

PREDIKSI HARGA CRYPTOCURRENCY BERDASARKAN ANALISIS SENTIMEN PADA TWITTER

Rafi Alvanzah, Herardita Cahyaning W., Dea Alifia Maharani S., Maisya Mutiara D. H., Fauzan Hilmy M.

Program Studi Matematika, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,

Universitas Indonesia

Kampus UI Depok, 16424, Indonesia

Abstrak

Dalam penelitian ini, kami menganalisis sentimen *tweet* pada Twitter untuk memprediksi fluktuasi harga dan tren *cryptocurrency* yang disebut Bitcoin (BTC) dan Ethereum (ETH). Kami mengekstrak *tweet* terkait masing-masing *cryptocurrency* dari 1 Januari 2020 hingga 31 Desember 2020 setiap *tweet* kemudian dihitung skor sentimen positif, netral, atau negatif dengan RoBERTa (*A Robustly Optimized BERT Pretraining Approach*) model. Kami kemudian mengklasifikasi *tweet* tersebut menjadi positif/netral/negatif berdasarkan skor sentimen *tweet* tertinggi dan menghitung jumlah kelas *tweet* hariannya. Lalu, kami membuat skor *engagement* harian menggunakan *likes* dan *retweet*. Kami juga mengekstrak volume *tweet* perhari terkait masing-masing *cryptocurrency* dengan periode yang sama. Ketiga fitur ini, di samping penjumlahan hariandari skor sentimen positif, negatif, dan netral dibandingkan dengan ~ 366 data harga harian untuk mengukur korelasinya dan didapat nilai korelasi antara 0,49 dan 0,8. Selanjut tiga fitur dengan korelasi tertinggi digunakan untuk melatih *Extreme Gradient Boosting Regression Tree Model* dan *Extreme Gradient Boosting Classification Tree Model*. Didapat bahwa model prediksi tren harga memiliki akurasi 59% untuk BTC dan 54% untuk ETH dengan skor *precision* 73% untuk BTC dan 61% untuk ETH. Dengan demikian, simpulkan bahwa model ini akan memberikan wawasan yang berguna untuk *trading cryptocurrency* harian. Untuk prediksi perubahan besaran harga memiliki skor RMSE 484,7 untuk BTC dengan besaran perubahan harga antara -1500 dan 2000 dan skor RMSE 16,7 untuk ETH dengan besaran perubahan harga antara -40 dan 40. Dengan hasil ini kami menyimpulkan bahwa sentimen *tweet* dapat digunakan untuk memprediksi harga *cryptocurrency*.

Kata Kunci: Analisis Sentimen, Klasifikasi, Regresi, RoBERTa, XGBoost, Twitter.

Abstract

In this study, we analyzed the sentiment of tweets on Twitter to predict price fluctuations and trends of cryptocurrencies called Bitcoin (BTC) and Ethereum (ETH). We extracted each cryptocurrency's related tweets from January 1, 2020, to December 31, 2020 each tweet and then calculated a positive, neutral, or negative sentiment score with the RoBERTa (A Robustly Optimized BERT Pretraining Approach) model. We then classify those tweets as positive/neutral/negative based on the highest tweet sentiment score and count the number of daily tweet classes. Then, we create a daily engagement score using likes and retweets. We also extract the volume of tweets per day related to each cryptocurrency with the same period. These three features, in addition to the daily summation of positive, negative, and neutral sentiment scores compared to ~ 366 of the daily price data to measure the correlation and obtained correlation values between 0.49 and 0.8. Furthermore, the three features with the highest correlation are used to train the XGBoost Regression (Extreme Gradient Boosting Regression Tree Model) and the XGBoost Classification (Extreme Gradient Boosting Classification Tree Model). It was found that the price trend prediction model has an accuracy of 59% for BTC and 54% for ETH with a precision score of 73% for BTC and 61% for ETH. Thus, conclude that this model will provide useful insights for day cryptocurrency trading. For the prediction of the price change has a score of RMSE 484.7 for BTC with a price change between -1500 and 2000 and a score of RMSE 16.7 for ETH with a price change between -40 and 40. With these results we conclude that tweet sentiment can be used to predict cryptocurrency prices.

Keywords: Sentiment Analysis, Classification, RoBERTa, XGBoost, Twitter.

1 PENDAHULUAN

Blockchain adalah sebuah *database* atau buku besar digital yang terdistribusi dalam jaringan *peer-to-peer* (desentralisasi). *Blockchain* pertama kali dipopulerkan oleh Satoshi Nakamoto dalam makalahnya yang terbit tahun 2008 berjudul *Bitcoin: A Peer-to-Peer Electronic Cash System* [paper bitcoin]. Pada Januari 2009, Nakamoto meluncurkan *source code* Bitcoin pada *SourceForge* [4]. Kemunculan Bitcoin memacu pertumbuhan dan penerapan teknologi *blockchain* di berbagai bidang, salah satunya adalah *cryptocurrency*. *Cryptocurrency* adalah mata uang digital yang dijamin keamanannya dengan kriptografi dan menerapkan teknologi *blockchain* [2].

Penerapan *blockchain* pada *cryptocurrency* mendapat perhatian yang masih dari media dan investor karena fiturnya yang inovatif, yaitu dapat mengatasi masalah *double-spending* tanpa melalui otoritas terpercaya. Dan harganya yang sangat fluktuatif. Pada saat penulisan penelitian dilakukan, Desember 2022, jumlah kapitalisasi pasar *cryptocurrency* mencapai \$844 milyar USD meningkat 341,9% dari \$191 milyar USD pada awal tahun 2020 dari <https://www.coingecko.com/> dan sempat mencapai puncaknya sebanyak \$3 triliun USD pada November 2021 atau menurun 71,9%.

Fluktuasi harga pada *cryptocurrency* ini menyebabkan ketidakpastian bagi para investor dan masyarakat yang ingin menggunakan *cryptocurrency* selain untuk investasi. Meskipun penyebab fluktuasi ini masih diperdebatkan, Peneliti Ladislav Kiroufek menemukan bahwa aset *cryptocurrency* memiliki perilaku yang mirip dengan aset keuangan pada umumnya, yaitu bersifat spekulatif [11]. Saat ini, media sosial telah menjadi alat utama untuk beropini dan bertukar pendapat. Salah satu media sosial yang populer digunakan untuk beropini terkait *cryptocurrency* adalah Twitter. Oleh karena itu, pada penelitian kali ini kami akan menganalisis pengaruh sentimen Twitter terhadap harga *cryptocurrency*.

1.1 Penelitian Terkait

Berikut adalah penelitian terdahulu terkait prediksi harga *cryptocurrency* dengan menggunakan beragam metode. Berdasarkan penelitian oleh McNally menggunakan LSTM dan RNN untuk memprediksi harga Bitcoin didapat akurasi lebih tinggi dengan menggunakan LSTM, yaitu 52.78%. Menurut penelitian ini, variansi data yang tinggi membuatnya sulit untuk mendapat nilai akurasi yang tinggi. Saran dari peneliti adalah perlu dilakukan *dropout* untuk mencegah overfitting dan mengoptimalkan model. Namun, hal itu tidak menjamin perubahan besar pada nilai akurasi.

Judul	Periode data	Jumlah hari	Model-akurasi (%)
McNally, Sean. (2016). Predicting the price of Bitcoin using machine learning	19/08/2013 – 19/07/2016	1065	LSTM – 52.78% RNN – 50.25%
Yan Li, Wei. (2019). Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model	30/12/2016 – 31/08/2018	600	(Precision) CNN – 64% LSTM – 58% CNN-LSTM – 64%
N. Uras and M. Ortu. (2021). Investigation of Blockchain Cryptocurrencies' Price Movements Through Deep Learning: A Comparative Analysis.	08/10/2015 – 03/10/2019	1456	(Precision) SVM – 53.7% CNN – 51.1% LSTM – 48.5%

Tabel 1 Penelitian terdahulu dan model

1.2 Rumusan Masalah

1. Apakah tren pada Twitter memiliki korelasi dengan harga *cryptocurrency*?
2. Apakah tren pada Twitter dapat memprediksi tren harga *cryptocurrency*?
3. Apakah tren pada Twitter dapat memprediksi perubahan harga *cryptocurrency*?

