# 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 terlimitasi pada dunia perindustrian, tetapi juga berdampingan juga dengan perkembangan teknologi pada dunia perekonomian dengan berkembangnya uang yang bermula dari fisik menjadi digital. *Cryptocurrency* atau mata uang kripto, dapat mengembangkannya lebih lanjut dengan menghilangkan peran bank sehingga suatu sistem dapat bekerja secara mandiri tanpa harus campur tangan pihak tengah (perantara).

Google trends menyatakan bahwa kata kunci “cryptocurrency” mencapai puncak popularitasnya pada bulan Mei 2021. Dengan kepopuleran mata uang kripto di mata masyarakat maka banyak pula investor investor yang melihat warna hijau di cakrawala. Ditambah lagi dengan melihat grafik yang disediakan oleh coinmarketcap kita dapat melihat harga mata uang kripto terpopuler bitcoin mengalami lonjakan sebanyak 405% selama tahun 2020 dan sebesar 161% selama tahun 2021.

Pasar mata uang kripto yang buka setiap saat adalah salah satu aspek menarik bagi para investor, berbeda dengan pasar *Foreign Exchange* (Forex) yang berhasil mencapai volume perdagangan harian sebesar 6.6 triliun dollar pada April 2019. Pasar forex buka setiap saat tapi hanya di hari kerja, hampir sama seperti pasar saham yang juga buka hanya pada hari kerja dan juga memiliki jam jam tertentu. Dengan pasar mata uang kripto yang buka setiap saat ini para investor dapat dengan leluasa menjual serta membeli aset mereka kapanpun.

Harga mata uang kripto yang melonjak serta pasar yang buka setiap saat memang sangat menarik bagi para investor, akan tetapi *volatility* atau harga yang naik turun secara drastis merupakan salah satu alasan investor memilih untuk menjauh dan takut akan masuk ke dalam pasar mata uang kripto. Apalagi harga bitcoin sempat menurun sebesar 19% pada 11 Januari 2021. Maka dari itu sangat dibutuhkan suatu model yang dapat memprediksi harga suatu kripto di masa depan untuk memaksimalkan keuntungan dan meminimalisir kerugian.

Penelitian yang dilakukan oleh Choi dan Varian dengan judul *Predicting the Present with Google Trends* serta Ettredge, dkk dengan judul *using web-based search data to predict macroeconomic statistics* mendapatkan kesimpulan bahwa kepopuleran atas suatu topik memiliki hubungan terhadap penjualan rumah bahkan sampai statistik macroeconomic. Ditambah dengan penelitian yang dilakukan oleh Eray Gemici dan Müslüm Polat mengenai *relationship between price and volume in the bitcoin market* yang menyimpulkan bahwa harga dan volume memiliki korelasi. Berangkat dari ketiga penelitian itu maka kepopuleran suatu mata uang kripto dan volumenya memiliki korelasi dengan harganya itu sendiri. Pada penelitian ini ditambahkan variabel harga serta sentimen sosial media untuk mendapatkan model terbarunya.

Kepopuleran model transformer sendiri pun semakin melonjak bersamaan dengan munculnya Chat GPT, atau Chat Generative Pre-Trained Transformer yang berhasil mengambil perhatian dunia dengan badai. Google trends menyatakan bahwa kata kunci “Transformer Deep Learning” dan “Transformer Model” mulai mengalami lonjakan dari awal tahun 2022 dan masing - masing mencapai puncak popularitas pada bulan Maret dan Juni tahun 2023.

## TINJAUAN PUSTAKA

Penelitian terkait peramalan bitcoin menggunakan berbagai macam teknik telah banyak dilakukan. Seperti penelitian yang dilakukan oleh Nasir, M. A., dkk (2019) yang menggunakan data google trends dan algoritma *machine learning* seperti *Vector Autoregression* serta *Random Forest* untuk meramal pergerakan harga Bitcoin berdasarkan harga dan kepopulerannya. Studi menunjukan bahwa banyak pencarian juga dapat menjadi variabel peramal untuk aset investasi. Bukan hanya dengan google trends, Alghamdi S., dkk (2022) menemukan relasi yang sangat kuat antara harga dengan sentimen dengan *Mean Absolute Error* (MAE) sebesar 0.245, *Mean Square Error* (MSE) 0.2528, dan *Root Mean Squared Error* (RMSE)0.5028. Ditambah dengan Gemici, E. dan Polat, M (2019) yang menyimpulkan bahwa harga dan volume memiliki kointegrasi.

Analisis menggunakan deret waktu multivariat sendiri bukanlah hal yang baru, Mehtab S. dan Sen J (2020) menemukan bahwa dalam memprediksi harga saham penggunaan multivariat terbukti lebih unggul dibandingkan univariat dengan membangun model yang lebih akurat. Hanus, S., dkk (2022) mengaplikasikan multivariat ke 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). Mengaplikasikan lima variabel yaitu harga tutup, harga buka, harga tertinggi, harga terendah, dan volume dari lima mata uang kripto yaitu Bitcoin, Ethereum, Cardano, Tether, serta Binance Coin. Mereka menemukan bahwa Bi-LSTM dan GRU memiliki performa yang serupa dalam rata rata *Mean Absolute Percentage Error* (MAPE) yaitu 0.0465712 untuk Bi-LSTM dan 0.0446512 untuk GRU, sedangkan LSTM memiliki rata rata MAPE sebesar 0.0529916. Walaupun LSTM unggul dalam performa di dataset USDT dan BNB, akan tetapi LSTM memiliki variansi yang lebih besar dibandingkan Bi-LSTM dan GRU

Model Transformer sendiri sudah mulai banyak digunakan bukan hanya untuk menentukan sentiment score dari suatu kalimat, akan tetapi untuk peramalan data deret waktu. Zhao H., dkk (2022) meneliti kapabilitas model transformer dalam meramal bitcoin dan ethereum menggunakan sentiment analysis dan membandingkannya dengan model LSTM. Mereka menemukan bahwa model LSTM lebih unggul dibandingkan model transformer, dengan model transformer dilatih menggunakan data bitcoin memiliki MSE 0.00037, MAPE 0.05816, dan MAE 0.01432. Dibandingkan dengan LSTM yang memiliki MSE 0.00032, MAPE 0.04613, dan MAE 0.01346. Dan yang membuat penelitian ini menarik adalah mereka menemukan bahwa model yang dilatih dengan data bitcoin memiliki memiliki peningkatan dalam memprediksi harga ethereum dibandingkan dengan model yang dilatih dengan data ethereum itu sendiri.

Terinspirasi dari penelitian yang telah disebutkan, penelitian ini diadakan untuk mengeksplorasi lebih jauh mengenai model transformer dalam memprediksi harga bitcoin dengan analisis sentimen serta kepopulerannya menggunakan 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 sentiment 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:



Dimana ✓ = digunakan, dan ✕ = tidak digunakan.

### 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.

### Sentimen Analisis

Bidirectional Encoder Representations from Transformer (BERT) menyediakan banyak model untuk mengklasifikasi sentimen dengan data teks, salah satunya adalah *Bert Base Multilingual Uncased Sentiment* dimana data latih yang digunakan untuk model ini berisi dari enam Bahasa yaitu Inggris, Belanda, Jerman, Perancis, Spanyol, dan Italia. Serta mengklasifikasikan sentiment sebuah teks mejadi skala satu sampai lima

## EKSPERIMEN

## HASIL DAN DISKUSI

## KESIMPULAN DAN SARAN