

COMPARISON OF DATA MINING ALGORITHM FOR FORECASTING BITCOIN CRYPTO CURRENCY TRENDS

Indri Tri Julianto^{*1}, Dede Kurniadi², Muhammad Rikza Nashrulloh³, Asri Mulyani⁴

^{1,2,3,4}Jurusan Ilmu Komputer, Institut Teknologi Garut, Indonesia

Email: ¹indritrijulianto@itg.ac.id, ²dede.kurniadi@itg.ac.id, ³rikza@itg.ac.id, ⁴asrimulyani@itg.ac.id

(Naskah masuk: 01 Maret 2022, Revisi : 04 Maret 2022, diterbitkan: 25 April 2022)

Abstract

The popularity of cryptocurrencies has been increasing in the approximately 10 years since their emergence in 2008. Bitcoin is the most popular and the most instrumental in the existence of cryptocurrencies. The price of coins in cryptocurrencies is the same as the price of shares in the capital market which always fluctuates and even tends to be more volatile than the stock market. This condition is very influential for actors in cryptocurrencies. This study aims to compare the Algorithm Forecasting so that it can be known the right algorithm in Forecasting the trend of Bitcoin. The algorithm used is Algorithm Supervised Learning that is Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, and polynomial Regression. Accuracy was measured using a 10 Fold Cross-validation model and evaluation is done by Root Mean Square Error (RMSE). The results showed that the Algorithm Neural Network is an Algorithm Forecasting best with RMSE value 277,237 +/- 74,736 (micro: 287,208 +/- 0.000) among other Algorithms so that Neural Network can be used for Forecasting cryptocurrency Bitcoin.

Keywords: Algorithm, Bitcoin, Cryptocurrencies, Forecasting, RMSE.

KOMPARASI ALGORITMA DATA MINING UNTUK FORECASTING TREN MATA UANG KRIPTO BITCOIN

Abstrak

Popularitas mata uang kripto semakin meningkat dalam kurun waktu kurang lebih 10 tahun sejak kemunculannya pada tahun 2008. *Bitcoin* adalah yang paling populer dan yang paling berperan dalam eksistensi mata uang kripto. Harga koin dalam mata uang kripto sama seperti harga pada saham di pasar modal yang selalu mengalami fluktuasi dan bahkan cenderung lebih fluktuatif dibandingkan pasar saham. Kondisi ini sangat berpengaruh bagi para pelaku di mata uang kripto. Penelitian ini bertujuan untuk membandingkan Algoritma Forecasting sehingga dapat diketahui Algoritma yang tepat dalam Forecasting tren *Bitcoin*. Algoritma yang digunakan adalah Algoritma Supervised Learning yaitu Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, dan Polynomial Regression. Akurasi diukur dengan menggunakan model validasi 10 Fold Cross-Validation dan evaluasi dilakukan dengan Root Mean Square Error (RMSE). Hasil penelitian menunjukkan bahwa Algoritma Neural Network merupakan Algoritma Forecasting terbaik dengan nilai RMSE 277.237 +/- 74.736 (mikro: 287.208 +/- 0.000) diantara Algoritma lainnya, sehingga Neural Network dapat digunakan untuk Forecasting mata uang kripto Bitcoin.

Kata kunci: Algoritma, Bitcoin, Forecasting, Kripto, RMSE.

1. PENDAHULUAN

Mata uang kripto atau mata uang yang (dienkripsi) merupakan jenis mata uang digital yang tidak dipengaruhi atau melibatkan Institusi Keuangan/Pemerintah melainkan ada di *Blockchain* atau disebut juga dengan teknologi pembukuan terdistribusi [1],[2]. *Bitcoin* merupakan salah satu mata uang kripto dimana transaksinya dijamin oleh teknik kriptografi yang sangat aman [2]. *Bitcoin* adalah mata uang kripto yang pertama kali muncul

dan yang masih populer hingga saat ini, nama *Bitcoin* menjadi identik dengan *Blockchain* itu sendiri. Kehadiran mata uang kripto merupakan jawaban bagi kebutuhan transaksi pada zaman digital sekarang yang mudah, cepat, transparan serta dapat diterima oleh kedua belah pihak yang melakukan transaksi [1].

