

PERBANDINGAN ALGORITMA *LINEAR REGRESSION*, *NEURAL NETWORK*, *DEEP LEARNING*, DAN *K-NEAREST NEIGHBOR (K-NN)* UNTUK PREDIKSI HARGA BITCOIN

Mikhael¹, Felix Andreas², Ultach Enri³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
email: mikhael.amos18155@student.unsika.ac.id¹, felix.andreas18119@student.unsika.ac.id²,
ultach@staff.unsika.ac.id³

Abstrak

Bitcoin merupakan mata uang digital yang menggunakan sistem kriptografi pertama di dunia. Tujuan utama diciptakannya Bitcoin adalah memungkinkan kedua belah pihak untuk melakukan transaksi secara langsung tanpa campur tangan pihak ketiga. Meskipun Bitcoin merupakan sebuah mata uang, banyak orang yang menggunakan Bitcoin sebagai alat untuk berinvestasi karena harganya cenderung naik cepat dalam waktu singkat. Namun, bukan berarti tidak memiliki risiko. Berinvestasi ke Bitcoin memiliki risiko yang tinggi karena volatilitas harganya sangat tinggi. Penelitian ini bertujuan untuk membandingkan algoritma yang digunakan untuk memprediksi harga Bitcoin. Dalam penelitian ini akan dilakukan prediksi terhadap harga Bitcoin dengan membandingkan empat model algoritma yaitu *Linear Regression*, *Neural Network*, *Deep Learning*, dan *k-Nearest Neighbor (k-NN)*. Tingkat akurasi dari tiap model algoritma akan diuji dengan metode validasi *K-Fold Cross Validation* dan dievaluasi menggunakan *Root Mean Square Error (RMSE)*. Hasil dengan uji T-Test dalam penelitian ini menyimpulkan bahwa model terbaik untuk memprediksi harga Bitcoin adalah model algoritma *Linear Regression* dan *Neural Network*, yaitu dengan hasil RMSE 296.227 +/- 60.125 (*micro average*: 301.655 +/- 0.000) dan 338.988 +/- 47.837 (*micro average*: 342.000 +/- 0.000).

Kata kunci: Perbandingan algoritma, Bitcoin, prediksi

Abstract

Bitcoin is the world's first digital currency that uses cryptographic system. The motive behind Bitcoin creation is allows two willing parties transact directly without the intervention of third party. Despite Bitcoin is a currency, there are many people who use Bitcoin as an investment as its price tends to rise quickly in short time. However, that doesn't mean Bitcoin have no risk. Investing in Bitcoin have a high-risk due to its price high volatility. This research aims to compare algorithms that used for predict Bitcoin price. In this research, Bitcoin price prediction will be made by comparing four algorithm models, namely Linear Regression, Neural Network, Deep Learning, and k-Nearest Neighbor (k-NN). Each algorithm accuracy will be tested using the K-Fold Cross Validation method and evaluated using Root Mean Square Error (RMSE). The results of T-Test interpretation in this research conclude that the best models for predict Bitcoin price are Linear Regression and Neural Network model, with the result of RMSE 296.227 +/- 60.125 (micro average: 301.655 +/- 0.000) and 338.988 +/- 47.837 (micro average: 342.000 +/- 0.000).

1. PENDAHULUAN

Bitcoin merupakan mata uang virtual yang menggunakan prinsip kriptografi yang berjalan dengan mekanisme *blockchain* dan *proof-of-work* dalam sistem terdistribusi dan terdesentralisasi, untuk itu Bitcoin sering disebut sebagai *cryptocurrency*. Bitcoin merupakan *cryptocurrency* pertama di dunia yang diajukan oleh Satoshi Nakamoto di pertengahan tahun 2008 dan awal tahun 2009 [1]. Tujuan utama diciptakannya Bitcoin adalah memungkinkan kedua belah pihak untuk melakukan transaksi secara langsung

tanpa campur tangan pihak ketiga [2].

Kenaikan harga Bitcoin yang drastis membuat orang menjadi tertarik untuk berinvestasi. Di awal tahun 2020 harga Bitcoin mencapai sekitar US\$ 7.000 dan naik hingga sekitar US\$ 29.000 di akhir tahun 2020. Di Indonesia sendiri, transaksi mata uang kripto mengalami lonjakan yang sangat tinggi. Dilansir dari situs berita CNBC Indonesia, investor di pasar kripto mengalami kenaikan hingga mencapai 6,5 juta investor, di mana tahun sebelumnya yaitu tahun 2020 hanya ada 2,5 juta investor di pasar kripto. Volume perdagangan di pasar kripto juga mengalami kenaikan yang sangat signifikan hingga mencapai Rp370 triliun di mana pada tahun 2020 hanya mencapai Rp 60 triliun [3].

Ada beberapa penelitian yang sebelumnya telah dilakukan menggunakan beberapa model algoritma untuk prediksi yang mendukung penelitian ini. Penelitian yang dilakukan oleh [4] telah melakukan prediksi harga *cryptocurrency* dengan algoritma *Deep Learning*, hasil terbaik pada prediksi harga Bitcoin dengan nilai RMSE 375,5336079.

