
ANALISA PERBANDINGAN *MODEL PREDICTION* DALAM PREDIKSI HARGA SAHAM MENGGUNAKAN METODE LINEAR REGRESSION, *RANDOM FOREST REGRESSION* DAN *MULTILAYER PERCEPTRON*

Evita Fitri[✉], Dwiza Riana

Universitas Nusa Mandiri, Jakarta, Indonesia

Email: evita.etv@nusamandiri.ac.id

DOI: <https://doi.org/10.46880/jmika.Vol6No1.pp69-78>

ABSTRACT

Long-term investments that are quite in demand by the public include stocks because this investment brings quite a large profit for its investors. However, in relation to this, stock price movements in general tend to be non-linear and non-stationary, this is because stock prices can be influenced by several factors whose results can change the pattern of stock price values either up or down, so this can make it difficult to stock prices prediction. In this study, a comparative analysis of prediction models was carried out in predicting stock prices using a technical approach based on past data, while the data used were historical stock prices by taking data samples from three issuers from the Indonesian capital market. There are three methods that were tested in this study, including Linear Regression (LR), Random Forest Regression (RFR) and Multilayer Perceptron (MLP). The test was carried out with two data modeling, namely partitioning which was validated with Cross Validation and data modeling with Cross Validation without partitioning. In this study, the prediction model with LR is able to produce a fairly low error prediction value with the lowest RMSE score of 0.010 and the highest RMSE of 0.012, the lowest MAPE of 1.2% and the highest of 1.9%, the lowest MAE of 0.006 and the highest. of 0.009, and the highest R2 value was 99.8% and the lowest was 99.6%. It can be concluded that in this study, the Linear Regression prediction model is able to predict historical data on stock prices better than the RFR and MLP models.

Keyword: *Stock Price Prediction, Linear Regression, Random Forest Regression, Multilayer Perceptron, Data Mining.*

ABSTRAK

Investasi jangka panjang yang cukup banyak diminati oleh kalangan masyarakat diantaranya ada saham, karena investasi ini mendatangkan keuntungan yang terbilang cukup besar bagi para investornya. Namun berkaitan dengan hal tersebut, pergerakan harga saham pada umumnya cenderung bersifat non linear dan non stasioner, hal ini dikarenakan harga saham dapat dipengaruhi oleh beberapa faktor yang hasil dapat merubah pola nilai harga saham baik naik atau turun, sehingga hal ini dapat mempersulit dalam memprediksi harga saham. Dalam penelitian ini, dilakukan analisa perbandingan model prediction dalam memprediksi harga saham menggunakan pendekatan teknikal berdasarkan data masa lampau, adapun data yang digunakan adalah historis harga saham dengan mengambil sampel data pada tiga emiten dari pasar modal Indonesia. Terdapat tiga metode yang diuji coba pada penelitian ini, diantaranya *Linear Regression* (LR), *Random Forest Regression* (RFR) dan *Multilayer Perceptron* (MLP). Pengujian dilakukan dengan dua pemodelan data yaitu *partitioning* yang divalidasi dengan *cross validation* dan pemodelan data dengan *cross validation* tanpa *partitioning*. Dalam penelitian ini, model prediksi dengan LR mampu menghasilkan nilai prediksi kesalahan yang cukup rendah dengan nilai RMSE sebesar 0.010 paling rendah dan RMSE tertinggi sebesar 0.012, MAPE terendah sebesar 1,2% dan tertinggi sebesar 1,9%, MAE terendah sebesar 0.006 dan tertinggi sebesar 0.009, serta nilai R2 tertinggi sebesar 99,8% dan terendah sebesar 99,6%. Hal ini dapat disimpulkan bahwa pada penelitian ini, model prediksi *Linear Regression* mampu memprediksi data historis harga saham lebih baik dibandingkan dengan model RFR dan MLP.

Kata Kunci: *Prediksi Harga Saham, Linear Regression, Random Forest Regression, Multilayer Perceptron, Data Mining.*

PENDAHULUAN

Terkait kondisi keuangan pada masa mendatang, salah satu hal yang dapat mengamankan

kondisi tersebut adalah dengan memiliki investasi jangka panjang. Pada pemanfaatannya, investasi dapat digunakan untuk memenuhi kebutuhan jangka panjang

diantaranya persiapan pensiun, sebagai biaya pendidikan maupun pernikahan yang cukup tinggi di masa depan ataupun sebagai salah satu alternatif dari terhindarnya masa inflasi.

Pada implementasinya seiring berjalannya waktu, investasi membuat dana tersebut semakin banyak tanpa kita sadari, bukan hanya itu melainkan sudah banyak layanan finansial berbasis teknologi canggih yang membantu penggunaanya dalam berinvestasi. Saham merupakan surat berharga yang dikeluarkan oleh sebuah perusahaan yang berbentuk Perseroan Terbatas (PT) atau yang biasa disebut emiten. Pada implementasinya, dengan memiliki saham maka menyatakan bahwa pemilik saham tersebut merupakan juga pemilik sebagian dari perusahaan tersebut (Maulana & Kumalasari, 2019).

Berbicara mengenai saham, investasi saham menjadi pilihan investasi yang cukup banyak diminati oleh kalangan dewasa, hal ini dikarenakan dalam berinvestasi saham dapat memberikan keuntungan yang cukup besar dan juga cepat bagi para investornya. Berkaitan dengan itu, terdapat pula beberapa faktor yang dapat mempengaruhi pola naik atau turunnya harga saham serta ketidakpastian sehingga dapat menyebabkan kerugian bagi para para investornya, diantaranya adalah terdiri dari faktor internal serta faktor eksternal, hal ini dikarenakan sifat harga saham yang fluktuatif.

