

# AUTO MACHINE LEARNING DENGAN MENGGUNAKAN H2O AUTOML UNTUK PREDIKSI HARGA BITCOIN

Geadalfa Giyanda<sup>1</sup> dan Siti Saidah<sup>2</sup>  
<sup>1,2</sup> Jurusan Sistem Informasi, Universitas Gunadarma  
Jl. Margonda Raya 100 Depok 16424

[1geadalfa@student.gunadarma.ac.id](mailto:geadalfa@student.gunadarma.ac.id),

[2sitisaidah@staff.gunadarma.ac.id](mailto:sitisaidah@staff.gunadarma.ac.id)

## ABSTRAK

Bitcoin merupakan salah satu penerapan konsep *cryptocurrency*, yang mengemukakan saran dan ide terhadap bentuk baru mata uang menggunakan kriptografi dengan fungsi untuk mengontrol pembuatan dan transaksi. Kesulitan memprediksi harga Bitcoin dapat terjadi, jika dilakukan secara manual, oleh karena itu dibutuhkan aplikasi untuk memprediksi harga Bitcoin dengan menggunakan bahasa pemrograman Python dinamis yang dilengkapi dengan manajemen memori otomatis. Python juga dapat berkolaborasi sebagai *library* dalam sebuah aplikasi salah satu contohnya adalah *library* H2O yang dibuat oleh H2O.ai. H2O telah memudahkan non-ahli untuk bereksperimen dengan *machine learning*, H2O AutoML dapat digunakan untuk mengotomatiskan alur kerja *machine learning*, yang mencakup pelatihan otomatis dan penyediaan banyak model dalam batas waktu yang ditentukan pengguna. Hasil evaluasi terhadap data uji dengan menggunakan model H2O AutoML dalam memprediksi harga pembukaan Bitcoin memperoleh nilai koefisien determinasi ( $R^2$ ) sebesar 0.968 dan nilai *error* sebesar 3.48%.

Kata kunci : Bitcoin, Machine Learning, Python, AutoML, H2O

## PENDAHULUAN

Perekonomian Indonesia dipengaruhi oleh fluktuasi kurs mata uang asing. Banyak pihak yang memiliki kepentingan khusus untuk menyiapkan langkah strategis, agar tidak mengalami kerugian yang besar. Penelitian dengan judul Analisis *Support Vector Regression* (SVR) Dalam Memprediksi Kurs Rupiah Terhadap Dollar Amerika Serikat mampu menghasilkan akurasi yang cukup baik pada kedua fungsi kernel yang digunakan, yaitu kernel linier dan kernel polynomial [1], penelitian yang mengamati Pemodelan Data Indeks Harga Saham Gabungan Menggunakan Regresi *Penalized Spline* menghasilkan nilai ketepatan model terbaik, dapat ditunjukkan dari besarnya nilai koefisien determinasi ( $R^2$ ) dan MAPE [2]. Peneliti berikut

memfokuskan Visualisasi Data untuk Memprediksi Pasar Saham dari Hasil Pengolahan Data Set S&P 500 dengan menggunakan R-Programming menyimpulkan kemampuan untuk meningkatkan ketepatan penyajian laporan transaksi [3].

Contoh lain transaksi yang dijalankan mirip dengan penjualan saham adalah mata uang digital atau virtual yang dirancang sebagai alat tukar, dikenal sebagai *Cryptocurrency* yang digunakan oleh sistem kriptografi untuk mengamankan dan memverifikasi setiap transaksi, serta untuk mengontrol pembuatan unit-unit (*token*) baru dari suatu *cryptocurrency* tertentu.

*Cryptocurrency* adalah entri yang terbatas dalam basis data yang tidak dapat diubah kecuali kondisi tertentu terpenuhi [4]. Bitcoin merupakan mata uang kripto yang

telah paling terkenal sebagai mata uang digital. Jumlah Bitcoin beredar secara terbatas, sehingga Bitcoin baru dibuat dengan tingkat yang dapat diprediksi dan menurun, dengan perkataan lain permintaan harus mengikuti besaran inflasi untuk menjaga harga tetap stabil. Pasar Bitcoin relatif kecil dibanding potensi yang bisa dicapai, oleh sebab itu tidak dibutuhkan jumlah uang yang besar untuk pergerakan naik-turunnya harga di pasaran, dengan demikian harga bitcoin relatif mudah berubah [5].

