

Pendekatan Algoritma Support Vector Regression Dalam Memprediksi Harga *Cryptocurrency* (Studi Kasus: *Binance*)

Nenni Mona Aruan^{*1}, Grace Widya Simanjuntak², Aldi Irvan Siagian³

^{1,2,3}Institut Teknologi Del; Toba, tel.: 0632-331 234; fax: 062 632 33111

e-mail: ^{*1}nenni.aruan@gmail.com, ²ifs18041@students.del.ac.id,

³ifs18007@students.del.ac.id

Abstrak

Salah satu jenis *Cryptocurrency* yang merupakan aset digital karena digunakan sebagai mata uang dalam bentuk digital adalah *Binance*. *Binance* merupakan jenis *Cryptocurrency* yang akan menjadi objek penelitian kali ini dikarenakan pergerakan harga *Binance* yang stabil dan masuk kedalam top 5 *cryptocurrency* dengan marketcap terbesar. Penelitian ini akan membantu para investor untuk memberikan informasi apakah mereka akan membeli atau menjual *Binance* mereka. Tujuan penelitian ini yaitu membangun model prediksi harga penutupan *Cryptocurrency* menggunakan algoritma *Support Vector Regression*. *Sentiment Analysis* juga dilakukan pada penelitian ini untuk mengidentifikasi dan mengekstraksi sentimen dari tiga akun yang berhubungan dengan *Binance* di media sosial twitter. *Sentiment Analysis* dilakukan menggunakan *Lexicon Based Approach* untuk memberikan label atau nilai sentimen terhadap setiap data. Model prediksi yang dibangun akan memperkirakan harga penutupan saham setiap harinya (*close*). Prediksi yang dilakukan ada dua percobaan, diantaranya adalah prediksi harga penutupan menggunakan dataset historis *Binance* dan dengan penambahan dataset sentimen. Hasil yang diperoleh pada penelitian ini yaitu bahwa prediksi menggunakan dataset historis *Binance* yang digabung dengan dataset sentimen memperoleh MSE yang lebih kecil yaitu 0.000830, sedangkan MSE untuk hasil prediksi menggunakan data historis *Binance* yaitu 0.00340. Dari hasil yang diperoleh, sentimen mampu menurunkan nilai *error* dan terbukti sentimen berpengaruh dalam menentukan harga *Cryptocurrency* untuk 1 hari kedepan.

Kata kunci— *Cryptocurrency*, Prediksi, *Support Vector Regression*, Analisis Sentimen

Abstract

Binance is kind of *Cryptocurrency* as virtual currency vause used traded through a cryptographic process. Every investor must try how to increase profits while minimizing risk, therefore *Cryptocurrency* price predictions are very important for everyone who invests. The objective of this research is to build a prediction model for the closing price of *cryptocurrencies* using the *Support Vector Regression* algorithm. *Sentiment Analysis* was also carried out in this study to identify and extract sentiments from three accounts related to *Binance* on Twitter social media. *Sentiment Analysis* is carried out using the *Lexicon-based Approach* to provide a label or sentiment value for each data. The prediction model built will estimate the closing price of the stock every day (*close*). The predictions carried out in this study consisted of two experiments, including the prediction of closing prices using the historical *Binance* dataset and the addition of the sentiment dataset. The results in this study are that predictions using the historical *Binance* dataset combined with the sentiment dataset obtain a smaller MSE of 0.000830, while the MSE for prediction results uses the historical data of *Binance*, which is

0.00340. From the results, sentiment can reduce the error value and prove that sentiment takes effect in Cryptocurrency in next day.

Keywords— Cryptocurrency, Prediction, Support Vector Regression, Sentiment Analysis

1. PENDAHULUAN

Pesatnya pertumbuhan teknologi informasi secara signifikan telah mempengaruhi banyak sektor termasuk keuangan. Digitalisasi produk dan proses keuangan bukanlah hal yang aneh saat ini. Seorang pakar ekonomi Ribert Reich dikutip dari cbsnews.com mengatakan, akan ada waktunya keberadaan uang fisik (uang kertas dan uang logam) akan tergantikan oleh uang jenis lainnya. *Cryptocurrency* telah menjadi salah satu alat transaksi pembayaran yang menggantikan uang fisik tersebut. *Cryptocurrency* adalah sebuah teknologi yang berbasis *blockchain* yang sering digunakan sebagai mata uang digital [1]. Mata uang digital tidak memiliki bentuk fisik uang layaknya mata uang kartal melainkan hanya sebuah *block* data yang diikat oleh *hash* sebagai validasinya. *Cryptocurrency* merupakan mata uang virtual dimana mata uang itu dihasilkan dan diperdagangkan melalui proses kriptografi. *Cryptocurrency* bersifat desentralisasi yang berarti bahwa mata uang tersebut beredar sepenuhnya tergantung kepada pasar dan tidak memiliki otoritas pusat yang dapat mengaturnya.

