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

Gilang Islamay Putra Djuharis

UIN Syarif Hidayatullah Jakarta

Abstract

Syalalalala

## PENDAHULUAN

Perkembangan teknologi tidak hanya terbatas pada sektor perindustrian, tetapi juga mencakup kemajuan teknologi dalam dunia ekonomi dengan adanya perkembangan mata uang digital yang dikenal sebagai cryptocurrency. Cryptocurrency memungkinkan pengembangan sistem yang bekerja secara mandiri tanpa melibatkan peran perantara seperti bank.

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. 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.

Salah satu hal menarik bagi para investor adalah pasar mata uang kripto yang buka sepanjang waktu, berbeda dengan pasar Foreign Exchange (Forex) yang hanya buka pada hari kerja dan memiliki jam tertentu. Hal ini memungkinkan para investor untuk melakukan transaksi jual-beli aset mereka dengan fleksibilitas.

Namun, volatilitas harga yang tinggi pada mata uang kripto menjadi salah satu alasan mengapa beberapa investor enggan masuk ke pasar ini. Terdapat fluktuasi harga yang drastis, contohnya harga Bitcoin turun sebesar 19% pada tanggal 11 Januari 2021. Oleh karena itu, diperlukan suatu model yang dapat memprediksi harga mata uang kripto di masa depan untuk membantu investor memaksimalkan keuntungan dan meminimalkan kerugian.

Beberapa penelitian, seperti yang dilakukan oleh Choi dan Varian dalam penelitian mereka yang berjudul "Predicting the Present with Google Trends", 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. Selain itu, penelitian oleh Eray Gemici dan Müslüm Polat tentang hubungan antara harga dan volume di pasar Bitcoin menunjukkan adanya korelasi antara harga dan volume. Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa popularitas dan volume mata uang kripto memiliki korelasi dengan harga. Dalam penelitian ini, kami akan menambahkan variabel harga dan sentimen media sosial untuk mengembangkan model prediksi yang lebih baru.

Selain itu, model transformer 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.

## TINJAUAN PUSTAKA

Sentimen analisis 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 sentiment analysis 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.

Walaupun ElKulako melatihnya dengan data training yang besar, tetapi 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.

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. 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. Selain itu, penelitian oleh Gemici dan Polat (2019) menyimpulkan adanya kointegrasi antara harga dan volume.

Penggunaan analisis deret waktu multivariat bukanlah hal baru. Mehtab dan Sen (2020) menemukan bahwa penggunaan model multivariat dalam memprediksi harga saham memberikan hasil yang lebih baik dibandingkan dengan pendekatan univariat. Hanus 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.

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.

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.

## DATA

### Pengambilan Data

Ketentuan baru yang dikeluarkan oleh twitter mengenai daftar biaya terbaru melimitasi ruang gerak data, maka dari itu data cuitan twitter dari 05/02/2021 hingga 09/01/2023 diambil dari Kaggle untuk penelitian kali ini. Dan *scraping* sosial media reddit dilakukan dengan mengambil komen yang terdapat pada daily discussion di r/bitcoin, ini dilakukan 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. Data yang didapatkan serta digunakan sebagai berikut:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Twitter | | Reddit | | Bitcoin | | Google Trends | |
| user\_name | ✕ |  |  | Open Time | ✕ |  |  |
| user\_location | ✕ | Title | ✕ | Open | ✕ |  |  |
| user\_description | ✕ | Score | ✕ | High | ✕ |  |  |
| user\_created | ✕ | Comments | ✓ | Low | ✕ |  |  |
| user\_followers | ✕ | Author | ✕ | Close | ✓ |  |  |
| user\_friends | ✕ | Flair | ✕ | Volume | ✓ | Day | ✓ |
| user\_favourites | ✕ | Post ID | ✕ | Close Time | ✓ | Bitcoin | ✓ |
| user\_verified | ✕ | URL | ✕ | Quote Asses Volume | ✕ |  |  |
| Date | ✓ | Timestamp | ✓ | Number of Trades | ✕ |  |  |
| Text | ✓ | Awards | ✕ | Taker Buy Base Asset Volume | ✕ |  |  |
| hastags | ✕ | Subreddit | ✕ | Taker Buy Quote Asset Volume | ✕ |  |  |
| Source | ✕ | Text | ✕ | Ignore | ✕ |  |  |
| is\_retweet | ✕ |  |  |  |  |  |  |

Dimana ✓ = digunakan, dan ✕ = tidak digunakan. Dan data google trends diambil secara manual dalam jangka tiap enam bulan dari 03/12/2017 hingga 30/06/2023 lalu data dinormalisasi dengan mencocokan tanggal yang berkesinambungan. Ini dilakukan guna mengambil data harian google trends.

