

IMPLEMENTASI ORANGE *DATA MINING* UNTUK PREDIKSI HARGA *BITCOIN*

Indriyanti¹, Nurul Ichsan², Haerul Fatah³, Tri Wahyuni⁴, Erni Ermawati⁵

¹ Universitas Bina Sarana Informatika
e-mail: indriyanti.iyt@bsi.ac.id

² Universitas Bina Sarana Informatika
e-mail: nurul.nrc@bsi.ac.id

³ Universitas Bina Sarana Informatika
e-mail: haerul.hef@bsi.ac.id

⁴ Universitas Bina Sarana Informatika
e-mail: tri.twy@bsi.ac.id

⁵ Universitas Bina Sarana Informatika
e-mail: erni.ert@bsi.ac.id

Abstrak

Bitcoin merupakan salah satu aset kripto yang biasa digunakan untuk bertransaksi secara virtual dan merupakan tiga teratas dari cryptocurrency yang paling banyak diperdagangkan. Harga *bitcoin* dapat berubah setiap waktu, untuk itu diperlukan suatu prediksi harga *bitcoin* dimasa mendatang. Penelitian ini bertujuan untuk menerapkan aplikasi orange data mining dalam memprediksi harga *bitcoin* dengan metode *K-Nearest Neighbor (K-NN)* dan *Support Vector Machine (SVM)*. Prediksi kinerja dari kedua metode yang digunakan dapat dilihat dari hasil Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE). Data *bitcoin* yang telah dikumpulkan akan dianalisa menggunakan aplikasi orange data mining, proses pengujian menerapkan *K-Fold Cross Validation (K=5)*, sedangkan proses perbandingan metode menggunakan *Test and Score*. Hasil perbandingan kedua model menunjukkan bahwa SVM karena memiliki nilai RMSE dan MAE paling kecil yaitu 0.010 dan 0.008, maka dari itu untuk prediksi harga *bitcoin* di masa mendatang model *Support Vector Machine (SVM)* dapat dijadikan rekomendasi terbaik dibandingkan model *K-NN*.

Kata Kunci: *Orange Tools, Data Mining, Prediksi, Bitcoin*

Abstract

Bitcoin is one of the most commonly used crypto assets for virtual transactions and is in the top three of the most traded cryptocurrencies. *Bitcoin* prices can change at any time, for that we need a *bitcoin* price prediction in the future. This study aims to apply orange data mining applications in predicting *bitcoin* prices using the *K-Nearest Neighbor (K-NN)* and *Support Vector Machine (SVM)* methods. The performance predictions of the two methods used can be seen from the results of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). *Bitcoin* data that has been collected will be analyzed using the orange data mining application, the testing process applies *K-Fold Cross Validation (K=5)*, while the method comparison process uses *Test and Score*. The results of the comparison of the two models show that SVM because it has the smallest RMSE and MAE values, namely 0.012 and 0.009, therefore for predicting future *bitcoin* prices the *Support Vector Machine (SVM)* model can be the best recommendation compared to the *K-NN* model.

Keywords: *Orange Tools, Data Mining, Prediction, Bitcoin*

1. Pendahuluan

Aset digital yang paling populer diberbagai kalangan saat ini yaitu *bitcoin*. *Bitcoin* merupakan salah satu aset kripto yang biasa digunakan untuk bertransaksi secara virtual. Kenaikan harga *bitcoin* menjadi salah satu faktor aset digital tersebut banyak diburu oleh orang banyak. Harga *bitcoin* di masa depan juga ada yang memprediksi akan mengalami penurunan bahkan bisa menjadi nol. Nilai *botcoin* yang senantiasa mengalami ketidakstabilan karena merupakan produk global yang dipengaruhi oleh peraturan di dunia, tidak seperti data aset lainnya (emas, indeks barang dan komoditas) (Saadah & Salsabila, 2021).

Menjaga agar aset *bitcoin* tetap aman dan tidak mengalami kerugian, maka para pemilik *bitcoin* perlu melakukan prediksi harga di masa yang akan datang. Prediksi harga *bitcoin* dapat dilakukan melalui beberapa metode, seperti pada penelitian mengenai prediksi harga *bitcoin* yang telah dilakukan sebelumnya.