Istilah *Cryptocurrency* sudah menjadi salah satu topik yang banyak diperbincangkan karena meningkatnya berbagai macam uang virtual seperti Bitcoin, Ethereum, Ripple, Cardano, *Binance* dan masih banyak lagi. *Cryptocurrency* terbaik saat ini selain Bitcoin adalah *Binance* yang juga masuk kedalam daftar top 5 koin *Cryptocurrency* dalam *marketcap* versi 2021. *Binance* merupakan salah satu platform yang digunakan untuk melakukan transaksi *Cryptocurrency* terbesar didunia yang membuat koin ini menjadi salah satu koin *Cryptocurrency* terbesar. Platform trading *Binance* menawarkan koin *Cryptocurrency* terbanyak, memiliki biaya transaksi rendah, fluktuasi harga *Binance* yang lebih stabil dibandingkan *Cryptocurrency* lainnya dan proses transaksi likuiditas *Binance* yang tinggi.

Penelitian ini akan memprediksi harga *Cryptocurrency* menggunakan data lampau dari *website investing* dan juga data analisis sentimen publik dari *tweets* (cuitan) di twitter. Didasari dari kenyataan yang menunjukkan harga *Cryptocurrency* dapat berganti dengan sangat cepat, sehingga dalam dunia pasar *Cryptocurrency* para investor harus cermat dalam pengambilan suatu keputusan kapan suatu *Cryptocurrency* sebaiknya dijual dan kapan suatu *Cryptocurrency* sebaiknya dipertahankan. Menurut perkiraan Forbes, beberapa milioner yang membangun kekayaan mereka di dunia mata uang crypto telah kehilangan USD15,5 miliar atau sekitar 217 triliun rupiah (kurs Rp14.000 per USD). Untuk itu, sangat penting suatu model untuk memprediksi harga penutupan *Cryptocurrency* untuk mempelajari pergerakan pasar *Cryptocurrency*.

Media sosial yang sangat digemari masyarakat dalam memberikan pendapat ataupun komentarnya saat ini adalah twitter [2]. Selain data *Cryptocurrency* pada *website coingecko*, sentimen publik mengenai pergerakan harga *Cryptocurrency* yang dimuat dari *tweets* (cuitan) yang terdapat pada media sosial twitter juga akan dianalisis. Sentimen Analisis yaitu proses memahami, mengekstrak dan mengolah data secara otomatis untuk mendapat informasi sentimen yang terdapat dari suatu opini [2]. Hal ini bertujuan untuk melihat bagaimana pengaruh sentimen positif dan negatif terhadap prediksi *Cryptocurrency* yang akan dilakukan. Salah satu contoh paling nyata adalah Elon Musk yang sering berkicau tentang mata uang krypto tertentu di akun Twitternya. Begitu besarnya influence Elon terhadap dunia krypto, dengan satu

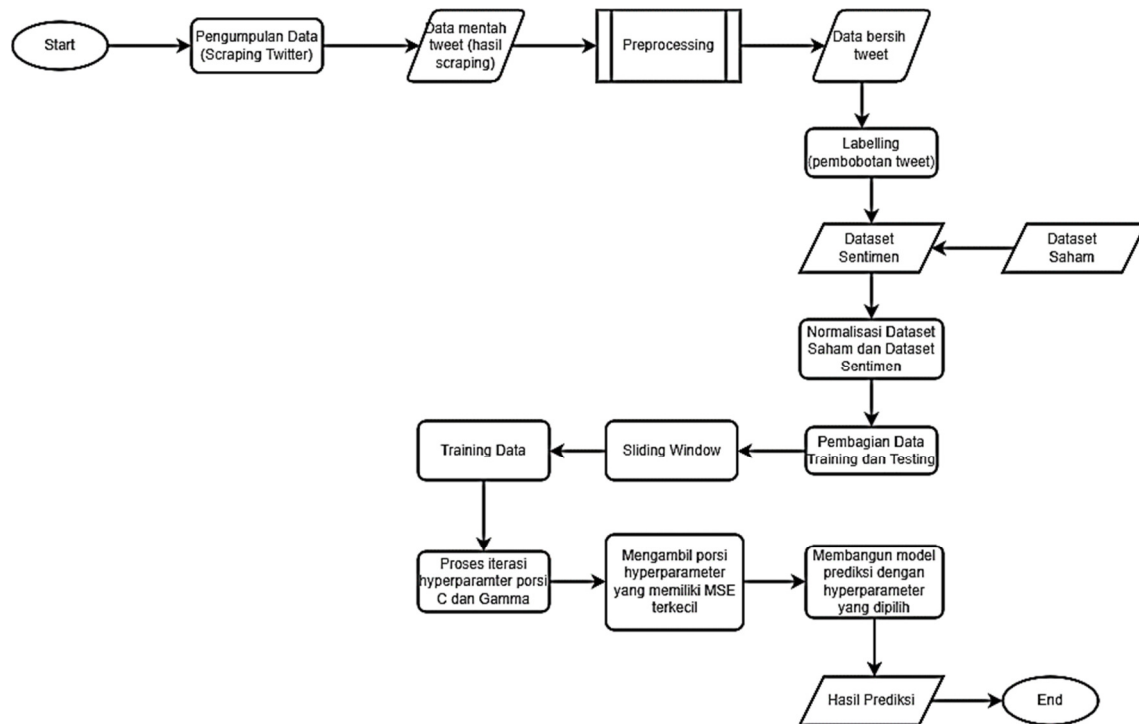
tweet saja sudah bisa melambungkan harga Dogecoin yang pernah masuk dalam tweet-nya. Adapun alasan dilakukannya penelitian ini yaitu untuk membantu para investor dalam prediksi indeks harga *Cryptocurrency* untuk mendapatkan keuntungan dari hasil prediksi tersebut.