### Persiapan Data

Dipanjan Sarkar dalam bukunya yang berjudul Text Analytics with Python menyatakan bahwa untuk mendapatkan hasil yang optimal dalam analisis teks dibutuhkan beberapa langkah. Yaitu tokenisasi teks dan normalisasi teks. 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.

Penelitian kali ini mengaplikasikan pra pemrosesan data tersebut 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.

Mata uang kripto yang melonjak kepopularitasannya juga menyebabkan banyaknya tweet yang bersifat spam dan dibuat oleh robot. Contoh dari tweet yang bersifat spam adalah sebagai berikut:

|  |  |
| --- | --- |
| **date** | **text** |
| **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%).  #Bitcoin #BTC https://t.co/5wDhcU31UB |
| **2021-10-19 09:16:20** | 🐳🚨 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  💬I’d take a fast nickel over a slow dime - buyerofblood |
| **2022-04-16 15:59:12** | 💰 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 |

### Label Manual

Pelabelan data manual telah dilakukan untuk melatih model, 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 sudah demikian, jumlah total data sebanyak 3.803 dapat diperbanyak lebih lanjut melalui teknik augmentasi. Salah satu metode augmentasi yang digunakan adalah dengan menerjemahkan teks ke suatu bahasa tertentu dan kemudian menerjemahkannya kembali ke bahasa asal. Menggunakan layanan API Google Translate, dengan berdasarkan *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, yaitu adalah sebagai berikut

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **languages** | **bleu2** | **bleu3** | **bleuMean** |
| **0** | Italian | 100 | 90 | 95 |
| **1** | French | 89 | 88 | 88,5 |
| **2** | Swedish | 85 | 86 | 85,5 |
| **3** | Danish | 84 | 82 | 83 |
| **4** | Portuguese | 75 | 91 | 83 |
| **5** | Indonesian | 81 | 82 | 81,5 |
| **6** | Polish | 79 | 84 | 81,5 |
| **7** | Croatian | 83 | 77 | 80 |
| **8** | Bulgarian | 79 | 80 | 79,5 |
| **9** | Finnish | 82 | 77 | 79,5 |
| **10** | Norwegian | 83 | 75 | 79 |
| **11** | Russian | 74 | 84 | 79 |
| **12** | Spanish | 78 | 80 | 79 |
| **13** | Dutch | 71 | 84 | 77,5 |
| **14** | Afrikaans | 71 | 83 | 77 |
| **15** | German | 72 | 81 | 76,5 |
| **16** | Slovak | 68 | 83 | 75,5 |
| **17** | Czech | 64 | 86 | 75 |
| **18** | Latvian | 73 | 77 | 75 |
| **19** | 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.

### Final Dataset

Data teks dikalkulasikan satu persatu menggunakan model BERT lalu diambil menjadi rata rata sentimennya perhari, lalu disatukan menjadi satu himpunan data besar berisikan tanggal, sentimen, harga bitcoin, volume bitcoin, dan trends dengan banyak 2035 data poin. Berikut merupakan visualisasi dari dataset

\*Grafik visualisasi dataset\*

## METODOLOGI

#### Transformer

Model RNN yang sudah ada sejak tahun 1986 dan LSTM yang ditemukan pada tahun 1997 sama sama memiliki satu kekurangan, yaitu mereka berdua memiliki waktu yang lama dalam pelatihan modelnya. Secara mereka adalah model *recurrent* yang mana melakukan pelatihan secara berurutan satu bersatu dan tidak mendukung paraparalelisasi.

*Attention is All You Need*, adalah judul karya ilmiah diberikan oleh delapan orang dari google. Mereka mengusulkan model terbaru yang merupakan peningkatan dari model berbasis *recurrent* untuk *Natural language Processing* (NLP) dengan nama Transformer dengan menggunakan mekanisme *attention*. Dengan mekanisme ini transformer dapat melihat hubungan antara satu kata dengan kata yang lainnya, dan paralelisasi dapat dilakukan karena mengkalkulasikan hubungan suatu kata dengan kata lainnya tidak perlu mengetahui nilai kata lainnya. *Positional encoding* juga diperkenalkan untuk mengetahui posisi suatu kata relatif dengan kata lainnya. Arsitektur yang menggunakan encoder dan decoder juga membuat paralelisasi dapat dilakukan untuk lebih jauh mempercepat pelatihan data dan membuat model ini superior dibandingkan model RNN lainnya.