Data Mining merupakan proses penambangan data untuk mencari pola yang menarik untuk menghasilkan pengetahuan dari sekumpulan data

yang jumlahnya sangat besar (*Big Data*) [3]–[6]. Terdapat dua model dalam *Data Mining*, yaitu *Supervised Learning* dan *Un-Supervised Learning*. Perbedaan keduanya terletak pada label atau target keluaran, dimana *Supervised Learning* itu memiliki label sedangkan *Un-Supervised Learning* tidak memiliki label [7].

Terdapat beberapa penelitian sebelumnya yang telah membahas mengenai *Forecasting*. Pertama adalah penelitian mengenai Analisis dan Perbandingan Algoritma Data Mining dalam Prediksi Harga Saham GGRM [8]. Kemudian penelitian kedua adalah mengenai *Forecasting* Harga *Cryptocurrency* Menggunakan Teknik *Data Mining* [2]. Penelitian ketiga mengenai Komparasi Model *Support Vector Machines* (SVM) dan *Neural Network* Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham [9]. Penelitian keempat mengenai *Forecasting* Model Penyakit Demam Berdarah *Dengue* di Provinsi DKI Jakarta Menggunakan Algoritma Regresi Linier Untuk Mengetahui Kecenderungan Nilai Variabel Prediktor Terhadap Peningkatan Kasus [7]. Adapun untuk lebih jelas mengenai perbandingan ke empat penelitian tersebut, maka disajikan dalam bentuk *Roadmap* Penelitian sebagaimana tampak pada Tabel 1.

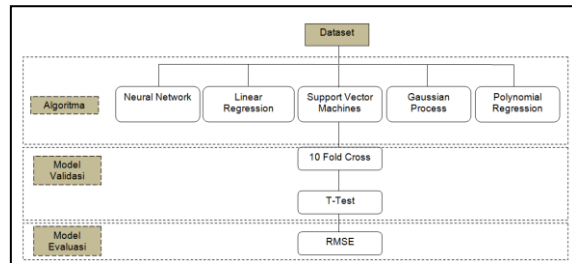
Tabel 1. *Roadmap* Penelitian

Penelitian	Algoritma	Tujuan Penelitian	Validasi & Evaluasi
1	<i>Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, Polynomial Regression</i>	<i>Forecasting</i>	<i>10 Fold Cross & RMSE</i>
2	<i>Random Forest, SVM, Naive Bayes</i>	<i>Forecasting</i>	ARIMA dan <i>Bayesian Structural Time Series (BSTS)</i>
3	<i>Neural Network, SVM</i>	<i>Forecasting</i>	<i>Performance prediction trend accuracy</i>
4	<i>Linear Regression</i>	<i>Forecasting</i>	<i>T-Test & RMSE</i>

Penelitian ini akan mengisi kesenjangan antara penelitian sebelumnya dengan penelitian sekarang dengan cara menggunakan Algoritma *Neural Network, Linear Regression, Support Vector Machine, Gaussian Process, dan Polynomial Regression* dengan tujuan untuk komparasi Algoritma *Forecasting* tersebut dengan Model Validasi *10 Fold Cross-Validation* dengan *T-Test* serta model Evaluasi dengan menggunakan *Root Mean Square Error (RMSE)*.

2. METODE PENELITIAN

Usulan kerangka penelitian disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 1, yang terdiri dari *Dataset*, Algoritma *Forecasting*, Model Validasi, dan Model Evaluasi.

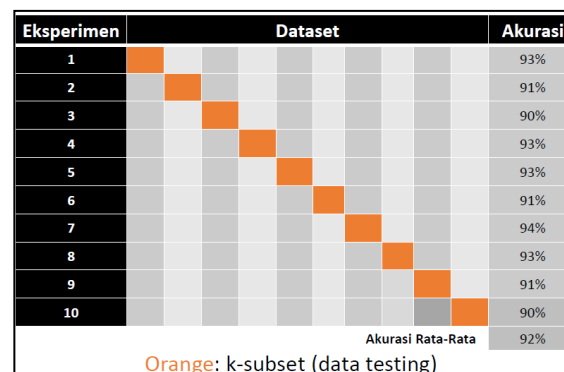


Gambar 1. Kerangka Penelitian [10]

Tahap pertama adalah pemilihan *Dataset*. *Dataset* yang digunakan merupakan data historis mata uang kripto [11]. *Dataset* ini mencakup *Bitcoin* dengan atribut seperti tanggal, harga pembukaan, harga penutupan, tertinggi, terendah, volume, dan kapitalisasi pasar yang didokumentasikan setiap hari dari 29 April 2013 hingga 6 Juli 2021.