Pada dasarnya melakukan prediksi *Cryptocurrency*, baik pergerakan (naik atau turun) ataupun harga di masa depan, menggunakan model pembelajaran mesin, model lainnya, merupakan bagian dari analisis teknikal karena penggunaan data historis. Salah satu analisis teknikal yang sering digunakan seorang investor adalah analisis *chart* dimana investor mengambil keputusan berdasarkan bentuk dari harga historis suatu saham, namun analisis ini memerlukan intuisi dari investor tersebut untuk mengekstraksi informasi dari bentuk *chart* tersebut.

Algoritma prediksi yang akan digunakan pada penelitian ini yaitu Support Vector Regression. SVR merupakan salah satu modifikasi SVM yang digunakan untuk pendekatan regresi. Harga *Cryptocurrency* dipengaruhi oleh banyak variabel maka perlu cara bagaimana variabel itu berhubungan. Dimana studi yang menyangkut masalah tersebut adalah analisis regresi dengan pendekatan algoritma SVR. Model SVR mempunyai keunggulan dibanding beberapa metode atau algoritma lainnya dalam hal pemanfaatan model nonlinear (perubahan yang relatif cepat di waktu ke waktu) seperti yang terjadi pada harga *Cryptocurrency* yang fluktuasinya berubah dengan cepat [3-5]. SVR merupakan teknik Machine Learning yang cocok untuk mengenali data time series [6-9]. Dalam penelitian ini, kami akan menggunakan data dari Investing.com dimana variabel yang akan digunakan dari data ini adalah harga pembukaan (*opening price*), harga tertinggi (*highest price*), harga terendah (*lowest price*), harga penutupan (*close*) dan total transaksi (*volume*). Alasan menggunakan variabel-variabel tersebut dalam penelitian ini karena penelitian ini termasuk dalam analisis teknikal dimana nantinya algoritma yang digunakan dalam penelitian ini akan mempelajari berdasarkan sejarah harga Binance yang membuat algoritma tersebut mampu mengumpulkan informasi tentang pola pergerakan harga berdasarkan sebelumnya. Serta menambahkan variabel tambahan yaitu cuitan pada media sosial twitter yaitu cuitan dari CEO Binance dan akun resmi Binance yang sangat berpengaruh dalam pergerakan harga. Hal ini diteliti seberapa pengaruh sentimen dalam menentukan harga *Cryptocurrency*.

2. METODE PENELITIAN

Desain percobaan yang dilakukan pada penelitian ini ada dua, yaitu percobaan prediksi tanpa data sentimen dan percobaan prediksi dengan menggunakan data sentimen.