Berlandaskan pengamatan di atas, peneliti tertarik untuk menerapkan konsep Auto Machine Learning Dengan Menggunakan H2O Automl Untuk Prediksi Harga *Bitcoin* menggunakan Bahasa Pemrograman Python, didukung oleh *software* Colaboratory oleh Google atau Jupyter Notebook yang bersifat *open source* dan bisa dijalankan pada semua komputer yang mempunyai sebuah *web browser* [6].

Batasan masalah penelitian ini adalah *Cryptocurrency* yang akan diprediksi adalah data historis *Bitcoin* dimulai dari 17 September 2014 – 26 Maret 2020 tergabung menjadi satu file csv terdiri dari 2018 baris dan 6 kolom. File ekstensi program berbentuk .ipynb (*Interactive Python Notebook*) hanya bisa dijalankan dengan Google Colaboratory atau Jupyter Notebook untuk menampilkan *Output* berupa grafik prediksi, koefisien determinasi ( $R^2$ ) dan MAE (*Mean Absolute Error*) dijalankan terhadap data *test* yaitu 20% dari keseluruhan data.

Penelitian ini bertujuan membuat aplikasi berekstensi .ipynb untuk memprediksi harga Bitcoin menggunakan Bahasa pemrograman

Python dan *library* H2O AutoML, dengan harapan mempermudah pengguna mengetahui pergerakan harga Bitcoin.

## TINJAUAN PUSTAKA

Mata uang kripto merupakan mata uang digital atau virtual yang menggunakan *cryptography* sebagai keamanan. *Cryptography* menyediakan mekanisme digunakan untuk mengamankan sistem dalam mata uang digital dengan cara menyandikan atau mengkodekan aturan dalam sistem mata uang kripto itu sendiri [7].

Bitcoin diluncurkan pada tahun 2009, pada Mei 2017, kapitalisasi pasar mata uang kripto aktif melampaui \$91 miliar. Financial Stability Board (2018) menyebutkan bahwa pada 8 Januari 2018, kapitalisasi pasar gabungan aset kripto naik hingga \$830 miliar, di mana sekitar 35% disebabkan oleh Bitcoin. Saat ini Bitcoin tetap mendominasi pasar, namun di sisi lain Bitcoin dihadapi oleh masalah teknis dan juga peningkatan teknologi mata uang kripto lainnya.

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat kegunaan kode [8].

Situs [research.google.com](https://research.google.com) Google Colaboratory atau "Google Colab" adalah produk dari Google Research yang memungkinkan siapa saja untuk menulis dan mengeksekusi kode python melalui *browser* dan sangat cocok untuk pembelajaran mesin (*Machine Learning*), analisis data, dan pendidikan. Secara teknis, Colab adalah layanan Jupyter Notebook yang dihosting yang tidak

memerlukan pengaturan untuk digunakan, sambil menyediakan akses gratis ke sumber daya komputasi termasuk *Graphical Processing Unit* (GPU).

Google Colab memiliki beberapa keuntungan yang bisa dimanfaatkan secara gratis, diantaranya adalah : *Graphical Processing Unit* (GPU), *Colaborate*, mudah terintegrasi dan fleksibel

H2O.ai adalah sebuah perusahaan yang bergerak dibidang *Artificial Intelligence* dan *Machine Learning* yang menyediakan berbagai *Library Open Source* yang bernama H2O juga dengan maksud untuk digunakan secara bebas untuk semua orang.

H2O AutoML dapat digunakan untuk mengotomatiskan alur kerja *machine learning*, yang mencakup pelatihan otomatis dan penyetalan banyak model dalam batas waktu yang ditentukan pengguna. Kumpulan model *stacked* atau sebuah model yang terbentuk dari tumpukkan beberapa model yang dilatih sebelumnya, lalu pada model terbaik dari masing-masing jenis akan secara otomatis dilatih pada koleksi masing-masing model untuk menghasilkan model ansambel yang sangat prediktif, dalam banyak kasus menjadi model dengan kinerja terbaik di *AutoML Leaderboard* [9].

## METODE PENELITIAN

Analisis untuk kebutuhan spesifikasi hardware terdiri dari *Personal Computer*, CPU AMD Ryzen 5 3600 6C/12T 3.6Ghz, RAM : 16GB DDR4, *Hardisk* : 2TB + 120GB SSD, GPU : Nvidia GTX 1070Ti, sedangkan spesifikasi perangkat lunak yang digunakan adalah Python 3.8 dan *Google Colab*.