Tahap kedua adalah menggunakan Algoritma *Data Mining Forecasting* yang terdiri dari Algoritma *Neural Network, Linear Regression, Support Vector Machines, Gaussian Process, dan Polynomial Regression*.

Tahap ketiga adalah Model Validasi. Model yang digunakan adalah *10 Fold Cross-Validation* merupakan *nasted operator* [12], dimana secara konsep validasi ini membagi data *training* menjadi 10 bagian yang sama dan kemudian melakukan pembelajaran sebanyak 10 kali. Setiap kali dipilih bagian lain dari *dataset* untuk pengujian dan sembilan data sisanya digunakan untuk pembelajaran. Kemudian, dihitung nilai rata-rata dan nilai penyimpangan dari 10 hasil pengujian berbeda [8]. Adapun tahapan model validasi ini disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 2.



Gambar 2. Model Validasi *10 Fold Cross*

Tahap keempat adalah Model evaluasi. Evaluasi yang digunakan dalam penelitian ini adalah *Root Mean Square Error (RMSE)*. Jika semakin kecil nilai *RMSE*, maka akurasi *forecasting* akan semakin baik.

3. HASIL DAN PEMBAHASAN

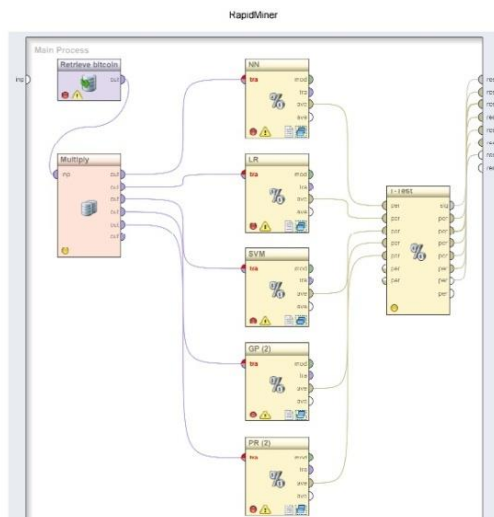
Sebelum dilakukan komparasi, dataset yang ada akan dilakukan tahapan *Preprocessing* dengan tujuan untuk membersihkan data [13],[14]. Tahapan ini dilakukan dengan cara manual, karena setelah diamati datanya cukup konsisten, tinggal memilah atribut mana saja yang akan digunakan. Adapun teknik dalam *Preprocessing* adalah sebagai berikut [15]:

1. Data *cleaning* bertujuan untuk membersihkan nilai yang kosong, tidak konsisten atau *missing value* dan *noisy*;
2. Data *integration* bertujuan untuk menyatukan tempat penyimpanan (data) yang berbeda dalam satu arsip.
3. Data *reduction* bertujuan untuk mengurangi jumlah atribut yang tidak digunakan, seperti halnya dalam *dataset* ini yang dihapus adalah atribut nama dan simbol.

Adapun tampilan *dataset* hasil *Preprocessing* disajikan dalam bentuk gambar, sebagaimana tampak pada Gambar 3.