Gambar 1. Flowchart Penelitian

Dari gambar 1 dapat dilihat bahwa proses yang terjadi dalam prediksi harga penutupan *Cryptocurrency* adalah yang pertama pengumpulan data (*data collecting*). Data historis *Binance* diperoleh dari website investing. Pengumpulan data sentimen yang dilakukan pada penelitian ini diperoleh dengan metode scraping, dimana data yang akan di scraping adalah tweet yang mengandung kata *Binance* dari tiga akun yang dianggap berpengaruh terhadap fluktuasi harga *Binance*. Data yang dihasilkan dalam format csv file. Setelah data diperoleh, data tersebut akan melalui tahap preprocessing sebelum digunakan untuk prediksi. Preprocessing untuk data sentimen tweet yang dilakukan ada empat tahapan yaitu *cleaning*, *case folding*, *tokenizing*, dan *stopword removal*. Hasil yang diperoleh pada tahap preprocessing adalah data bersih tweet yang tidak mengandung hashtag, mention, URL dan karakter unik lainnya. Selanjutnya, data sentimen akan dilabeli dan diberi bobot masing-masing. Kedua data kemudian akan di normalisasi. Normalisasi data yang dilakukan pada penelitian ini bertujuan untuk menyeragamkan data, agar data yang digunakan untuk prediksi adalah data yang seragam sehingga mudah untuk diproses. Setelah dinormalisasi, tahap selanjutnya adalah prediksi harga *Cryptocurrency* menggunakan dua data, yaitu data historis *Binance* digabung dengan data sentimen.

Selanjutnya data melalui proses *Sliding window*, yaitu manipulasi data dengan menggeser data satu periode sebelumnya. Hal ini dilakukan karena data yang diolah yaitu time series atau runtun waktu. Perubahan data ataupun pergerakan nilai variabel pada suatu periode (t) akan berdampak kepada nilai tukar pada waktu (t+1). Setelah *sliding window* dilakukan selanjutnya dilakukan *training* data dengan *support vector regression*, kemudian dilakukan proses iterasi untuk pemilihan porsi *hyperparameter* C dan gamma [7]. Setelah iterasi selesai,

maka iterasi yang memiliki MSE terkecil akan diambil sebagai iterasi terbaik dan selanjutnya porsi *hyperparameter* tersebut digunakan untuk membangun model prediksi.

2.1 Pengumpulan Data

Eksperimen yang dilakukan pada penelitian ini terdiri dari dua eksperimen, yaitu eksperimen prediksi harga penutupan *Cryptocurrency* tanpa data sentimen, yang kedua eksperimen prediksi harga penutupan *Cryptocurrency* dengan menggunakan data sentimen. Pengumpulan data historis *Binance* diperoleh dari website investing dengan kurun waktu dari tanggal 01 September 2019-14 Maret 2022. Pengumpulan data sentimen yang dilakukan pada penelitian ini diperoleh dengan metode scraping. Data diambil dari cuitan tiga akun yang memberi pengaruh terhadap fluktuasi harga *Binance* yaitu akun CEO *Binance* (*cz_Binance*), akun resmi *Binance* (*Binance*) dan *Binance* smart chain (*BinanceChain*). Jumlah data sentimen yang diperoleh adalah 19.813 *tweets* dengan kurun waktu dari tanggal 01 September 2019 sampai tanggal 14 Maret 2022 kemudian data tersebut di bersihkan terlebih dahulu dengan cara *preprocessing*. Data yang dihasilkan dalam format csv file. Setelah data diperoleh, data tersebut akan melalui tahap *preprocessing* sebelum digunakan untuk prediksi. Berikut ini contoh sentimen twitter yang telah di *preprocessing*.

Tabel 1. Sentimen Twitter Setelah Dilakukan *Preprocessing*

Input Today's topic will be focused on developing blockchain games. https://t.co/T62e9MovLm
Output Today's topic focus is developing a blockchain game

Preprocessing yang dilakukan bertujuan untuk mengelola data agar memiliki keseragam dan kemudahan sebelum data tersebut diolah melalui tahapan-tahapan berikut.

2.2 Data Preprocessing

Data preprocessing merupakan proses pengolahan data, sehingga data yang hendak digunakan merupakan data yang baik dan bersih. Pada penelitian ini kedua data yang digunakan yaitu data historis *Binance* dan data sentimen *tweet* akan di-*preprocessing*, dimana tahap *preprocessing* yang dilakukan pada data historis *Binance* hanya normalisasi data, sedangkan untuk data sentimen tahapan *preprocessing* yang dilakukan adalah *cleaning*, *case folding*, *tokenizing*, dan *stopword removal* [10]. Penjelasan dari setiap tahapan tersebut adalah sebagai berikut.

2.2.1 Cleaning

Tahapan pertama yang dilakukan pada *preprocessing* adalah *cleaning*, dimana semua karakter yang bukan alphabet akan dihilangkan, seperti simbol maupun angka karena simbol dan angka dianggap perlu dihilangkan karena tidak mengandung sentimen di dalam suatu teks.

2.2.2 Case Folding

Pada tahap ini semua huruf kapital akan diubah menjadi huruf kecil, dan semua karakter yang bukan merupakan alphabet akan dihilangkan.