Aplikasi dibuat dengan menampilkan data harga Bitcoin mentah, dimaksudkan harga yang ditampilkan belum diolah untuk selanjutnya diubah menjadi tabel, sehingga bisa divisualisasikan dengan baik menggunakan grafik, agar *user* dapat melihat harga Bitcoin yang fluktuatif dengan mudah, proses berikutnya data dipisahkan menjadi data *train* dan data *test* dengan tujuan dapat dilakukan pembuatan model dan proses *training* terhadap model dengan menggunakan *library* H2O AutoML, model yang terbentuk diprediksi terhadap data *test* dan visualisasi ditampilkan dengan grafik.

Tahapan penelitian meliputi tahap data *input*, data *preprocessing* meliputi proses data *cleaning*, data *preparation* dan data *splitting*, lalu melakukan konversi data *train* dan data *test* ke H2O *DataFrame*, dilanjutkan ke proses *training* dan *testing*, menghitung Koefisien Determinasi ( $R^2$ ) dan Mean Absolute Percentage Error (MAPE) terhadap tingkat kesalahan hasil prediksi, rancangan tampilan visualisasi dan tahap akhir dilakukan visualisasi hasil prediksi terhadap nilai aktual berdasarkan periode yang diujikan terhadap model prediksi.

### 1. Tahap Data Input

Data Bitcoin harian yang diolah ada pada periode 17 September 2014 sampai dengan 26 Maret 2020, data tersebut diperoleh dari situs *Yahoo! Finance*. Data Bitcoin tersebut terdiri dari beberapa kolom yaitu : Date, Open, High, Low, seperti pada Tabel 1 dan Tabel 2 sebagian dari hasil dari *input* data menjadi *Pandas DataFrame*.

Tabel 1. Harga Bitcoin Open, High dan Low

Date	Open	High	Low
9/17/2014	465.86	468.17	452.42
9/18/2014	456.86	456.86	413.10
9/19/2014	424.10	427.83	384.53
9/20/2014	394.67	423.30	389.88
9/21/2014	408.08	412.43	393.18

Tabel 1 yang dimaksud dengan *Open* adalah harga pembukaan dalam bursa perdagangan Bitcoin pada tanggal 17 September 2014, harga pembukaan Bitcoin adalah 465.86 dalam satuan Dolar Amerika Serikat (“US\$”) yang artinya harga 1 Bitcoin adalah \$465.86. Harga *High* adalah harga tertinggi pada saat bursa perdagangan, *Low* adalah harga terendah dalam bursa perdagangan, terdapat 3 harga lain yang juga mempengaruhi terhadap transaksi jual beli harga Bitcoin, yaitu harga *Close*, *Adj.Close* dan *Volume*, seperti tampilan yang ada pada Tabel 2

Tabel 2. Harga Bitcoin Close, Adj Close dan Volume

Date	Close	Adj Close	Volume
9/17/2014	457.33	457.33	21056800
9/18/2014	424.44	424.44	34483200
9/19/2014	394.80	394.80	37919700
9/20/2014	408.90	408.90	36863600
9/21/2014	398.82	398.82	26580100

Tabel 2 menjelaskan Bitcoin *Close* adalah harga saat penutupan bursa perdagangan, *Adj Close* adalah harga penutupan yang disesuaikan dengan aksi korporasi seperti *right issue*, *stock split* atau *stock reverse* dan *Volume* adalah jumlah transaksi jual atau beli yang dilakukan selama bursa perdagangan buka.

## 2. Tahap Data Preprocessing

Tahap data *preprocessing* dilakukan untuk membentuk data Bitcoin yang masih “mentah” menjadi bentuk data yang dapat diterima dan dipelajari polanya (*pattern*) oleh model yang akan dibuat. Tahap ini terdiri dari tiga proses yaitu data *cleaning*, data *preparation* dan data *splitting*. Hasil yang diperoleh adalah data *input* yang terbagi menjadi data *train* dan *test*, untuk selanjutnya akan dikonversi ke H2O DataFrame dan siap digunakan untuk data latih, pengujian model dan visualisasi hasil prediksi.

### 2.1 Data Cleaning

Data Bitcoin yang diperoleh dari situs *Yahoo! Finance* mempunyai beberapa sampel yang null atau “NaN”, hal ini dikarenakan setiap tahun terdapat hari libur bursa perdagangan *cryptocurrency* yang mengikuti tanggal hari libur di Amerika. Data *cleaning* untuk membersihkan sampel tersebut dilakukan, karena model prediksi tidak bisa menerima masukkan yang tidak mempunyai nilai atau “NaN”.