Harga *cryptocurrency* dapat diprediksi menggunakan metode *K-Nearest Neighbours (K-NN)* dengan menggunakan parameter nilai $K=3$ (Fatah et al., 2018). Penelitian dilakukan pada 3 jenis *cryptocurrency* yaitu dataset *bitcoin*, *ethereum* dan *ripple*. *Neural Network* dengan *backpropagation* juga dapat digunakan untuk memprediksi harga *bitcoin* (Pinudom et al., 2019). Metode lain yang sering digunakan untuk memprediksi harga *bitcoin* antara lain *Forecasting* (Derbentsev et al., 2021) dan *Long Short Term Memory (LSTM)* (Rizkilloh & Widiyanesti, 2022).

Hasil terbaik dari suatu prediksi dapat diperoleh apabila kumpulan data selalu diperbarui (Swamy et al., 2021). *Artificial Intelligent* model untuk memprediksi harga *bitcoin* dan untuk membandingkan kinerja suatu metode dilihat dari nilai *RMSE* dan *MSE* sebagai alat evaluasi dan perkiraan (Septiarini et al., 2020). Metode regresi lain yang dapat digunakan untuk memprediksi harga *bitcoin* yaitu *Support Vector Machine (SVM)* (Indriyanti & Subekti, 2018). Penelitian lain yang menggunakan metode metode SVM untuk memprediksi harga *bitcoin* yaitu penelitian yang dilakkan oleh Greaves dkk, dalam penelitian tersebut dilakukan perbandingan metode antara SVM dan ANN dimana metode ANN memiliki akurasi sebesar 55% dengan regular ANN. Peneliti juga menyimpulkan bahwa prediksi pada

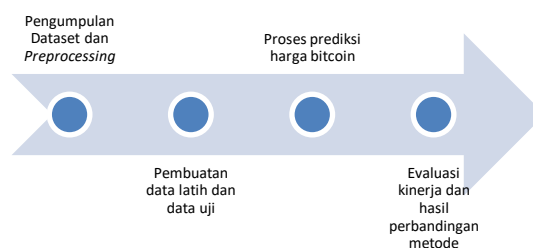
data yang diteliti memiliki keterbatasan karena harga secara teknis didikte oleh bursa yang perilakunya terletak di luar wilayah data tersebut (Greaves, Alex and Au, 2015).

Tools yang biasanya digunakan untuk menilai kinerja suatu metode antara lain WEKA dan Rapid Miner. Kinerja suatu metode untuk melakukan prediksi dapat dilakukan juga dengan menggunakan orange tools. Orange data mining merupakan aplikasi *open source* yang mampu membantu peneliti dalam menganalisa suatu data (Alim, 2021).

Penelitian ini bertujuan untuk mendapatkan metode terbaik untuk memprediksi harga *bitcoin*. Pada penelitian ini akan menerapkan orange tools untuk memprediksi harga *bitcoin* dengan menggunakan metode K-NN dan SVM. Kedua metode tersebut akan di evaluasi dan dibandingkan kinerjanya sehingga dapat diketahui metode yang palig baik untuk memprediksi harga *bitcoin* berdasarkan nilai RMSE dan nilai MAE dari masing-masing metode. Nilai tersebut dapat dilihat menggunakan widget *Test and Score* pada aplikasi Orange.

2. Metode Penelitian

Tujua dari penelitian ini untuk melakukan analisa perbandingan metode K-NN dan SVM untuk memprediksi harga *bitcoin* dengan menggunakan Orange Tools untuk simulasinya.



Gambar 1. Tahapan Penelitian
Sumber: Peneliti (2022)

Tahapan pada penelitian ini terdiri dari 4 tahap yaitu pengumpulan dataset, preprocessing, pembuatan data latih dan data uji, proses prediksi harga *bitcoin*, evaluasi kinerja dan hasil perbandingan metode.

Preprocessing Data

Dataset yang digunakan merupakan data *bitcoin* dengan rentang waktu 30 September 2021 sampai dengan 30 Desember 2021 dari laman <https://investing.com/crypto/bitcoin/historical-data>. fitur yang digunakan pada penelitian ini antara lain:

- Date : tanggal transaksi bitcoin
- Price : harga bitcoin pada tanggal terkait
- Open : harga awal bitcoin pada tanggal terkait
- High : harga tertinggi bitcoin pada tanggal terkait
- Low : harga terendah bitcoin pada tanggal terkait
- Volume : total transaksi perdagangan bitcoin pada tanggal terkait
- Change %: perubahan nilai harga bitcoin pada tanggal terkait terhadap harga bitcoin pada tanggal sebelumnya.