2. 2.3 Tokenizing

Pada tahap ini, tweets akan dibagi menjadi beberapa kata berdasarkan spasi dan simbol yang tidak relevan akan dihapus. Token hanya terdiri dari semua karakter alfa, karakter alfanumerik atau hanya karakter numerik saja.

2. 2.4 Stopword Removal

Stopword removal merupakan tahapan untuk menghilangkan kata-kata yang kurang memiliki makna. Kata-kata yang tidak memiliki emosi atau makna apapun disebut sebagai stopwords. Misalnya kata penghubung seperti 'di', 'ke', 'dari', 'oleh', dan sebagainya. Data yang digunakan sebagai masukan pada tahap ini adalah data hasil tokenisasi. Setiap kalimat stopwords yang ditemukan pada masing-masing data akan dihapus.

2. 3 Normalisasi Data

Data historis yang digunakan sebagai dataset pada penelitian ini diperoleh dari website investing, kemudian data historis *Binance* tersebut akan di *preprocessing*, adapun tahapan *preprocessing* yang dilakukan adalah normalisasi. Normalisasi yang dilakukan menggunakan *Min Max Normalization*, berikut ini rumus Min Max [11].

$$nv = f(v) = \frac{v - \min(v)}{(v) - \min(v)} \quad (1)$$

2. 4 Data Labeling

Pelabelan data pada penelitian ini dilakukan dengan menggunakan *lexicon based approach*, yang bertujuan untuk mengelompokkan sentimen apakah sentimen tersebut termasuk sentimen positif, negatif atau netral, sehingga dapat dilihat bagaimana pengaruhnya terhadap prediksi harga penutupan *Cryptocurrency*. Untuk pembobotan data sentimen dilakukan menggunakan sentiwordnet, dimana sentimen positif akan diberi nilai 1, untuk sentimen negatif diberi bobot -1 sedangkan untuk sentimen yang bersifat netral akan diberi nilai 0. *Lexicon-based approach* atau pendekatan berbasis leksikon digunakan untuk menentukan sebuah opini apakah opini tersebut mengekspresikan sentimen positif atau negatif dengan menggunakan kamus sentimen. Menurut hasil penelitian David Moeljadi [12] akan dihitung total skor dari sentimen dan mengklasifikasikan sentimen yang terdapat pada setiap *tweet*. Total skor akan dijumlahkan dengan perhitungan persamaan berikut [13].

$$S_{positive} = \sum_{i \in t}^n positive\ score_i \quad (2)$$

$$S_{negative} = \sum_{i \in t}^n negative\ score_i \quad (3)$$

Setelah kalimat opini digolongkan maka akan diubah menjadi numerik. Pada penelitian ini, metode pembelajaran mesin yang digunakan untuk meakukan prediksi adalah *Support Vector Regression*. Hasil keluaran dari proses Data Labeling adalah seperti pada tabel 2 berikut:

Tabel 2. Contoh Hasil *Data Labeling Sentiment*

<i>Tweet</i>	<i>Date</i>	<i>Positive</i>	<i>Negative</i>	<i>Neutral</i>	<i>Compound</i>	<i>Sentimen</i>
<i>say picture worth thousand word might true saved image nft usually worth nothing</i>	01/09/2021 23:15	0,511	0	0,489	0,8126	<i>Positive</i>
<i>live talking binancesmartchainst anniversary conference</i>	01/09/2021 13:53	0	0	1	0	<i>Neutral</i>
<i>binance add farm mbox amp tribe isolated margin</i>	02/09/2021 10:00	0	0,247	0,753	-0,3182	<i>Negative</i>

2. 4 *Support Vector Regression*

Support Vector Regression (SVR) adalah algoritma yang dikembangkan dari algoritma *Support Vector Machine* (SVM) digunakan untuk kasus regresi [14]. Algoritma SVR ini digunakan untuk menemukan suatu fungsi $f(x)$ yang mempunyai deviasi \square paling besar dari nilai aktual y untuk semua data *train*. Jika telah diperoleh nilai $\square = 0$ maka ditemukanlah suatu regresi yang sempurna [15]. Untuk menggunakan SVR maka hal yang perlu dilakukan adalah mencari fungsi *hyperplane optimal* melalui parameter yang dapat memisahkan kumpulan data yang telah diinput sesuai dengan yang diharapkan. Suatu model memiliki performa yang baik jika model tidak mengalami *overfitting* yaitu jika data yang digunakan di pelatihan dan yang digunakan untuk tes menghasilkan akurasi yang berbeda (akurasi berkurang). Model juga tidak boleh *underfitting* yaitu data pelatihan tidak mewakili data keseluruhan data yang akan digunakan sehingga menyebabkan performa yang buruk pada model.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, evaluasi model dilakukan untuk mendapatkan model terbaik untuk peramalan harga cryptocurrency Binance yang dilakukan dengan menggunakan data sentimen dan tanpa sentimen.