Hal ini dibuktikan dalam penelitian (Rusyida & Pratama, 2020) yang membahas prediksi harga saham pada PT. Garuda Indonesia, Tbk pada masa pandemi Covid-19. Pada penelitian tersebut diketahui bahwa pada harga saham PT. Garuda Indonesia cukup mengalami penurunan pada saat pandemi Covid-19. Namun tidak semua emiten cenderung mengalami penurunan, sebagian emiten atau perusahaan pun ada yang mengalami perubahan berupa peningkatan penghasilan, sebagai contoh yakni industri obat-obatan yang cenderung lebih dibutuhkan daripada sebelumnya terkait faktor yang mempengaruhi harga saham seperti adanya sebuah wabah. Terkait hal tersebut, pada penelitian ini penulis membahas mengenai analisis perbandingan model prediksi harga saham pada beberapa sektor industri di Indonesia berdasarkan data-data terdahulu, adapun data-data yang akan digunakan oleh penulis merupakan data-data historis harga saham harian yang bersifat data *Time Series* dengan periode tahun 2015 sampai dengan 2020.

Pada penelitian (Wang, 2020) yang membahas mengenai prediksi harga saham berdasarkan pendekatan machine learning dengan menggunakan metode LR dan metode SVM, kedua metode tersebut

dinyatakan efektif untuk digunakan dalam memprediksi harga saham yang mana pada penelitian tersebut diambil sampel pada pasar modal China, pada penelitian ini pun ditemukan bahwa model SVM sudah optimal dengan nilai 17,13% serta nilai maksimal drawdown 0.32 sehingga model ini dapat membantu dalam strategi memprediksi pasar modal di China.

Selanjutnya penelitian peramalan dalam prediksi harga saham pun dilakukan oleh (Saifuddin & Hermawan, 2019) dengan menerapkan algoritma Artificial Neural Network dan menggunakan empat atribut yaitu nilai *open*, *high*, *low* sebagai prediktor dan *close* sebagai *class* yang berfokus pada penentuan nilai RMSE. Adapun hasil dengan mengoptimalkan suatu nilai-nilai parameter dan pemilihan untuk ukuran hidden layer memberikan hasil yang cukup baik yakni dengan nilai akurasi dan RMSE masing-masing sebesar 0.266 +/- 0.000.

Pada penelitian (Maulana & Kumalasari, 2019) melakukan analisa perbandingan model algoritma data mining pada prediksi harga saham GGRM menggunakan model Linear Regression, *Neural Network*, *SVM*, *Gaussian Process*, dan *Polynomial Regression* dengan lima atribut yaitu *date*, *open*, *high*, *low* dan *close* dengan hasil penelitian menyimpulkan bahwa data harga saham GGRM dapat diprediksi dengan menggunakan model *algoritma Neural Network* dengan hasil akurasi dan nilai RMSE sebesar 612.474 +/- 89.402 (mikro: 618.916 +/- 0.000) paling kecil dibandingkan dengan model algoritma lainnya, sehingga dengan prediksi hal ini dapat membantu dalam memprediksi harga saham GGRM di pasar modal.

Berdasarkan hal-hal di atas, maka peneliti bertujuan melakukan perbandingan model prediksi dalam prediksi harga saham pada tiga sektor industri di Indonesia (INDF, KAEF dan TLKM) dengan tujuan berfokus pada pencapaian minimnya tingkat *error* diantaranya *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) dan tingkat nilai R2 menggunakan pendekatan *Machine Learning* dengan model algoritma Linear Regression, *Random Forest Regression* dan *Multilayer Perceptron*.

LANDASAN TEORI Pasar Modal

Definisi yang dikemukakan oleh (Rahmah, 2019) yang menyatakan bahwa pasar modal merupakan tempat bertemunya pihak pihak yang memiliki kelebihan dalam hal kapasitas modal untuk para investor dengan pihak yang membutuhkan tambahan

modal baik modal jangka panjang maupun jangka pendek.

Saham

Pada pasar modal, transaksi jual beli saham selalu dilakukan oleh banyak pihak-pihak terkait. Saham adalah sebuah tanda bukti sebagai penyertaan kepemilikan sebuah modal pada suatu perusahaan yang bersangkutan.

Saham juga merupakan bentuk kertas yang tercantum dengan jelas berisi beberapa informasi diantaranya nilai nominal, nama perusahaan dan diikuti dengan hak dan kewajiban yang dijelaskan kepada setiap pemegangnya atau pemilik saham, selain itu saham pun menjadi salah satu aset yang dimiliki oleh pemegang saham yang nantinya siap untuk dijual (Handini, 2020) Bagi investor, investasi saham merupakan salah satu investasi yang paling banyak diminati, hal ini dikarenakan dapat memberikan tingkat return yang lebih tinggi dibandingkan dengan obligasi dan reksadana. Pendapatan return ini yang nantinya diharapkan bagi para investor, adapun pendapatan return ini terdiri dari dividen dan *capital gain* (Angraini & Yusra, 2019).

Harga Saham

Menurut para ahli terkait harga saham diantaranya oleh (Azis, Mintarti, & Nadir, 2015) yang mendefinisikan harga saham adalah menunjukkan harga pada pasar riil dan merupakan harga yang paling mudah ditentukan karena merupakan harga suatu saham pada pasar modal yang sedang berjalan atau berlangsung, harga saham pun dapat dikatakan harga penutupnya jika pasar tutup. Adapun harga saham sebagai nilai saham, yang mana dijelaskan bahwa nilai tersebut terdiri dari tiga nilai harga saham yaitu nilai buku, nilai pasar dan nilai intrinsik (Jogiyanto, 2016).

Pada penelitian (Christian & Frecky, 2019) menyatakan terdapat faktor-faktor yang mempengaruhi harga saham (per lembar saham) yakni adanya hukum permintaan dan penawaran, berita umum, indeks harga saham tingkat suku bunga, dana asing di bursa efek, mata uang dolar Amerika serta adanya kondisi fundamental.