Sampling dataset yang digunakan dapat dilihat pada Tabel 1 di bawah ini.

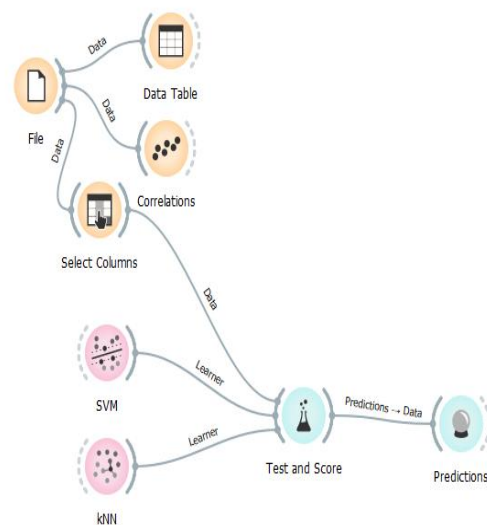
Tabel 1. Sampling Dataset

Date	Price	Open	High	Low	Vol.	Change %
12/30/2021	4674 0.7	4647 0.7	4692 2	4600 3	69.6 5K	0.006
12/29/2021	4646 1.7	4754 8.4	4812 1.7	4612 7.8	63.9 2K	-0.0228
12/28/2021	4754 5.2	5070 3.4	5070 3.8	4734 5.7	74.3 9K	-0.0618
12/27/2021	5067 8.2	5078 3.6	5201 6.3	5045 9	43.9 0K	-0.002
12/26/2021	5077 9.9	5040 4	5124 7	4952 7	32.3 4K	0.0074
12/25/2021	5040 6.4	5078 9.6	5114 0.2	5019 6.8	26.6 7K	-0.0075
12/24/2021	5078 6.1	5080 7.4	5179 5.3	5042 8.4	45.9 7K	-0.0003
12/23/2021	5080 1	4859 9.6	5135 5.6	4805 3.4	56.2 4K	0.0453
12/22/2021	4859 9.7	4889 8.7	4954 0.3	4844 9.7	40.8 9K	-0.0062
12/21/2021	4890 2	4690 0.1	4930 6.1	4665 4	57.8 8K	0.0428
12/20/2021	4689 5.2	4670 3.4	4748 8.9	4557 7.3	57.5 3K	0.0041
12/19/2021	4670 4.5	4685 5.7	4827 0.2	4643 8.4	44.0 5K	-0.0032
12/18/2021	4685 6.2	4617 1.8	4733 0.2	4551 3.4	36.9 3K	0.0148
12/17/2021	4617 2.2	4763 0.4	4799 7.2	4552 3.8	67.5 0K	-0.0306
12/16/2021	4762 8.2	4887 6.6	4938 7.8	4754 3.9	48.2 0K	-0.0254

Pada tahap ini dilakukan perubahan value Date yang bertipe *string* menjadi *numeric* dan karena pada data ini tidak ada yang mengalami *missing value* sehingga hanya dilakukan pemilihan data saja.

Data Mining Process

Model prediksi harga *bitcoin* dapat dianalisa menggunakan orange tool untuk memilih metode terbaik. Model prediksi dataset harga *bitcoin* dapat dilihat pada Gambar 2. dibawah ini.



Gambar 2. Design model prediksi harga *bitcoin*
Sumber: Peneliti (2022)

Setelah dilakukan preprocessing data, maka selanjutnya dataset diproses kedalam model prediksi pada software orange dengan menggunakan metode K-NN dan SVM.

K-Nearest Neighbor (K-NN)

Algoritma K-NN memiliki langkah-langkah sebagai berikut ((Fatah et al., 2018):

- Tentukan parameter K (jumlah banyaknya tetangga terdekat)
- Hitung jarak antara sample data uji dan seluruh sample data penelitian
- Urutkan jaraknya
- Ambil tetangga terdekat
- Kumpulkan/tentukan kategori tetangga terdekat
- Gunakan mayoritas kategori sederhana/tentukan kategori yang

pakings ering muncul (mayoritas) sebagai nilai prediksi dari data baru.