1.3 Batasan Masalah

Sentimen hanya dikumpulkan dari media sosial Twitter karena banyak digunakan oleh pengguna global untuk penyebaran informasi di berbagai aspek kehidupan manusia serta memiliki API yang dapat diakses untuk pengumpulan data dengan mudah. *Cryptocurrency* yang digunakan sebagai objek penelitian adalah Bitcoin dan Ethereum karena termasuk *cryptocurrency* yang populer dan sering dibicarakan oleh masyarakat di Twitter. Periode prediksi harga *cryptocurrency* mulai 1 Januari 2020 sampai 31 Desember 2020.

1.4 Tujuan

1. Melakukan dan mengukur sentimen *tweet* pada Twitter terkait *cryptocurrency*
2. Mengukur korelasi sentimen *tweet* dan volume *tweet* terhadap harga *cryptocurrency*
3. Menentukan tren harga *cryptocurrency* dengan menggunakan analisis sentimen pada Twitter
4. Mengukur perubahan harga *cryptocurrency* dengan menggunakan analisis sentimen pada Twitter

2 TINJAUAN PUSTAKA

2.1 Blockchain

Teknologi *blockchain* adalah *database* rantai blok sistematis yang saling terkait dan berisi riwayat transaksi serta data lainnya yang berkaitan dengan pengguna. *Database* ini diperiksa dan dikelola oleh jaringan global disertai bantuan pengawasan yang dilakukan oleh komunitas luas sehingga tidak ada satu orang pun yang memiliki kekuasaan atas *database* tersebut; tidak ada yang bisa mengubah ataupun menghapus riwayat transaksi. Dengan kata lain, *database* konvensional yang tersentralisasi berada dan ditempatkan di satu server, sementara data yang ada pada *blockchain* didistribusikan ke pengguna *software*. Oleh karena itu, apabila dibandingkan dengan *database* konvensional yang tersentralisasi, informasi pada teknologi *blockchain* tidak dapat dimanipulasi karena strukturnya terdistribusi dan diawasi oleh komunitas seperti yang telah disebutkan sebelumnya.

Sistem ini menyediakan transaksi keuangan yang aman secara kriptografis dan bersifat anonim antara node pengguna jaringan sehingga transaksi yang masuk lebih memungkinkan untuk dievaluasi dan diterima oleh semua pengguna karena aspek transparasinya. Setiap kali seseorang melakukan transaksi, riwayat transaksi akan disimpan di jaringan dan algoritma komputer digunakan untuk menentukan apakah transaksi tersebut autentik atau tidak. Setelah diverifikasi, transaksi baru ini ditautkan dengan transaksi sebelumnya yang berujung pada pembentukan 'rantai' transaksi. Rantai inilah yang selanjutnya disebut sebagai *blockchain*.

Blockchain adalah teknologi terobosan yang mendapat banyak perhatian karena kegunaannya dalam *cryptocurrency* digital yang dapat digunakan untuk lebih dari sekadar transaksi; tetapi juga dapat sebagai daftar aset dan kepemilikan. Teknologi *blockchain* sangat menarik untuk digali pengaplikasiannya, tetapi juga sangat rumit dan sifatnya sangat transdisipliner. Pemahaman mendasar mengenai elemen ekonomi, ilmu komputer, dan *cryptography* dibutuhkan untuk memahami teknologi ini secara keseluruhan. Meskipun *blockchain* rentan terhadap sejumlah masalah keamanan, tetapi terdapat peluang yang baik bagi *blockchain* untuk dikembangkan penggunaannya dalam sistem transaksi *online*.

2.2 Cryptocurrency

Cryptocurrency masuk ke dalam kelompok mata uang virtual yang dapat dipertimbangkan sebagai media pertukaran digital. Cryptocurrency bekerja berdasarkan prinsip-prinsip kriptografi yang memungkinkan kinerja transaksi ekonomi yang aman, terdesentralisasi, dan terdistribusi. Cryptocurrency menggabungkan konsep kriptografi dengan uang virtual elektronik. Aturan dasar cryptocurrency adalah bahwa tidak ada orang (atau organisasi) yang diizinkan untuk mempercepat atau mengeksploitasi produksi mata uang tertentu karena sistem cryptocurrency secara keseluruhan telah diatur sedemikian rupa sehingga hanya menghasilkan jumlah cryptocurrency tertentu yang telah ditentukan. Nilai yang ditentukan sebelumnya dan terkenal mengontrol tingkat produksi. Jaringan komputasi *peer-to-peer* yang digunakan oleh cryptocurrency memungkinkan pertukaran unit cryptocurrency (yang juga dikenal sebagai koin) antarklien dapat dilakukan tanpa biaya.

2.3 Analisis Sentimen

Analisis sentimen merupakan suatu metode yang digunakan untuk mengidentifikasi bagaimana sentimen yang terkandung dalam suatu teks opini. Analisis sentimen dapat digunakan untuk mendapatkan persentase sentimen positif dan sentimen negatif terhadap seseorang, perusahaan, institusi, produk atau pada sebuah kondisi

tertentu. Nilai dari analisis sentimen digunakan untuk mengategorikan teks tersebut sebagai sentimen positif, sentimen negatif, ataupun sentimen netral.

2.4 Twitter Scraping

Tahap pertama dan paling mendasar dalam banyak bidang analisis data, seperti *natural language processing*, analisis sentimen, dan *machine learning*, adalah dengan mengambil informasi, yang salah satunya bisa didapatkan dari jejaring sosial. Dalam hal ini, proyek rancangan analisis data mengandalkan pengumpulan data yang telah ada sebelumnya untuk menghasilkan prediksi yang lebih akurat. Untuk memberikan hasil yang relevan, penting untuk mengumpulkan data yang tepat untuk percobaan serta pengujian algoritma pada bidang sains data. Platform jejaring sosial *online* digunakan untuk mengekstraksi data pengguna dan mengubah data tersebut menjadi input yang dapat dibaca mesin, yang kemudian digunakan oleh berbagai algoritma untuk menghasilkan output yang diharapkan dapat bersifat prediktif.

Peneliti data telah mulai menggunakan Twitter sebagai pendekatan ilmiah karena Twitter dapat memfasilitasi proses pengumpulan data dalam jumlah besar yang dapat dilakukan dalam waktu singkat. Twitter dengan cepat menjadi jejaring sosial pilihan untuk pengumpulan data karena pengguna dapat memposting pesan singkat, juga dikenal sebagai *tweet*, secara real-time.

Twitter Scraping adalah teknik untuk mengekstraksi data dari Twitter dan menyimpannya ke *file* sistem atau basis data untuk dilakukan analisis data. Banyak metode scraping dari Twitter yang dapat dilakukan, yaitu menggunakan, *tweepy*, *twint*, dan *snsrape*. Pada proyek ini, digunakan module *snsrape* pada jupyter notebook. Berbeda dengan metode pada umumnya, module *snsrape* tidak menggunakan API dalam pengambilan data *tweet*. Selain itu, dengan *snsrape*, data yang diekstraksi dapat di spesifikasi berdasarkan jumlah like, komen, repost, bahasa yang digunakan dalam *tweet*, hingga tanggal suatu *tweet* diunggah. *Snsrape* juga tidak memiliki batasan maksimal *tweet* dalam pengambilan data sehingga membuat metode ini lebih efektif dibandingkan metode lain.