Data Time Series

Data *Time Series* merupakan susunan observasi berturut berdasarkan waktu, adapun data *Time Series* sendiri yaitu diumpamakan sebagai serangkaian data pengamatan yang asalnya sendiri yaitu dari satu sumber bersifat tetap adapun terjadinya berdasarkan indeksasi waktu secara statis atau berurutan dengan adanya

interval waktu yang bersifat tetap. Data time series menggunakan notasi untuk menyatakan nilai numerik dari suatu pengamatan, dengan menunjukkan adanya periode waktu suatu terjadinya pengamatan. Adapun pada proses prediksi data *Time Series* tidak adanya keterlindatan dengan suatu variabel bebas lain selain indeks waktu itu tersebut. Data tersebut didasarkan pada urutan yang memiliki jarak yang sama sama dalam segi waktu seperti mingguan atau bulanan, kuartalan, dan lain sebagainya. (Izzah, 2017).

Algoritma Linear Regression

Algoritma Linear Regression sendiri merupakan jenis aturan classification dan Regression pada kegiatan data mining, selain Linear Regression yang termasuk pada golongan ini, terdapat pula *Support Vector Machine*, *Logistic Regression* dan lain sebagainya. Analisis regresi linier adalah teknik data mining untuk menentukan bahwa terdapat hubungan antara variabel yang ingin diramalkan dalam hal ini adalah variabel tidak bebas dengan variabel lain dalam hal ini adalah variabel bebas (Yanto, 2018).

Adapun selanjutnya peramalan tersebut didasarkan pada sebuah asumsi mengenai pola pada pertumbuhan data sebuah historis yang bersifat linier, walaupun sebenarnya tidak mencapai 100% linier. Dan pola pertumbuhan ini didekati dengan suatu model yang menggambarkan hubungan-hubungan yang terkait dalam suatu keadaan. Selanjutnya dalam regresi linier, pemodelan data dilakukan dalam bentuk grafik, dengan bentuk garis continues berjumlah dua dimensi. Karena hal tersebut dibutuhkan variabel X dan Y, adapun dalam regresi linier variabel Y adalah sebagai variabel respon dan variabel X disebut variabel prediktor. Adapun formula dari kedua variabel tersebut dapat dilihat dengan rumus sebagai berikut:

$$Y = a + bx$$

Keterangan:

Y = Variabel respon (variabel *dependent*)

a = Konstanta

b = Koefisien regresi (besaran respon yang dihasilkan oleh *predictor*)

x = Variabel prediktor (variabel bebas atau *independent*)

Adapun diperolehnya nilai a dan b dapat dihitung dengan rumus sebagai berikut:

$$a = \frac{(\sum y) (\sum x^2) - (\sum x)(\sum xy)}{n (\sum X^2) - (\sum X)^2}$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

Random Forest Regression

Metode *Random Forest* yang digunakan untuk pemodelan regresi disebut juga *Random Forest Regression* atau RFR. *Random Forest* dikenalkan oleh Breiman pada tahun 2001. *Random Forest* sendiri membangun tree menggunakan sampel bootstrap data yang berbeda dan mengubah cara regresi membangun pohon atau tree. Pada pohon standar, setiap node dibagi menggunakan split terbaik di antara semua variabel, sedangkan pada *Random Forest* setiap node dibagi menggunakan yang terbaik diantara subset prediktor yang dipilih secara acak pada node tersebut. *Random Forest* mempunyai dua buah parameter, yaitu jumlah variabel dalam subset acak pada setiap node dan jumlah pohon (Tresnawati, Kusuma, Wijaya, & Hasibuan, 2019).

Pada pemodelan ini banyak digunakan dalam pemodelan prediktif. Adapun metode ini memadukan percontohan ulang *bootstrap* dan seleksi peubah untuk mengurangi ragam galat prediksi akibat masalah prediktor berkorelasi tinggi dan meningkatkan ketepatan prediksi peubah respon dari metode regresi pohon.

Random Forest Regression sendiri membangun banyak pohon regresi lalu menghitung nilai rata-rata hasil prediksi peubah respon dari semua pohon regresi tersebut. Gugus data asli yang terdiri dari p vektor prediktor x_1, x_1, \dots, x_p dan satu vektor peubah respon (y) dinotasikan dengan Z gugus data baru yang dibuat dengan menggunakan percontohan ulang *bootstrap* dari Z dinotasikan dengan Z^* , dan fungsi pohon regresi pada gugus data hasil *bootraping* ke- b dinotasikan dengan $T_b(x^*)$. Pada algoritma RFR, untuk memprediksi peubah respon dari percontohan ulang yang dilakukan sebanyak B adalah sebagai berikut [7]:

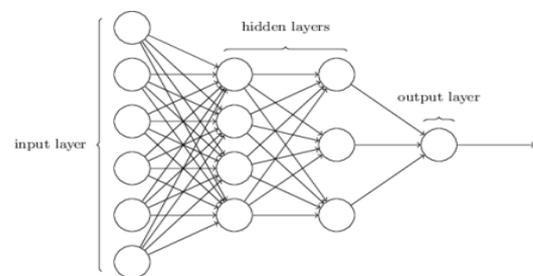
1. Pada percontohan ulang *bootstrap* ke b , dengan $b = 1, 2, \dots, B$, maka lakukan yang pertama, buat gugus data Z^* dari gugus asli menggunakan metode percontohan ulang *bootstrap*, dengan Z^* adalah gugus data yang terdiri dari $p/3$ prediktor (x^*) dan n amatan (jika $p/3$ menghasilkan nilai desimal maka dibulatkan ke atas. Dan kedua buat pohon *Random Forest*, $T_b(x^*)$ dari Z^* menggunakan metode pohon regresi.
2. Prediksi peubah respon dari metode RFR dengan B kali percontohan ulang *bootstrap*, $\hat{f}_{rfr}^B(x)$ diperoleh dengan, $\hat{f}_{rfr}^B(x) = (1/B) \sum_{b=1}^B T_b(x^*)$.

Multilayer Perceptron

Bentuk yang cukup umum digunakan pada peramalan jaringan syaraf tiruan adalah Multilayer Perceptron. MLP sendiri merupakan ANN dari perceptron adapun hidden layer pada ANN feedforward berjumlah satu atau bahkan lebih. Serta pada jaringan ini biasanya terdiri dari satu lapisan neuron yang bersifat komputasi keluaran. Pada MLP, setiap lapisan memiliki fungsi khusus diantaranya lapisan masukan atau input layer berfungsi menerima sinyal atau vektor masukan dari luar dan mendistribusikannya ke semua neuron dalam hidden layer.