#### Arsitektur Transformer



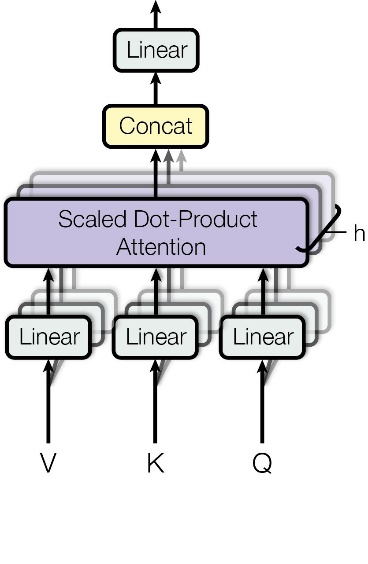
1. *Positional Encoding*

Sebelum model dimasukan kedalam encoder, karena transformer bukanlah model *recurrent* maka tidak diketahui jarak antara satu kata relatif dengan kata lainnya, maka dilakukan *positional encoding* terlebih dahulu untuk mendeterminasi posisi kata kata yang dimasukan menggunakan fungsi sin dan cos berikut.

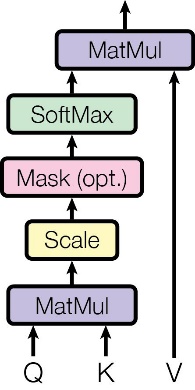
Dimana pos adalah posisi dan i adalah dimensi. Karena sin dan cos merupakan fungsi periodik, penggunaan sin dan cos dalam *positional encoding* dilakukan untuk membantu model dalam menangkap hubungan antara data meskipun jaraknya jauh.

1. *Multi-Head Attention*

Sama halnya seperti *attention* yang sudah dijelaskan, *Multi-Head Attention* memberikan nilai kedalam suatu kata relatif dengan posisi kata lainnya. Dengan *Multi-Head Attention*, transformer dapat memparalelisasi *attention* tersebut dengan menumpukannya. Rumus dan visualisasi dari *attention* adalah sebagai berikut:



Dimana *Scaled Dot-Product Attention* dapat dijelaskan lebih lanjut dari gambar berikut



Dimana Q adalah Query, Dk adalah key dari dimensi, dan V adalah value dari dimensi.

1. *Add & Norm*

Output lalu ditambahkan dengan data yang sebelumnya dan dinormalisasikan. Ini dilakukan guna menangkap informasi terbaru sembari tidak melupakan informasi yang telah ada dalam data itu sendiri, dan dinormalisasi untuk mendapatkan data yang lebih konvergen

1. *Feed Foward*

Data yang telah melewati *Add & Norm* lalu melalui *Feed Foward Neural Network* dengan aktivasi ReLU

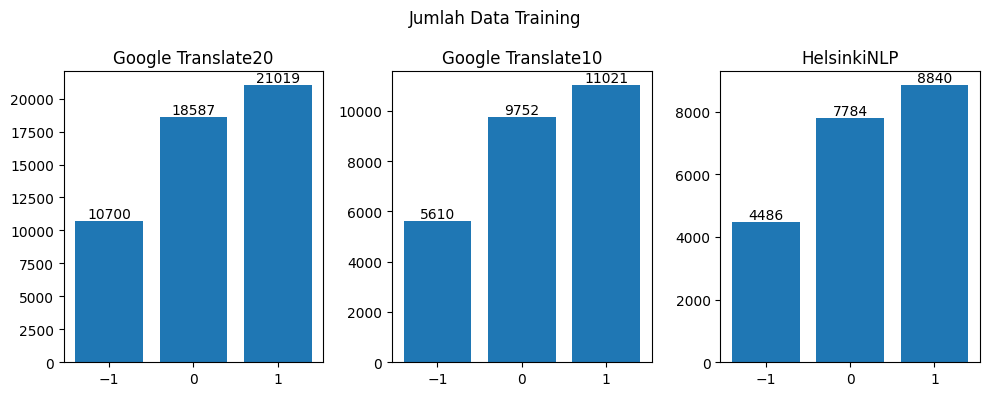
Ini dilakukan untuk mendapatkan komponen hubungan yang kompleks dan non-linear dari data itu sendiri.