Penelitian lain yang dilakukan oleh [5] membandingkan lima algoritma untuk memprediksi harga saham GGRM yaitu *Neural Network*, *Linear Regression*, *Support Vector Machine*, *Gaussian Process*, dan *Polynomial Regression*. Hasil prediksi terbaik ditunjukkan oleh model algoritma *Neural Network* dengan nilai RMSE 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000). Sementara itu di posisi kedua terdapat algoritma *Linear Regression* dengan nilai RMSE 659.806 +/-76.480 (mikro: 664.360 +/-0.000)

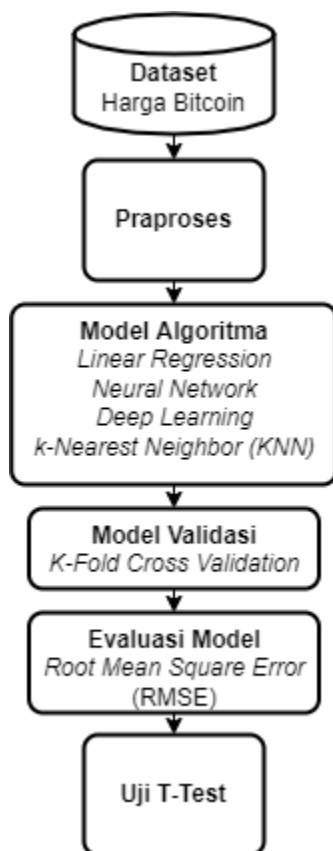
Penelitian lain dilakukan oleh [6] dengan membandingkan metode jaringan saraf tiruan dan pohon keputusan untuk memprediksi *churn*. Dalam penelitian ini model jaringan saraf tiruan menghasilkan hasil evaluasi terbaik dengan akurasi mencapai 86%.

Penelitian yang dilakukan oleh [7] tentang prediksi potensi hujan harian menggunakan algoritma k-Nearest Neighbor (KNN). Hasil pengujian didapatkan hasil RMSE 12.493+/-0.000.

Dalam penelitian ini akan dilakukan prediksi terhadap harga Bitcoin dengan membandingkan empat model algoritma yaitu *Linear Regression*, *Neural Network*, *Deep Learning*, dan *k-Nearest Neighbor (k-NN)*. Alasan penulis membandingkan keempat model algoritma tersebut karena berdasarkan referensi pada penelitian ini, keempat algoritma tersebut memiliki akurasi yang cukup akurat untuk melakukan prediksi suatu fenomena. Adapun tujuan penelitian ini adalah untuk mengetahui model algoritma (*Linear Regression*, *Neural Network*, *Deep Learning*, dan *k-Nearest Neighbor*) yang paling akurat untuk memprediksi harga Bitcoin.

2. METODE PENELITIAN

Tahapan penelitian ini dimulai dengan mengambil *dataset* harga Bitcoin kemudian melakukan pra-proses, pembuatan model algoritma, pembuatan model validasi, dan evaluasi model dengan *Root Mean Square Error (RMSE)* seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Dataset

Dalam penelitian ini *dataset* harga Bitcoin per hari diambil dari tanggal 1 Januari 2015 sampai tanggal 30 Oktober 2021 sebanyak 2496 data. Dataset dapat diperoleh dengan mengakses <https://coinmarketcap.com/currencies/bitcoin/historical-data/> [8]. Data terdiri dari 7 atribut dan satu atribut dipilih menjadi kelas. Atribut-atribut tersebut antara lain:

1. Date: atribut yang menyatakan rentetan waktu (*time series*)
2. Open: harga pembukaan Bitcoin
3. High: harga tertinggi Bitcoin selama satu hari
4. Low: harga terendah Bitcoin selama satu hari
5. Close: harga penutupan Bitcoin selama satu hari
6. Volume: volume perdagangan Bitcoin dalam US\$
7. Market Cap: nilai kapitalisasi pasar Bitcoin

2.2 Praproses

Tahapan praproses merupakan tahap awal dari pengolahan data. Praproses bertujuan agar data yang diuji akan berkualitas baik dan tidak ada kecacatan. Adapun teknik dalam praproses data [9] yaitu:

1. Data *cleaning*, yaitu membersihkan nilai yang kosong, tidak konsisten atau tupel kosong seperti *missing value* dan *noisy*.
2. Data *integration*, yaitu menyatukan data ke dalam satu arsip.
3. Data *reduction*, penghapusan atribut yang tidak diperlukan.

2.3 Model Algoritma

Pemodelan algoritma merupakan tahap pemilihan algoritma yang digunakan untuk memprediksi harga Bitcoin. Pada penelitian ini algoritma yang diterapkan untuk prediksi adalah *Linear Regression* (LR), *Neural Network* (NN), *Deep Learning* (DL), dan *k-Nearest Neighbor* (k-NN).