2.5 RoBERTa

RoBERTa (A Robustly Optimized BERT Pretraining Approach) adalah suatu model transformasi yang telah dilatih sebelumnya pada data besar berbahasa Inggris dengan self-supervised learning. Model ini dilatih hanya dengan teks mentah tanpa label apa pun sehingga dapat dilatih menggunakan banyak data yang tersedia untuk umum dan kemudian dengan proses otomatis untuk menghasilkan input dan label dari teks tersebut.

Model klasifikasi yang merupakan perluasan dari BERT ini dilatih dengan objektif *masked language modeling* (MLM). Dari suatu kalimat yang diambil, model secara acak akan melakukan *mask* 15% dari kata-kata dalam input dan menjalankan seluruh kalimat tersebut dengan model yang kemudian akan memprediksi kata-kata yang di-*mask*. Dengan begitu, model akan mempelajari representasi dalam bahasa Inggris yang kemudian dapat digunakan untuk mengekstraksi fitur yang berguna untuk *downstream tasks*. RoBERTa telah dilatih untuk pemrosesan *Natural Language Processing* (NLP) dengan lebih dari 124 juta *tweet* sehingga sangat baik digunakan untuk klasifikasi analisis sentimen pada data *tweet*.

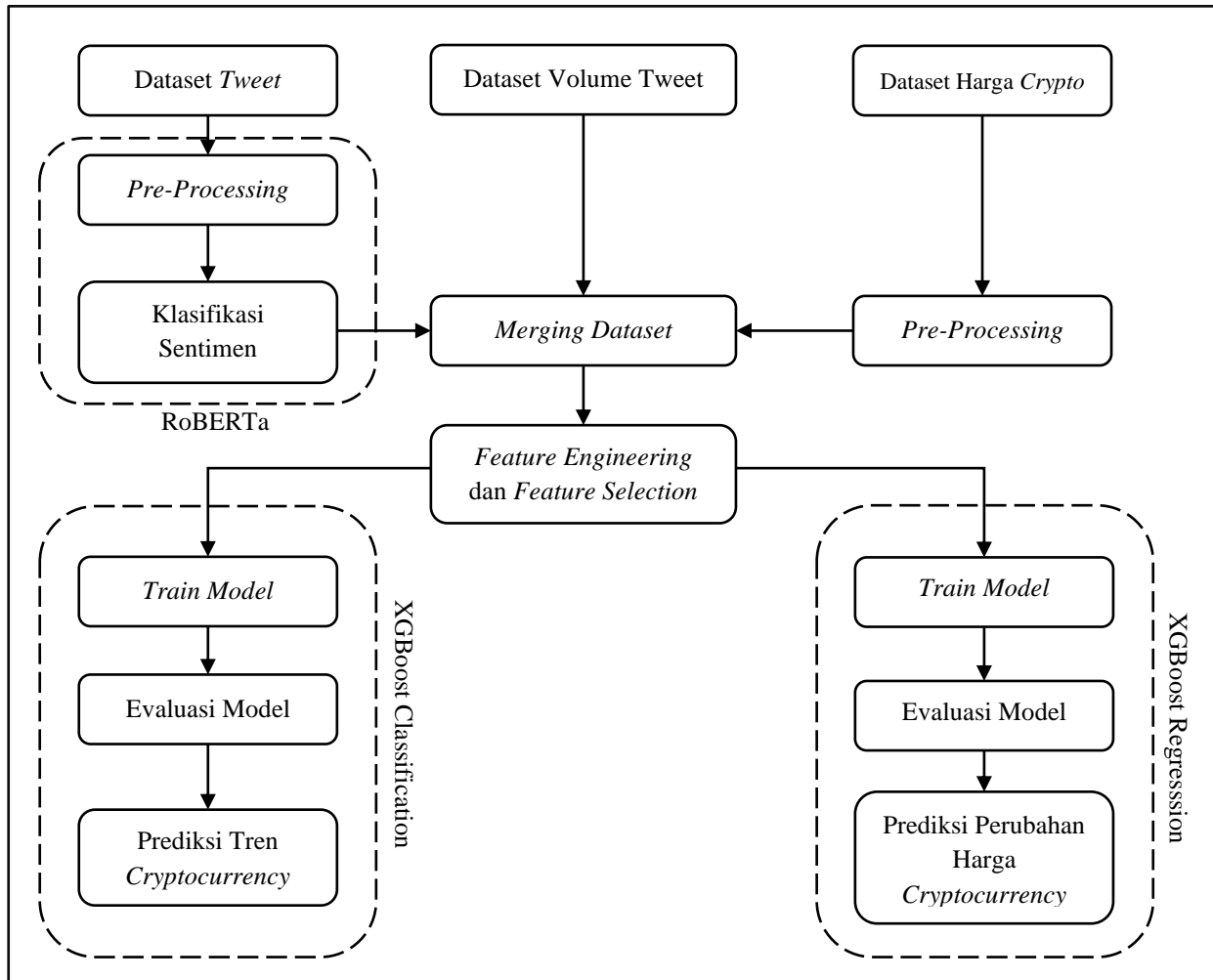
2.6 XGBoost

Algoritma *gradient boosting* dikembangkan untuk mampu menghasilkan tingkat prediktif yang sangat tinggi. Namun, karena kebutuhan untuk menghasilkan satu *decision tree* pada satu waktu untuk mengurangi kesalahan semua *tree* sebelumnya yang ada dalam model, adaptasi dari *gradient boosting* menjadi terbatas. Akibatnya, untuk menguji atau melatih model yang ‘kecil’ pun akan sangat membutuhkan banyak waktu. *Gradient boosting* kemudian dilakukan secara berbeda karena dikembangkannya metode baru yang disebut eXtreme Gradient Boosting (XGBoost).

Metode XGBoost (eXtreme Gradient Boosting) merupakan pengembangan dari *gradient boosting*, yaitu algoritma yang dapat menemukan solusi yang optimal untuk berbagai masalah, khususnya pada regresi, klasifikasi dan ranking. Konsep dasar dari algoritma ini adalah menyesuaikan parameter pembelajaran secara berulang untuk menurunkan loss function. Pada XGBoost, data disimpan tersusun untuk menghemat waktu pencarian sehingga berhasil mengurangi jumlah waktu yang dibutuhkan untuk melatih model, yang berdampak pada peningkatan kinerja model tersebut. Fitur perangkat lunak dan perangkat keras XGBoost dapat menjadi kombinasi yang ideal untuk meningkatkan metode *boosting* saat ini secara akurat dan cepat. Selain itu, perlu digarisbawahi pula bahwa

XGBoost telah menunjukkan keunggulan daripada metode lain dalam masalah klasifikasi, regresi, dan pemodelan prediktif menggunakan dataset terstruktur atau tabular.

3 METODE PENELITIAN



Gambar 1

3.1 Data Preparation dan Pre-Processing

Preprocessing merupakan tahap awal dalam pemrosesan datasets. Data yang digunakan terkadang terdapat berbagai permasalahan yang dapat mengganggu hasil seperti diantaranya adalah *missing value*, *data redundant*, *outliers*, ataupun format data yang tidak sesuai dengan sistem. Oleh karenanya untuk mengatasi permasalahan tersebut dibutuhkan tahap *Preprocessing*.

3.2 Cleaning Dataset Twitter

Proses *cleaning* yang dilakukan meliputi mengubah *mention* '@username_pengguna' menjadi '@user', mengubah URL atau *link* apapun menjadi 'http'. Hal ini dilakukan karena model tidak menganalisis nama pengguna dan konten situs web.