Fungsi dari *library* Pandas untuk melihat jumlah nilai NaN pada sampel. Fungsi “.isna()” untuk mencari nilai NaN dan fungsi “.sum()” untuk menjumlahkan berapa sampel yang mengandung nilai NaN tersebut, cara menghapus sampel tersebut dengan menggunakan fungsi

“dropna()” yang telah tersedia dari *library* Pandas lalu melihat kembali apakah masih ada nilai sampel yang “NaN” atau tidak dengan menggunakan tiga fungsi “print” yang di dalamnya terdapat beberapa fungsi dari *library* Pandas yaitu fungsi “df.tail()” untuk melihat 5 sampel terbawah dataset, “df.isna().sum()” untuk melihat dan menjumlahkan nilai NaN yang ada pada dataset yang telah dimasukkan ke dalam variabel df dan “df.shape” untuk melihat berapa baris dan kolom yang ada pada dataset, lalu dapat diketahui bahwa data Bitcoin sudah tidak memiliki nilai NaN dan memiliki 2017 sampel/baris.

## 2.2. Data Preparation

Dataset Bitcoin yang telah dibersihkan, dilanjutkan ke tahap data *preparation* untuk mendapatkan model distribusi *log*. Analisis regresi yang melibatkan data runtun waktu, jika model regresi memasukan tidak hanya nilai variabel bebas saat ini atau  $X_t$  tapi juga nilai variabel bebas masa lalu pada waktu  $t - 1$ ,  $t - 2$  dan seterusnya [10].

Model regresi seperti ini disebut model *distributed log*, ketika distribusi *log* telah dilakukan pada dataset Bitcoin, maka kolom pada dataset tersebut bertambah menjadi 34 kolom. Semua index yang ada, kecuali “Open” akan dipakai sebagai variabel  $X$  untuk proses *training* pada model dan variabel – variabel tersebut tidak perlu untuk diolah, karena akan diolah secara otomatis dengan menggunakan model H2O AutoML

Perbedaan antara kolom harga *Open* yang asli dengan kolom yang lain setelah *log* terdistribusi adalah nilai pada kolom “Open\_log1” adalah nilai yang ada pada baris kedua pada kolom “Open” begitu juga nilai yang

ada pada “Open\_log2” adalah nilai dari baris kedua pada kolom “Open\_log1” begitu pula sampai kepada “Open\_log3”. Lalu, nilai yang ada di dalam kolom “Open\_avg\_window\_length2” adalah rata-rata dari hasil penjumlahan kolom “Open” dan “Open\_log1” per baris, begitu juga dengan kolom “Open\_avg\_window\_length” lainnya.

## 2.3. Data Splitting

Dataset Bitcoin selanjutnya dibagi menjadi data *train* dan data *test*. Pembagian data ini berguna untuk menentukan data yang ingin dilatih dan data yang akan menjadi evaluasi model untuk menentukan ke akuratan prediksi dan tingkat kesalahan (*error*). Pembagian dilakukan dengan presentasi data *train* 80% dan data *test* 20%, sehingga diperoleh dataset Bitcoin dataset *train* sebanyak 1611 sampel/baris dan 34 kolom, sedangkan dataset *test* memiliki 403 sampel/baris dan 34 kolom.

## 3. Tahap Konversi DataFrame

Data *train* dan data *test* berada pada format Pandas *DataFrame*, namun untuk menggunakan model H2O AutoML Pandas *DataFrame* tidak bisa digunakan sebagai input. H2O mewajibkan penggunaan H2O *DataFrame* sebagai *input* agar bisa menggunakan model H2O AutoML.

Hasil dari proses konversi Pandas *DataFrame* ke H2O *DataFrame* hampir tidak berbeda dengan Pandas *DataFrame* tetapi terlihat jelas bahwa H2O *DataFrame* tidak memiliki garis pembatas antar sel dan tidak memiliki kolom indeks yang biasa terletak paling kiri yang memiliki indeks baris berbentuk tanggal.

### 3.1. Permodelan dan Prediksi

Semua model H2O mengharuskan untuk memasukkan variabel X (“*Predictors*”) dan Y (“*Response*”). Variabel *predictor* adalah sebuah variabel masukan atau *input* yang bersifat independen/bebas biasanya dilambangkan dengan huruf ‘X’ sedangkan *response* adalah variabel keluaran atau *output* yang bersifat dependen/bergantung kepada variabel masukannya, biasanya variabel *response* dilambangkan dengan huruf ‘Y’, dengan menggunakan model H2O AutoML tidak perlu sulit untuk menentukan variabel X dan y karena dengan *auto machine learning*, model tersebut yang akan menentukan dan mengolah variabel *predictor* itu sendiri, oleh sebab itu pengguna bisa memasukkan semua kolom ke dalam variabel *predictor* kecuali kolom yang ingin diprediksi, dalam kasus ini kolom “Open” akan menjadi variabel *response* atau variabel ‘y’ sehingga kolom “Open” tidak akan dimasukkan ke dalam variabel X atau *predictor*.