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama, diantaranya *euclidean distance* dan *manhattan distance (city block distance)*, yang paling sering digunakan adalah *euclidean distance*.

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah pengklasifikasi diskriminatif yang menghasilkan hyperplane pemisah. Toleransi kesalahan disertakan untuk membuat hyperplane pemisah menjadi kuat jika terjadi data kelas yang tidak dapat dipisahkan (Greaves, Alex and Au, 2015). Dalam Algoritma SVM ada trik kernel dimana ada SVM *linear* dan SVM *nonlinear*. SVM *hyperplane linear* bekerja hanya pada data yang dapat dipisahkan dengan cara *linear*. SVM *Non Linear* yaitu data yang berdistribusi pada kelas yang tidak linear sering digunakan pendekatan kernel pada fitur awal set (Drajana, 2017).

Kernel dapat diartikan sebagai suatu fungsi yang memetakan fitur data yang memiliki dimensi awal rendah fitur lainnya yang berdimensi lebih tinggi bahkan jauh lebih tinggi.

Root Mean Square Error (RMSE)

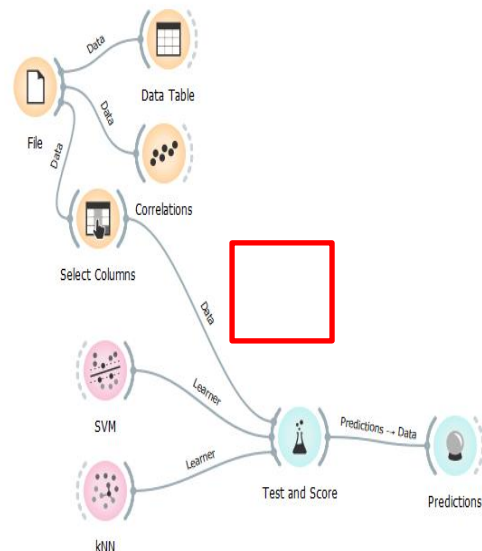
Keakuratan dari sebuah peramalan diperlukan evaluasi terhadap data yang sebenarnya. Ada beberapa metode yang sering digunakan untuk melakukan perhitungan kesalahan peramalan, salah satunya adalah *root mean square error (RMSE)*. Metode ini sering digunakan dalam mengevaluasi hasil peramalan. Ukuran yang sering digunakan dari perbedaan antara nilai-nilai diprediksi oleh model atau estimator dan nilai-nilai yang benar-benar diamati yaitu RMSE (Drajana, 2017).

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Hasil *Mean Absolute Error* menunjukkan nilai rata-rata kesalahan mutlak atau absolut dari nilai sebenarnya dengan nilai peramalan. Keakuratan dari hasil perhitungan prediksi menggunakan metode *Mean Absolute Error (MAE)* yang digunakan untuk mengukur tingkat

keakuratan suatu model peramalan (Suryanto, 2019).

Proses Evaluasi Hasil Perbandingan Model Prediksi



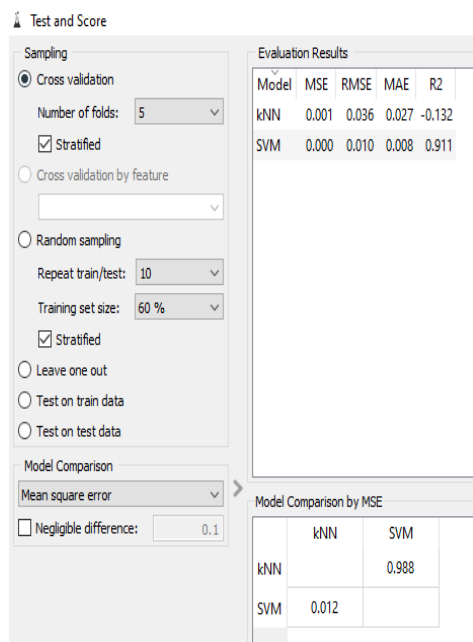
Gambar 3. Design untuk menghitung keberhasilan model prediksi
Sumber: Peneliti (2022)

Pada Gambar 3. Merupakan proses evaluasi hasil perbandingan dari model prediksi yang diuji. Proses perhitungan keberhasilan model prediksi pada orange tool dapat menggunakan *widget Test and Score* dan hasil prediksinya dapat dilihat menggunakan *widget Predictions*.