Lalu pada lapisan keluaran menerima sinyal keluaran (atau dengan kata lain stimulus pola) dari hidden layer dan memunculkannya sinyal atau nilai atau kelas keluaran dari keseluruhan pada jaringan (Hadimarta, Muhima, & Kurniawan, 2020).

Arsitektur pada *Multilayer Perceptron* secara visual dapat dilihat pada gambar berikut:



Gambar 1. Arsitektur Multilayer Perceptron

Pada visualisasi Gambar diatas dapat dijelaskan bahwa (Fajar & Padjadjaran, 2018) dalam melakukan peramalan, *output* yang dihasilkan hanya satu nilai dengan penggunaan dua *hidden layer* (adapun dalam penggunaannya, *hidden layer* dapat berjumlah lebih dari satu), yang kemudian pada setiap *hidden layer* terdapat 4 node untuk *hidden layer* pertama dan 3 node pada *hidden layer* kedua. Adapun pada *input layer* peramalan *one step ahead*, \hat{x}_{t+1} dihitung dengan menggunakan *input* yang berasal dari variabel lag k yang diteliti (x_{t-k}) dan atau variabel eksogen lainnya.

Pada penerapan MLP, terdapat banyaknya jumlah algoritma pelatihan yang disediakan atau tersedia, namun yang cukup populer adalah *Backpropagation*. Adapun terkait proses pelatihan yang dilakukan oleh algoritma *Backpropagation* dapat dikatakan sama dengan *Perceptron*. Diberikannya pada jaringan sejumlah data latih hal tersebut sebagai data masukan. Perhitungan keluaran pada jaringan dilakukan dan bila ada kesalahan pada kasus ini dilihat dari yang didapatkan dan di ingin perihal perbedaan

antara target dan nilai keluaran, adapun bobot suatu jaringan akan bersifat baru untuk mengurangi kesalahan atau *error*. Adapun persamaannya sebagai berikut (Hadimarta et al., 2020) :

$$v = \sum_i^r x_i \cdot w_i$$

Keterangan :

v = Nilai keluaran *hidden layer*.

x_i = Nilai *input* atau fitur.

w_i = Nilai bobot.

Adapun fungsi aktivasi yang dapat digunakan terdapat dua fungsi yaitu *sigmoid biner* atau *sigmoid bipolar*. Persamaan dalam fungsi aktivasi sebagai berikut:

$$y = \frac{1}{1 + e^{-v}}$$

Keterangan :

y = Nilai sigmoid.

e = Eksponen.

Untuk merambatkan atau menyalurkan sinyal bersifat *error*, dimulai dari layer luar dan kembali ke *hidden layer*. Persamaan sinyal *error* sebagai berikut:

$$e_k(P) = y_{dk}(P) - y_k(P)$$

Keterangan:

e_k = Nilai selisih / *error*.

y_{dk} = Nilai Sebenarnya.

y_k = Nilai prediksi.

Adapun aturan yang digunakan untuk memperbaharui nilai bobot pada koneksi suatu *hidden layer* menuju *output layer* sebagai berikut:

$$w_{jk}(p + 1) = w_{jk}(p) + \Delta w_{jk}$$

Keterangan:

Δw_{jk} = Koreksi nilai bobot

w_{jk} = Nilai bobot

Adapun jika mengalami masukan *neuron* pada *output layer* bersifat berbeda dari masukan *neuron* pada suatu *input layer* x_i . Oleh sebab itu untuk menghitung koreksi suatu nilai bobot adalah untuk menggantikan x_i sinyal *output* suatu *neuron* j pada *hidden layer* y_j . Persamaan sebagai berikut :

$$w_{jk}(p + 1) = w_{jk}(p) + \Delta w_{jk}$$

Keterangan:

η = *Learning rate*.

$\delta_j(p)$ = Gradien *error*.

p = Iterasi

$\Delta w_{jk}(p)$ = Koreksi bobot.

Pada nilai η adalah merupakan laju suatu pembelajaran, lalu $\delta_j(p)$ adalah sebuah gradien *error* pada sebuah *neuron* k dalam layer luaran pada iterasi menuju p .

METODE PENELITIAN

Sumber Data

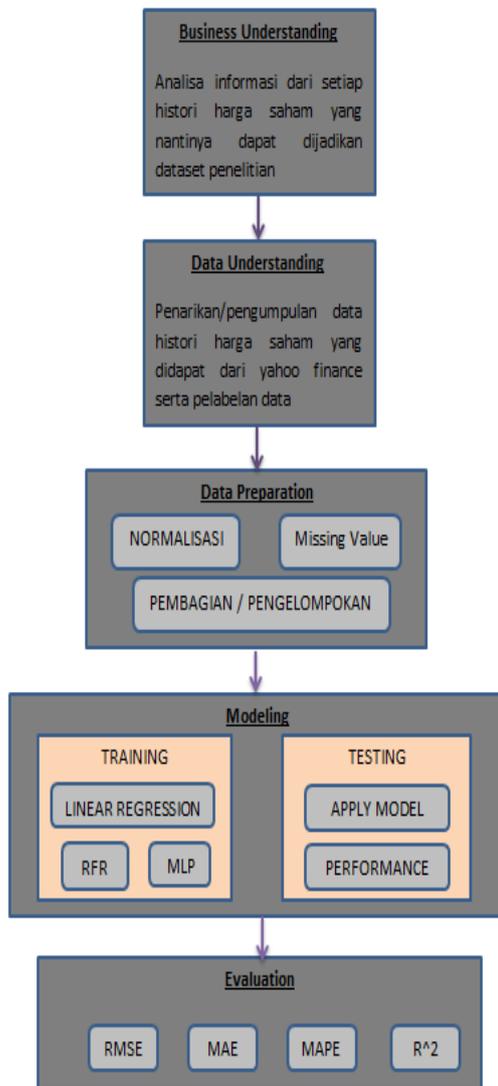
Sumber data yang digunakan adalah menggunakan data sekunder berupa data bersifat *Time Series* dari historis harga saham pada tiga perusahaan yang didapat dari website <https://finance.yahoo.com> dengan jumlah periode pengambilan data selama 5 tahun yaitu 01 September 2015 hingga 01 September 2020, adapun detail dataset yang digunakan dapat dilihat pada gambar berikut:

RINGKASAN DATASET - STOCK PRICE PREDICTION						
Periode Dataset Di Ambil (Dikumpulkan) - 01 September 2015 s.d 01 Septembe 2020 (Rentang Waktu 5 Tahun)						
No	Sektor Industri	Nama Perusahaan	Kode Saham	Source Dataset	Atribut	Jumlah Instance
1	Industri Barang Konsumsi	PT Indofood Sukses Makmur Tbk	INDF.JK	https://finance.yahoo.com/quote/INDF.JK/?p=INDF.JK	Open, High, Low, Close	1263
2	Transportasi & Infrastruktur	PT Telekomunikasi Indonesia Tbk	TLKM.JK	https://finance.yahoo.com/quote/TLKM.JK/	Open, High, Low, Close	1263
3	Industri Bahan Dasar & Kimia	PT Kimia Farma Tbk	KAEF.JK	https://finance.yahoo.com/quote/KAEF.JK/historical?p=KAEF.JK	Open, High, Low, Close	1263
Total Instance Seluruh Dataset						3789

Gambar 2. Detail Sumber Dataset

Metode Usulan

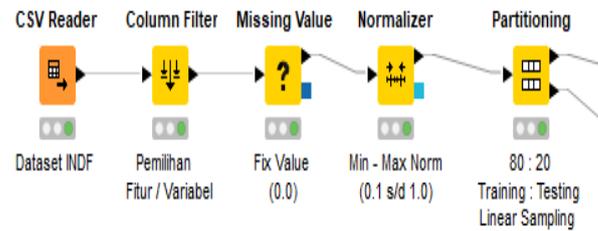
Metode penelitian yang digunakan dalam eksperimen ini menggunakan salah satu metodologi standar dalam penelitian Data Mining yaitu model Cross-Standard Industry for Data Mining (CRISP-DM) yang terdiri dari 6 tahapan diantaranya adalah tahapan Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation dan Deployment. Namun pada penelitian ini dilakukan proses dimulai dari Data Understanding hingga proses Evaluation. Berikut ini merupakan tahapan metode penelitian yang diusulkan:



Gambar 3. Metode Usulan Penelitian

Metode Pengolahan Dataset

Setelah data didapatkan maka dilakukan pengecekan data awal dengan proses preprocessing guna membersihkan data harga saham yang memiliki isi tidak sempurna seperti data yang hilang atau data yang tidak valid, pada penelitian ini, tahap ini disebut sebagai proses missing value data. Selain itu, pada tahap preprocessing ini dilakukan juga normalisasi data dari masing-masing historis harga saham. Adapun tahapan yang dilakukan dalam proses preprocessing pada penelitian yang dilakukan ini yaitu terlihat pada gambar berikut:



Gambar 4. Proses Pengolahan Dataset

Pada proses pengolahan dataset awal dapat dijelaskan sebagai berikut:

- File Reader, proses ini dilakukan sebagai tahap awal pembacaan file dataset bertipe CSV dengan menggunakan node CSV Reader
- Column Filter, digunakan dalam proses preprocessing sebagai proses penyaringan atau pemilihan atribut yang akan digunakan dalam uji coba penelitian, adapun atribut yang digunakan yaitu open, high, low dan close.
- Missing Value, digunakan sebagai proses pembersihan data dari data record yang tidak lengkap atau atribut yang tidak relevan.
- Normalisasi, proses normalisasi dimaksudkan untuk menyeimbangkan nilai antara fitur-fitur yang ada, tehnik yang digunakan ialah Min-Max Normalization.
- Partitioning, pada pengolahan dataset dilakukan partitioning data dengan jumlah pembagian data taining dan data testing sebesar 80 : 20

HASIL DAN PEMBAHASAN

Dataset

Data harga saham yang didapatkan dari website Yahoo Finance diantaranya data date, open price, high price, low price dan close price dengan periode 5 tahun dari 01 September 2015 hingga 01 September 2020 berjumlah 3.789 (jumlah data dari tiga emiten yaitu emiten indofood, kimia farma dan telkom). Adapun dataset yang dikumpulkan dari masing-masing emiten yang digunakan dalam penelitian ini sebagai berikut:

Date	Open	High	Low	Close
2015-09-01	5300	5300	5125	5200
2015-09-02	5125	5250	5075	5200
2015-09-03	5275	5350	5175	5325
2015-09-04	5400	5400	5250	5275
2015-09-07	5175	5200	4925	4940
2015-09-08	4950	4950	484	4900
2015-09-09	4950	5000	4895	4965
2015-09-10	4900	5000	4890	4960
2015-09-11	5000	5000	4960	5175
2015-09-14	5000	5325	5150	5325
2015-09-15	5250	5325	5175	5300
2015-09-16	5325	5375	5150	5250
2015-09-17	5300	5350	5150	5350
2015-09-18	5300	5350	5225	5325
2015-09-21	5175	5325	5150	5225
2015-09-22	5300	5325	5200	5300
2015-09-23	5200	5225	5100	5175
2015-09-25	5075	5000	5050	5150
2015-09-28	5000	5150	4960	5100
2015-09-29	5125	5400	4990	5400
2015-09-30	5325	5525	5325	5325
2015-10-01	5525	5625	5425	5550
2015-10-02	5400	5550	5275	5275
2015-10-05	5325	5550	5325	5525
2015-10-06	5450	5525	5425	5525
2015-10-07	5525	5650	5500	5575

Gambar 5. Dataset Emiten INDF

Date	Open	High	Low	Close
2015-09-01	705	765	700	730
2015-09-02	720	730	690	725
2015-09-03	725	730	710	715
2015-09-04	710	720	695	705
2015-09-07	680	695	670	695
2015-09-08	685	700	670	690
2015-09-09	690	710	690	705
2015-09-10	695	710	685	690
2015-09-11	690	700	645	665
2015-09-14	665	675	660	665
2015-09-15	665	680	660	665
2015-09-16	670	675	660	675
2015-09-17	670	710	670	710
2015-09-18	720	745	685	700
2015-09-21	690	700	660	690
2015-09-22	685	700	670	675

Gambar 6. Dataset Emiten KAEF

Date	Open	High	Low	Close
2015-09-01	2865	2875	2830	2830
2015-09-02	2790	2810	2765	2775
2015-09-03	2805	2820	2785	2800
2015-09-04	2800	2820	2800	2810
2015-09-07	2760	2765	2695	2700
2015-09-08	2680	2740	2680	2730
2015-09-09	2750	2755	2730	2730
2015-09-10	2720	2785	2705	2765
2015-09-11	2800	2800	2760	2760
2015-09-14	2800	2800	2765	2785
2015-09-15	2755	2770	2710	2710
2015-09-16	2720	2745	2700	2720
2015-09-17	2700	2710	2655	2680
2015-09-18	2690	2720	2675	2690

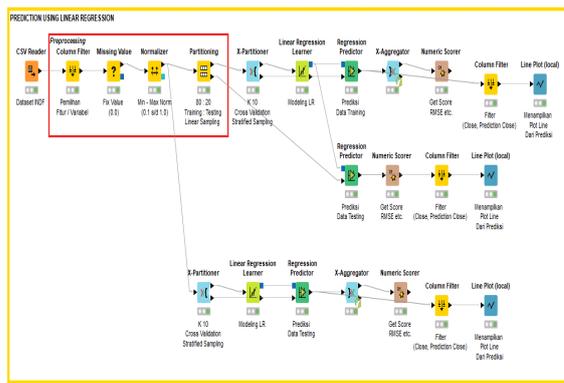
Gambar 7. Dataset Emiten TLKM

PENGUJIAN

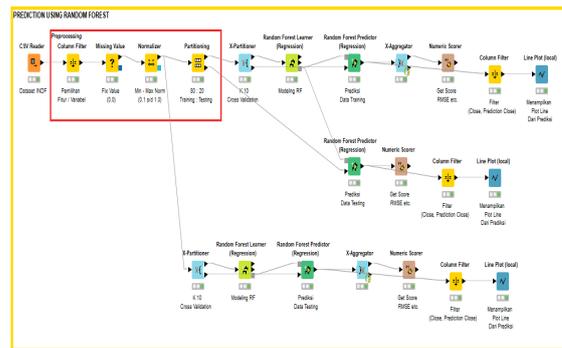
Adapun pemodelan pengujian pada penelitian ini menggunakan desain pemodelan *partitioning* dengan *cross validation* dan desain pemodelan dengan *cross validation* (tanpa *partitioning*).

Design Pemodelan Partitioning Dengan Cross Validation

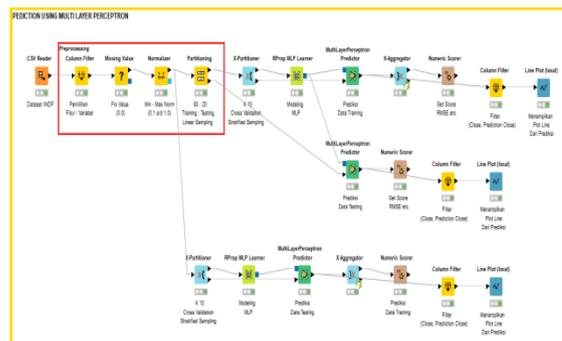
Pada penelitian ini, uji coba dilakukan dengan dua model uji coba, diantaranya uji coba pertama dengan *partitioning* data training dan testing sebesar 80 : 20 setelahnya dilakukan proses validasi dengan 10 K-fold Cross Validation, dan kedua dilakukan dengan cara pengujian model prediksi tanpa *partitioning*. Berikut gambar di bawah ini adalah salah satu desain model pengolahan data historis harga saham yang mewakili hasil desain model yang diuji cobakan pada ketiga dataset emiten dengan menggunakan metode LR, RFR dan MLP sebagai berikut:



Gambar 8. Model Pengolahan Data Menggunakan Linear Regression



Gambar 9. Model Pengolahan Data Menggunakan RFR



Gambar 10. Model Pengolahan Data Menggunakan MLP

Pengujian dataset diatas merupakan pengujian yang melakukan *partitioning* dataset dengan jumlah 80:20 setelahnya dilakukan validasi *cross validation* pada setiap model pengujian (LR, RFR dan MLP).

Adapun pada ujicoba MLP, dilakukan percobaan dalam jumlah hidden layer sebanyak 1 hingga 10 hidden layer, dan didapatkan hasil nilai hidden layer 3,4 dan 5 sebagai nilai percobaan hidden layer yang menghasilkan nilai error terkecil. Adapun pada uji coba setiap model prediksi yang dilakukan didapatkan hasil sebagai berikut:

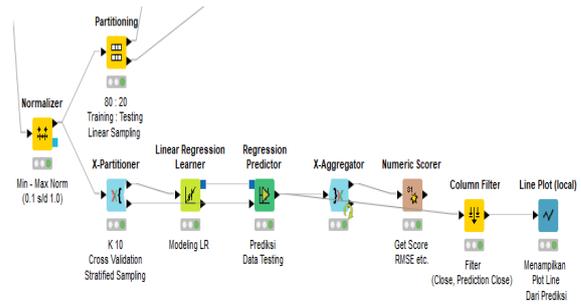
Data	Metode	Data Training				Data Testing				Keterangan	
		RMSE	MAPE	MAE	R2	RMSE	MAPE	MAE	R2		
INDF	LR	0.012	0.019	0.009	0.996	0.012	0.018	0.009	0.996		
	RFR	0.016	0.025	0.012	0.993	0.017	0.025	0.012	0.992		
	MLP		0.020	0.032	0.014	0.990	0.020	0.032	0.015	0.990	5 Hidden Layer
			0.020	0.031	0.014	0.990	0.020	0.031	0.015	0.990	6 Hidden Layer
			0.021	0.033	0.015	0.988	0.020	0.033	0.015	0.990	7 Hidden Layer
KAEF	LR	0.011	0.013	0.007	0.998	0.010	0.012	0.006	0.998		
	RFR	0.013	0.017	0.008	0.997	0.012	0.016	0.007	0.997		
	MLP		0.017	0.025	0.011	0.994	0.016	0.027	0.010	0.995	5 Hidden Layer
			0.017	0.024	0.011	0.994	0.015	0.022	0.009	0.996	6 Hidden Layer
			0.019	0.026	0.011	0.993	0.014	0.023	0.009	0.996	7 Hidden Layer
TLKM	LR	0.011	0.016	0.008	0.997	0.010	0.015	0.007	0.997		
	RFR	0.014	0.021	0.010	0.995	0.012	0.020	0.009	0.996		
	MLP		0.019	0.030	0.014	0.991	0.018	0.028	0.013	0.992	5 Hidden Layer
			0.019	0.030	0.014	0.991	0.017	0.030	0.013	0.993	6 Hidden Layer
			0.020	0.033	0.015	0.990	0.019	0.031	0.014	0.991	7 Hidden Layer

Gambar 11. Hasil Perbandingan RMSE, MAPE, MAE dan R2 Dari Model Prediksi LR, RFR dan MLP

Hasil rangkuman perbandingan score error pada Gambar 11 di atas menghasilkan kesimpulan dari percobaan dengan tiga model prediksi (LR, RFR, MLP) menggunakan model validasi partitioning sebagai pembagian data training dan data testing yang divalidasi kan dengan Cross Validation yaitu bahwa pada ketiga model algoritma yang diuji cobakan pada setiap dataset ditemukan nilai error terendah ada pada model prediksi Linear Regression sebesar 0.012 untuk emiten INDF, 0.011 untuk emiten KAEF dan TLKM, dan selanjutnya pemodelan dengan metode RFR menghasilkan RMSE sebesar 0.013 pada emiten KAEF, sebesar 0.014 pada dataset TLKM dan sebesar 0.016 pada dataset INDF, hasil prediksi tersebut cukup baik berupa nilai error yang cukup rendah dibandingkan dengan metode MLP.

Design Pemodelan Dengan Cross Validation (Tanpa Partitioning)

Pada penelitian ini, uji coba kedua dilakukan dengan cara pengujian model prediksi tanpa partitioning. Dari keseluruhan jumlah instance dalam dataset (1263 instance per emiten) dimodelkan dengan LR, RFR dan MLP tanpa partitioning (pembagian data training dan testing) melainkan langsung pengujian dengan validasi K-10 Fold menggunakan Cross Validation (node X-Partitioner dan X-Aggregator). Adapun berikut merupakan desain pengujian model menggunakan Cross Validation dengan metode LR, RFR dan MLP sebagai berikut:



Gambar 12. Sampel Desain Model Dengan Cross Validation (INDF, KAEF, TLKM)

Proses di atas menunjukkan beberapa tahapan, bila uji coba pertama dilakukan dengan proses partitioning, untuk uji coba kedua ini setelah proses data transformasi maka selanjutnya dimodelkan dengan Cross Validation menggunakan node X-Partitioner. Setelah dilakukannya evaluasi model menggunakan validasi data dengan Cross Validation pada keseluruhan dataset berjumlah 1263 per emiten selanjutnya yaitu pemodelan learner menggunakan node Learner dari masing-masing model prediksi (Linear Regression, Random Forest Regression dan MLP). Adapun konfigurasi pada tahap Cross Validation menggunakan nilai K atau Number of Validation 10 dengan teknik Stratified Sampling berdasarkan atribut yang menjadi class prediksinya (close). Pada pengujian model prediksi MLP juga digunakan hidden layer 1 hingga 10 sebagai ujicoba untuk menghasilkan nilai tingkat error yang lebih rendah.

Adapun didapatkan hasil dari uji coba setiap model prediksi yang dilakukan sebagai berikut:

DATASET	METODE	DATA TESTING				KET	
		RMSE	MAPE	MAE	R ²		
INDF	LR	0.012	0.019	0.009	0.996		
	RFR	0.018	0.034	0.013	0.991		
	MLP		0.020	0.033	0.015	0.990	5 Hidden Layer
			0.020	0.032	0.015	0.989	6 Hidden Layer
			0.021	0.033	0.015	0.989	7 Hidden Layer
KAEF	LR	0.011	0.013	0.007	0.998		
	RFR	0.013	0.017	0.008	0.997		
	MLP		0.017	0.025	0.011	0.995	5 Hidden Layer
			0.017	0.023	0.011	0.995	6 Hidden Layer
			0.017	0.026	0.011	0.994	7 Hidden Layer
TLKM	LR	0.010	0.016	0.008	0.997		
	RFR	0.014	0.021	0.010	0.995		
	MLP		0.018	0.029	0.013	0.992	5 Hidden Layer
			0.019	0.029	0.014	0.991	6 Hidden Layer
			0.019	0.030	0.014	0.991	7 Hidden Layer

Gambar 13. Hasil Perbandingan RMSE, MAPE, MAE dan R2 Dari Model Prediksi LR, RFR dan MLP (Tanpa partitioning data)

Pada hasil rangkuman perbandingan score error di atas menghasilkan kesimpulan dari percobaan

dengan tiga model prediksi (LR, RFR, MLP) menggunakan model validasi dengan Cross Validation tanpa adanya pembagian data melalui partitioning yaitu bahwa pada ketiga model algoritma yang diuji cobakan pada setiap dataset emiten ditemukan nilai error terendah ada pada model prediksi Linear Regression sebesar 0.010 untuk emiten TLKM, 0.011 untuk emiten KAUF dan RMSE sebesar 0.012 pada emiten INDF, dan selanjutnya pemodelan dengan metode RFR menghasilkan RMSE sebesar 0.013 pada emiten KAUF, sebesar 0.014 pada dataset TLKM dan sebesar 0.018 pada dataset INDF, hasil prediksi tersebut cukup baik karena nilai error yang cukup rendah dibandingkan dengan metode MLP.

KESIMPULAN

Berdasarkan pembahasan dan hasil penelitian yang telah didapatkan, maka dapat disimpulkan diantaranya pada tiga model prediction yang telah digunakan dalam memprediksi histori harga saham yaitu model Linear Regression, Random Forest Regression dan Multilayer Perceptron mampu memberikan nilai error prediksi yang cukup baik pada uji coba dataset pada setiap jenis emitennya dengan rata-rata nilai Root Mean Square Errornya (RMSE) di range 0.010 sebagai nilai terendah hingga sebesar 0.020 sebagai nilai tertingginya.

Dalam keseluruhan uji coba pada dataset dengan model prediksi LR, RFR, MLP serta kedua model validasi yang dilakukan pada setiap model prediksi. Dapat dilihat dari rangkuman hasil penelitian, bahwa metode atau model prediksi Linear Regression mampu memberikan nilai RMSE yang cukup rendah pada setiap jenis dataset emitennya dibandingkan dengan metode atau model prediksi RFR dan juga MLP, dengan nilai RMSE sebesar 0.010 paling rendah dan 0.012 paling tinggi. Adapun error score lainnya yaitu nilai MAPE terendah sebesar 1,2% dan tertinggi sebesar 1,9 %, nilai MAE terendah 0.006 dan tertinggi sebesar 0.009, serta nilai koefisien determinasi (R²) tertinggi sebesar 99,8% dan terendah sebesar 99,6% pada model prediksi LR.

Pada model prediksi RFR memiliki nilai error yang cukup kecil dibandingkan dengan MLP. Pada model prediksi RFR nilai RMSE terkecil sebesar 0.012 dan tertinggi sebesar 0.018, nilai MAPE terendah sebesar 1,6% dan tertinggi sebesar 3,4%, nilai MAE terkecil sebesar 0.007 dan tertinggi sebesar 0.013, nilai R² tertinggi sebesar 99,7% dan terendah sebesar 99,1%. Sedangkan pada model prediksi MLP nilai RMSE terkecil sebesar 0.014 dan tertinggi sebesar 0.020, nilai MAPE terendah sebesar 2,2% dan tertinggi

sebesar 3,3%, nilai MAE terkecil sebesar 0,009 dan tertinggi sebesar 0,015, nilai R² tertinggi sebesar 99,6% dan terendah sebesar 98,9%.

DAFTAR PUSTAKA

- Angraini, I., & Yusra, I. (2019). Pendekatan Data Panel Terhadap Return Saham : Studi Kasus Pada Perusahaan LQ45. *Japanese Journal of Allergology*, 49.
- Azis, M., Mintarti, S., & Nadir, M. (2015). *Manajemen Investasi Fundamental, Teknikal, Perilaku Investor dan Return Saham*. Yogyakarta: Deepublish (Group Penerbit CV Budi Utama).
- Christian, N., & Frecky, F. (2019). Analisis Pengaruh Faktor-Faktor Yang Mempengaruhi Harga Saham Pada Perusahaan Yang Terdaftar Di Bursa Efek Indonesia. *Jurnal Benefita*, 1(1), 115. <https://doi.org/10.22216/jbe.v1i1.3417>
- Fajar, M., & Padjadjaran, U. (2018). *Meningkatkan Akurasi Peramalan dengan Menggunakan Metode Hybrid Singular Spectrum Analysis-Multilayer Perceptron Neural Networks*. (February). <https://doi.org/10.13140/RG.2.2.34999.01443>
- Hadimarta, T. F., Muhima, R. R., & Kurniawan, M. (2020). Implementasi Multilayer Perceptron Pada Jaringan Saraf Tiruan Untuk Memprediksi Nilai Valuta Asing. *INTEGER: Journal of Information Technology*, 5(1), 56–63. <https://doi.org/10.31284/j.integer.2020.v5i1.909>
- Handini, S. (2020). *BUKU AJAR : MANAJEMEN KEUANGAN* (M. Dr. Sri Handayani, ed.). Surabaya: Scopindo Media Pustaka.
- Izzah, A. (2017). Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier. *Kinetik*, 2(3), 141–149. <https://doi.org/10.22219/kinetik.v2i3.268>
- Jogiyanto, H. (2016). *Teori Portofolio dan Analisis Investasi*. Yogyakarta: BPFE-Yogyakarta.
- Maulana, R., & Kumalasari, D. (2019). Analisis Dan Perbandingan Algoritma Data Mining Dalam Prediksi Harga Saham Ggrrm. *Jurnal Informatika Kaputama (JIK)*, 3(1), 22–28.
- Rahmah, M. (2019). *Hukum Pasar Modal*. Rawamangun, Jakarta: Prenada Media Group.
- Rusyida, W. Y., & Pratama, V. Y. (2020). Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA. *Square : Journal of Mathematics and Mathematics Education*, 2(1), 73. <https://doi.org/10.21580/square.2020.2.1.5626>
- Saifuddin, & Hermawan, A. (2019). *Prediksi Data Historis Saham PT. Bank Rakyat Indonesia Tbk (BBRI) Menggunakan Model Algoritma Artificial Neural Network*. 139–144.
- Tresnawati, L. H., Kusuma, W. A., Wijaya, S. H., & Hasibuan, L. S. (2019). Asosiasi Single

- Nucleotide Polymorphism pada Diabetes Mellitus Tipe 2 Menggunakan Random Forest Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(4), 320.
<https://doi.org/10.22146/jnteti.v8i4.531>
- Wang, H. (2020). Stock price prediction based on machine learning approaches. *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/3414274.3414275>
- Yanto, R. (2018). Implementasi Data Mining Estimasi Ketersediaan Lahan Pembuangan Sampah menggunakan Algoritma Simple Linear Regression. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(1), 361–366.
<https://doi.org/10.29207/resti.v2i1.282>