Date	High	Low	Open	Close	Volume	Marketcap
4/29/2013 23:59	147.488	134	134.444	144.54	0	1603768865
4/30/2013 23:59	146.93	134.05	144	139	0	1542813125
5/1/2013 23:59	139.89	107.72	139	116.99	0	1298954594
5/2/2013 23:59	125.6	92.2819	116.38	105.21	0	1168517495
5/3/2013 23:59	108.128	79.1	106.25	97.75	0	1085995169
5/4/2013 23:59	115	92.5	98.1	112.5	0	1250316563
5/5/2013 23:59	118.8	107.143	112.9	115.91	0	1288693176
5/6/2013 23:59	124.663	106.64	115.98	112.3	0	1249023060
5/7/2013 23:59	113.444	97.7	112.25	111.5	0	1240593600
5/8/2013 23:59	115.78	109.6	109.6	113.566	0	1264049202
5/9/2013 23:59	113.46	109.26	113.2	112.67	0	1254535382
5/10/2013 23:59	122	111.551	112.799	117.2	0	1305479080
5/11/2013 23:59	118.679	113.01	117.7	115.243	0	1284207489
5/12/2013 23:59	117.449	113.435	115.64	115	0	1281982625
.....

Gambar 3. Dataset Harga Bitcoin



Gambar 4. Proses Model Validasi dan T-Test

Gambar 3 merupakan *Dataset* yang telah siap digunakan kemudian akan dimasukkan ke aplikasi *Data Mining* yaitu Rapidminer [16]. Tahap yang dilakukan selanjutnya adalah membuat model pada Rapidminer dengan memasukkan Algoritma *Forecasting*, kemudian dilakukan Validasi 10 *Fold Cross*, lalu dilakukan Uji beda menggunakan *T-Test*. Adapun model yang dibangun, disajikan sebagaimana tampak pada Gambar 4.

Kemudian hasil pertama yang didapatkan adalah nilai RMSE. Adapun nilai yang dihasilkan, disajikan dalam bentuk tabel, sebagaimana tampak pada Tabel 2.

Tabel 2. Nilai RMSE

No	Algoritma	RMSE
1	Neural Network	277.237 +/- 74.736 (mikro: 287.208 +/- 0.000)
2	Linear Regression	375.126 +/- 94.506 (mikro: 386.839 +/- 0.000)
3	Support Vector Machines	10083.984 +/- 1147.922 (mikro: 10149.239 +/- 0.000)
4	Gaussian Process	13095.632 +/- 1073.000 (mikro: 13139.262 +/- 0.000)
5	Polynomial Regression	69352814509.404 +/- 25546303501.967 (mikro: 73917429049.054 +/- 0.000)

Tabel 2 menunjukkan bahwa Algoritma *Neural Network* memiliki nilai RMSE yang paling rendah diantara Algoritma lainnya. Selanjutnya ditampilkan hasil *T-Test*, sebagaimana tampak pada Gambar 5.

T-Test Significance	NN	LR	SVM	GP	PR
277.237 +/- 74.736	0.001	0.000	0.000	0.000	0.000
375.126 +/- 94.506	0.000	0.000	0.000	0.000	0.000
10083.984 +/- 1147.922	0.000	0.000	0.000	0.000	0.000
13095.632 +/- 1073.000	0.000	0.000	0.000	0.000	0.000
69352814509.404 +/- 25546303501.967	0.000	0.000	0.000	0.000	0.000

Probabilities for random values with the same result.
Bold values are smaller than alpha=0.050 which indicates a probably significant difference between the actual mean values!

Gambar 5. Hasil T-Test

Gambar 5 merupakan hasil *T-Test* menunjukkan bahwa nilai yang tampil untuk tiap Algoritma adalah lebih kecil dari nilai Alpha yaitu 0,050, sehingga terjadi perbedaan yang signifikan antar Algoritma. Maka dari itu untuk urutan Algoritma terbaik disajikan pada Tabel 3.

Tabel 3. Nilai RMSE

No	Algoritma	RMSE	T-Test
1	Neural Network	277.237 +/- 74.736 (mikro: 287.208 +/- 0.000)	Significant Difference
2	Linear Regression	375.126 +/- 94.506 (mikro: 386.839 +/- 0.000)	Significant Difference
3	Support Vector Machines	10083.984 +/- 1147.922 (mikro: 10149.239 +/- 0.000)	Significant Difference
4	Gaussian Process	13095.632 +/- 1073.000 (mikro: 13139.262 +/- 0.000)	Significant Difference
5	Polynomial Regression	69352814509.404 +/- 25546303501.967 (mikro: 73917429049.054 +/- 0.000)	Significant Difference

Tabel 3 menunjukkan bahwa antara kelima algoritma tersebut memiliki hasil nilai *T-Test* yang berbeda secara signifikan, sehingga urutannya *Neural Network* merupakan algoritma terbaik.