3.3 Tokenizing Dataset Twitter

Proses ini memecah kalimat dan memotong kata menjadi *term-term* berdasarkan spasi yang disebut token. Pada proses ini digunakan model *Robustly Optimized BERT Pre-training Approach (RoBERTa)* untuk memasang kata dengan token tertentu.

3.4 Pengukuran Skor Sentimen dengan Roberta

Pada tahapan ini setiap *tweet* mendapat perlakuan pengukurnya skor sentimen atau *scoring* secara otomatis dengan menggunakan model *roberta-base*. Tujuan dari proses ini adalah untuk mendapatkan label/sentiment dari setiap *term/kata* yang ada pada *tweet*. Keluaran dari proses ini adalah skor *roberta positif*, *negative*, dan *netral* untuk setiap *tweet*.

3.5 Klasifikasi Sentimen Analisis

Berdasarkan hasil *tokenizing* dan pengukuran skor sentimen dengan model *roberta-base*, suatu *tweet* dapat diklasifikasikan menjadi tiga kelas yaitu, 0 untuk *tweet* netral, 1 untuk *tweet* positif, dan -1 untuk *tweet* negatif.

3.6 Preprocessing Harga Cryptocurrency

Pada tahapan ini akan dibuat fitur baru yang mengukur perbedaan harga antara hari-H dan hari-H+1 yang dinamakan Price Diff dan mengukur tren harga *cryptocurrency* antara hari-H dan hari-H+1 yang dinamakan +1D.

3.7 Merging Dataset

Dari ketiga dataset yang telah di *preprocessing* (Dataset *Tweet* dan Dataset Harga *Crypto*) dan Dataset Volume *Tweet* lalu di gabungkan menjadi satu dataset. Dataset *Tweet* yang sudah di *preprocessing* di jumlahkan agregat perhari untuk skor *roberta positif*, *negative*, *netral*, dan klasifikasi sentimen *tweet* yang dibagi menjadi antara *tweet* dengan kelas negatif dan positif. Dibentuk pula kolom baru yang dinamakan Impact yang disusun dengan Equation 1 untuk mengukur *engagement* dari *tweet* lalu digabungkan dengan dataset Harga *Crypto* dan Dataset Volume *Tweet*. Sehingga didapat dataset sesuai dengan Tabel 2.

Date	Impact	Classes_-1	Classes_1	roberta negatif	roberta netral	roberta positif	Tweet volume	+1D BTC/ETH	Price Diff	BTC/ETH close
------	--------	------------	-----------	--------------------	-------------------	--------------------	-----------------	----------------	------------	------------------

Tabel 2 Merge dataset

$$Impact = 2 \times Retweet \times classes + likes \times classes$$

Equation 1

3.8 Fitur dan Label

Berikut fitur-fitur yang digunakan dalam proses prediksi dibawah:

1. Date: Waktu saat *tweet* di post dalam Twitter
2. Impact: *Engagement* dari *tweet*
3. classes_-1: Jumlah kelas *tweet* negatif pada waktu tertentu
4. classes_1: Jumlah kelas *tweet* positif pada waktu tertentu
5. roberta_neg: Skor sentimen negatif dari *tweet* tertentu
6. roberta_neu: Skor sentimen neutral dari *tweet* tertentu
7. roberta_pos: Skor sentimen positif dari *tweet* tertentu
8. tweet_volume: Jumlah *tweet* yang di-mining di waktu tertentu
9. +1D BTC/ETH: Menunjukkan status kenaikan harga BTC/ETH
10. Price Diff BTC/ETH: Menunjukkan perbedaan harga BTC/ETH dari waktu sebelumnya
11. BTC/ETH Volume: Jumlah transaksi BTC/ETH
12. BTC/ETH_close: Harga BTC/ETH terakhir di waktu tertentu

3.9 Koefisien Korelasi

Koefisien korelasi bertujuan untuk mengukur seberapa kuat hubungan linier antara dua variabel. Ini berguna karena memungkinkan kita untuk mengetahui seberapa erat hubungan antara dua variabel dan seberapa besar pengaruh satu variabel terhadap yang lain. Misalnya, jika kita ingin tahu seberapa besar pengaruh pendapatan seseorang terhadap tingkat kebahagiaannya, kita dapat menggunakan koefisien korelasi untuk menentukan seberapa kuat hubungan antara kedua variabel tersebut.

Pada penelitian ini digunakan korelasi pearson. Nilai berkisar antara -1 sampai 1, dengan nilai yang mendekati 1 menunjukkan korelasi positif yang kuat, nilai yang mendekati 0 menunjukkan tidak ada korelasi, dan nilai yang mendekati -1 menunjukkan korelasi negatif yang kuat.

3.10 Data Train dan Data Test

Pembagian data latih dan data uji yang akan dilakukan pada penelitian ini adalah dengan rasio 80:20, di mana 80% (data dengan periode 1 Januari 2020 - 30 September 2020) digunakan sebagai data latih dan 20% (data dengan periode 1 Oktober 2020 - 31 Desember 2020) digunakan sebagai data uji.

3.11 Evaluasi Model

Tujuan evaluasi model adalah untuk mengukur seberapa baik model tersebut bekerja pada data yang belum dilihat sebelumnya, yang biasanya disebut data uji atau data validasi. Ini bertujuan untuk mengetahui apakah model tersebut generalisasi dengan baik, atau apakah hanya bekerja dengan baik pada data yang telah digunakan untuk melatihnya (data pelatihan). Evaluasi performa model yang baik sangat penting untuk memastikan bahwa model tersebut dapat digunakan secara efektif untuk memprediksi output pada data baru di dunia nyata.

Pada penelitian ini digunakan metrik, seperti *confusion matrix*, *accuracy*, *precision*, *recall*, *F1-core*, dan *root mean squared error* (RMSE).

3.11.1 Confusion Matrix

Confusion matrix merupakan suatu metode yang digunakan untuk mengukur kinerja model klasifikasi pada suatu data uji yang sebenarnya diketahui nilainya [5]. Metode pengukuran yang sering disebut sebagai *error matrix* ini memberikan informasi perbandingan klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya dengan tujuan mengetahui seberapa baik performa model yang digunakan.

Confusion matrix pada dasarnya disusun oleh nilai aktual dan nilai prediksi. Terdapat empat istilah sebagai representasi dari hasil proses klasifikasi, yaitu:

1. *True Positive (TP)*: Data yang bernilai positif dan diprediksi benar sebagai positif.
2. *False Negative (FN)*: Data yang bernilai positif, tetapi diprediksi sebagai negatif.
3. *False Positive (FP)*: Data yang bernilai negatif, tetapi diprediksi sebagai positif.
4. *True Negative (TN)*: Data yang bernilai negatif dan diprediksi benar sebagai negatif.