3. Hasil dan Pembahasan

3.1. Hasil simulasi 2 model prediksi

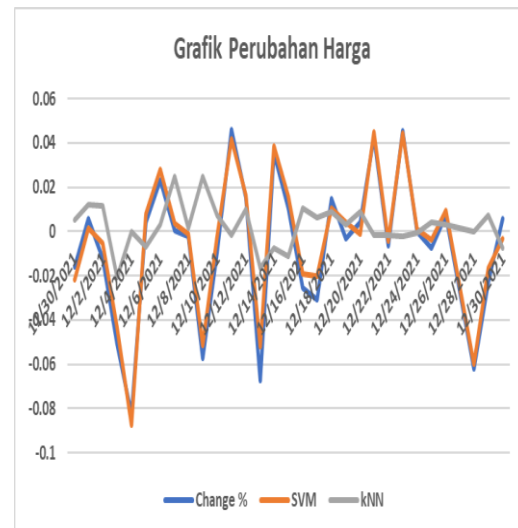
Data yang telah melalui tahap preprocessing selanjutnya dilakukan pengujian untuk mendapatkan model prediksi terbaik. Model prediksi yang telah dilakukan pengujian dan evaluasi dengan menggunakan kumpulan data uji pada aplikasi orange dimana 1 atribut sebagai targetnya didapatkan hasil simulasi seperti pada Gambar 4.



Gambar 4. Hasil *Test and Score*
Sumber: Peneliti (2022)

Gambar 4 merupakan penampakan dari *widget Test and Score* aplikasi Orange. Pada penelitian ini proses pengujian menerapkan K-Fold Cross Validation (K=5) yang dapat diatur pada *widget Test and Score* seperti yang terlihat pada gambar. Pada *widget* tersebut juga diperlihatkan hasil evaluasi kedua metode, dimana hasil perbandingan dari kedua model prediksi harga *bitcoin* ditunjukkan dengan nilai MSE, RMSE, MAE dan R2.

Penelitian kali ini hanya memperhatikan nilai MAE dan RMSE dalam melihat kinerja dari model prediksi yang analisa. Hasil evaluasi menunjukkan bahwa kedua model prediksi harga *bitcoin* memiliki nilai akurasi yang berbeda. *Support Vector Machine (SVM)* memiliki nilai MAE dan RMSE lebih kecil dibandingkan dengan *K-NN* yaitu 0.008 dan 0.010. Hal tersebut menunjukkan bahwa metode terbaik untuk prediksi harga *bitcoin* yaitu *Support Vector Machine (SVM)*.



Gambar 5. Grafik Perubahan Harga Hasil Prediksi
Sumber: Peneliti (2022)

Grafik pada gambar 5 menunjukkan perubahan harga *bitcoin* hasil prediksi menggunakan 2 metode. Perubahan harga hasil prediksi metode *Support Vector Machine (SVM)* dan *K-Nearest Neighbor (K-NN)* menunjukkan secara jelas bahwa prediksi perubahan harga *bitcoin* menggunakan metode SVM yang paling baik dibanding metode K-NN, dikarenakan tingkat *error* / selisih nilai metode K-NN sangat tinggi dan terlalu jauh hasil prediksinya dengan perubahan harga *bitcoin* sesungguhnya. Hasil prediksi perubahan harga *bitcoin* menggunakan metode *Support Vector Machine (SVM)* mendekati perubahan harga *bitcoin* sesungguhnya/ tingkat *error*nya kecil.

3.2. Hasil Prediksi Metode SVM

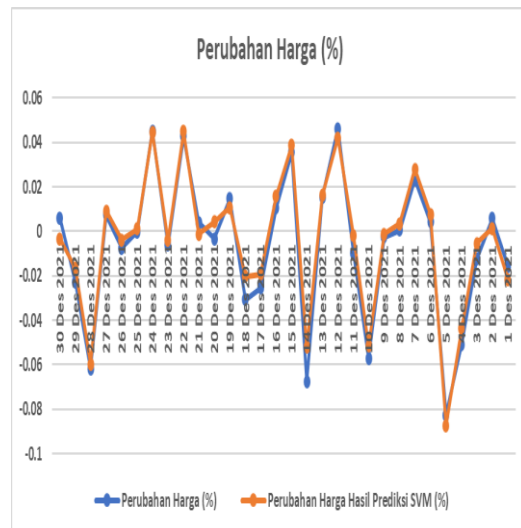
Tabel 2. Hasil Prediksi menggunakan Metode SVM