### 3.2. Pemodelan H2O AutoML

Pembuatan model H2O AutoML sangat sederhana, pengguna hanya memasukkan fungsi “H2OAutoML()” ke dalam sebuah variabel yang bernama “aml”, di dalam fungsi model tersebut terdapat parameter yang bisa dikosongkan karena sudah menjadi fitur sebuah model *auto machine learning* akan berjalan secara otomatis, namun peneliti memasukkan dua parameter yaitu parameter “*max\_runtime\_secs*” untuk mengatur waktu *running* model dan “*seed*” untuk memastikan bahwa angka acak yang dihasilkan oleh algoritma *machine learning* selalu sama setiap kali dijalankan [11]

### 3.3. Proses Pelatihan H2O AutoML

Tahap proses melatih atau *training* model H2O AutoML merupakan tahap yang paling lama untuk dijalankan, tetapi dengan menggunakan parameter “*max\_runtime\_secs*” pengguna bisa mengubah waktu maksimal yang digunakan untuk melatih model di dalam model H2O AutoML.

Variabel “aml” yang di dalamnya telah dimasukkan model H2O AutoML dilatih dengan menggunakan fungsi “*train*” yang di dalamnya terdapat 4 (empat) buah parameter yaitu parameter variabel masukan (x), parameter variabel keluaran (y), *DataFrame* untuk melatih model (*training\_frame*) dan *DataFrame* untuk memvalidasi model (*leaderboard\_frame*). Proses pelatihan model berjalan hingga maksimal 600 detik atau 10 menit karena telah dideklarasikan saat pembuatan model.

### 3.4. Pemilihan Model Terbaik

Proses pelatihan model H2O AutoML selama sepuluh menit akan menghasilkan 10 model terbaik dengan berbagai macam algoritma dan pengaturan dari setiap model yang disesuaikan untuk menghasilkan tingkat kesalahan (“*Error*”) yang paling kecil sehingga pengguna tidak perlu menkonfigurasi model satu per satu untuk mendapatkan model terbaik dari dataset saat ini.

Model terbaik yang dihasilkan diurut berdasarkan 5 macam kesalahan diantaranya adalah *Mean\_Residual\_Deviance*, RMSE (“*Root Mean Squared Error*”), MSE (“*Mean Squared Error*”), MAE (“*Mean Absolute Error*”) dan RMSLE (“*Root Mean Squared Logarithmic Error*”) [12].

Kesalahan yang terjadi berguna untuk mengevaluasi model yang telah menggunakan dataset Bitcoin. Apabila dilihat secara baik-baik, sepuluh model terbaik tersebut memiliki `model_id` yang sama contohnya seperti `StackedEnsemble`, `XGBoost` dan `GBM`. Namun, di dalam setiap model terdapat konfigurasi yang berbeda-beda yang sudah dioptimasi secara otomatis, hal itu bisa dilihat dari nama belakangnya yang menggunakan angka atau keterangan seperti `model_8` pada model `XGBoost` diurutkan ke 3 model terbaik.

Pengamatan sepuluh model terbaik yang dibuat secara otomatis oleh AutoML maka sudah diputuskan bahwa model H2O AutoML dengan `model_id` `“StackedEnsemble_BestOfFamily_AutoML_20200617_063034”` sebagai model terbaik karena memiliki tingkat kesalahan terkecil pada 5 macam kesalahan (`“Error”`). Model teratas dari sepuluh model terbaik ini lalu dimasukkan ke dalam variabel `“lead”` secara otomatis dengan menggunakan fungsi `“leader”` yang telah ada secara default pada library H2O AutoML.

### 3.5. Prediksi Harga Pembukaan Bitcoin

Model terbaik H2O AutoML yang telah dimasukkan ke dalam variabel `“lead”` lalu digunakan untuk memprediksi harga pembukaan Bitcoin dari data *train* dan data *test*. Tujuan memprediksi dari kedua dataset ini untuk memastikan bahwa tingkat *error* model lebih besar terhadap data *test* dan lebih kecil terhadap data *train*.