## HASIL DAN PEMBAHASAN

5.1 Analisis Sentimen

5.1.1 Persiapan Data Latih

Data latih yang tersedia setelah melakukan augmentasi adalah sebagai berikut

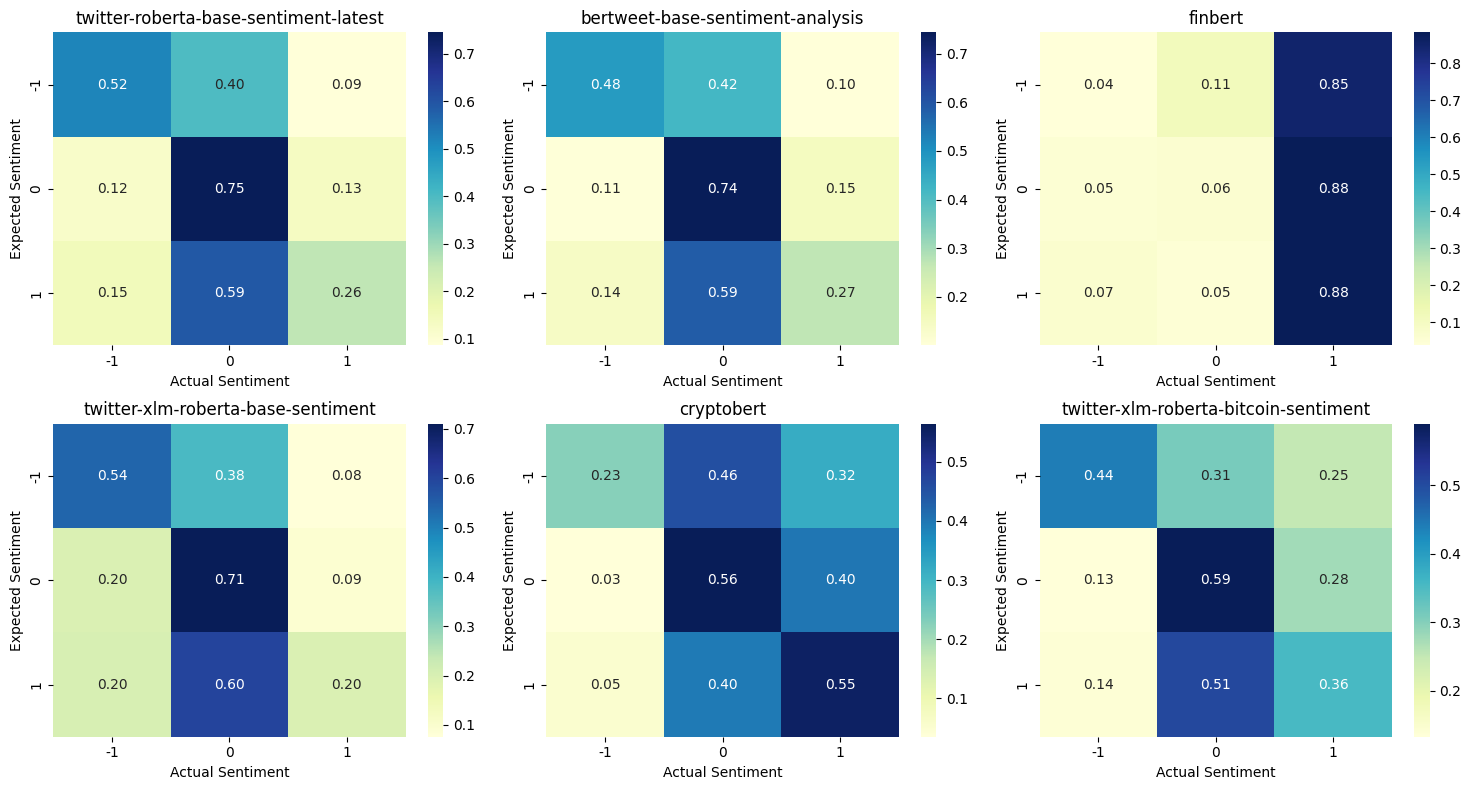


Akan tetapi, diperlukan data yang seimbang agar pelatihan model tidak condong ke satu sentimen. Maka dilakukan *undersampling* atau mengambil jumlah data sebanyak data yang paling minoritas. Maka didapatkan jumlah data training sebagai berikut