		Harga Prediksi	
		Naik (Output = 1)	Turun (Output = 0)
Harga Ril	Naik (Output = 1)	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	Turun (Output = 0)	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Tabel 3 Confusion matrix

3.11.2 Accuracy

Accuracy menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Maka, *accuracy* merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data. Dengan kata lain, *accuracy* merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya) [6]. Nilai *accuracy* dapat diperoleh dengan persamaan berikut:

$$Accuracy = \frac{\sum(True\ Positive) + \sum(True\ Negative)}{\sum(Total\ Populasi)}$$

Equation 2

3.11.3 Recall

Recall menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Maka, *recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif [6]. Nilai *recall* dapat diperoleh dengan persamaan berikut:

$$Recall = \frac{\sum(True\ Positive)}{\sum(False\ Negative) + \sum(True\ Positive)}$$

Equation 3

3.11.4 Precision

Precision menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Maka, *precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah di prediksi dengan benar, berapa banyak data yang benar-benar positif [6]. Nilai *precision* dapat diperoleh dengan persamaan berikut:

$$Precision = \frac{\sum(True\ Positive)}{\sum(False\ Positive) + \sum(True\ Positive)}$$

Equation 4

3.11.5 F1-Score

F1-Score merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan [7]. Nilai terbaik *F1-Score* adalah 1.0 dan nilai terburuknya adalah 0. Secara representasi, jika *F1-Score* punya skor yang baik mengindikasikan bahwa model klasifikasi kita punya *precision* dan *recall* yang baik [8]. Nilai *F1-Score* dapat diperoleh dengan persamaan berikut:

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

Equation 5

4 HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Dataset yang digunakan merupakan gabungan antar tiga dataset, yaitu dataset *tweet* dari API Twitter, dataset volume *tweet* dari <https://bitinfocharts.com> dan dataset harga *cryptocurrency* (*Bitcoin* dan *Ethereum*) dari Kaggle. Dataset yang dikumpulkan memiliki periode waktu yang berbeda-beda, sehingga saat melakukan *merging* dataset, data harga *cryptocurrency* dan data volume *tweet* dicocokkan berdasarkan data hasil analisis sentimen *tweet*. Setelah dilakukan *merging*, dataset memiliki total 366 sampel dari 1 Januari 2020 hingga 31 Desember 2020.

4.1.1 Dataset Tweet

Dataset *Tweet* diambil dari periode 1 Januari 2020 hingga 3 Februari 2021 dengan *tweet* berbahasa Inggris, dan minimal 100 likes per *tweet*. Kata kunci untuk pencarian yang digunakan adalah 'btc', 'bitcoin', '\$btc', 'eth', 'ether', 'ethereum', '\$eth', 'cryptocurrency', 'crypto', dan 'blockchain'. Keluaran dari sistem ini berupa 108.806 sampel data *tweet* yang terdiri dari fitur *date*, *retweet*, *reply*, *likes*, dan *tweet content*. Gambar 2 dan Gambar 3 adalah sampel *tweet* dari API Twitter yang diperoleh menggunakan metode data mining dan modul SNScrape.

Date	Retweet	Reply	Likes	Tweet
2020-05-27 15:08:34+00:00	81	13	247	We've also released the new Enjin SDK for @jav...
2020-07-18 17:20:11+00:00	10	3	261	You ever find someone hot and then find out th...
2021-01-18 16:53:36+00:00	22	12	183	15,000 views 24-hrs\n\nOrange Pill [OP24] - Bi...
2020-12-08 22:03:49+00:00	46	9	173	Today is "Thank You Tuesday!" Every Tuesday we...
2020-07-01 21:28:26+00:00	22	12	226	Boom!\n\n@APompliano just guested on @billburr...

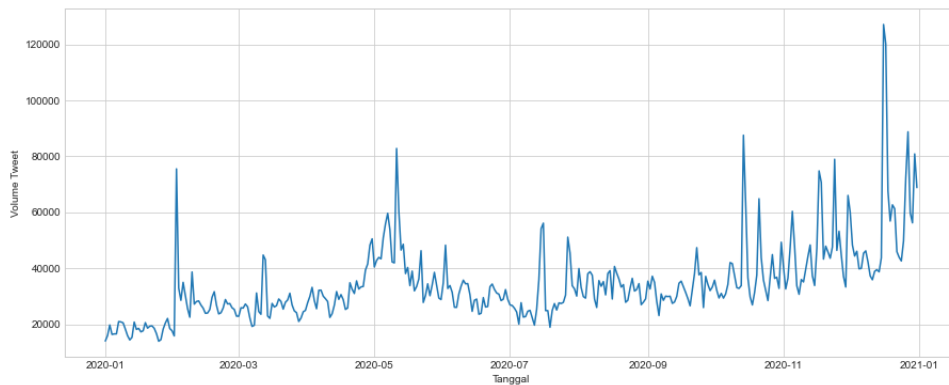
Gambar 2 Sampel tweet terkait BTC

Date	Retweet	Reply	Likes	Tweet
2020-06-25 12:56:33+00:00	22	66	364	I'm expecting a HEAVY retracement for both sto...
2020-04-05 01:06:13+00:00	8	2	406	I have nothing to promote except my boii crypt...
2020-10-18 18:22:01+00:00	26	4	216	Absolutely outstanding listen from @tferriss a...
2020-05-31 16:48:04+00:00	62	54	1060	Before crypto, I was psyched if I made 10% on ...
2020-09-13 04:43:00+00:00	15	13	256	The innovation and competition happening in #c...

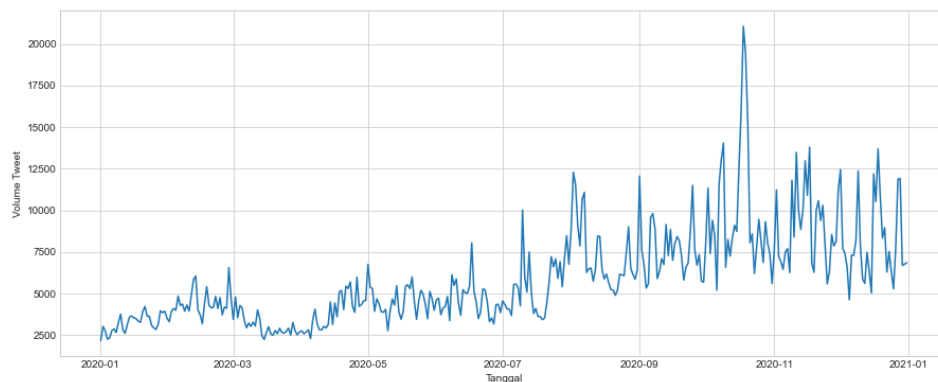
Gambar 3 Sampel tweet terkait ETH

4.1.2 Dataset Volume Tweet

Dataset Volume Tweet diambil dari periode 16 Maret 2016 hingga 26 Desember 2022. Keluaran dari sistem ini berupa 2477 data terkait jumlah *tweet* harian terkait BTC dan ETH. Dari periode keseluruhan dataset, hanya diambil sampel data dari tanggal 1 Januari 2020 sampai 31 Desember 2020. Gambar 5 dan Gambar 6 adalah grafik dari Volume tweet perhari terkait BTC dan ETH. Dataset diperoleh menggunakan metode data mining modul BeautifulSoup4 dari situs <https://bitinfocharts.com>.



Gambar 6 Volume tweet terkait BTC perhari



Gambar 5 Volume tweet terkait ETH perhari

	Date	Bictoin - Tweets
2100	2020/01/08	20820
2101	2020/01/09	20500
2102	2020/01/10	18266
2103	2020/01/11	15900
2104	2020/01/12	14383

Gambar 4 Sampel dataset volume tweet terkait BTC

4.1.3 Dataset Harga *Bitcoin* dan *Ethereum*

Dataset diperoleh dari situs Kaggle berjudul Cryptocurrency Price. Dataset berisi harga harian *Bitcoin* dan *Ethereum*, dimana dalam satu hari tersebut harga terdiri dari harga pembukaan dan penutupan, harga tertinggi dan terendah, dan volume *Crypto* yang diperjualkan pada dari periode 29 April 2013 hingga 6 Juli 2021 untuk BTC dan periode 8 Agustus 2015 hingga 6 Juli 2021 untuk harga ETH. Dari periode keseluruhan dataset, hanya diambil sampel data dari tanggal 1 Januari 2020 sampai 31 Desember 2020. Gambar 7 dan Gambar 8 menunjukkan pertumbuhan nilai tukar BTC dan ETH terhadap USD.