4. KESIMPULAN

Adapun kesimpulan dari penelitian ini adalah sebagai berikut, hasil penelitian menunjukkan bahwa Algoritma *Neural Network* merupakan Algoritma *Forecasting* terbaik dengan nilai RMSE 277.237 +/- 74.736 (mikro: 287.208 +/- 0.000), lebih kecil dibandingkan dengan Algoritma lainnya. Selain itu hasil *T-Test* menunjukkan perbedaan signifikan antar Algoritma. Algoritma *Neural Network* sangat tepat digunakan untuk *Forecasting* mata uang kripto *Bitcoin*

DAFTAR PUSTAKA

- [1] R. C. Noorsanti, H. Yulianton, and K. Hadiono, "Blockchain - Teknologi Mata Uang Cryptocurrency," *Pros. SENDI_U 2018*, pp. 306–311, 2018.
- [2] A. P. Singh and S. Malani, "Understanding and Predicting Trends In Cryptocurrency Prices Using Data Mining Techniques," *IIIT Hyderabad*, pp. 1–7, 2018.
- [3] D. T. Larose and C. D. Larose, *Discovering Knowledge In Data An Introduction to Data Mining Second Edition Wiley Series on Methods and Applications in Data Mining*. Canada: John Wiley & Sons, Inc, 2014.
- [4] C. C. Aggarwal, *Data Mining: The Textbook*. New York: Springer, 2015.
- [5] M. Ardiansyah Sembiring, M. Fitri Larasati Sibuea, and A. Saptia, "Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar," *J. Sci. Soc. Res.*, vol. 1, no. February, pp. 73–79, 2018, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>.
- [6] A. A. Argasah and D. Gustian, "Data Mining Analysis To Determine Employee Salaries According To Needs Based On The K-Medoids Clustering Algorithm Analisis Data Mining Untuk Menentukan Gaji Karyawan Sesuai Penilaian Kemampuan Menggunakan Algoritma K-Medoids," *JUTIF*, vol. 3, no. 1, pp. 29–35, 2022.
- [7] A. R. Muhajir, E. Sutoyo, and I. Darmawan, "Forecasting Model Penyakit Demam Berdarah Dengue Di Provinsi DKI Jakarta Menggunakan Algoritma Regresi Linier Untuk Mengetahui Kecenderungan Nilai Variabel Prediktor Terhadap Peningkatan Kasus," *Fountain Informatics J.*, vol. 4, no. 2, pp. 33–40, 2019.
- [8] R. Maulana and D. Kumalasari, "Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham GGRM," *J. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://finance.yahoo.com/quote/GGRM.J>.
- [9] R. H. Kusumodestoni and Sarwido, "Komparasi Model Support Vector Machines (SVM) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham," *J. Inform. UPGRIS*, vol. 3, no. 1, pp. 1–9, 2017, doi: 10.26877/jiu.v3i1.1536.
- [10] P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study : Internet Selebgram Rachel Vennya Escape From Quarantine) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *JUTIF*, vol. 3, no. 1, pp. 141–145, 2022.
- [11] S. Kumar, "Cryptocurrency Historical Prices Dataset," *Kaggle*, 2018. <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory/%0Aversion/13>.
- [12] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M. Arifin, "Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification," *J. Phys. Conf. Ser.*, vol. 1402, no. 6, pp. 1–6, 2019, doi: 10.1088/1742-6596/1402/6/066055.
- [13] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Technique*. San Francisco: Morgan Kaufmann, 2011.
- [14] N. Ye, *Data Mining: Theories, Algorithms, and Examples*. New York: Taylor & Francis Group, 2014.
- [15] Han and Kamber, *Data Mining Concepts and Technique*. San Francisco: Diane Cerra, 2006.
- [16] K. Fatmawati and A. P. Windarto, "Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 3, no. 2, pp. 173–178, 2018.