3. 1 *Hasil Prediksi tanpa menggunakan data sentimen*

Dataset yang digunakan adalah data lampau *Binance* dengan kurun waktu dua tahun terakhir. Dalam menggunakan metode SVR perlu dilakukan pemilihan *hyperparameter* yang optimal untuk mendapatkan model prediksi yang baik. Setiap *hyperparameter* diberikan batas nilai di bawah dan di atas memiliki kelipatan secara eksponensial karena akan mendapatkan hasil yang lebih maksimal. Kemudian untuk memperoleh model terbaik, dilakukan kembali tuning untuk setiap parameternya. Sebelumnya, penulis melakukan beberapa percobaan untuk setiap nilai parameter dan menyeleksi kombinasi parameter yang menghasilkan *error* yang minimum. Nilai parameter yang ditetapkan setelah nilai *error* yang ditunjukkan sudah minimal dimana tidak terdapat lagi perubahan yang signifikan pada iterasinya. Berikut parameter yang dipilih untuk dilanjutkan ke tahap SVR.

Tabel 3. Pemilihan Parameter SVR

C	$10^0, 10^1, 10^2, 10^3, 10^4, 10^5, 10^6$
Gamm	$10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1,$
a	$10^2, 10$

Berdasarkan nilai batas bawah dan atas hyperparameter C dan gamma yang telah ditentukan sebelumnya, hasil error MSE dari setiap kombinasi nilai parameter yang dilakukan pada data training akan ditunjukkan pada Tabel 4.

Tabel 4. Hasil MSE Untuk Setiap Kombinasi Dengan Dataset *Binance*

Kombinasi	C	Gamma	MSE
1	10^0	10^1	0.005187
2	10^1	10^1	0.005185
3	10^2	10^0	0.003430
4	10^3	10^{-2}	0.004700
5	10^4	10^{-1}	0.003530
6	10^5	10^{-1}	0.002919
7	10^6	10^{-1}	0.004472

Berdasarkan hasil evaluasi pada Table42 didapatkan nilai error MSE terkecil yaitu pada kombinasi ke 3 dengan nilai MSE sebesar 0.003430 dimana nilai hyperparameter C sebesar 100 dan gamma 1. Sehingga, kombinasi ke 3 dengan nilai hyperparameter C = 100 dan hyperparameter gamma = 1 yang akan dipilih dan digunakan pada model peramalan SVR.

Berdasarkan nilai MSE yang dihasilkan kecil yaitu 0.003430, MSE ini merupakan nilai error terkecil yang diperoleh dari dataset yang ada di karenakan sudah melalui eksperimen untuk mencari model terbaik.

3. 2 Hasil Prediksi menggunakan data sentimen

Dataset yang digunakan adalah data lampau *Binance* dan dataset sentimen dengan kurun waktu dua tahun terakhir. Pada bab ini akan dibahas apakah sentimen dapat mempengaruhi harga penutupan (*close*). Dalam menggunakan metode SVR perlu dilakukan pemilihan hyperparameter yang optimal untuk mendapatkan model prediksi yang baik. Terlebih dahulu perlu diberikan nilai batas bawah dan batas atas pada setiap hyperparameter. Nilai-nilai batas ini didapatkan dengan melihat hasil penelitian sebelumnya yang menyatakan bahwa performa nilai hyperparameter akan semakin maksimal apabila nilai batas yang diberikan untuk masing – masing hyperparameter memiliki kelipatan secara eksponensial. Untuk memperoleh parameter yang paling optimum sebagai model terbaik maka dilakukan tuning untuk setiap parameter. Sebelumnya, penulis melakukan beberapa percobaan untuk setiap nilai parameter dan menyeleksi kombinasi parameter yang menghasilkan *error* yang minimum. Berikut adalah hasil dari perhitungan nilai error MSE dari setiap kombinasi parameter yang telah ditentukan sebelumnya.