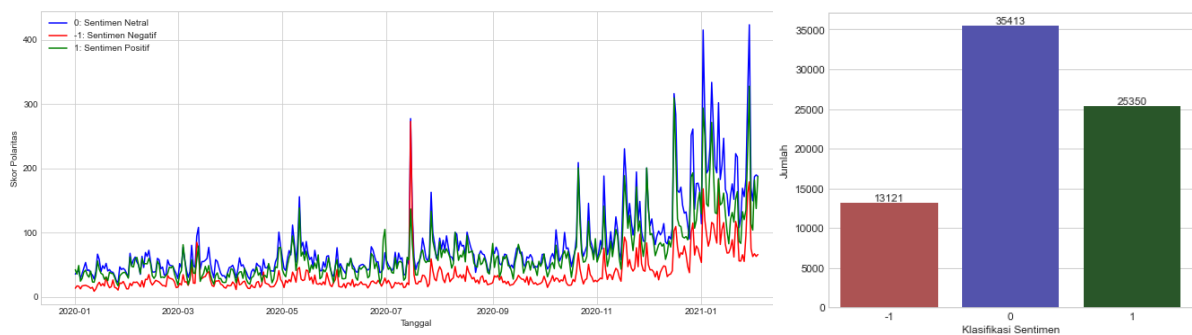
	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
Date										
2020-01-01 23:59:59	2439	Bitcoin	BTC	2020-01-01 23:59:59	7254.330611	7174.944153	7194.891971	7200.174393	1.856566e+10	1.305808e+11
2020-01-02 23:59:59	2440	Bitcoin	BTC	2020-01-02 23:59:59	7212.155253	6935.269972	7202.551122	6985.470001	2.080208e+10	1.266994e+11
2020-01-03 23:59:59	2441	Bitcoin	BTC	2020-01-03 23:59:59	7413.715099	6914.995908	6984.428612	7344.884183	2.811148e+10	1.332334e+11
2020-01-04 23:59:59	2442	Bitcoin	BTC	2020-01-04 23:59:59	7427.385794	7309.514012	7345.375275	7410.656566	1.844427e+10	1.344425e+11
2020-01-05 23:59:59	2443	Bitcoin	BTC	2020-01-05 23:59:59	7544.496872	7400.535561	7410.451694	7411.317327	1.972507e+10	1.344695e+11

Gambar 7 Sampel Dataset Harga *Bitcoin*

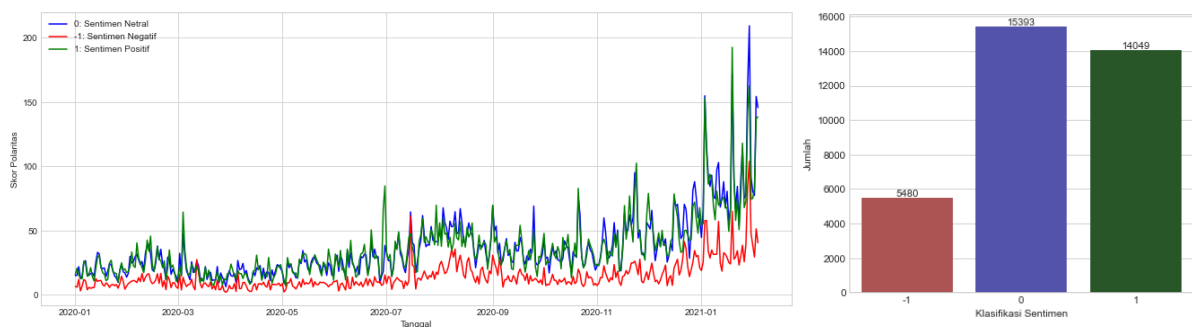
	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
Date										
2020-01-01 23:59:59	1608	Ethereum	ETH	2020-01-01 23:59:59	132.835357	129.198285	129.630656	130.802008	7.935230e+09	1.427106e+10
2020-01-02 23:59:59	1609	Ethereum	ETH	2020-01-02 23:59:59	130.820038	126.954911	130.820038	127.410182	8.032709e+09	1.390251e+10
2020-01-03 23:59:59	1610	Ethereum	ETH	2020-01-03 23:59:59	134.554014	126.490020	127.411261	134.171712	1.047685e+10	1.464213e+10
2020-01-04 23:59:59	1611	Ethereum	ETH	2020-01-04 23:59:59	136.052714	133.040551	134.168523	135.069371	7.430905e+09	1.474195e+10
2020-01-05 23:59:59	1612	Ethereum	ETH	2020-01-05 23:59:59	139.410202	135.045629	135.072094	136.276777	7.526675e+09	1.487557e+10

Gambar 8 Sampel Dataset Harga *Ethereum*

4.2 Klasifikasi Sentimen



Gambar 9 Analisis sentimen tweet terkait *BTC*



Gambar 10 Analisis sentimen tweet terkait *ETH*

Date	Tweets	Roberta_neg	Roberta_neu	Roberta_pos	Classes
2020-01-01 18:23:45+00:00	@user @user Shrek 5: Invest all your mon...	0.035846	0.865769	0.098386	0
2020-01-01 02:20:38+00:00	Bitcoin the worst decision i made this decade	0.914383	0.077082	0.008533	-1
2020-01-01 15:12:05+00:00	BREAKING: Bitcoin's price just set a new...	0.002858	0.214380	0.782763	1
2020-01-01 02:12:10+00:00	\$BTC\n\nI'm pretty sure we're gonna bouc...	0.002113	0.070673	0.927214	1
2020-01-01 23:58:54+00:00	I bought my wife some Bitcoin to help...	0.514108	0.369716	0.116176	-1

Tabel 4 Klasifikasi sentimen tweet

Dataset *tweet* yang sebelumnya telah diperoleh dari proses data *mining*, lalu diukur skor polaritasnya dengan RoBERTa yang telah di-*tuning* secara khusus untuk mendeteksi sentimen dari *tweet*. Skor polaritas yang didapat untuk setiap kategori sentimen (negatif, netral, positif) yang tertera pada Roberta_neg, Roberta_neu, dan Roberta_pos lalu dibandingkan antara satu sama lain dan skor tertinggi menjadi klasifikasi untuk *tweet* tersebut sentimen yang tertara pada Classes. Tabel 4 menunjukkan sampel klasifikasi sentimen yang dilakukan terhadap *tweet*. Gambar 9 dan Gambar 10 menunjukkan jumlah agregat sentimen *tweet* perharinya dan jumlah total hasil klasifikasi *tweet*.

4.3 Merging Dataset dan Features Engineering

Setelah dilakukan *preprocessing* pada Dataset Tweet sehingga didapat dataset dengan fitur sesuai dengan Tabel 4, akan dilakukan *grouping* perhari dengan penjumlahan agregat dari skor sentimen (data numerik) dan klasifikasi sentimen (data kategorik) sehingga didapat skor sentimen dan jumlah klasifikasi sentimen harian seperti pada Tabel 5. Dibentuk pula fitur baru bernama Impact dengan Equation 1 dari dataset Tweet.

Untuk Dataset Harga *Crypto* Gambar 7 dan Gambar 8 akan dibentuk pula fitur baru yang bernama +1D BTC/ETH dan Price Diff BTC/ETH untuk melihat perubahan harga dan tren harga *crypto* di keesokan harinya. Akan dihapus pula fitur-fitur yang tidak digunakan seperti High, Low, Open, SNo, Volume, dan Marketcap sehingga didapat Tabel 6 (serupa untuk ETH).

Lalu ketiga dataset Tabel 5, Tabel 6, dan Dataset Volume *Tweet* digabung menjadi satu dataset sehingga bisa menjadi *input* untuk model XGBoost sehingga didapat Tabel 7.