Data	Perubahan Harga (%)	Perubahan Harga Hasil Prediksi SVM (%)
1	0.006	-0.00340201
2	-0.0228	-0.016615
3	-0.0618	-0.0597342
4	0.0074	0.00930472
5	-0.0075	-0.0038535
6	-0.0003	0.00106415
7	0.0453	0.0445473

Data	Perubahan Harga (%)	Perubahan Harga Hasil Prediksi SVM (%)
8	-0.0062	-0.00421735
9	0.0428	0.0449605
10	0.0041	-0.00120672
11	-0.0032	0.00427539
12	0.0148	0.010638
13	-0.0306	-0.0201877
14	-0.0254	-0.0192562
15	0.0104	0.0157007
16	0.0356	0.0386921
17	-0.0674	-0.0521063
18	0.0155	0.0161676
19	0.0462	0.0419993
20	-0.0097	-0.0021335
21	-0.057	-0.0515659
22	-0.0024	-0.00127377
23	0.0007	0.00352398
24	0.0234	0.0280475
25	0.0043	0.007646
26	-0.0827	-0.0872493
27	-0.0507	-0.0435935
28	-0.0125	-0.0052806
29	0.0058	0.0012716
30	-0.0158	-0.0215032

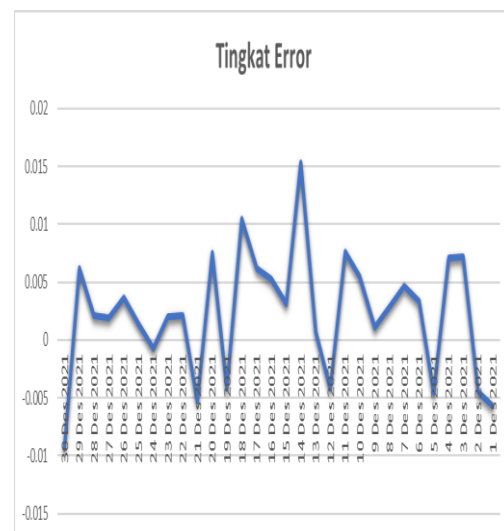
Tabel 2 di atas menunjukkan *sampling* perubahan nilai harga *bitcoin* dan hasil prediksi perubahan nilai harga *bitcoin* menggunakan metode *Support Vector Machine (SVM)*. Data yang ditampilkan hanya beberapa saja, dikarenakan terlalu banyak data yang digunakan.

Dari 30 data yang ditampilkan selisih perubahan nilai harga *bitcoin* tidak terlalu jauh antara perubahan harga *bitcoin* sebenarnya dan perubahan harga *bitcoin* berdasarkan hasil prediksi dari metode *Support Vector Machine*.



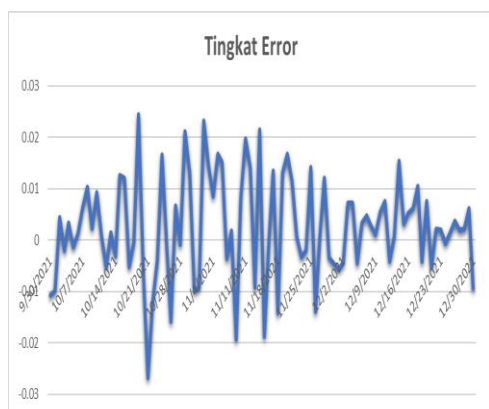
Gambar 6. Grafik Hasil Prediksi Perubahan Harga
Sumber: Peneliti (2022)

Grafik pada gambar 6 menunjukkan bahwa selisih nilai/error antara perubahan harga sesungguhnya dan perubahan harga hasil prediksi metode *Support Vector Machine (SVM)* mendekati nilai sesungguhnya, namun ada beberapa nilai yang selisihnya jauh dari perubahan harga *bitcoin* sesungguhnya. Data yang disajikan pada grafik di atas hanya 30 data saja sebagai sampel.



Gambar 7. Grafik Tingkat Error
Sumber: Peneliti (2022)

Grafik tingkat error dataset bitcoin di atas dapat dilihat bahwa tingkat error tertinggi sebesar 0,15 dan nilai tingkat error terendah sebesar -0,01.



