

Perbandingan Algoritma Extreme Gradient Boosting Dan Random Forest Untuk Memprediksi Harga Terendah Saham Dengan Index ISSI

Francisco Ready Permana¹ Iin Ernawati²

Program Studi S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450

fransiscorp@upnvj.ac.id¹, iinernawati@upnvj.ac.id²

Abstrak. Investasi saham merupakan sebuah kegiatan yang cukup aman dan menjanjikan jikalau tau cara menganalisisnya. Cara paling awal untuk menganalisis saham adalah melihat saham berdasarkan indeksinya. Saham dengan indeks ISSI (Indeks Saham Syariah Indonesia) adalah saham yang bisa dijadikan pilihan untuk berinvestasi karena saham ini memiliki tingkat stabilitas yang baik dibandingkan dengan indeks saham lainnya. Maka dari itu, penelitian ini ingin membuat sebuah model *machine learning* yang bisa memprediksi harga terendah saham ISSI sebagai nilai ambang bawah serta membandingkan dua algoritma andal yaitu algoritma *random forest* dan *extreme gradient boosting* menggunakan data saham yang diambil dari situs *google finance*. Tahapannya meliputi identifikasi masalah, studi literatur, menyiapkan data, *load* dataset, *exploratory data analysis*, *preprocessing*, pembagian data, pelatihan data dan evaluasi model. Untuk mengetahui algoritma mana yang lebih baik, kedua algoritma tersebut dibandingkan menggunakan tiga metrik penilaian seperti *Mean Squared Error* (MSE), *Mean Absolute Percentage Error* (MAPE), dan R^2 . Hasilnya, didapatkan bahwa *random forest* memberikan performa yang lebih baik dengan nilai rata-rata MSE uji sebesar sebesar 0.6458, rata-rata MAPE uji sebesar 0.0033, dan nilai rata-rata R^2 sebesar 0.9985.

Kata Kunci: Harga Saham, Saham, ISSI, Perbandingan, Algoritma, *Random Forest*, *Extreme Gradient Boosting*,

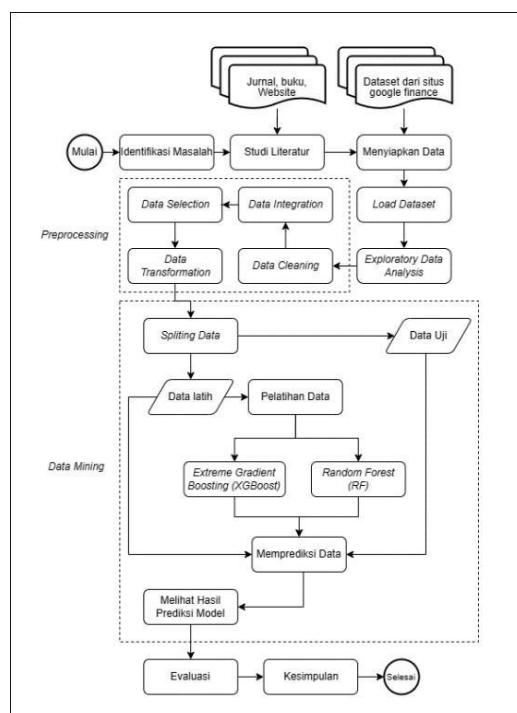
1 Pendahuluan

Menurut refrensi [1] ISSI adalah indeks cukup bagus dan memiliki memiliki tingkat stabilitas yang lebih baik dibanding IHSG. Hal ini juga didukung oleh artikel pada Buletin Keuangan, Investasi, dan ekonomi yang diterbitkan oleh Direktorat SMI pada 15 Juli 2022 menunjukkan bahwa saham dengan indeks ISSI berada pada titik yang menarik perhatian. Sebab walau pada periode tersebut kelima indeks saham memiliki tren yang melemah, fluktuasi saham dengan indeks ISSI masih ada diangka positif yaitu (0,09%) [2]. Hal ini tentunya menarik perhatian untuk dilakukan peramalan atau prediksi.

Untuk memprediksi harga saham, mempelajari pola fluktuasi harga berdasarkan waktu pada hari-hari sebelumnya sangat penting [3]. Machine learning adalah pendekatan yang bisa digunakan untuk mempelajari pola dari data karena pada prosesnya, tidak diperlukan penulisan kode eksplisit[4]. Dalam peramalan harga saham menggunakan machine learning, pemrosesan data timeseries dengan algoritma ensemble learning seperti random forest dan extreme gradient boosting adalah metode yang dapat diandalkan [5]. Penelitian [5] menunjukkan bahwa random forest dapat memprediksi harga saham dengan baik. Sementara itu menurut penelitian [6], extreme gradient boosting juga memiliki performa yang baik dalam prediksi harga saham dan dapat bersaing dengan pendekatan deep learning yang membutuhkan komputasi yang berat. Oleh karena itu, penelitian ini akan membandingkan algoritma-algoritma tersebut untuk meramalkan harga terendah saham dengan menggunakan indeks ISSI sebagai ambang batas bawah berdasarkan metrik tertentu.

Beberapa penelitian serupa telah dilakukan sebelumnya. Penelitian oleh Fitri juga memprediksi harga saham tiga perusahaan (Indofood, Kimia Farma, dan Telkom) dan membandingkan algoritma linear regression, random forest regression, dan multi-layer perceptron. Algoritma random forest memberikan hasil yang baik dengan R^2 sebesar 99.7% [7]. Selain itu, penelitian oleh Chen menggunakan data empat perusahaan (Amazon, Microsoft, Vanguard Total Stock Market ETF, dan Google). Algoritma extreme gradient boosting memberikan nilai MAPE terbaik, yaitu 1.49%, dibandingkan dengan algoritma ARIMA dan LSTM [8].

2 Metode Penelitian



Gambar. 1. Tahapan Metode Penelitian

2.1 Identifikasi Masalah

Investasi saham adalah bentuk investasi yang aman dan menjanjikan, namun memiliki risiko. Untuk mengurangi risiko, diperlukan analisis saham. Salah satu metode yang mudah adalah melihat saham berdasarkan indeks saham. ISSI bisa menjadi pilihan sebab cenderung lebih stabil dibandingkan indeks saham lainnya. Tujuan penelitian ini adalah membuat model *machine learning* untuk memprediksi harga terendah saham ISSI dan membandingkan algoritma extreme gradient boosting dan random forest. Evaluasi dilakukan menggunakan metrik penilaian MSE, MAPE, dan R2.

2.2 Studi Literatur

Pada tahapan ini peneliti melakukan studi literatur untuk mendapatkan teori-teori yang valid terkait dengan penelitian yang dilakukan sebelumnya. Sumber literatur yang digunakan untuk dijadikan referensi antara lain jurnal, artikel, dan situs internet.

2.3 Menyiapkan Data

Setelah data telah dikumpulkan akan dilanjutkan dengan memasukkan data kedalam kode program dan melakukan konfigurasi data dengan mengatur kolom *Date* menjadi indeks serta mengatur *delimiter* yang memisahkan setiap kolom pada data.

2.4 Load Dataset

Data awal yang digunakan untuk penelitian ini adalah data ringkasan saham ISSI yang diambil dari situs *google finance* dari bulan Mei 2011 sampai Juni 2022. Data tersebut berbentuk *timeseries* yang memiliki enam kolom dan 2683 baris yang berisi catatan nilai harga pembuka (*Open*), nilai harga penutup (*Close*), nilai harga tertinggi (*High*), nilai terendah (*Low*), dan nilai *Volume* yang dicatat setiap hari selama sebelas tahun.

2.5 Exploratory Data Analysis (EDA)

Exploratory Data Analysis adalah tahapan sebelum melakukan pelatihan data. Tujuan dari tahapan ini adalah

untuk mengeksplorasi data sehingga dapat mengetahui secara singkat informasi apa yang ada pada data. Ada beberapa hal yang dilakukan pada tahapan ini yaitu melihat statistik deskriptif, dan melihat bentuk persebaran data.

2.5.1 Melihat Statistik Deskriptif

Tahapan awal pada proses EDA adalah melihat statistik deskriptif pada setiap kolom. Tujuannya adalah untuk melihat informasi mengenai nilai rata-rata, standar deviasi dan *interquartile* pada setiap kolom. Selain itu, tahapan ini juga bertujuan untuk melakukan pengecekan apakah terdapat anomali pada data yang dimiliki.

2.5.2 Melihat Bentuk Persebaran Data

Tahap selanjutnya adalah melihat visualisasi data. Tahapan ini bertujuan untuk menganalisis data agar lebih mudah dipahami. Untuk proses visualisasi data, penelitian ini akan menggunakan fungsi dari *package seaborn*, *pandas*, dan *matplotlib*. Tujuannya, adalah untuk menampilkan bentuk persebaran data serta melihat pola atau tren yang ada pada data sehingga data lebih mudah dipahami.

2.6 Preprocessing

Tahapan selanjutnya adalah preprocessing data yang bertujuan untuk mempersiapkan data pelatihan agar bersih, sesuai, dan siap digunakan. Preprocessing mengikuti tahapan KDD (*Knowledge Discovery Data*) seperti *data cleaning*, *integration*, *selection*, dan *transformation*.

2.6.1 Data Cleaning

Tahapan pertama pada *preprocessing* adalah *data cleaning*. Proses *data cleaning* yang dilakukan adalah menghapus kolom yang tidak digunakan dan menghapus data dengan *missing value*. Tujuannya untuk memastikan kolom yang dilatih nanti benar-benar kolom yang bersih.

2.6.2 Data Integration

Tahapan berikutnya adalah *data integration*. Tahapan ini merupakan proses yang mengacu pada kegiatan untuk menggabungkan data yang akan digunakan pada penelitian ini.

2.6.3 Data Selection

Selanjutnya adalah tahapan *data selection*. Pada Tahapan ini akan dilakukan proses pengeleksian data. Memilih dan memilah data mana saja yang relevan untuk dilakukan pemrosesan selanjutnya. Tujuannya adalah untuk menyeleksi atau memilih data-data mana saja yang benar-benar relevan untuk dilakukannya pemrosesan selanjutnya.

2.6.4 Data Transformation

Data yang telah terseleski kemudian akan dilakukan proses transformasi data. Pada tahapan ini akan dilakukan proses untuk mengubah format data menjadi data yang bisa diolah dalam proses *data mining*. Tujuannya adalah untuk menyesuaikan ataupun untuk mentransformasikan data yang terpilih dari proses sebelumnya ke dalam bentuk *mining procedure*.

2.7 Data mining

Setelah tahapan *preprocessing*, dilanjutkan dengan proses *data mining*. Pada proses ini, terdiri dari empat tahapan yaitu proses pembagian data, proses pelatihan data dengan dua algoritma yang ingin dibandingkan yaitu algoritma *random forest* dan algoritma *extreme gradient boosting*, memprediksi data dengan kedua model yang telah dibangun sebelumnya, dan melihat hasil prediksi dari model yang telah dibuat menggunakan diagram visualisasi.

2.7.1 Data Splitting

Dalam data mining, tahap awal adalah pembagian data menjadi kolom fitur dan kolom target. Dalam penelitian ini, kolom target adalah "Low", sementara kolom lainnya adalah fitur. Data kemudian dibagi menjadi tiga bagian menggunakan k-fold cross validation. Pembagian dilakukan secara acak untuk mewakili seluruh data.

2.7.2 Pelatihan Data

Tahap berikutnya adalah pelatihan data, di mana dua algoritma, yaitu XGBoost dan *random forest*, akan diimplementasikan dan dibandingkan performanya. Dalam penelitian ini, proses pelatihan data dilakukan sebanyak tiga kali menggunakan tiga set data latih dan data uji yang dihasilkan dari pembagian data tersebut dengan *k-fold cross validation*. *Hyperparameter* yang digunakan pada kedua algoritma ini mengikuti nilai default dari *package* masing-masing, seperti *n_estimator* sebesar 100, *learning_rate* atau *eta* sebesar 0.3, *reg_lambda* sebesar 1, dan *gamma* sebesar 0.

2.7.3 Memprediksi Data

Pada tahapan ini, dilakukan proses memprediksi data latih dan data uji dari ketiga set data hasil *k-fold cross validation* menggunakan model dengan dua algoritma yang ingin dibandingkan yaitu *random forest* dan XGBoost.

2.7.4 Melihat Hasil Prediksi Model

Tahapan selanjutnya adalah melihat hasil prediksi dari kedua model yang telah dibangun. Pada tahapan ini, dibuat diagram visualisasi agar dapat melihat hasil prediksi model secara keseluruhan dengan lebih mudah.

2.8 Evaluasi

Setelah semua tahapan selesai, dilakukan evaluasi performa model menggunakan algoritma XGBoost dan *random forest*. Evaluasi akan menggunakan data latih dan data uji untuk mendapatkan metrik penilaian seperti MSE, MAPE, dan R2. Metrik-metrik ini digunakan untuk mengevaluasi performa kedua model tersebut dan digunakan sebagai pertimbangan dalam membuat kesimpulan.

3 Hasil dan Pembahasan

Dalam penelitian ini, data yang digunakan adalah data saham dengan indeks ISSI dari tanggal 13 Mei 2011 hingga 1 Juli 2022. Data ini diperoleh dari situs Google Finance. Data tersebut terdiri dari enam kolom dan 2683 baris. Enam kolom tersebut yakni *Date*, *Open*, *High*, *Low*, *Close*, dan *Volume*. Penjelasan lebih rinci tentang masing-masing kolom dapat ditemukan dalam Tabel 1.

Tabel 1. Atribut variabel dataset.

<i>Date</i>	Kolom yang menampung informasi mengenai tanggal saat dilakukannya pencatatan harga saham
<i>Open</i>	Kolom yang menampung informasi harga saham saat pembukaan atau saat suatu saham mulai diperdagangkan pada hari tersebut
<i>High</i>	Kolom yang menampung informasi mengenai harga saham tertinggi saat diperdagangkan pada hari dilakukannya pencatatan
<i>Low</i>	Kolom yang berisi informasi mengenai harga saham terendah saat diperdagangkan pada hari dilakukannya pencatatan
<i>Close</i>	Kolom yang berisi tentang informasi harga saham saat penutupan atau saat pasar saham ditutup ditutup pada hari dilakukannya pencatatan
<i>Volume</i>	kolom yang berisi informasi mengenai jumlah lembar saham yang diperjualbelikan pada hari dilakukannya pencatatan

3.6 Exploratory Data Analysis (EDA)

3.6.1 Melihat Statistik Deskriptif

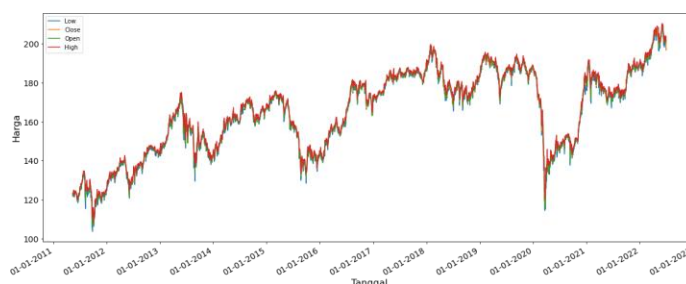
Tahapan pertama pada EDA adalah melihat statistik deskriptif dari data. Statistik deskriptif dari data dapat dilihat pada tabel 2 berikut.

Tabel 2. Statistik Deskriptif Data.

	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>
<i>count</i>	2683	2683	2683	2683	2540
<i>mean</i>	164,717738	165,603645	163,695785	164,68262	904493112.391
<i>std</i>	21,05626	20,998957	21,102087	21,01580	1622197893.152
<i>min</i>	106,64	108,03	103,69	106,64	0
25%	147,31	148,08	146,45	147,325	0
50%	169,13	170,31	168,26	169,18	0
75%	182,46	183,515	181,55	182,53	1023097424
<i>max</i>	209,75	210,32	208,75	209,31	9467918584

Dalam Tabel 4.3, terlihat adanya anomali yang mencolok pada data, yaitu pada kolom Volume. Kolom Volume memiliki banyak nilai nol, mulai dari nilai terkecil (*min*) hingga nilai *median* (50%). Selain itu, dari Tabel 2, dapat dilihat bahwa selisih antara nilai mean dan median (kuartil 50%) dari keempat kolom yang digunakan (Low, High, Close, Open) tidak terlalu besar. Hal ini menunjukkan bahwa tidak ada outlier (nilai yang sangat besar atau sangat kecil) pada data yang digunakan [9].

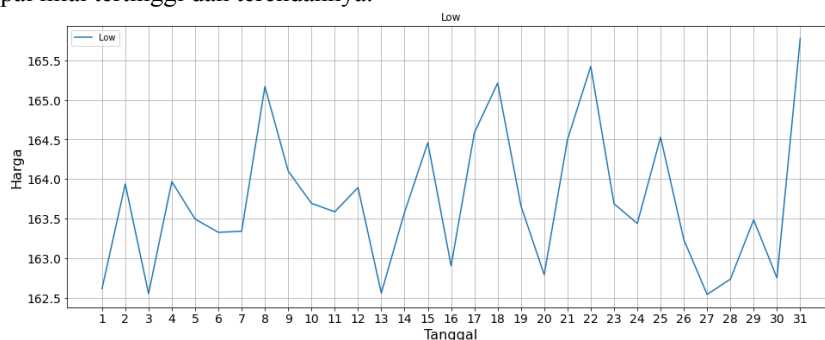
3.6.2 Melihat Bentuk Persebaran Data



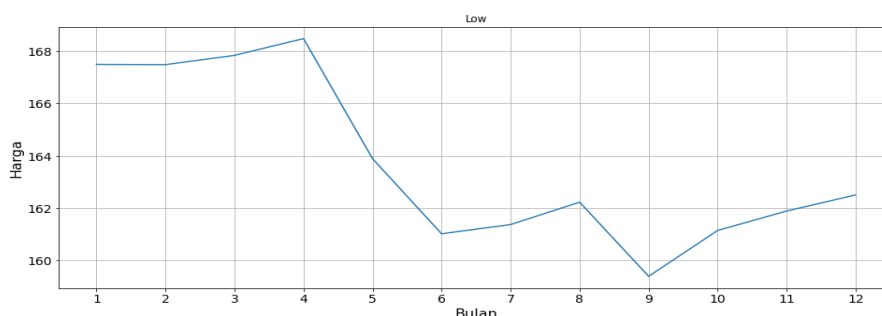
Gambar 2. Fluktuasi Harga Saham Indeks ISSI

Grafik di atas menunjukkan bahwa saham ISSI mencapai nilai terendah pada September 2011 dan nilai tertinggi pada Juni 2022. Terlihat juga adanya penurunan yang signifikan pada tahun 2019-2021, dipengaruhi oleh pandemi covid-19 yang mempengaruhi perekonomian global. Penelitian pada jurnal [10] menunjukkan bahwa ISSI mengalami penurunan selama pandemi, tetapi masih lebih baik daripada saham non-syariah. Namun, setelah periode COVID-19, fluktuasi saham ISSI mulai membaik dan mencapai puncaknya pada Juni 2022.

Selain itu, penelitian juga dilanjutkan dengan melakukan observasi mengenai pada tanggal dan bulan berapa saja saham ini mencapai nilai tertinggi dan terendahnya.



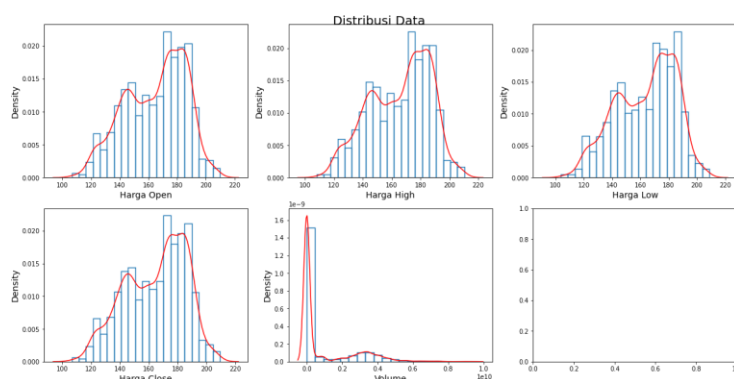
Gambar 3. Fluktuasi Harga Saham Indeks ISSI Berdasarkan Tanggal



Gambar 4. Fluktuasi Harga Saham Indeks ISSI Berdasarkan Bulan

Berdasarkan gambar 3, dan gambar 4, ditemukan bahwa jika data dikelompokkan berdasarkan tanggal, secara rata-rata saham dengan indeks ISSI cenderung mencapai titik terendahnya pada tanggal 27 dan titik tertingginya pada tanggal 31. Sementara itu, jika data dikelompokkan berdasarkan bulan, secara rata-rata saham dengan indeks ISSI cenderung mencapai titik terendahnya pada bulan September dan titik tertingginya pada bulan April.

Selanjutnya, adalah melakukan observasi mengenai bentuk distribusi data. Tujuannya adalah untuk melihat bagaimana pola persebaran dari data yang dimiliki. Untuk itu, dibuat diagram visualisasi untuk mempermudah proses observasi. Gambar 5 adalah histogram yang dibuat dengan *package* seaborn.



Gambar 5. Histogram Bentuk Persebaran Data

Histogram pada Gambar 5 memberikan informasi bahwa data pada kolom *Open*, *Low*, *Close*, dan *High* memiliki distribusi yang baik. Namun, pada kolom *Volume*, distribusi datanya condong ke sebelah kiri. Selain itu, histogram tersebut juga menunjukkan bahwa nilai-nilai *Open*, *Low*, *Close*, dan *High* banyak berkumpul di rentang nilai 130 hingga 190, sedangkan kolom *Volume* didominasi oleh nilai nol.

Penelitian dilanjutkan dengan menganalisis korelasi antara fitur-fitur. Diagram visualisasi heatmap pada Gambar 6 menunjukkan hubungan antara variabel-variabel dalam data penelitian ini.



Gambar 6. Heatmap Korelasi Variabel pada Data

Dalam Gambar 6, terlihat bahwa kolom Open, High, Low, dan Close memiliki korelasi yang kuat, ditunjukkan oleh warna kuning di *heatmap*. Hal ini menunjukkan hubungan erat antara kolom-kolom tersebut, sehingga perubahan nilai pada satu kolom akan berdampak pada kolom lainnya. Namun, kolom Volume tidak memiliki korelasi signifikan dengan kolom lain. Nilai korelasi antara Volume dengan kolom lain hanya sekitar negatif 0.24, menunjukkan bahwa perubahan nilai Volume tidak mempengaruhi kolom lain secara signifikan, begitu pula sebaliknya.

3.7 Preprocessing

Preprocessing yang dilakukan pada penelitian ini mengacu pada alur proses pada *data mining* atau biasa disebut juga *knowledge discovery data* (KDD). Proses pada KDD meliputi *data cleaning*, *data integration*, *data selection*, dan *data transformation*.

3.7.1 Data Cleaning

Tahap pertama dalam preprocessing adalah data cleaning. Pada tahapan ini, dilakukan observasi dan penghapusan missing value pada kolom data yang digunakan. Ditemukan bahwa hanya kolom Volume yang memiliki 143 baris data dengan missing value. Namun, proses pembersihan data akan dilakukan setelah peninjauan ulang mengenai relevansi kolom Volume.

3.7.2 Data Integration

Tahap berikutnya adalah *data integration*, di mana data dari beberapa sumber digabungkan. Namun, dalam penelitian ini, data hanya berasal dari situs *Google Finance*. Oleh karena itu, proses *data integration* hanya melibatkan pengambilan data dari situs *Google Finance* yang telah dilakukan sebelumnya.

3.7.3 Data Selection

Tahap selanjutnya adalah data selection, di mana kolom Volume dipertimbangkan untuk dihapus berdasarkan hasil EDA dan data cleaning. Studi sebelumnya [11] menunjukkan bahwa kolom Volume tidak relevan dan memiliki missing value, nilai anomali, dan tidak berkorelasi dengan kolom lainnya. Oleh karena itu, kolom Volume tidak digunakan dalam penelitian ini, sehingga data yang digunakan tidak mengandung kolom Volume.

3.7.4 Data Transformation

Selanjutnya adalah data transformation. Pada penelitian ini, tidak perlu dilakukan proses transformasi data karena data saham yang diperoleh sudah dalam format CSV yang dapat langsung diolah oleh bahasa pemrograman Python yang digunakan.

3.8 Data Mining

Setelah data dilakukan *preprocessing*, data siap untuk memasuki tahapan selanjutnya yaitu proses *data mining*. *Data mining* yang dilakukan pada penelitian ini adalah untuk membuat sebuah model *machine learning* dengan dua algoritma yang ingin dibandingkan yaitu *random forest* dan *extreme gradient boosting*. Kemudian, membandingkan performa kedua algoritma tersebut dengan beberapa metrik penilaian. Tahapan ini terdiri dari proses pembagian data, pelatihan data, dan memprediksi data. Untuk nilai *hyperparameter* yang digunakan pada kedua algoritma yang ingin dibandingkan adalah nilai bawaan dari *package* masing-masing.

3.8.1 Pembagian Data

Pada tahap ini, data dibagi menjadi kolom fitur dan kolom target. Kolom *Low* menjadi kolom target, sedangkan kolom *High*, *Close*, dan *Open* menjadi kolom fitur. Selanjutnya, data dibagi menjadi data latih dan data uji menggunakan metode *k-fold Cross Validation* dengan nilai *k* sebesar tiga. Metode ini membagi data menjadi beberapa bagian untuk melihat performa model pada set data latih dan uji yang berbeda. Tujuannya adalah untuk menganalisis apakah model cenderung *overfitting* atau tidak. Dalam penelitian ini, data akan dibagi menjadi tiga bagian dengan pengambilan acak untuk memastikan representasi keseluruhan data.

Tabel 3. Salah Satu Contoh Data Latih Hasil *K-Fold Cross Validation*

	Fitur			Target
	<i>Open</i>	<i>High</i>	<i>Close</i>	<i>Low</i>
13/05/2011 15:00	122.95	123.28	122.95	122.61
16/05/2011 15:00	122.03	122.94	122.03	121.59
18/05/2011 15:00	123.46	123.59	123.46	122.07
19/05/2011 15:00	123.81	124.16	123.81	123.44
20/05/2011 15:00	124.84	124.84	124.84	123.82

3.8.2 Pelatihan Data

Setelah data dibagi menjadi data latih dan data uji, langkah selanjutnya adalah melatih data menggunakan

algoritma yang akan dibandingkan, yaitu *random forest* dan XGBoost. Dalam penelitian ini, jumlah pelatihan data menyesuaikan jumlah k pada k -fold yang ditentukan yaitu 3. Selanjutnya, performa kedua algoritma akan diuji menggunakan tiga metrik penilaian, yaitu MSE, MAPE, dan R2. Nilai metrik penilaian dari ketiga hasil pelatihan akan diambil rata-ratanya sebagai evaluasi performa algoritma.

3.8.3 Memprediksi Data

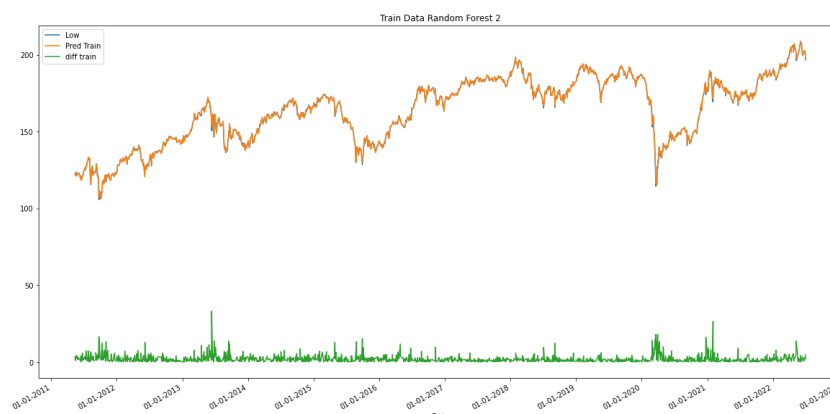
Setelah melatih data dengan tiga set data latih dan data uji menggunakan k -fold cross validation dan dua algoritma yang dibandingkan (*random forest* dan *extreme gradient boosting*), langkah selanjutnya adalah menggunakan model yang telah dibangun untuk memprediksi data menggunakan data latih dan data uji. Tabel 4 menunjukkan contoh hasil prediksi dari model yang telah dibuat dengan menggunakan lima baris data yang diambil secara acak.

Tabel 4. Contoh Lima Data Hasil Prediksi Model

	Fitur			Aktual	Prediksi	
	Open	High	Close	Low	RF	XGB
27/04/2021	176.150	176.550	174.780	174.250	174.371	174.528
09/03/2017	174.620	174.950	174.830	174.330	174.243	174.256
18/06/2014	160.440	161.180	160.440	160.310	159.863	159.841
30/09/2019	189.120	189.360	188.930	188.040	188.075	188.065
28/02/2019	194.160	194.160	191.200	190.850	190.680	190.817

3.8.4 Melihat Hasil Prediksi Model

Setelah melakukan prediksi data menggunakan kedua algoritma yang telah dilatih sebelumnya, tahap selanjutnya adalah mengevaluasi hasil prediksi dari kedua algoritma pada ketiga set data latih dan data uji hasil pembagian data dengan k -fold cross validation. Tujuannya adalah untuk mengetahui sejauh mana perbedaan antara nilai prediksi dari model yang telah dibuat dengan nilai aktual pada keseluruhan data. Untuk itu, digunakan diagram visualisasi yang menampilkan nilai aktual, nilai prediksi, dan selisih antara keduanya. Selisih nilai dikalikan dengan sepuluh untuk memberikan gambaran yang lebih jelas tentang tingkat perbedaan antara nilai aktual dan nilai prediksi. Garis biru menunjukkan nilai aktual, garis oranye menunjukkan nilai prediksi, dan garis hijau menunjukkan selisih antara nilai aktual dan nilai prediksi yang dikalikan sepuluh. Gambar 7 merupakan salah satu contoh diagram hasil prediksi dari model yang telah dibuat.



Gambar 6. Contoh visualisasi Hasil Prediksi Model

3.9 Evaluasi

Tabel 5. Nilai Metrik Penilaian Model

		Rata-Rata nilai metrik		
		MSE	MAPE	R2
Latih	Random Forest	0.0927	0.0013	0.9998
	XGBoost	0.0808	0.0013	0.9998

Uji	Random Forest	0.6458	0.0033	0.9985
	XGBoost	0.8019	0.0037	0.9982

Model *random forest* dan *extreme gradient boosting* memiliki performa yang baik, tetapi *random forest* memberikan performa lebih baik. Hal ini terlihat dari nilai MSE rata-rata pada tabel 5. Random forest memiliki MSE latih 0.0927 dan MSE uji 0.6458, sedangkan XGBoost memiliki MSE latih 0.0808 dan MSE uji 0.8019. Random forest lebih baik dalam metrik MSE karena XGBoost cenderung overfitting dengan perbedaan tingkat MSE yang lebih besar antara latih dan uji.

Untuk metrik penilaian MAPE, pemenangnya adalah *random forest* karena nilai rata-rata MAPE latih *random forest* dari ketiga *k-fold* ada diangka 0.0013 dan nilai rata-rata MAPE uji *random forest* dari ketiga *k-fold* ada diangka 0.0033. Disisi lain, MAPE latih XGBoost dari ketiga *k-fold* ada diangka 0.0013 dan nilai MAPE uji XGBoost dari ketiga *k-fold* ada diangka 0.0037. Hal itu berarti bahwa *random forest* menang tipis atas XGBoost pada nilai MAPE uji.

Random forest mengungguli XGBoost pada metrik penilaian R2. Nilai R2 latih keduanya sama, yaitu 0.9998. Namun, pada R2 uji, random forest sedikit unggul dengan nilai 0.9985, sedangkan XGBoost memiliki nilai 0.9982. Dengan demikian, dapat disimpulkan bahwa random forest memberikan performa yang lebih baik daripada XGBoost. Meskipun pada saat pelatihan data ditemukan bahwa random forest memerlukan waktu pelatihan yang lebih lama, sekitar 3.2 detik, dibandingkan dengan XGBoost yang hanya memerlukan waktu 1.8 detik.

4 Kesimpulan dan Saran

4.6 Kesimpulan

Random forest dan XGBoost keduanya memberikan performa yang baik dalam memprediksi harga terendah saham dengan indeks ISSI. Random forest memiliki waktu pelatihan data sekitar 3.2 detik, sedangkan XGBoost lebih cepat dengan waktu sekitar 1.8 detik. Namun, berdasarkan ketiga metrik penilaian pada tabel 5, random forest memberikan nilai MSE dan MAPE yang lebih rendah serta nilai R2 yang lebih tinggi dibandingkan XGBoost. Dengan demikian, dapat disimpulkan bahwa random forest lebih akurat dalam memprediksi harga terendah saham dengan indeks ISSI dibandingkan dengan XGBoost.

4.7 Saran

Saran untuk penelitian selanjutnya adalah melakukan tuning hyperparameter dari kedua algoritma yang digunakan, yaitu random forest dan XGBoost, untuk meningkatkan performa. Selain itu, dalam penelitian ini, kolom Volume tidak digunakan karena memiliki nilai anomali, missing value, tidak berkorelasi dengan kolom lain, dan tidak relevan berdasarkan penelitian terkait. Oleh karena itu, untuk penelitian selanjutnya, disarankan untuk mencari data dengan kualitas yang lebih baik untuk kolom Volume dan mencari cara untuk memprosesnya agar memberikan informasi tambahan yang berguna.

5 Referensi

- [1] F. Anggun Pratitis and T. Andre Setiyono, “Komparasi Indeks Saham Syariah Indonesia (ISSI) Sebelum dan Saat Pandemi Covid-19 Volume 1 Number 1 2021,” 2021. [Online]. Available: <http://e-journal.iainpekalongan.ac.id/index.php/jief/issue/current>
- [2] T. A. J. Barata, P. Santos, and A. Aulia, “Buletin Investasi, Keuangan, dan Ekonomi,” 2022.
- [3] Q. Qi, “Analysis and Forecast on the Price Change of Shanghai Stock Index,” *Journal of Economics, Business and Management*, vol. 10, no. 1, pp. 72–78, 2022, doi: 10.18178/joebm.2022.10.1.676.
- [4] J. Wira and G. Putra, “Pengenalan Konsep Pembelajaran Mesin dan Deep Learning Edisi 1.4 (17 Agustus 2020),” 2020.
- [5] V. U. K, V. U. Gowda, M. K. Y, and K. Chandrashekar Assistant Professor, “Stock Price Trend Forecasting Using Machine Learning,” 2021. [Online]. Available: www.ijisrt.com
- [6] A. B. Gumelar *et al.*, “Boosting the accuracy of stock market prediction using XGBoost and long short- term memory,” in *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, iSemantic 2020*, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 609–613. doi: 10.1109/iSemantic50169.2020.9234256.
- [7] E. Fitri and D. Riana, “Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Menggunakan Metode Linear Regression, Random Forest Regression Dan Multilayer Perceptron,” *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 6, no. 1, pp. 69–78, Apr. 2022, doi: 10.46880/jmika.Vol6No1.pp69- 78.
- [8] L. Chen, “Stock Price Prediction using Adaptive Time Series Forecasting and Machine Learning Algorithms,” 2020.
- [9] J. Yang, S. Rahardja, and P. Fränti, “Mean-shift outlier detection and filtering,” *Pattern Recognit*, vol. 115, Jul. 2021, doi: 10.1016/j.patcog.2021.107874.
- [10] A. S. Utomo, “Kinerja Saham Syariah Pada Masa Pandemi Covid 19,” 2022, doi: 10.37729/sjmb.v18i1.7662.