Date	Impact	Roberta_neg	Roberta_neu	Roberta_pos	classes_-1	classes_1
2020-01-01	4255	6.701138	15.148669	16.150193	6	14
2020-01-02	9203	6.006404	15.207750	20.785846	4	21
2020-01-03	1228	12.083549	21.996594	15.919856	12	15

Tabel 5 Dataset Tweet

Date	+1D BTC	Price Diff BTC	BTC_close
2020-01-01	0	-214.704392	7200.174393
2020-01-02	1	359.414183	6985.470001
2020-01-03	1	65.772383	7344.884183

Tabel 6 Dataset Harga BTC

Date	Impact	Roberta_neg	Roberta_neu	Roberta_pos	classes_-1	classes_1	tweet_volume	+1D BTC	Price Diff BTC	BTC_close	+1D BTC
2020-01-01	4255	6.70	15.14	16.15	6	14	2161.0	0	-214.70	7200.17	0
2020-01-02	9203	6.00	15.20	20.78	4	21	3019.0	1	359.41	6985.47	1
2020-01-03	1228	12.08	21.99	15.91	12	15	2750.0	1	65.77	7344.88	1

Tabel 7 Merge Dataset

4.4 Korelasi Fitur

Hubungan antara setiap fitur dengan harga BTC berdasarkan korelasi pearson dapat dilihat pada Tabel 8. Dapat dilihat bahwa korelasi fitur dengan harga BTC memiliki rentang nilai antara 0.49 sampai 0.71 hal ini menandakan bahwa fitur-fitur tersebut memiliki hubungan linear (korelasi) yang cukup kuat dengan harga BTC. Fitur yang memiliki korelasi paling kuat dengan harga BTC adalah Roberta_neu, Roberta_pos, dan Classes_1 untuk itu fitur tersebut yang akan digunakan dalam memprediksi tren dan perubahan harga BTC.

	Koefisien Korelasi
Impact	0.54
Classes_-1	0.49
Classes_1	0.7
Roberta_neg	0.58
Roberta_neu	0.69
Roberta_pos	0.71
tweet volume	0.6

Tabel 8 Korelasi Fitur BTC

Hubungan antara setiap fitur dengan harga ETH berdasarkan korelasi pearson dapat dilihat pada Tabel 9. Dapat dilihat bahwa korelasi fitur dengan harga ETH memiliki rentang nilai antara 0.52 sampai 0.83 hal ini menandakan bahwa fitur-fitur tersebut memiliki hubungan linear (korelasi) yang cukup kuat dengan harga ETH. Fitur yang memiliki korelasi paling kuat dengan harga ETH adalah Roberta_neu, Roberta_pos, dan tweet_volume untuk itu fitur tersebut yang akan digunakan dalam memprediksi tren dan perubahan harga ETH.

	Koefisien Korelasi
Impact	0.52
Classes_-1	0.53
Classes_1	0.71
Roberta_neg	0.63
Roberta_neu	0.76
Roberta_pos	0.73
tweet volume	0.83

Tabel 9 Korelasi Fitur ETH

4.5 Model Prediksi Tren Harga Cryptocurrency

Untuk memprediksi tren harga *cryptocurrency* dipilih tiga fitur dengan korelasi tertinggi dari Tabel 8 dan Tabel 9 sebagai input dan +1D BTC/ETH sebagai variabel target yang dapat direpresentasikan sebagai masalah klasifikasi biner dengan label 0 menandakan turunnya harga di keesokan harinya dan label 1 menandakan kenaikan harga dikeesokan harinya. Agar performa model menjadi lebih baik dilakukan pula *tuning* dengan 3-fold *cross validation* sebanyak 500 iterasi. Didapat model dengan performa seperti pada Tabel 12, dari hasil tersebut didapat bahwa model dapat memprediksi tren kenaikan harga *cryptocurrency* dengan cukup baik jika dilihat dari nilai *precision* yang cukup tinggi, yaitu 0.73 untuk BTC dan 0.61 untuk ETH dan F1-score yang cukup tinggi pula, yaitu 0.66 untuk BTC dan 0.64 untuk ETH. Namun, model cukup kesulitan untuk memprediksi tren penurunan harga *cryptocurrency* jika dilihat dari nilai akurasi yang menurun jika dibandingkan dengan *precision* dan F1-score yang lebih tinggi dari *accuracy*.

		Harga Prediksi	
		Naik (Output = 1)	Turun (Output = 0)
Harga Real	Naik (Output = 1)	37	24
	Turun (Output = 0)	14	17

Tabel 10 Confusion matrix model prediksi tren harga BTC

		Harga Prediksi	
		Naik (Output = 1)	Turun (Output = 0)
Harga Real	Naik (Output = 1)	37	18
	Turun (Output = 0)	24	13

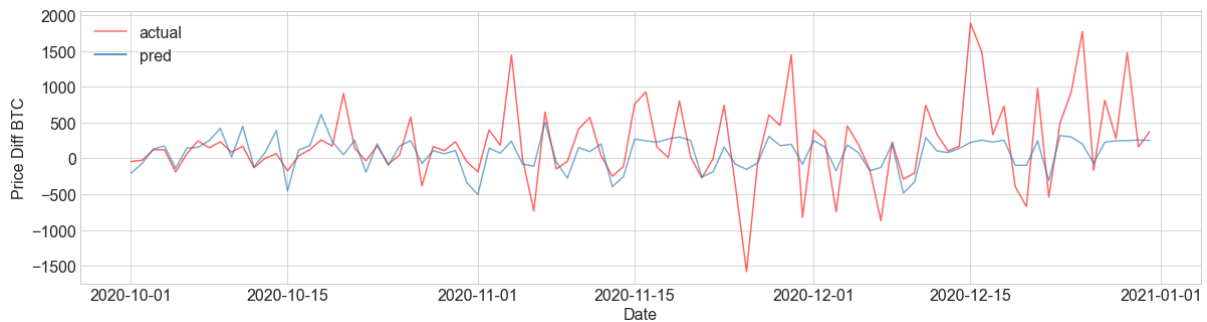
Tabel 11 Confusion matrix model prediksi tren harga ETH

	BTC	ETH
Accuracy	0.59	0.54
Precision	0.73	0.61
Recall	0.61	0.67
F1 Score	0.66	0.64
Mean Cross-Validation Score	0.48	0.51
K-fold CV Average Score	0.46	0.50

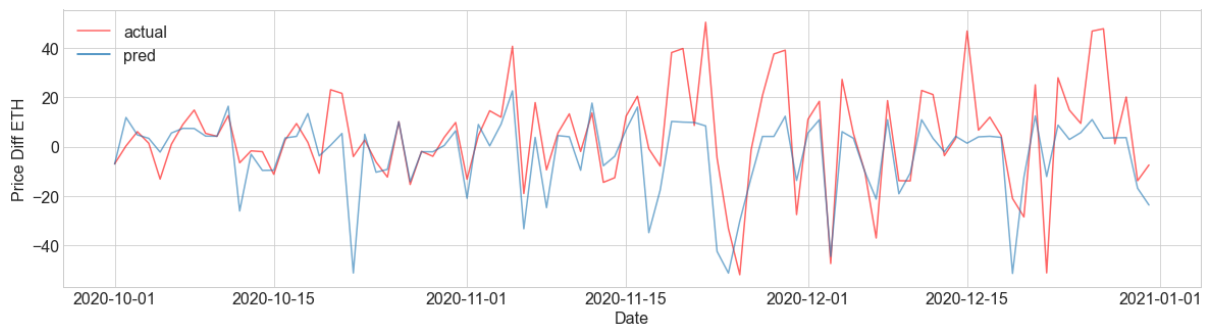
Tabel 12 Performa model prediksi tren harga cryptocurrency

4.6 Model Prediksi Perubahan Harga *Cryptocurrency*

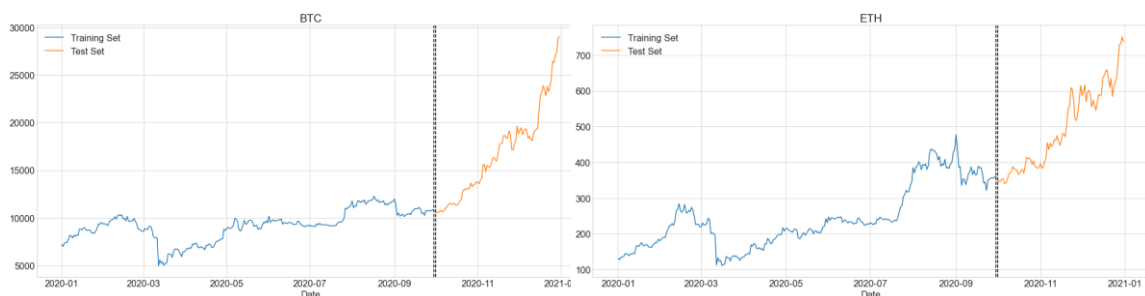
Untuk memprediksi perubahan harga *cryptocurrency* dipilih tiga fitur dengan korelasi tertinggi dari Tabel 8 dan Tabel 9 sebagai input dan Price Diff BTC/ETH sebagai variabel target yang dapat direpresentasikan sebagai masalah regresi. Sama dengan model prediksi tren harga *crypto* dilakukan pula *tuning* dengan cara yang sama dengan 3-cross fold validation sebanyak 500 iterasi. Didapat model dengan RMSE 484.7 untuk prediksi perubahan harga BTC dan RMSE 16.7 untuk prediksi perubahan harga ETH. Gambar 11 dan Gambar 12 adalah perbandingan antara perubahan harga aktual dan prediksi yang dilakukan oleh model. Dapat dilihat bahwa model tidak bisa memprediksi perubahan harga yang besar hal ini mungkin terjadi karena model di latih pada periode 1 Januari 2020 hingga 30 September 2020 dimana perubahan harga yang terjadi tidak terlalu besar dibandingkan data uji pada periode 1 Oktober 2020 hingga 31 Desember 2020 yang bisa dilihat pada Gambar 13, sehingga model kurang sensitif terhadap perubahan harga yang besar.



Gambar 11 Prediksi perubahan harga BTC



Gambar 12 Prediksi perubahan harga ETH



Gambar 13 data splitting untuk data latih dan data uji

5 KESIMPULAN

Berdasarkan koefisien korelasi fitur yang pada 4.4, fitur *Classes_1*, *Robert_neu*, dan *Roberta_pos* memiliki koefisien korelasi pada kisaran 0.69 sampai 0.73, sehingga menurut sumber [16] ketiga fitur tersebut memiliki korelasi positif kuat dengan harga *cryptocurrency*. Fitur *Impact*, *Classes_-1*, dan *Roberta_neg* memiliki koefisien korelasi pada kisaran 0.49 sampai 0.63, sehingga ketiga fitur tersebut memiliki korelasi positif sedang dengan harga *cryptocurrency*. Sedangkan fitur *tweet_volume* memiliki korelasi positif kuat pada dataset harga Bitcoin, tetapi berkorelasi positif sedang pada dataset harga Ethereum. Sehingga dapat disimpulkan bahwa sentimen *tweet* memiliki korelasi yang kuat terhadap harga *cryptocurrency* sehingga dapat digunakan untuk memprediksi harga *cryptocurrency*.

Tiga fitur dengan korelasi tertinggi lalu dipilih sebagai *input* untuk melakukan prediksi tren dan perubahan harga *cryptocurrency*. Untuk model prediksi tren harga *cryptocurrency* didapat model dengan performa pada Tabel 12 sehingga dapat dikatakan hasil akhir model cukup baik dengan akurasi 0.59 untuk BTC dan 0.54 untuk ETH. Selain itu model juga cukup akurat untuk melakukan *trading* harian *cryptocurrency* jika dilihat dari nilai *precision* yang cukup tinggi, 0.73 untuk BTC dan 0.61 untuk ETH.

Sedangkan untuk model prediksi perubahan harga *cryptocurrency* dengan input yang sama dengan model prediksi tren harga *cryptocurrency* didapat model dengan performa yang cukup baik dengan RMSE 484.7 dan *range* harga [-1500, 2000] untuk BTC dan RMSE 16.7 dan *range* harga [-40, 40] untuk ETH. Meskipun demikian, model kurang sensitif untuk memprediksi perubahan harga yang besar sehingga model perlu dilatih lagi dengan periode yang lebih panjang dengan perubahan harga yang lebih beragam.

Dengan demikian, didapat hasil akhir konstruksi model yang menjawab tujuan awal dari penelitian.

DAFTAR PUSTAKA

- [1] Hayes, Adam. (2022). Blockchain Facts: What Is It, How It Works, and How It Can Be Used. Diakses pada 8 Oktober 2022, dari <https://www.investopedia.com/terms/b/blockchain.asp>
- [2] Frankenfield, Jake. (2022). Cryptocurrency Explained with Pros and Cons for Investment. Diakses pada 8 Oktober 2022 dari <https://www.investopedia.com/terms/c/cryptocurrency.asp>
- [3] Prasetya, A., Ferdiansyah, Kunang, Y. N., Negara, E.S., & Chandra, W. (2021). Sentiment Analisis Terhadap Cryptocurrency Berdasarkan Comment Dan Reply Pada Platform Twitter. *Journal of Information Systems and Informatics*, 268-277
- [4] Davis, J. The crypto-currency: bitcoin and its mysterious inventor. The New Yorker. (2014, October). Available online at: <https://www.newyorker.com/magazine/2011/10/10/the-crypto-currency> (accessed April 30, 2018).
- [5] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*, 1 ed. Yogyakarta: Andi Offset, 2012.
- [6] <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>
- [7] <https://rey1024.medium.com/mengenal-accuracy-precision-recall-dan-specificity-serta-yang-diprioritaskan-b79ff4d77de8>
- [8] <https://stevkarta.medium.com/membicarakan-precision-recall-dan-f1-score-e96d81910354>
- [9] McNally, Sean. (2016). Predicting the Price of Bitcoin Using Machine Learning. Diakses pada November 2022 dari <https://norma.ncirl.ie/2496/1/seanmcnally.pdf>
- [10] N. Uras and M. Ortu. (2021). Investigation of Blockchain Cryptocurrencies' Price Movements Through Deep Learning: A Comparative Analysis, 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), pp. 715-722, doi: 10.1109/SANER50967.2021.00091.
- [11] Kristoufek, L.: What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. PLOS ONE 10(4) (04 2015) 1–15
- [12] Parlika, R., Pradika, S., Hakim, A., & N M, K. (2020). ANALISIS SENTIMEN TWITTER TERHADAP BITCOIN DAN CRYPTOCURRENCY BERBASIS PYTHON TEXTBLOB. *Jurnal Ilmiah Teknologi Informasi Dan Robotika*, 2(2), 33-37.
- [13] Sarmah, S. (2018). Understanding Blockchain Technology. *Computer Science and Engineering*, 8(2):23-29. DOI: 10.5923/j.computer.20180802.02
- [14] Haque, B., & Rahman, M. (2020). Blockchain Technology: Methodology, Application and Security. *International Journal of Computer Science and Network Security*, 20(2).
- [15] Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina1, K., Martinez-Hernandez, V., Sanchez, V., & Perez-Meana, H. (2018). *A Web Scraping Methodology for Bypassing Twitter API*.
- [16] Vejacka, M. (2014). Basic Aspects of Cryptocurrencies. *Journal of Economy, Business and Financing*, 2(2).
- [17] Ramraj, S., Uzir, N., Sunil, R., and Banerjee, S. (2016). Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets. *International Journal of Control Theory and Applications* (9):40. ISSN: 0974–557