Gambar 8. Grafik Tingkat Error keseluruhan Dataset

Sumber: Peneliti (2022)

Gambar 8 merupakan grafik tingkat error dari perubahan harga bitcoin antara perubahan harga sebenarnya dengan perubahan harga hasil prediksi metode *Support Vector Machine (SVM)*. Dari data tingkat error didapatkan rata-rata tingkat error perubahan harga bitcoin yang telah diprediksi menggunakan metode *Support Vector Machine (SVM)* sebesar 0,0025.

4. Kesimpulan

Hasil implementasi orange data mining untuk prediksi harga *bitcoin* dapat dilihat pada *widget Test and Score*, dimana diperoleh hasil nilai *RMSE* dan *MAE* dari masing-masing model. Nilai *RMSE* dan *MAE* dari metode *K-Nearest Neighbor (K-NN)* sebesar 0.036 dan 0.027. Sedangkan metode *Support Vector Machine* memiliki nilai *RMSE* dan *MAE* sebesar 0.010 dan 0.008.

Berdasarkan nilai *RMSE* dan *MAE* tersebut dapat disimpulkan bahawa metode *Support vector Machine (SVM)* merupakan metode terbaik untuk memprediksi perubahan harga *bitcoin* dibandingkan dengan metode *K-Nearest Neighbor (K-NN)*. Metode terbaik pada penelitian ini dapat diteliti kembali untuk penelitian yang akan mendatang. Penelitian selanjutnya disarankan dapat membandingkan metode *SVM* dengan metode lainnya untuk memprediksi harga bitcoin menggunakan tools orange atau dapat menerapkan tools lainnya.

Referensi

- Alim, S. (2021). *IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR*, 6(2), 133–144.
- Derbentsev, V., Babenko, V., Khrestalev, K., Obruch, H., & Khrestalova, S. (2021). *Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices*. 34(01), 140–148.
- Drajana, I. C. R. (2017). Metode *Support Vector Machine* Dan *Forward Selection* Prediksi Pembayaran Pembelian Bahan Baku Kopro. *ILKOM Jurnal Ilmiah*, 9(2), 116–123. <https://doi.org/10.33096/ilkom.v9i2.134.116-123>
- Fatah, H., Subekti, A., Komputer, S. I., & Komputer, I. (2018). *K-NEAREST NEIGHBOURS*. 14(2), 137–144.
- Greaves, Alex and Au, B. (2015). Using the bitcoin transaction graph to predict the price of bitcoin. *Quoted*, 3, 22. http://snap.stanford.edu/class/cs224w-2015/projects_2015/Using_the_Bitcoin_Transaction_Graph_to_Predict_the_Price_of_Bitcoin.pdf
- Indriyanti, I., & Subekti, A. (2018). *Pemodelan Prediktif Konsumsi Energi Bangunan Gedung Komersial Dengan Algoritma Support Vector Machine*. *Jurnal Pilar Nusa Mandiri*, 14(2), 225. <https://doi.org/10.33480/pilar.v14i2.882>
- Pinudom, B., Tansuchat, R., Reza, M., Fauzi, R., Nasution, M., Marisa, W., Parlita, R., & Pratama, A. (2019). *Backpropagation neural network prediction for cryptocurrency bitcoin prices* *Backpropagation neural network prediction for*. <https://doi.org/10.1088/1742-6596/1339/1/012060>
- Rizkilloh, M. F., & Widiyanesti, S. (2022). *Prediksi harga Cryptocurrency Menggunakan Algoritma Long Short term Memory (LSTM)*. 5, 1–2.
- Saadah, S., & Salsabila, H. (2021). *Jurnal Politeknik Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest*. 7(1), 24–32.

- Septiarini, T. W., Taufik, M. R., & Afif, M. (2020). *A comparative study for Bitcoin cryptocurrency forecasting in period A comparative study for Bitcoin cryptocurrency forecasting in period 2017-2019*.
<https://doi.org/10.1088/1742-6596/1511/1/012056>
- Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, 11(1), 78–83.
<https://doi.org/10.32764/saintekbu.v11i1.298>
- Swamy, U. R., Rupini, D., Shravya, G., & Akhila, V. (2021). *Crypto-Currency Price Prediction Using Machine Learning Techniques*. 8(6), 212–215.