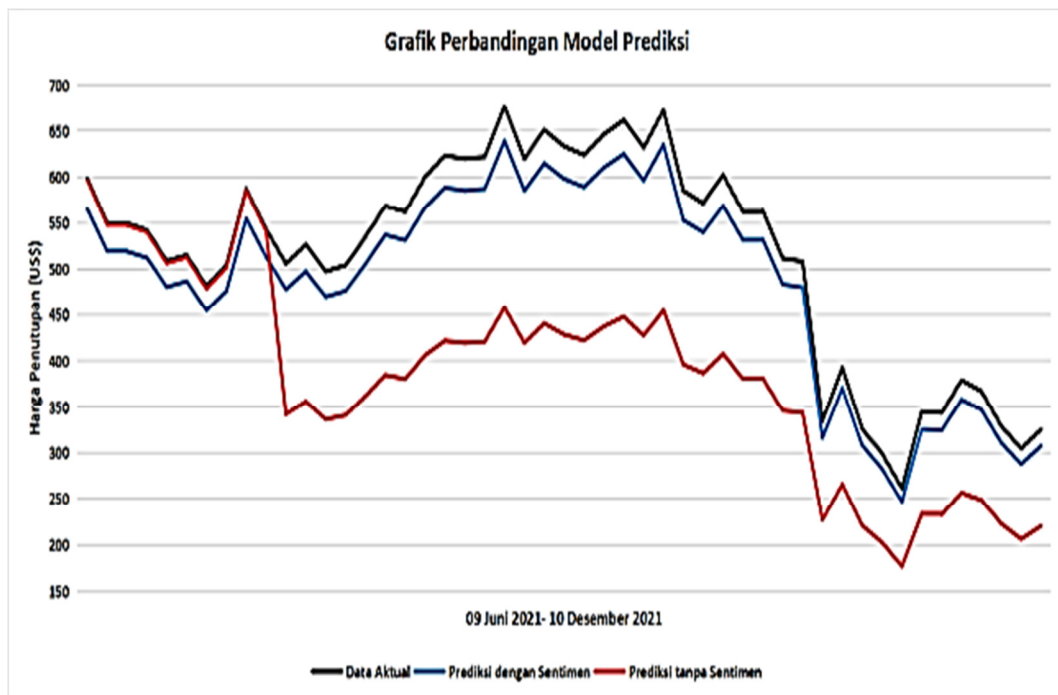
Tabel 5. Hasil MSE Untuk Setiap Kombinasi Dengan Dataset Sentimen

Kombinasi	C	Gamma	MSE
1	10^0	10^2	0.004961
2	10^1	10^2	0.004961

3	10^2	10^0	0.000830
4	10^3	10^0	0.004773
5	10^4	10^{-3}	0.002468
6	10^5	10^{-3}	0.001857
7	10^6	10^{-1}	0.001058

Berdasarkan hasil evaluasi pada Tabel 5 tersebut, kombinasi 3 terpilih sebagai model peramalan SVR sekaligus sebagai model terbaik karena menghasilkan nilai MSE terkecil dibandingkan kombinasi lainnya dengan nilai *hyperparameter* C dan gamma 1 masing-masing 100 dan 1.

Dari nilai MSE pada Tabel 3 di atas menunjukkan bahwa model yang di bangun dengan dataset sentimen menghasilkan nilai MSE yang lebih kecil yang berarti model yang dibangun dengan dataset sentimen lebih baik daripada tidak menggunakan sentimen, hal itu disebabkan dengan di tambah nya variabel sentimen, model untuk prediksi harga penutupan jadi lebih baik karena pola yang diberikan oleh sentimen twitter membantu membuat model peramalan mendapat pola yang lebih baik. Penambahan nilai sentimen terhadap data historis *Binance* memiliki pengaruh terhadap akurasi dengan berkurangnya nilai error sebesar 0.002600 poin. Nilai error tersebut merupakan nilai kesalahan rata-rata dari seluruh data, yang mana berarti bahwa terjadi penurunan tingkat perbedaan rata-rata data aktual dengan data prediksi dengan menggunakan data sentimen sebesar 0.002600 poin. Maka dapat disimpulkan dengan ditamhkannya dataset sentimen yang diperoleh dari media sosial twitter memberi pengaruh terhadap model prediksi yaitu mampu menghasilkan error yang lebih kecil.



Gambar 2. Grafik Perbandingan Model Prediksi

Berdasarkan Gambar 2 diatas dapat kita lihat bahwa grafik model peramalan menggunakan penambahan dataset sentimen menghasilkan bentuk grafik yang lebih rapat dengan data aktual dibandingkan dengan grafik tanpa penambahan data sentimen.