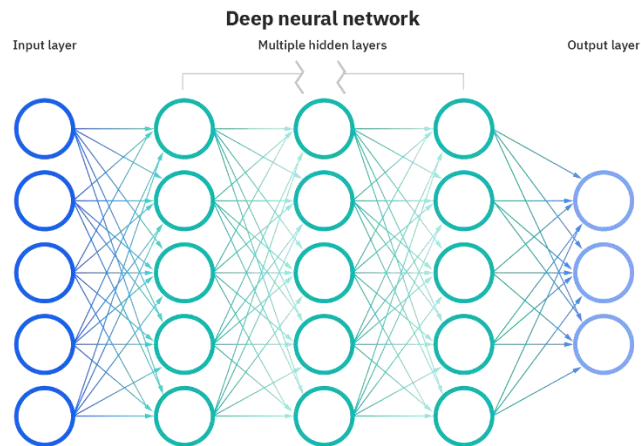
1) *Linear Regression*

Model regresi merupakan teknik untuk menganalisis estimasi nilai variabel dependen ‘y’ dengan rentang nilai variabel independen ‘x’. Regresi linear multivariat merupakan teknik statistika untuk memprediksi hasil dari variabel jawaban dengan menggunakan sejumlah variabel penjelas [10]. formula dasar untuk model regresi linear multivariat adalah seperti persamaan (1).

$$y = \beta_0 + \beta_1x_1 + \dots + \beta_mx_m + \varepsilon \quad \dots\dots\dots(1)$$

2) *Neural Network*

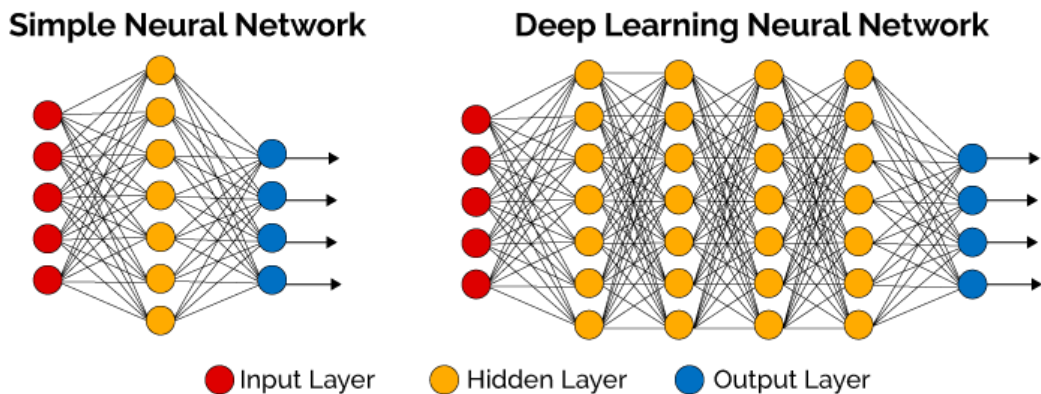
Neural Network sering disebut juga sebagai *artificial neural networks* (ANNs) adalah subset dari *machine learning*. *Neural Network* terinspirasi berdasarkan cara kerja otak manusia, meniru neuron biologi mengirim sinyanya satu sama lain. *Neural Network* tersusun atas lapisan-lapisan (*layers*) yang saling terhubung, terdiri atas *input layer*, *hidden layer*, dan *output layer*. Model struktur *Neural Network* dapat dilihat pada Gambar 2.



Gambar 2. Struktur *Neural Network* [11]

3) *Deep Learning*

Deep Learning merupakan perkembangan dari *backpropagation neural network*, *recurrent neural network* (RNN) karena ketidakmampuan RNN untuk menyimpan memori jang panjang terutama untuk data historis [4]. Sama halnya dengan *Neural Network*, model *Deep Learning* terinspirasi berdasarkan cara kerja otak manusia. Perbedaan antara *Neural Network* dan *Deep Learning* adalah biasanya *Neural Network* berisi atas satu atau dua *hidden layer*, sedangkan *Deep Learning* berisi banyak *hidden layer* antara *input layer* dan *output layer* [12]. Perbedaan antara *Neural Network* dan *Deep Learning* ditunjukkan pada Gambar 3.



Gambar 3. Neural Network vs Deep Learning [13]

4) *K-Nearest Neighbor* (k-NN)

k-NN merupakan algoritma yang menggunakan teknik *lazy learning* dengan metode melakukan klasifikasi terhadap objek berdasarkan data pembelajaran dengan jarak terdekat dari objek tersebut. Model prediksi k-NN dilakukan dengan mencari kelompok *k* objek dalam data *training* yang paling dekat dengan data *testing* [5]. Untuk mendefinisikan jarak antara dua titik (*x* dan *y*), digunakan rumus *Euclidean Distance* seperti pada persamaan (2).

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \dots\dots\dots(2)$$

Keterangan:

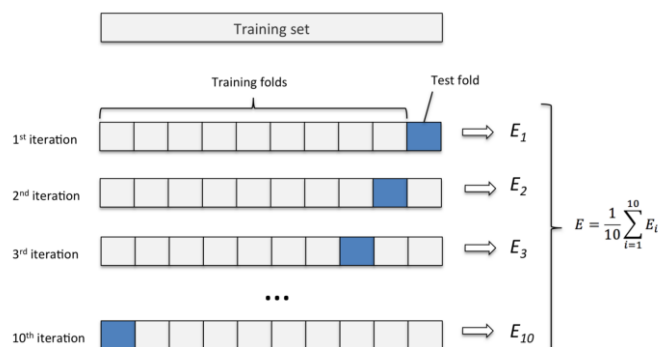
x = titik pada data *training*

y = titik pada data *testing*

D = jarak antara data *training* *x* dan data *testing* *y*

2.4 Model Validasi

Model validasi yang digunakan dalam penelitian ini adalah *K-Fold Cross Validation* (KCV). KCV akan mempartisi data *k* bagian dan dilakukan sebanyak *k* iterasi. Setiap kali dipilih bagian dari *dataset*, *k* – 1 pertama digunakan sebagai data *training* sementara sisanya digunakan sebagai data *testing*. Proses ini akan diulang sebanyak *k* kali kemudian dihitung nilai rata-rata penyimpangan (*error*) dari *k* hasil pengujian berbeda. Gambar 4 menjelaskan ilustrasi KCV.



Gambar 4. Diagram K-Fold Cross Validation dengan *k* = 10 [14]

2.5 Evaluasi Model

Evaluasi model untuk mengukur keakuratan dari prediksi algoritma dalam penelitian ini adalah menggunakan *Root Mean Square Error* (RMSE). Nilai RMSE bertujuan untuk mengetahui seberapa jauh nilai kesalahan suatu model terhadap garis regresi linier [15]. Semakin tingkat akurasi yang dihasilkan ditunjukkan dengan nilai RMSE yang kecil [16]. Nilai RMSE dapat dihitung dengan persamaan (3).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \dots\dots\dots(3)$$

Di mana:

- n = jumlah data yang diamati
- x_i = nilai data yang diamati
- \hat{x}_i = nilai prediksi

2.6 Uji T-Test

T-Test merupakan metode pengujian hipotesis kepada objek yang memiliki perlakuan yang berbeda. Meskipun menggunakan objek yang sama, tetapi sampel tetap terbagi menjadi dua perlakuan: data dengan perlakuan pertama dan data dengan perlakuan kedua [17]. Tujuan diperlukan adanya uji T-Test adalah untuk mengetahui perbandingan rata-rata dari dua grup yang tidak saling berkaitan satu sama lain. Tidak saling berkaitan yang dimaksud adalah pengujian dilakukan pada dua populasi yang berbeda [18]. Singkatnya, uji T-Test diperlukan untuk membedakan rata-rata dua populasi berdasarkan hipotesis yang telah diajukan selama penelitian dilakukan. Misal pengujian T-Test dilakukan dengan tingkat signifikansi sebesar 0.05 ($\alpha = 0.05$) dengan hipotesis awal (H_0) “Tidak ada perbedaan signifikan antar variabel independen” maka kriteria pengujian T-Test adalah sebagai berikut [19]:

1. Jika $\alpha > 0.05$, maka hipotesis awal (H_0) diterima dan hipotesis alternatif (H_1) ditolak. Hal ini menyimpulkan bahwa variabel-variabel independen tersebut tidak memiliki perbedaan yang signifikan.
2. Jika $\alpha < 0.05$, maka hipotesis awal (H_0) ditolak dan hipotesis alternatif (H_1) diterima. Hal ini menyimpulkan bahwa variabel-variabel independen tersebut memiliki perbedaan yang signifikan.

3. HASIL DAN PEMBAHASAN

3.1 Praproses

Data awal memiliki 7 atribut yaitu Date, Open, High, Low, Close, dan Volume, dan Market Cap yang dapat dilihat pada Tabel 1.

Tabel 1 Data Awal

Date	Open	High	Low	Close	Volume	Market Cap
01/01/2015	320,43	320,43	314,00	314,25	8.036.550,00	4.297.535.768,00

02/01/2015	314,08	315,84	313,57	315,03	7.860.650,00	4.309.551.126,00
03/01/2015	314,85	315,15	281,08	281,08	33.054.400,00	3.846.269.872,00
04/01/2015	281,15	287,23	257,61	264,20	55.629.100,00	3.616.320.975,00
05/01/2015	265,08	278,34	265,08	274,47	43.962.800,00	3.758.098.008,00
...
27/10/2021	60.352,00	61.435,18	58.208,19	58.482,39	43.657.076.893,00	1.102.757.767.323,00
28/10/2021	58.470,73	62.128,63	58.206,92	60.622,14	45.257.083.247,00	1.143.165.999.365,00
29/10/2021	60.624,87	62.927,61	60.329,97	62.227,96	36.856.881.767,00	1.173.512.442.368,00
30/10/2021	62.239,36	62.330,15	60.918,39	61.888,83	32.157.938.616,00	1.167.178.054.167,00
31/10/2021	61.850,49	62.406,17	60.074,33	61.318,96	32.241.199.927,00	1.156.485.845.063,00

Kemudian atribut-atribut pada data awal akan diperiksa korelasinya menggunakan Matriks Korelasi (*Correlation Matrix*). Hal ini bertujuan untuk memeriksa apakah tiap atribut memiliki hubungan yang kuat. Menurut [20] hubungan variabel dikatakan kuat jika bernilai lebih dari 0.71. Hasil Matriks Korelasi dapat dilihat pada Gambar 5.

Attributes	Close	Date	Open	High	Low	Volume	Market Cap
Close	1	?	0.999	0.999	0.999	0.751	1.000
Date	?	1	?	?	?	?	?
Open	0.999	?	1	0.999	0.999	0.753	0.998
High	0.999	?	0.999	1	0.999	0.756	0.999
Low	0.999	?	0.999	0.999	1	0.746	0.999
Volume	0.751	?	0.753	0.756	0.746	1	0.751
Market Cap	1.000	?	0.998	0.999	0.999	0.751	1

Gambar 5. Matriks Korelasi

Berdasarkan Gambar 5, dapat disimpulkan bahwa tiap atribut memiliki hubungan yang kuat karena memiliki nilai di atas 0.71. Namun demikian, jika dilihat pada Tabel 1, atribut Market Cap memiliki rentang nilai yang sangat besar. Hal ini akan menjadi masalah jika atribut Market Cap diikuti sertakan selama proses prediksi (proses *training*) karena rentang nilai yang sangat jauh dibandingkan dengan atribut lain akan mempengaruhi performa model algoritma [21].

Hasil dari praproses data yang sudah dipilih 6 dari 7 atribut di mana atribut Market Cap tidak diikuti sertakan. Atribut yang dipilih adalah Date, Open, High, Low, Close, dan Volume. Adapun atribut Close dijadikan sebagai label selama proses pelatihan (*training*), hal ini dikarenakan atribut Close merepresentasikan harga final Bitcoin yang diperdagangkan di akhir hari. Hasil praproses data ditunjukkan pada Tabel 2.

Tabel 2 Data hasil praproses

Date	Open	High	Low	Close	Volume
01/01/2015	320,43	320,43	314,00	314,25	8.036.550,00
02/01/2015	314,08	315,84	313,57	315,03	7.860.650,00
03/01/2015	314,85	315,15	281,08	281,08	33.054.400,00
04/01/2015	281,15	287,23	257,61	264,20	55.629.100,00
05/01/2015	265,08	278,34	265,08	274,47	43.962.800,00
...
27/10/2021	60.352,00	61.435,18	58.208,19	58.482,39	43.657.076.893,00
28/10/2021	58.470,73	62.128,63	58.206,92	60.622,14	45.257.083.247,00
29/10/2021	60.624,87	62.927,61	60.329,97	62.227,96	36.856.881.767,00
30/10/2021	62.239,36	62.330,15	60.918,39	61.888,83	32.157.938.616,00
31/10/2021	61.850,49	62.406,17	60.074,33	61.318,96	32.241.199.927,00

3.2 Pengujian dengan *Linear Regression*

Pengujian dengan *Linear Regression* dilakukan dengan dua tahap, yaitu tahap pertama dengan mengeliminasi fitur kolinear (*eliminate collinear feature*) dan tahap kedua tanpa mengeliminasi fitur kolinear. Tiap tahapan dilakukan juga dengan memilih fitur seleksi (*feature selection*), fitur seleksi tersebut antara lain *None* (tanpa fitur seleksi), *M5 Prime*, *Greedy*, *T-Test*, dan *Iterative T-Test*. Hasil pengujian menggunakan algoritma *Linear Regression* dapat dilihat pada Tabel 3.

Tabel 3 Hasil pengujian dengan algoritma *Linear Regression*

<i>Eliminate Collinear Feature?</i>	Fitur Seleksi	RMSE
Ya	None	514.479 +/- 51.756 (micro average: 516.806 +/- 0.000)
	M5 Prime	497.349 +/- 138.457 (micro average: 514.332 +/- 0.000)
	Greedy	509.136 +/- 87.131 (micro average: 515.729 +/- 0.000)
	T-Test	510.834 +/- 73.566 (micro average: 515.645 +/- 0.000)
	Iterative T-Test	506.928 +/- 99.210 (micro average: 515.639 +/- 0.000)
Tidak	None	299.964 +/- 49.404 (micro average: 303.609 +/- 0.000)
	M5 Prime	298.986 +/- 54.010 (micro average: 303.380 +/- 0.000)
	Greedy	300.081 +/- 52.875 (micro average: 304.280 +/- 0.000)
	T-Test	296.227 +/- 60.125 (micro average: 301.655 +/- 0.000)
	Iterative T-Test	299.397 +/- 55.640 (micro average: 304.010 +/- 0.000)

Berdasarkan hasil pengujian pada Tabel 3, dapat disimpulkan bahwa hasil prediksi terbaik dengan algoritma *Linear Regression* adalah dengan tidak mengeliminasi fitur kolinear dan menggunakan fitur seleksi T-Test yaitu dengan nilai RMSE 296.227 +/- 60.125 (micro average: 301.655 +/- 0.000).

3.3 Pengujian dengan *Neural Network*

Pengujian dengan *Neural Network* dilakukan dengan dua tahap: pertama dengan satu *hidden layer* dan kedua dengan dua *hidden layer*, masing-masing nilai masukan dari ukuran *hidden layer* berkisar antara 1 sampai 10. Adapun siklus pelatihan (*training cycles*) yang dilakukan pada pengujian ini adalah sebanyak 200 kali. Metode yang dilakukan pada pengujian ini adalah *Trial-and-error* yaitu dengan meng-*input* satu per satu ukuran *hidden layer* hingga mendapatkan nilai RMSE terkecil. Hasil pengujian algoritma *Neural Network* ditunjukkan pada Tabel 4 dan Tabel 5.

Tabel 4 Hasil pengujian *Neural Network* dengan satu *hidden layer*

Ukuran Hidden Layer	RMSE
1	406.099 +/- 58.029 (micro average: 409.797 +/- 0.000)
2	347.467 +/- 51.582 (micro average: 350.917 +/- 0.000)
3	349.310 +/- 43.294 (micro average: 351.745 +/- 0.000)
4	343.151 +/- 39.662 (micro average: 345.217 +/- 0.000)
5	345.647 +/- 47.116 (micro average: 348.536 +/- 0.000)
6	348.978 +/- 38.734 (micro average: 350.920 +/- 0.000)
7	338.988 +/- 47.837 (micro average: 342.000 +/- 0.000)
8	363.468 +/- 62.700 (micro average: 368.262 +/- 0.000)
9	359.981 +/- 96.201 (micro average: 371.416 +/- 0.000)
10	341.002 +/- 44.730 (micro average: 343.613 +/- 0.000)

Tabel 5 Hasil pengujian *Neural Network* dengan dua *hidden layer*

LAYER 1 \ LAYER 2	1	2	...	10
...
3	439.638 +/- 75.834 (micro average: 445.456 +/- 0.000)	387.702 +/- 44.987 (micro average: 390.026 +/- 0.000)	...	396.216 +/- 55.626 (micro average: 399.720 +/- 0.000)

4	428.533 +/- 65.805 (micro average: 433.008 +/- 0.000)	368.026 +/- 47.055 (micro average: 370.738 +/- 0.000)	...	414.308 +/- 56.990 (micro average: 417.783 +/- 0.000)
...
10	435.988 +/- 58.438 (micro average: 439.500 +/- 0.000)	369.199 +/- 56.941 (micro average: 373.135 +/- 0.000)	...	403.383 +/- 48.295 (micro average: 405.980 +/- 0.000)

Berdasarkan hasil pengujian pada Tabel 4 dan Tabel 5, prediksi dengan *Neural Network* terbaik dengan satu *hidden layer* dengan ukuran *hidden layer* = 7 yaitu dengan nilai RMSE 338.988 +/- 47.837 (micro average: 342.000 +/- 0.000).

3.4 Pengujian dengan *Deep Learning*

Tahapan pengujian dengan *Deep Learning* memiliki tahapan yang sama seperti pada pengujian dengan *Neural Network* yaitu dengan satu *hidden layer* dan dua *hidden layer*. Pengujian ini menggunakan aktivasi fungsi Tanh dan rentang ukuran *hidden layer* berkisar antara 10 sampai 100 (kelipatan 10). Adapun jumlah iterasi yang telah dilakukan pada masing-masing pengujian ini adalah 50 kali (*epochs* = 50). Hasil pengujian dapat dilihat pada Tabel 6 dan Tabel 7.

Tabel 6 Hasil pengujian *Deep Learning* dengan satu *hidden layer*

Epochs	Ukuran Hidden Layer	RMSE
50	10	534.346 +/- 278.324 (micro average: 595.737 +/- 0.000)
	20	425.931 +/- 59.875 (micro average: 429.687 +/- 0.000)
	30	406.659 +/- 72.835 (micro average: 412.516 +/- 0.000)
	40	417.221 +/- 90.987 (micro average: 426.023 +/- 0.000)
	50	401.980 +/- 71.410 (micro average: 407.631 +/- 0.000)
	60	402.321 +/- 82.707 (micro average: 409.854 +/- 0.000)
	70	420.442 +/- 86.149 (micro average: 428.275 +/- 0.000)
	80	411.801 +/- 77.376 (micro average: 418.269 +/- 0.000)
	90	402.637 +/- 69.582 (micro average: 408.009 +/- 0.000)
	100	403.630 +/- 59.218 (micro average: 407.523 +/- 0.000)

Tabel 7 Hasil pengujian *Deep Learning* dengan dua *hidden layer*

LAYER 1 LAYER 2	10	...	90	100
...
30	435.547 +/- 68.780 (micro average: 440.344 +/- 0.000)	...	431.654 +/- 75.037 (micro average: 437.456 +/- 0.000)	433.159 +/- 58.697 (micro average: 436.692 +/- 0.000)
40	444.022 +/- 74.638 (micro average: 449.572 +/- 0.000)	...	425.389 +/- 84.891 (micro average: 432.906 +/- 0.000)	421.116 +/- 43.427 (micro average: 423.114 +/- 0.000)
...
100	463.124 +/- 120.123 (micro average: 476.892 +/- 0.000)	...	388.987 +/- 57.361 (micro average: 392.777 +/- 0.000)	399.381 +/- 46.531 (micro average: 401.819 +/- 0.000)

Berdasarkan hasil pengujian pada Tabel 6 dan Tabel 7, dapat disimpulkan bahwa model prediksi *Deep Learning* terbaik adalah dengan menggunakan dua *hidden layer* saat *input Layer 1* = 90 dan *input Layer 2* = 100 dengan nilai RMSE 388.987 +/- 57.361 (micro average: 392.777 +/- 0.000).

3.5 Pengujian dengan k-Nearest Neighbor (k-NN)

Pengujian dengan k-NN dilakukan dengan *trial-and-error* di mana nilai *k* akan dimasukkan satu per satu dengan rentang nilai *k* adalah 1 sampai 10 kemudian hasil diteliti untuk mendapatkan nilai RMSE terkecil. Adapun hasil dari pengujian prediksi harga Bitcoin dengan algoritma k-NN dapat dilihat pada Tabel 8.

Tabel 8 Hasil pengujian dengan algoritma k-NN

K	RMSE
1	17586.591 +/- 1846.632 (micro average: 17674.380 +/- 0.000)
2	1831.954 +/- 654.234 (micro average: 1934.758 +/- 0.000)

3	1854.017 +/- 578.465 (micro average: 1934.078 +/- 0.000)
4	1880.167 +/- 566.364 (micro average: 1955.931 +/- 0.000)
5	1928.314 +/- 560.062 (micro average: 2000.674 +/- 0.000)
6	1985.260 +/- 557.342 (micro average: 2055.001 +/- 0.000)
7	2048.469 +/- 554.735 (micro average: 2115.563 +/- 0.000)
8	2099.916 +/- 525.651 (micro average: 2158.852 +/- 0.000)
9	2101.816 +/- 522.744 (micro average: 2160.083 +/- 0.000)
10	2126.448 +/- 542.584 (micro average: 2188.433 +/- 0.000)

Berdasarkan hasil pengujian pada Tabel 8, model terbaik didapat pada saat nilai $k = 2$ dengan perolehan nilai RMSE 1831.954 +/- 654.234 (micro average: 1934.758 +/- 0.000).

3.6 Uji T-Test

Berdasarkan hasil pengujian dari tiap algoritma, maka hasil perbandingan algoritma dapat dilihat pada Tabel 9.

Tabel 9 Hasil perbandingan algoritma

Algoritma	RMSE
Linear Regression	296.227 +/- 60.125 (micro average: 301.655 +/- 0.000)
Neural Network	338.988 +/- 47.837 (micro average: 342.000 +/- 0.000)
Deep Learning	388.987 +/- 57.361 (micro average: 392.777 +/- 0.000)
k-Nearest Neighbor (k-NN)	1831.954 +/- 654.234 (micro average: 1934.758 +/- 0.000)

Uji T-Test dilakukan untuk mengetahui apakah nilai akurasi tiap algoritma memiliki perbedaan yang signifikan. Adapun taraf signifikansi (α) sebesar 5% (0.05), artinya jika hasil uji $t < 0.05$ maka perbandingan algoritma memiliki perbedaan yang signifikan (H1). Hasil uji T-Test dapat dilihat pada Gambar 6.

A	B	C	D	E
	296.227 +/- 60.125	338.988 +/- 47.837	388.987 +/- 57.361	1831.954 +/- 654.234
296.227 +/- 60.125		0.095	0.002	0.000
338.988 +/- 47.837			0.048	0.000
388.987 +/- 57.361				0.000
1831.954 +/- 654.234				

Gambar 6. Hasil uji T-Test

Keterangan:

B : Linear Regression

C : Neural Network

D : Deep Learning

E : k-NN

Berdasarkan hasil uji T-Test pada Gambar 6, perbandingan antara algoritma *Linear Regression* dan *Neural Network* tidak memiliki perbedaan yang signifikan (H0). Namun, *Linear Regression* dan *Neural Network* memiliki perbedaan yang signifikan jika dibandingkan dengan *Deep Learning* dan k-NN (H1). Dengan demikian, algoritma *Linear Regression* dan *Neural Network* merupakan algoritma dengan performa terbaik dalam memprediksi harga Bitcoin.

4. Kesimpulan

Berdasarkan hasil perbandingan antara model algoritma *Linear Regression*, *Neural Network*, *Deep Learning*, dan *k-Nearest Neighbor* (k-NN), model algoritma *Linear Regression* dan *Neural Network* memegang hasil prediksi terbaik dalam prediksi harga Bitcoin, dengan hasil akurasi prediksi RMSE 296.227 +/- 60.125 (micro average: 301.655 +/- 0.000) dan 338.988 +/- 47.837 (micro average: 342.000 +/- 0.000).

Saran untuk penelitian berikutnya adalah perlu dilakukan perbandingan dengan algoritma terbaru. Khusus untuk *Neural Network* dan *Deep Learning* perlu diteliti lebih lanjut dengan menggunakan kombinasi *hidden layer* lain, bisa juga dilakukan dengan menambah *hidden layer*.

DAFTAR PUSTAKA

- [1] Chowdhury N. Inside Blockchain, Bitcoin, and Cryptocurrencies. First Edition. Auerbach Publications. 2019.
- [2] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008.
- [3] CNBC Indonesia. *Bikin Kaget, Transaksi Kripto di Indonesia Tembus Rp 370 T*. Available from: <https://www.cnbcindonesia.com/tech/20210721112800-37->

- 262411/bikin-kaget-transaksi-kripto-di-indonesia-tembus-rp-370-t [Accessed 1st January 2022]
- [4] Karno ASB, Hastomo W, Arif D, Moreta ES. Optimasi Portofolio Dan Prediksi Cryptocurrency Menggunakan Deep Learning Dalam Bahasa Python. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K. 2020. 4(1): 193-202.
- [5] Maulana R, Kumalasari D. Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham GGRM. Jurnal Informatika Kaputama. 2019. 3(1): 22-28.
- [6] Irmada HN, Astriratma R, Afrizal S. Perbandingan Metode Jaringan Syaraf Tiruan Dan Pohon Keputusan Untuk Prediksi Churn. JSI: Jurnal Sistem Informasi (*E-Journal*). 2019;11(2).
- [7] Harun R, Chandra Pelangi K, Lasena Y. Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN). MISI (Jurnal Manajemen informatika & Sistem Informasi) [Internet]. 2020;3(1):8–15.
- [8] CoinMarketCap. Available from: <https://coinmarketcap.com/currencies/bitcoin/historical-data/>. [Accessed 1st January 2022].
- [9] Han J, Kamber M, Pei J. Data Mining Concepts and technique. San Francisco: Diane Cerra. 2006.
- [10] Maulud D, Abdulazeez AM. A Review on Linear Regression Comprehensive in Machine Learning. Journal of Applied Science and Technology Trends. 2020. 1(4): 140-147.
- [11] IBM Cloud Education. Neural Networks. Available from: <https://www.ibm.com/cloud/learn/neural-networks> [Accessed 3rd January 2022].
- [12] Mostafa B, El-Attar N, Abd-Elhafeez S, Awad W. Machine and Deep Learning Approaches in Genome: Review Article. Alfarama Journal of Basic & Applied Sciences. 2021. 2(1): 105-113.
- [13] Global Engage. Deep Learning In Digital Pathology. Available from: <http://www.global-engage.com/life-science/deep-learning-in-digital-pathology/>. [Accessed 03 January 2022].
- [14] Karl. Introduction to Support Vector Machines and Kernel Methods. K-fold cross-validation. Weblog. Available from: <http://karlrosaen.com/ml/learning-log/2016-06-20/> [Accessed 1st January 2022].
- [15] Hastomo W, Satyo A, Sudjiran. Long Short Term Memory Machine Learning Untuk Memprediksi Akurasi Nilai Tukar IDR Terhadap USD. Seminar Nasional Teknologi Informasi dan Komunikasi STI&K. 2019. 3(1): 115-124.
- [16] Hastomo W, Karno ASB, Kalbuana N, Nisfiani E, ETP L. Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi Covid-19. Jurnal Edukasi dan Penelitian Informatika. 2021. 7(2): 133-140.
- [17] Ardiansyah, Rahayuningsih PA, Maulana R. Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner. Jurnal Khatulistiwa Informatika. 2018. 6(1): 20-28.
- [18] Palupi R, Yulianna DA, Winarsih SS. Analisa Perbandingan Rumus Haversine Dan Rumus Euclidean Berbasis Sistem Informasi Geografis Menggunakan Metode

- Independent Sample t-Test. *JITU: Journal Informatic Technology And Communication*. 2021 Jul 29;5(1):40–7.
- [19] Magdalena R, Krisanti MA. Analisis Penyebab dan Solusi Rekonsiliasi Finished Goods Menggunakan Hipotesis Statistik dengan Metode Pengujian Independent Sample T-Test di PT.Merck, Tbk. *Jurnal TEKNO*. 2019;16(1):35–48.
- [20] Sellar T, Arulrajah AA. The Role of Social Support on Job Burnout in the Apparel Firm. *International Business Research*. 2018 Dec 24;12(1):110.
- [21] Givari MR, Sulaeman MR, Umaidah Y. Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. *Jurnal Nuansa Informatika [Internet]*. 2022;16(1):141–9.