5.1.2 Pemilihan Model Analisis Sentimen

Setelah semua data telah dibersihkan, pertama yang dilakukan adalah membangun model analisis sentimennya. Dikarenakan membangun model dari awal membutuhkan kekuatan komputasi yang besar, maka dilakukan pelatihan dari model yang sudah ada. Dengan memilih model yang paling besar akurasinya terhadap data yang telah dilabeli manual. Berikut merupakan confusion matrix dan juga akurasi kandidat model



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | True Negative | True Neutral | True Positive | Overall Accuracy | Time Cost |
| cardiffnlp/twitter-roberta-base-sentiment-latest | 0,52 | 0,75 | 0,26 | 0,49 | 68,86 |
| finiteautomata/bertweet-base-sentiment-analysis | 0,48 | 0,74 | 0,27 | 0,49 | 66,36 |
| ProsusAI/finbert | 0,04 | 0,06 | 0,88 | 0,39 | 64,45 |
| cardiffnlp/twitter-xlm-roberta-base-sentiment | 0,54 | 0,71 | 0,2 | 0,46 | 63,13 |
| ElKulako/cryptobert | 0,23 | 0,56 | 0,55 | 0,48 | 74,19 |
| svalabs/twitter-xlm-roberta-bitcoin-sentiment | 0,44 | 0,59 | 0,36 | 0,46 | 66,28 |

Berdasarkan data yang terdapat dalam heatmap dan tabel di atas, dapat diketahui terdapat 3 kandidat model dengan tingkat akurasi terbesar (diwarna biru terang).

5.1.3 Pembangunan Model Analisis Sentimen

*grid search* dilakukan dengan 3 hyperparameter, yaitu weight\_decay dengan rentang 0 sampai 0.3, learning\_rate dengan rentang 1e-5 sampai 5e-5, serta warmup\_steps antara 0, 250, dan 500.

5.1.4. Hasil Model Analisis Sentimen

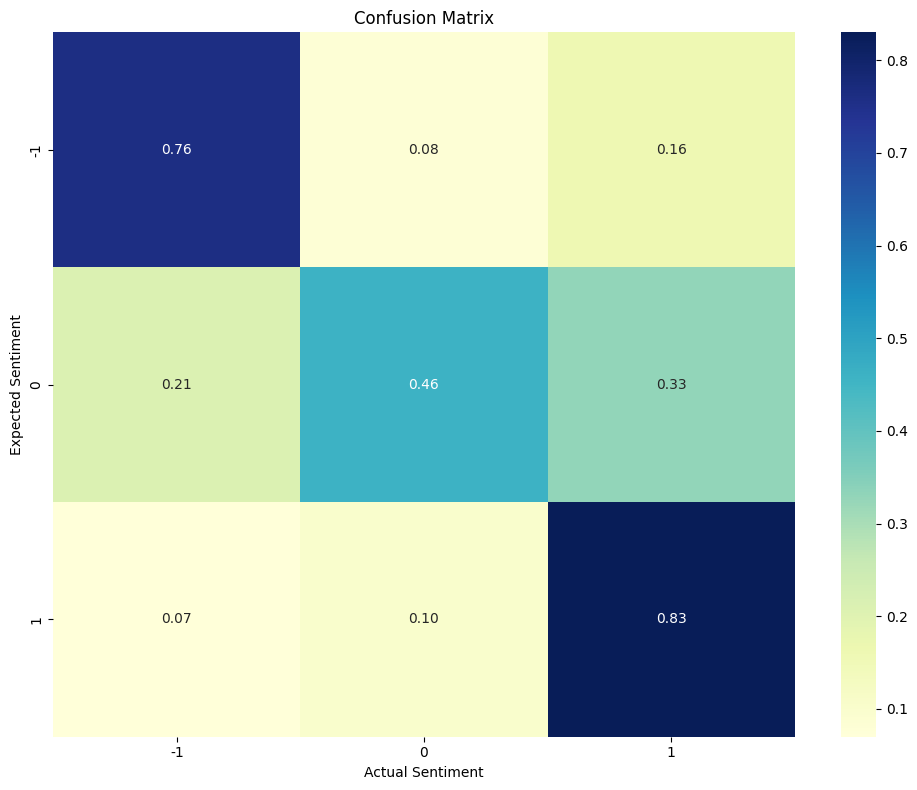
Akurasi paling akurat yang didapatkan setelah melakukan grid search memiliki hyperparameter learning rate sebesar 2.164916560792168e-05, per\_device\_eval\_batch\_size sebesar 32, per\_device\_train\_batch\_size sebesar 16, warmup steps sebesar 500, dan weight decay sebesar 0.1295835055926347. Serta pelatihan dilakukan sebanyak 600 steps (0.57 epoch). Didapatkan hasil evaluasi dan Confusion Matrix pada data uji sebagai berikut

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **eval\_accuracy** | **eval\_f1** | **eval\_loss** | **eval\_precision** | **eval\_recall** |
| 0,676666667 | 0,676594863 | 0,858020842 | 0,678267182 | 0,676666667 |

5.2 Analisis Data Eksplorasi Sentimen Analisis

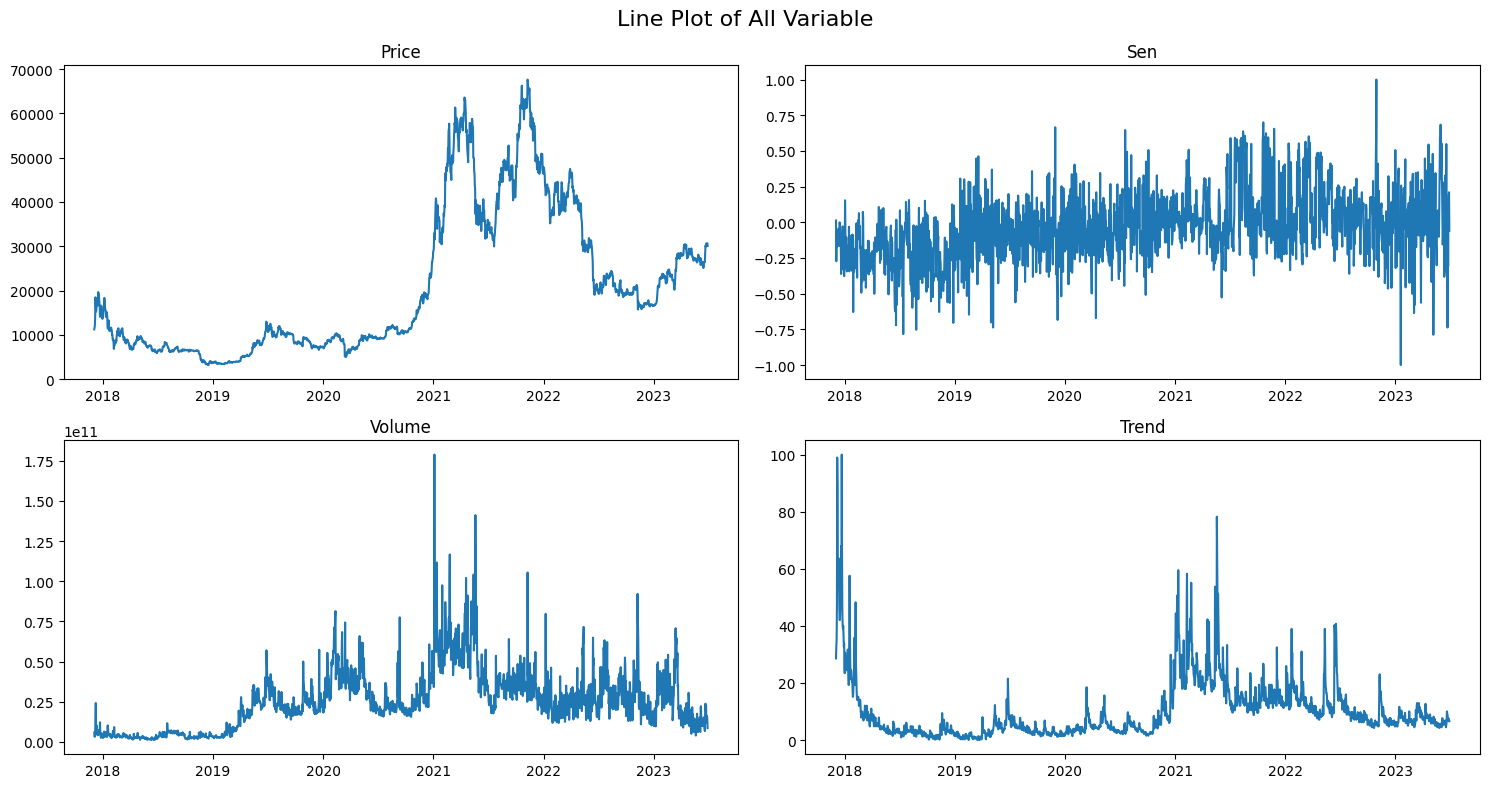
5.2.1. Data Sentimen Analisis

Hasil confusion matrix dari model saat di uji pada data uji adalah sebagai berikut



5.2.2. Data Rangkaian Waktu

Berikut adalah data timeseries dari seluruh variabel, menskalakan data sentimen menjadi -1 hingga 1 dan data Trend menjadi 0 – 100 juga dilakukan untuk memvisualisasikan masing masing sentiment dari negatif, netral, hingga positif. Dan tren dari 0 hingga 100 dalam skala kepopuleran itu sendiri.



5.1.2.9 *Outlier*

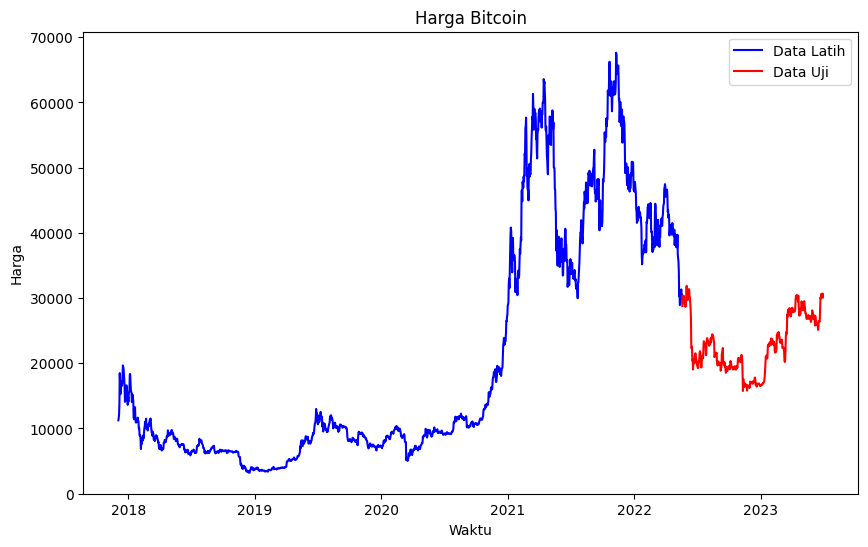
5.1.2.9 *Outlier*

5.1.2.9 *Outlier*

5.1.2.9 *Outlier*

5.3 Peramalan Deret Waktu

Data deret waktu harga bitcoin dibagi menjadi dua, yaitu 80% data latih dan 20% data uji untuk memvalidasi peramalan harga bitcoin.



5.2.1. Pembangunan Model Peramalan Deret Waktu

*grid search* mendapatkan model paling akurat dengan *hyperparameter* input\_chunk\_length sebesar, output\_chunk\_length sebesar, d\_model sebesar, nhead sebesar, num\_encoder\_layers sebesar, num\_decoder\_layers sebesar, dim\_feedforward sebesar, dan dropout sebesar.

5.2.2. Hasil Model Peramalan Deret Waktu

Akurasi paling akurat yang didapatkan setelah melakukan grid search dengan data validasi sebesar dan data uji coba sebesar . Dengan prediksi harga dibandingkan dengan harga asli dalam data uji coba sebagai berikut

\*Gambar grapik time series bang\*

## KESIMPULAN DAN SARAN

Saran

1. Menggunakan model yang memiliki maksimum token lebih tinggi. Dikarenakan X (dahulunya twitter) yang dahulu melimitasi 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.

2. Membandingkan model HelsinkiNLP atau model translasi lainnya dengan Google Translate menggunakan bahasa yang sama dan juga dengan parameter yang sama.

3. Menggunakan arsitektur transformer yang lebih mutahkir. Dikarenakan model yang digunakan pada penelitian ini masih naif, yaitu tidak memanfaatkan hasil dari model itu sendiri untuk menjadi argumen konsiderasi dalam memprediksi hasil selanjutnya.