Variabel `“preds”` dan `“preds1”` digunakan untuk menyimpan hasil dari prediksi model H2O AutoML

terbaik, variabel `“preds”` untuk menyimpan hasil prediksi dari data *test* dan `“preds1”` menyimpan hasil prediksi dari data *train*. Hasil prediksi tersebut diubah menjadi variabel `“df_results”` untuk data *test* dan `“df_results1”` untuk data *train* dengan menggunakan format Pandas DataFrame, untuk menampilkan masing-masing 5 data nilai asli (`“real”`) dan hasil prediksinya terhadap kedua dataset tersebut.

### 3.6. Evaluasi Model

Evaluasi dibutuhkan untuk mengetahui keakuratan sebuah model *machine learning*, evaluasi mampu menghitung akurasi, performa dan tingkat kesalahan (`“error”`). Hasil prediksi dari model H2O AutoML bisa dievaluasi ke akuratnya dengan menggunakan `“Koefisien Determinasi”` yang dilambangkan dengan *R Squared* ( $R^2$ ) dan tingkat *error* model dapat dievaluasi dengan menggunakan metode *Mean Absolute Percentage Error* (MAPE), karena metode ini sangatlah sesuai untuk masalah prediksi statistik yang ada pada dataset Bitcoin.

Koefisien Determinasi ( $R^2$ ) dihitung untuk mengetahui sejauh mana kemampuan sejumlah variabel bebas yang ada dalam model persamaan regresi linier berganda secara bersamaan mampu menjelaskan variabel tidak bebasnya. Nilai  $R^2$  berada di rentang 0 sampai 1. Nilai dikatakan ‘baik’ bila di atas angka 0,5, sedangkan nilai  $R^2$  dibilang ‘tidak baik’ bila di bawah 0,5 [13].

Secara umum mengacu dari hasil penghitungan koefisien determinasinya, maka sebuah model regresi yang dihasilkan dari H2O AutoML bisa dikatakan layak dipakai bila nilai  $R^2$  lebih dari 0,5. Hal ini

karena sebagian besar variabel terikatnya mampu dijelaskan dengan baik oleh variabel bebasnya. Sebaliknya, model regresi linier dianggap tidak layak pakai apabila nilai  $R^2$ -nya di bawah 0,5. Rumus koefisien determinasi adalah sebagai berikut :

$$R^2 = \frac{SSR}{SST} \quad \dots (1)$$

Keterangan :

$R^2$  = Koefisien Determinasi

SSR = *Regression Sum of Squares*

SST = *Total Sum of Squares*

*Mean Absolute Percentage Error* (MAPE) atau rata-rata persentase kesalahan absolut, juga dikenal sebagai rata-rata persentase absolut, adalah ukuran akurasi prediksi metode peramalan dalam statistik, misalnya dalam estimasi tren, juga digunakan sebagai fungsi kerugian ("*loss function*") untuk masalah regresi dalam *machine learning*. Rumus dari MAPE adalah sebagai berikut :

$$MAPE = \frac{1}{n} \sum \left| \frac{Actual - Forecast}{Actual} \right| * 100 (2)$$

Rumus di atas, jika *Actual* sama dengan *Forecast* maka MAPE berharga nol, yang artinya semakin nilai MAPE mendekati 0 maka model sangat akurat dan tidak ada kesalahan dalam memprediksi [12].

Perhatikan gambar 1 merupakan nilai akurasi MAPE dan koefisien determinan.

```
Akurasi R2 dan Error prediksi vs test
R2 = 0.9681189512219353
MAPE: 3.48%

Akurasi R2 dan Error prediksi vs train
R2 : 0.9992115633449767
MAPE: 1.38%
```

Gambar 1. Nilai Akurasi MAPE dan Koefisien Determinan

Hasil dari evaluasi model telah didapatkan untuk kedua dataset yaitu data *train* dan data *test* adalah nilai

koefisien determinasi ( $R^2$ ) sebesar 0.968 untuk data uji, 0.999 untuk data latih dan nilai MAPE sebesar 3.48% untuk data uji dan 1.38% untuk data latih.

### 3.7. Perancangan Tampilan Program

Nilai akurasi dan *error* tidak dapat mudah dipahami oleh sebagian orang, sehingga dibutuhkan visualisasi dalam bentuk grafik pada program untuk mengatasi masalah tersebut [14].

Pembuatan tampilan program bisa dilakukan dengan beberapa cara salah satunya bisa menggunakan coding dengan bahasa pemrograman Python, karena Python memiliki beberapa *library* yang dibuat khusus untuk menampilkan grafik dari data asli dan data yang telah diprediksi oleh model H2O AutoML. Tampilan program yang digunakan adalah grafik plot korelasi dan grafik plot garis, karena dataset yang berbentuk tabel dan mempunyai sifat *time series* sangat cocok dan valid apabila ditampilkan dengan menggunakan kedua plot tersebut.

Langkah pengkodean dengan menggunakan bahasa pemrograman Python yaitu menentukan panjang dan lebar plot visualisasi, penentuan jenis plot visualisasi dan data yang ingin ditampilkan serta warna dari plot itu sendiri, pada kasus ini plot yang akan digunakan adalah *scatter plot* atau diagram pencar yaitu plot visualisasi dengan bentuk titik-titik yang berfungsi untuk melakukan pengujian terhadap seberapa kuatnya hubungan antara 2 (dua) variabel serta menentukan jenis hubungan dari 2 (dua) variabel tersebut apakah hubungan positif, hubungan negatif ataupun tidak ada hubungan sama sekali. Data yang digunakan adalah

dataset asli dari data latih dan data uji serta data prediksi untuk tiap-tiap dataset yang akan terbagi menjadi dua plot.

Langkah pengkodean untuk menampilkan plot visualisasi garis tidak jauh berbeda dengan langkah untuk membuat plot visualisasi korelasi. Pengkodean untuk tampilan plot garis hanya berbeda pada jenis plotnya yaitu *line plot*, lalu untuk membandingkan dengan kedua data prediksi dan kedua dataset maka dalam setiap plot harus dibuat 2 *line plot* dengan dataset, label dan warna yang berbeda agar pembaca atau orang yang melihatnya dapat membedakan dan membandingkan keakuratan data prediksi terhadap data aslinya.

Visualisasi hasil prediksi model H2O AutoML merupakan *output* atau keluaran dari program yang peneliti buat agar memudahkan siapapun yang melihatnya. Pengkodean yang telah dibuat mempunyai *output* yang berbentuk diagram pencar atau disebut juga dengan *scatter plot*. Plot korelasi pada sumbu x yang memiliki nilai data prediksi dan sumbu y yang memiliki nilai asli dari data uji menunjukkan bahwa persebaran titik-titik atau dot-dot yang ada di dalam plot dipengaruhi oleh hasil prediksi terhadap data uji, apabila hasil prediksi tidak sama bahkan jauh dari nilai asli data uji maka titik-titik tersebut akan menyebar ke berbagai tempat dan tidak bisa membentuk sebuah garis lurus yang rapat, hal ini disebabkan karena model memprediksi nilai atau *value* yang tidak ada pada saat proses *training*.

Plot korelasi yang menunjukkan titik-titik atau dot-dot

yang sangat rapat dan membentuk garis lurus, berbeda dengan plot korelasi terhadap data uji karena model H2O AutoML dapat memprediksi dengan tepat dan nilai yang diprediksi tidak jauh berbeda dengan nilai asli pada data latih. Hal ini disebabkan oleh model memprediksi nilai atau *value* yang menjadi *input* untuk melakukan proses pelatihan ("*training*") model.

Plot garis pada hasil prediksi dari model H2O AutoML ditunjukkan oleh warna merah muda dan nilai asli dari data uji ditunjukkan oleh warna hijau. Terlihat bahwa hasil prediksi mengikuti pola yang ada pada data uji namun tidak persis sama dengan data uji, hal ini sudah dibuktikan pada evaluasi model yang menghasilkan tingkat akurasi model terhadap data uji yang mendapatkan nilai ( $R^2$ ) 0.968 dan nilai *error* 3.48%.

Plot garis menunjukkan bahwa model H2O AutoML yang telah dibuat memiliki akurasi yang sangat tinggi dan *error* yang kecil sehingga data prediksi hampir sama persis terhadap data latih. Hal ini sudah dibuktikan pada evaluasi model yang menghasilkan tingkat akurasi model terhadap data latih yang mendapatkan nilai ( $R^2$ ) 0.999 dan nilai *error* 1.38% yang disebabkan karena model memprediksi nilai atau *value* yang menjadi *input* untuk melakukan proses pelatihan ("*training*") model sehingga model dapat memprediksi dengan mudah nilai data latih. Namun plot garis ini juga menunjukkan bahwa model digunakan *overfitting* terhadap data yang digunakan. *Overfitting* adalah suatu keadaan dimana data yang digunakan untuk pelatihan itu adalah yang "terbaik", sehingga apabila dilakukan tes dengan

menggunakan data yang berbeda dapat mengurangi akurasi. Terjadinya *overfitting* saat memprediksi nilai pada data latih adalah hal wajar dan sering digunakan untuk membandingkan dengan model yang memprediksi terhadap data uji, *overfitting* yang tidak boleh dibiarkan adalah saat model memprediksi data uji karena model tersebut akan memiliki tingkat akurasi dan *error* yang sangat tinggi apabila menggunakan data yang berbeda atau belum ada.

### **Kesimpulan**

Model H2O AutoML telah berhasil dijalankan dengan model\_id: StackedEnsemble\_BestOfFamily\_AutoML\_20200617\_063034 merupakan model yang baik. Berdasarkan hasil dari nilai statistik pada hasil memprediksi harga pembukaan Bitcoin model tersebut memperoleh nilai koefisien determinasi ( $R^2$ ) sebesar 0.968 dan nilai *error* sebesar 3.48% terhadap data uji dan nilai koefisien determinasi ( $R^2$ ) 0.999 dan nilai *error* sebesar 1.38% terhadap data latih, dapat dilihat bahwa model H2O AutoML memberikan tingkat akurasi yang cukup baik karena nilai  $R^2$  cenderung mendekati 1 dan *error* hampir 0%.

### **Saran**

Saran untuk pengembangan lebih lanjut untuk menyempurnakan dan memperluas penggunaan aplikasi yaitu dengan mencoba menggunakan dataset dari *cryptocurrency* lainnya, memperbarui dataset dan pengotomatisan aplikasi supaya dapat memprediksi harga Bitcoin atau uang digital lainnya setiap hari secara terus menerus.

### **DAFTAR PUSTAKA**

- [1] Amanda, Rizky dan Yasin, Hasbi dan Prahutama, Alan. "Analisis Support Vector Regression (SVR) dalam Memprediksi Kurs Rupiah terhadap Dollar Amerika Serikat". *Jurnal Gaussian*, vol. 3., nomor 4, pp 849 sd 858, 2014
- [2] Agustina, Novia dan Suparti, Suparti dan Mukid, Moch Abdul. "Pemodelan Data Indeks Harga Saham Gabungan Menggunakan Regresi Penalized Spline". *Jurnal Gaussian*, vol. 4, nomor 3, pp 603 sd 612, 2015.
- [3] Ramadhan, Bagaskara. "Visualisasi Data Untuk Memprediksi Pasar Saham Dari Hasil Pengolahan Data Set S&P 500 Dengan Menggunakan *R-Programming*". *Penulisan Ilmiah*, Universitas Gunadarma, 2018.
- [4] Coinvestasi : Apa itu Cryptocurrency? <https://coinvestasi.com/belajar/apa-itu-cryptocurrency/> (Online, diakses 10 April 2020)
- [5] Situs resmi Bitcoin. Pertanyaan yang Sering Diajukan. <https://bitcoin.org/id/faq> (Online, diakses 10 April 2020).
- [6] Adam, Rian. Mengenal Google Colab. <https://structilmy.com/2019/05/mengenal-google-colab/> (Online, diakses 14 Mei 2020)
- [7] Narayanan, Arvind dan Bonneau, Joseph dan Felten, Edward dan Miller, Andrew dan Goldfeder, Steven. "Bitcoin dan teknologi mata uang kripto: pengantar yang komprehensif". *Princeton University Press*. 2016.
- [8] <https://www.python.org/about/> (Online, diakses 14 Mei 2020).

[9] Dokumentasi H2O.ai. AutoML: Automatic Machine Learning. <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html>  
(Online, diakses 15 Mei 2020)

[10] Ghofur, AM. Koefisien Determinasi dalam Analisis Regresi. <https://maglearning.id/2019/02/24/koefisien-determinasi-dalam-analisis-regresi/>  
(Online, diakses 18 Juni 2020)

[11] Shahebaz, Mohammad. 2018. Gentle Introduction to AutoML from H2O.ai. <https://medium.com/analytics-vidhya/gentle-introduction-to-automl-from-h2o-ai-a42b393b4ba2>  
(Online, diakses 11 April 2020).

[12] Supardi. 2013. *Aplikasi Statistika dalam Penelitian Konsep Statistika yang Lebih Komprehensif*. Jakarta: Change Publication.

[13] Purwati, Jeny. “Likuiditas dan Efisiensi Pasar pada Mata Uang Kripto”. *Skripsi. Fakultas Ekonomi, Manajemen, Universitas Islam Indonesia, Yogyakarta*. 2019.

[14] Nawawi, Hadari. “Analisis Regresi dengan MS Excel 2007 dan SPSS 17”. *PT. Elex Media Komputindo, Jakarta*, 2010.