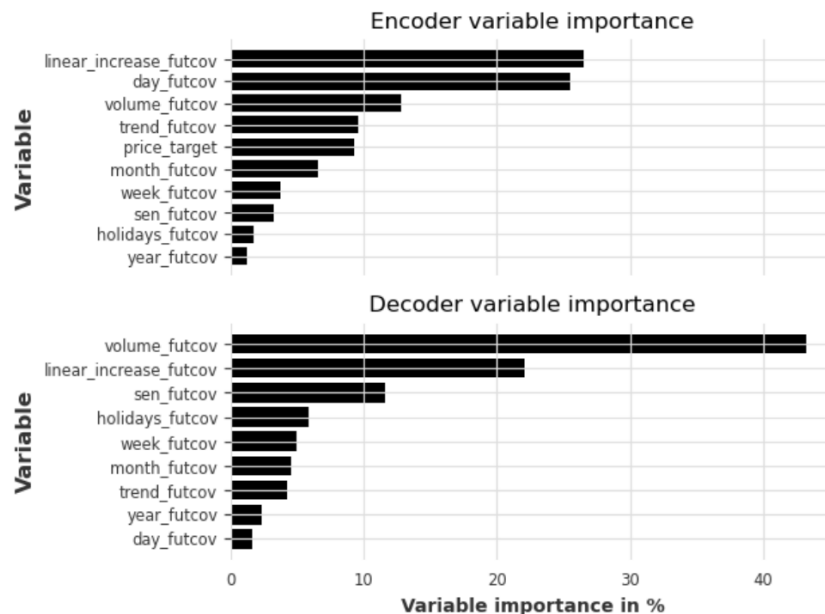
Included Covariate	RMSE	sMAPE	MAPE
Trend, Sentiment, and Volume	0,0260	3,0199	3,0499
Sentiment and Volume	0,0265	3,0321	3,0610
Trend and Sentiment	0,0259	2,9919	3,0036
Trend and Volume	0,0260	3,0488	3,0754
Price	0,0264	3,0406	3,0717
Sentiment	0,0263	3,0409	3,0674
Trend	0,0256	3,0428	3,0565
Volume	0,0263	3,0442	3,0782

Tabel 8. Hasil Peramalan Model *Long-Short Term Memory*

Included Covariate	RMSE	sMAPE	MAPE
Trend, Sentiment, and Volume	0,0263	3,0236	3,0500
Sentiment and Volume	0,0263	2,9956	3,0196
Trend and Sentiment	0,0262	3,1219	3,1545
Trend and Volume	0,0262	3,0615	3,0887
Price	0,0264	3,0189	3,0414
Sentiment	0,0263	3,0128	3,0287
Trend	0,0263	3,0272	3,0553
Volume	0,0264	3,0238	3,0506

Berdasarkan tabel 7, dapat disimpulkan bahwa model TFT sendiri sudah mendapatkan hasil yang sangat memuaskan yaitu memiliki nilai RMSE sekitar 0.026

baik dimasukan variable seperti trend, volume, ataupun sentiment. Dapat disimpulkan bahwa arsitektur seleksi variabel yang terdapat di TFT berfungsi dengan baik dengan tidak mempertimbangkan variabel yang insignifikan dalam perhitungan prediksi. Tabel 8 juga memperlihatkan bahwa model TFT kebanyakan lebih unggul dibandingkan dengan model dasar LSTM. Dengan nilai RMSE terendah model TFT sebesar 0.0256 dan LSTM sebesar 0.0262.



Gambar 14. *Autoencoder Variable Importance*

Tabel 9. *Nilai Autoencoder Variabel Importance*

type	encoder	decoder
year_futcov	1,20	2,30
holidays_futcov	1,70	5,80
sen_futcov	3,20	11,60
week_futcov	3,70	4,90
month_futcov	6,50	4,50
price_target	9,30	-
trend_futcov	9,60	4,20
volume_futcov	12,80	43,20
day_futcov	25,50	1,60
linear_increase_futcov	26,50	22,10

Variabel variabel yang penting di dalam Encoder terlihat terdistribusi secara rata dari variabel harga keatas sembari menghiraukan variabel tahun, libur, sentimen, bulan, serta minggu. Sedangkan untuk Decoder sendiri sangat bergantung dengan variabel volume dengan nilai kepentingan lebih dari 40%, dengan variabel pendukung yaitu linear_increase dan sentimen.

D. Simpulan

Meledaknya mata uang kripto ke masyarakat umum meningkatkan urgensi untuk pembuatan model yang kokoh dan akurat. Dengan menggunakan model

paling mutakhir yaitu transformers dalam *natural language processing* dan *Time Series Forecasting*, dilakukan beberapa eksperimen.

Ditemukan bahwa menggunakan 1464 data untuk melatih model cryptobert mendapatkan peningkatan akurasi sebesar 0.16% dan ketika dilakukan augmentasi data dengan *back-translation* menggunakan 10 bahasa yang memiliki BLEU score terbesar oleh google translate, didapatkan peningkatan lagi sebesar 0.01%. Oleh karena itu disimpulkan data latih yang banyak bukan berarti akan menghasilkan model yang terbaik akan tetapi kualitas dari data latih itu juga harus diperhatikan, dan juga *Easy Data Augmentation* dilakukan akan tetapi tidak membuahkan peningkatan akurasi.

Pengintegrasian sentimen, volume, dan trend kedalam model prediksi harga bitcoin mendapatkan peningkatan hasil dibandingkan hanya menintegrasikan kovariat waktu, dari RMSE sebesar 0,0264 menjadi 0,0260. Dan juga model TFT lebih unggul dibandingkan dengan model LSTM, dengan LSTM mendapatkan hasil RMSE terbaik sebesar 0,0262 dan TFT terbaik sebesar 0,0256.

Saran untuk penelitian selanjutnya adalah dengan memperbesar token dari model yang digunakan. Dikarenakan X (dahulunya twitter) yang dahulu memlimitasi sebanyak 280 karakter, per-9 Februari 2023 sudah bisa melebihi dari 4000 karakter. Ditambah data reddit dan sumber sumber lain memiliki maksimal karakter yang tidak terhingga. Serta perbandingan model translasi dengan Google Translate dapat dilakukan menggunakan bahasa yang sama untuk mengevaluasi model itu sendiri.

Data yang digunakan untuk meramal deret waktu juga dapat diperluas. Keterbatasan saat ini menyebabkan penelitian hanya memanfaatkan data dari 03/12/2017 hingga 29/06/2023, sedangkan Bitcoin telah tersedia sejak tahun 2009.

E. Ucapan Terima Kasih

Ucapan terima kasih tidak lupa diberikan kepada pihak pihak yang telah membantu penelitian ini. Yang antara lain adalah Bapak Djuharis Rasul dan Ibunda Yanawati untuk dukungannya selama 4.5 tahun masa kuliah, Risky Amalia Marhariyadi yang membantu melabeli dataset manual, SURGE AI telah memberikan dataset manual, Izzudin yang telah memberikan insightnya, Farrel yang telah meminjamkan GTX 1050 untuk mempercepat komputasi, serta Nindi, Sheila, dan teman teman sekalian yang telah memberikan dukungan secara mental.

F. Referensi

- [1] L. Zhao, "The function and impact of cryptocurrency and data technology in the context of financial technology: introduction to the issue," *Financ. Innov.*, vol. 7, no. 1, 2021, doi: 10.1186/s40854-021-00301-w.
- [2] A. H. Al-Nefaie and T. H. H. Aldhyani, "Bitcoin Price Forecasting and Trading: Data Analytics Approaches," *Electron.*, vol. 11, no. 24, 2022, doi: 10.3390/electronics11244088.
- [3] U. Rahardja, Q. Aini, E. Purnamaharap, and R. Raihan, "GOOD, BAD AND DARK BITCOIN: A Systematic Literature Review," *Aptisi Trans. Technopreneursh.*, vol. 3, no. 2, pp. 1–5, 2021, doi: 10.34306/att.v3i2.175.
- [4] D. Higdon, J. Nelson, and J. Ibarra, JuanAbraham, "Cryptocurrency price

- prediction using tweet volumes and sentiment analysis," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 1, 2018.
- [5] M. A. Nasir, T. L. D. Huynh, S. P. Nguyen, and D. Duong, "Forecasting cryptocurrency returns and volume using search engines," *Financ. Innov.*, vol. 5, no. 1, 2019, doi: 10.1186/s40854-018-0119-8.
 - [6] S. Alghamdi, S. Alqethami, T. Alsubait, and H. Alhakami, "Cryptocurrency Price Prediction using Forecasting and Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, pp. 891–900, 2022, doi: 10.14569/IJACSA.2022.01310105.
 - [7] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
 - [8] G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A Transformer-based Framework for Multivariate Time Series Representation Learning," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2114–2124, 2021, doi: 10.1145/3447548.3467401.
 - [9] E. Gemici and M. Polat, "Relationship between price and volume in the Bitcoin market," *J. Risk Financ.*, vol. 20, no. 5, pp. 435–444, 2019, doi: 10.1108/JRF-07-2018-0111.
 - [10] M. Ettredge, J. Gerdes, and G. Karuga, "Using web-based search data to predict macroeconomic statistics," *Commun. ACM*, vol. 48, no. 11, pp. 87–92, 2005, doi: 10.1145/1096000.1096010.
 - [11] H. Choi and H. Varian, "Predicting the Present with Google Trends," *Econ. Rec.*, vol. 88, no. SUPPL.1, pp. 2–9, 2012, doi: 10.1111/j.1475-4932.2012.00809.x.
 - [12] S. Mehtab and J. Sen, "Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Timeseries," 2020, doi: 10.36227/techrxiv.15088734.v1.
 - [13] S. Hansun, A. Wicaksana, and A. Q. M. Khaliq, "Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00601-7.
 - [14] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Appl. Syst. Innov.*, vol. 4, no. 1, pp. 1–30, 2021, doi: 10.3390/ASI4010009.
 - [15] B. Lim, S. Arik, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecast.*, vol. 37, no. 4, pp. 1748–1764, 2021, doi: 10.1016/j.ijforecast.2021.03.012.
 - [16] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, 2018, doi: 10.1145/3236009.
 - [17] D. Sarkar, *Text Analytics with Python*, vol. 32, no. 1. 2016. doi: 10.1140/epja/i2006-10279-1.
 - [18] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006. doi: 10.1017/cbo9780511546914.
 - [19] and G. E. H. A. Krizhevsky, I. Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
 - [20] W. Wang, B. Li, D. Feng, A. Zhang, and S. Wan, "The OL-DAWE Model: Tweet

- Polarity Sentiment Analysis with Data Augmentation," *IEEE Access*, vol. 8, pp. 40118–40128, 2020, doi: 10.1109/ACCESS.2020.2976196.
- [21] S. Edunov, M. Ott, M. Auli, and D. Grangier, "Understanding back-translation at scale," *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 489–500, 2018, doi: 10.18653/v1/d18-1045.
- [22] M. Bayer, M. A. Kaufhold, and C. Reuter, "A Survey on Data Augmentation for Text Classification," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–44, 2022, doi: 10.1145/3544558.
- [23] J. Ma and L. Li, "Data Augmentation for Chinese Text Classification Using Back-Translation," *J. Phys. Conf. Ser.*, vol. 1651, no. 1, 2020, doi: 10.1088/1742-6596/1651/1/012039.
- [24] S. T. Aroyehun and A. Gelbukh, "Aggression Detection in Social Media: Using Deep Neural Networks, Data Augmentation, and Pseudo Labeling," *COLING 2018 - 1st Work. Trolling, Aggress. Cyberbullying, TRAC 2018 - Proc. Work.*, pp. 90–97, 2018.
- [25] M. Aiken, "An Updated Evaluation of Google Translate Accuracy," *Stud. Linguist. Lit.*, vol. 3, no. 3, p. p253, 2019, doi: 10.22158/sll.v3n3p253.
- [26] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 6382–6388, 2019, doi: 10.18653/v1/d19-1670.
- [27] M. Frohmann, M. Karner, S. Khudoyan, R. Wagner, and M. Schedl, "Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting," *Big Data Cogn. Comput.*, vol. 7, no. 3, 2023, doi: 10.3390/bdcc7030137.
- [28] K. Bandara, R. Hyndman, and C. Bergmeir, "MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns," *Int. J. Oper. Res.*, vol. 1, no. 1, p. 1, 2022, doi: 10.1504/ijor.2022.10048281.
- [29] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 759–762, 2009, doi: 10.1109/IEDM.2009.5424230.
- [30] M. Mudassir, S. Bennbaia, D. Unal, and M. Hammoudeh, "Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach," *Neural Comput. Appl.*, vol. 6, 2020, doi: 10.1007/s00521-020-05129-6.
- [31] C. Jeenanunta, K. D. Abeyrathna, and M. H. M. R. S. Dilhani, "Time Series Outlier Detection for Short-Term Electricity Load Demand Forecasting | INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND TECHNOLOGY (ISJET)," vol. 2, no. 1, pp. 37–50, 2018, [Online]. Available: <https://ph02.tci-thaijo.org/index.php/isjet/article/view/175908>
- [32] F. R. Johnston, J. E. Boyland, M. Meadows, and E. Shale, "Some Properties of a Simple Moving Average when Applied to Forecasting a Time Series," *J. Oper. Res. Soc.*, vol. 50, no. 12, p. 1267, 1999, doi: 10.2307/3010636.
- [33] R. Mushtaq, "TESTING TIME SERIES DATA FOR STATIONARITY," pp. 1–19.
- [34] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *Forecasting*, vol. 5, no. 1, pp. 196–209, 2023, doi:

10.3390/forecast5010010.