4. KESIMPULAN

Berikut beberapa poin yang bisa disimpulkan dari hasil penelitian diatas:

1. Model dengan algoritma SVR tanpa menggunakan dataset sentimen menghasilkan MSE (*Mean Square Error*) dengan nilai 0.003430 dengan C sebesar 100 dan gamma 1, sedangkan model algoritma SVR dengan menggunakan dataset sentimen menghasilkan MSE (*Mean Square Error*) dengan nilai 0.000830 dengan C sebesar 100 dan gamma 1.
2. Berdasarkan hasil implementasi yang telah diperoleh, penambahan nilai sentimen memiliki pengaruh terhadap model prediksi. Penambahan nilai sentimen terhadap data historis *Binance* memiliki pengaruh terhadap nilai error dengan berkurangnya nilai kesalahan sebesar 0.002640 poin. Maka dapat disimpulkan dengan ditambahkan dataset sentimen mampu menghasilkan *error* yang lebih kecil.

5. SARAN

Beberapa masukan yang bisa dilakukan untuk penelitian sejenis yang akan dilakukan adalah:

1. Menambahkan algoritma optimasi *machine learning* lainnya untuk *tuning* parameter *Support Vector Regression* yang bertujuan untuk mendapatkan nilai hyperparameter yang tepat.
2. Menggunakan algoritma *time series* lainnya yang seperti ARIMA dan algoritma Fuzzy untuk mendapatkan perbandingan model peramalan terbaik.
3. Sebaiknya menemukan konsep pemodelan menggunakan *multivariate time series* yang sudah terbukti mampu menggunakan konsep *multivariate time series* dalam meramalkan harga penutupan (*close*) untuk hari yang akan datang.

DAFTAR PUSTAKA

- [1] Bhintara, I. B. P. 2018. *Teknologi Blockchain Cryptocurrency di Era Revolusi Digital. Prosiding Seminar Nasional Pendidikan Teknik Informatika.*
- [3] Buntoro, G. A., Adji, T. B., & Purnamasari, A. E. (2014). *Sentiment Analysis Twitter dengan Kombinasi Lexicon dan Double Propagation. CITEE*, 39- 43.
- [4] P. Meesad and R. I. Rasel. 2013. *Predicting Stock Market Price Using Support Vector Regression. International Conference on Informatics, Electronics and Vision, ICIEV 2013* (2013)
- [5] N. D. Maulana, B. D. Setiawan, & C. Dewi, 2019. *Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery).* Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol.3, No.3, Maret 2019, hlm, 2986-2995.

-
- [6] R. U. Fadilah, D. Agfiannisa, Y. Azhar. 2020. *Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine*. Fountain of Informatics Journal. Vol. 5, No.2, November 2020, ISSN (Online)
- [7] M. Usmani, S. H. Adil, K. Raza, S. S. A. Ali. 2016. *Stock Market Prediction Using Machine Learning Techniques*. 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 - Proceedings (2016)
- [8] Yudhawan, D. (2020). *Implementasi Support Vector Regression Untuk Peramalan Harga Saham Perusahaan Pertambangan di Indonesia (Studi Kasus: PT Adaro Energy Tbk, PT Bukit Asam Tbk, dan PT Indo Tambangraya Megah Tbk)*. (Skripsi, Universitas Islam Indonesia, 2020) Diakses dari <https://dspace.uui.ac.id/123456789/28082>
- [9] Sembiring. A., Ginting. H.(2019). *Prediksi Nilai Tukar Mata Uang Dolar Amerika Serikat Terhadap Rupiah Indonesia dengan Algoritma Support Vector Regression*. (Skripsi, Institut Teknologi Del, 2019).
- [10] Ventiano, E. Djunaedy, Amaliyah. 2019. *Perhitungan Intensitas Radiasi Matahari berdasarkan Pola Sebaran Awan menggunakan Metode Support Vector Regression (SVR)*. e-Proceeding of Engineering, Vol. 6, No.2, Agustus 2019. ISSN: 2355-9365
- [11] A. S. Eesa and W. K. Arabo, "Normalization Methods for Backpropagation: A Comparative Study," *Science Journal of University of Zakho*, pp. 319-323, 2017.
- [12] T. A. Le, D. Moeljadi, Y. Miura and T. Ohkuma, "Sentiment Analysis for Low Resource Languages: A Study on Informal Indonesian Tweets," *The 12th Workshop on Asian Language Resources*, 2016.
- [13] E. W. Pamungkas and D. G. P. Putri, "An Experimental Study of Lexicon-based Sentiment Analysis on Bahasa Indonesia," *International Annual Engineering Seminar (InAES)*, 2016.
- [14] Sanusi, Buono, Imas, Faqih. 2014. *Downscaling Modeling Using Support Vector Regression for Rainfall Prediction*. TELKOMNIKA Indonesian Journal of Electrical Engineering. Vol. 12, No.8, Agustus 2014, pp. 6423-6430. DOI: 10.11591/telkomnika.v12i8.6195
- [15] Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu